POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

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

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

Author: **Andriana Takić**

Student ID: 941155
Advisor: Prof. Augusto Sarti
Co-advisors: Sebastian Gonzalez, Alessandra Calcagno
Academic Year: 2021-22

Abstract

Music is often related to emotions, as it has the ability to affect, shape, and even manipulate emotional states of individuals. Understanding the relationship between music and emotions can help the development of affective computing systems that incorporate emotions into human computer interaction (HCI). The results of research in this field can be applied to various technical innovations, such as software that provide recognition of emotions induced by music or automatic music synthesis. 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, supported by self-report questionnaires. A listening experiment was designed and conducted with 10 recruited subjects who were asked to wear non-invasive equipment 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 analysis of various statistical features was performed, exploring the correlation between the manually annotated tension in music and the responses captured by the recorded physiological data. The correlation between subjects' physiological and self-report responses was also analyzed, as well as the differences between statistical measures of each subject. The main contributions of this work are the created experimental protocol with detailed instructions, the acquired dataset containing data for the four physiological signals for 10 subjects, as well as their responses to self-report questionnaires, then the proposed pipeline for data processing and analysis, and the preliminary results. The limitations concerning the number of subjects affected the acquired dataset size and thus the robustness of different statistical measures and the applicability of the conclusions that were reached. 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 and responses of individuals.

Keywords: physiological signals, induced emotions, listening experiment, MIR, EEG

Abstract in lingua italiana

La musica è spesso correlata alle emozioni, poichè è in grado di generare, formare o addirittura manipolare lo stato emotivo degli individui. Comprendere la relazione tra la musica e le emozioni può aiutare lo sviluppo di sistemi di Affective Computing che incorporino le emozioni nell'Interazione Uomo-Computer (HCI). I risultati della ricerca in questo campo possono essere applicati a varie innovazioni tecnologiche, come i software di riconoscimento di emozioni indotte dalla musica, o la composizione musicale automatizzata. L'obiettivo del presente studio è l'esplorazione delle risposte emotive umane a diverse caratteristiche temporali nella musica, in particolare sulla tensione armonica, tramite l'acquisizione di vari parametri fisiologici, supportata dall'utilizzo di questionari. Un esperimento di ascolto è stato impostato e condotto con 10 soggetti, con indosso della strumentazione non invasiva, a cui sono stati monitorati e registrati i segnali fisiologici mentre ascoltavano certi stimoli musicali. I segnali scelti per l'acquisizione erano elettroencefalogramma (EEG), elettrocardiogramma (ECG), attività elettrodermica (EDA), e respiratoria (RSP). L'analisi di vari indicatori statistici è stata condotta, esplorando la correlazione tra la tensione musicale annotata manualmente e le risposte estrapolate dai dati fisiologici registrati. Inoltre, sono state analizzate anche la correlazione tra le risposte dei soggetti e le differenze tra misure statistiche su ciascun soggetto. I contributi principali di questo lavoro sono i protocolli sperimentali, con istruzioni dettagliate, i dataset con i dati dei 4 segnali fisiologici per 10 soggetti, le loro risposte ai questionari, la catena di processamento, e analisi e i risultati preliminari. La limitatezza del numero di soggetti ha influenzato la dimensione del dataset, e di conseguenza, la robustezza di diverse misure statistiche, nonchè l'affidabilità delle conclusioni ottenute. L'esplorazione di più indicatori e delle loro connessioni, nonchè la combinazione con dati fMRI o fNIRS, potrebbe dare adito a sistemi di Affective Computing più avanzati , utilizzabili in ambito sanitario, terapeutico, dell'intrattenimento o altri campi dove è richiesto di monitorare o influenzare lo stato emotivo o le risposte emotive degli individui.

Parole chiave: segnali fisiologici, emozioni indotte, esperimento di ascolto, MIR, EEG

Contents

1 | Introduction

1.1. Motivation

The recent advance in computer science, engineering, psychology, and neuroscience has motivated the development of systems and devices that aim to recognize, interpret, and process human emotions. Particularly, a new branch of artificial intelligence - *affective computing* - has emerged, aiming to explore the ability to give machines emotional intelligence, including modeling emotional states, sympathy, empathy, and more.

Emotions play a vital role in our daily lives since they reflect the way in which we experience our environment; they are associated with thoughts, behavioral responses, and a degree of pleasure or displeasure [19, 27, 38, 105]. While there are questions on whether emotions cause changes in humans' behavior, their physiology is closely linked to the arousal of the nervous system; according to neurobiology, emotions are considered to be elaborations of arousal patterns in which neurochemicals increase or decrease the brain's activity level. Therefore, modeling and recognizing emotions helps further understanding of another layer of human nervous system, as well as how our surroundings, personal habits, etc. affect our emotional responses.

Due to their complexity and subjectivity, recognizing and evaluating emotions is not such a straight-forward task; it is usually based on three kinds of evidence: *self-report*, *behavior*, and *physiological reaction*. The most common and simple approach to measure emotional responses to music is through a first-person perspective or self-report – either verbal or non-verbal (such as moving a slider, pressing a bar, drawing a picture) [68]. Apart from self-report methods, further research of physiological signals as responses to external stimuli could help building a better understanding of human emotions.

Music is often referred to as a language of emotions and is considered to be a powerful tool for arousing emotions in humans coming from all cultures. According to a study by Juslin and Laukka [68], majority of people report that their primary motivation to listen to music is the emotional effect it has on them. It can be inferred that music has an ability to affect, shape, and even manipulate emotional states of a person; for instance, people tend to listen to specific types of music in order to affect their mood or state of mind [134]. Music therapy is also used as a tool to deal with different psychological and behavioural diseases [26].

The relationship between music and emotions has been a subject of research in various disciplines, such as philosophy, musicology, and sociology. There are numerous theories on how musical expectations are created, maintained, confirmed, or disrupted. Music can communicate and induce a range of powerful emotions [67], which has been the subject of intense scientific investigation. 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* (where the music seems to express a particular emotion with a certain level or agreement among listeners).

Understanding the relationship between music and emotions can also help the development of affective computing systems that incorporate emotions into human computer interaction (HCI) [111]. The results of research in this field can be applied to various technical innovations, such as software that provide recognition of emotions induced by music or automatic music synthesis.

1.2. Related work

Several research methods have been used to explore the relation between music and psycho-physiological responses of humans, most frequently used ones being self-reporting, biological, music-analytic, clinical, individual, and cultural related [34]. Some studies focused on analyzing human emotional states use physical signals such as facial expression [37], speech, posture and more, while others measure the internal physiological signals [54]. Among the latter category, the most commonly used signals are Electroencephalogram (EEG), Functional near-infrared spectroscopy (fNIRS), Electrocardiogram (ECG), Blood volume pressure (BVP), Electromyogram (EMG), Electrodermal activity (EDA), and Respiratory activity (RSP) [34].

The nature of music stimuli used for similar listening experiments can be classified into natural music characterized by changes in its dynamics, tempo, and performer's expression, and computerized music, which involves synthesized sound without expression [74, 75]. The latter is therefore easier to describe, control and measure, but also more limited in the aspect of musical expression. The music genres used in the field are mainly classical and commercial music. The music stimuli mainly include short excerpts of few seconds, but also whole music pieces that are several minutes long [34]. Some studies focus on temporal analysis of the acquired data [74, 75, 89], while some others compare elicited emotions to diverse genres, therefore averaging responses over the whole music excerpt [15, 71, 88, 120].

1.3. Summary of this study

This study aims to explore the human emotional reactions to temporal features in music, focusing on harmonic tension, by means of acquiring physiological data of subjects and their responses to self-report questionnaires. The first version of the study started as "Sound Resonance Project" back in 2019, in a collaboration between Politecnico di Milano – Department of Electronics, Information and Bioengineering (DEIB) and Johns Hopkins University (USA) - International Arts and Mind Lab, Brain Science Institute. The goal was to acquire physiological signals of the attendees of a polyphonic choir's concert, while they listen to the performance. For that purpose, the company Empatica Srl (Milan, Italy) [78] provided a number of wearable sensors in form of wristbands - Empatica E4, which measures electrodermal activity (EDA), heart rate variability (HRV), and several other signals. The preliminary results did not show significant correlation among the subjects, nor between the chosen music features and the physiological data, possibly due to the low sampling rate of the sensors (4 Hz). The project was thus soon suspended, also due to the pandemics of the SARS-CoV-2 virus.

Nevertheless, it was considered important to continue the research on this topic; therefore, as a continuation of the "Sound Resonance Project", this study was initiated. The aim of this study was to explore different methods to recognize and interpret induced emotions using physiological signals, provide preliminary assessment on the topic, as well as directions for the future research. This time, the study was done as an internal collaboration between the two laboratories of Politecnico di Milano - Department of Electronics, Information and Bioengineering (DEIB): Image and Sound Processing Lab and Brain Lab. A listening experiment was designed and conducted where subjects were asked to wear non-invasive equipment in order to have their physiological signals monitored and recorded, while they listen to certain music stimuli. The set of music stimuli contained four choral pieces recorded from the previously described concert, as well as 12 simple chord sequences recorded with a MIDI keyboard. According to the research on the state of the art, the exploration of emotions elicited by choral pieces as music stimuli represents a novelty in this field.

The physiological signals chosen to be acquired in the listening experiment are Electroencephalogram (EEG), Electrocardiogram (ECG), Electrodermal activity (EDA), and Respiratory activity (RSP). It is worth noting that, although using physiological signals, this study is used only for the purpose of scientific research and has no diagnostic or medical purposes. In addition to the listening part, the subjects filled several self-report questionnaires that aimed to acquire data related to subjects' lifestyle, musical expertise and preference, and emotional competencies, as well as the elicited emotions while listening to the music stimuli. The questionnaires were added to this study with the assumption that the collected data could be correlated with the subjects' reactions and therefore their physiological responses to certain music stimuli. As the study aims to give preliminary analysis, a limited number of 10 subjects were recruited for the study, all of them being university students of different nationalities and cultural backgrounds. The physiological data from all subjects, as well as their responses to questionnaires were collected and processed. The analysis of various statistical features was performed, exploring the correlation between the manually annotated tension in music and the responses captured by the recorded physiological data. The correlation between subjects' physiological and self-report responses was also analyzed, as well as the differences between statistical measures of each subject. The main contributions of this study are the created experimental protocol, the proposed pipelines for data acquisition and data processing, the preliminary results and elaborations, and suggestions for further work, as well as the acquired dataset of physiological signals and responses to questionnaires.

1.4. Thesis outline

The document is organized as follows: Chapter 2 provides background for the topics explored in this work. Later, Chapter 3 summarizes the work that is done in the field. Chapter 4 provides details on the whole method, including the experimental setup, data acquisition, data processing, and more. Chapter 5 shows and discussed preliminary results of the work. Finally, Chapter 6 contains the main conclusions of the study, explains contributions, and provides suggestions for possible further developments.

2 | Theoretical background

2.1. Emotions

Even though there is no scientific consensus on a universal definition of emotions, they can be referred to as subjective and conscious mental experiences accompanied by particular neurophysiological responses or changes. Emotions are associated with thoughts, feelings, behavioral responses, and a degree of pleasure or displeasure [19, 27, 38, 105]. The emotions also play a big role in natural human interaction [37]. Research on emotion has included many fields such as psychology, medicine, sociology of emotions, and computer science, aiming to explain the origin and function of emotions from different aspects.

2.1.1. Theories of emotions and possible applications

Emotions are complex; they are considered to involve several components, such as subjective experience, cognitive processes, expressive behavior, psychophysiological changes, and instrumental behavior. There are numerous theories about relations of these components; and while many of them question whether emotions cause changes in human's behavior or not, their physiology is closely linked to arousal of the nervous system. For instance, James [59] argued that the perception of what he called an "exciting fact" directly led to a physiological response, known as "emotion". According to James, a stimulus triggers activity in the autonomic nervous system, which in turn produces an emotional experience in the brain. Neurobiology explains human emotions as pleasant or unpleasant mental states organized in the limbic system of the mammalian brain; they are considered to be elaborations of vertebrate arousal patterns in which neurochemicals increase or decrease the brain's activity level. A group of researchers [18, 86, 107] suggested that emotion is related to a group of structures in the center of the brain called the limbic system, which includes the hypothalamus, cingulate cortex, hippocampi, and other structures.

The recent developments in fields of computer science, engineering, and neuroscience has motivated the emersion of a new field called *affective computing*. Affective computing is considered a branch of artificial intelligence that aims at designing systems and devices that can recognize, interpret, and process human emotions [111]. It is considered that such systems could bring notions of empathy and sympathy into artificially intelligent machinery. The applications of these systems could span across healthcare, entertainment, e-learning, marketing, human monitoring, and security [54].

2.1.2. Models of emotions

Various theories of emotion have been introduced in studies of music; however, it is difficult to define a universal representation of emotions to be used in scientific research. Although there is currently no dominating theoretical paradigm in research on music emotion, most researchers implicitly oriented their work to one of the existing models [68]. The two most widely accepted and used emotional models are Discrete Emotional Model (DEM) and Affective Dimensional Model (ADM), while sometimes the Binary Emotional Model (BEM) is also used due to its simplicity [35, 111].

Discrete Emotional Model (DEM)

In his work, Paul Ekman [35–38] has supported the view that emotions are discrete, measurable, and physiologically distinct. Ekman also found that certain emotions seem to be universally recognized and and shared across languages and cultures. His research [37] examined six basic emotions: anger, disgust, fear, happiness, sadness and surprise, represented in Figure 2.1. Plutchik [112], who fundamentally agreed with Ekman's perspective, developed the so called *wheel of emotions*, suggesting eight primary emotions grouped on a positive or negative basis. The wheel of emotions is represented in Figure 2.2.

Figure 2.1: Six basic emotions proposed by Ekman [37], taken from [53].

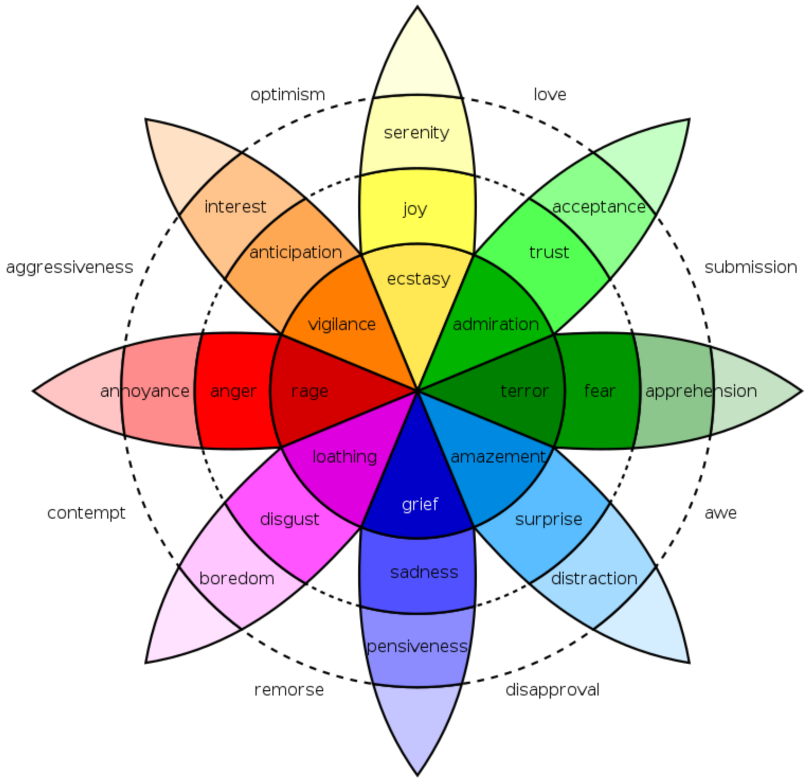

Figure 2.2: Robert Plutchik's wheel of emotions, with eight primary emotions grouped on a positive or negative basis

Coming from those findings, the Discrete Emotional Model (also known as categorical model) proposes distinction of emotions into a number of basic discrete states. The assumption behind the model is that all the other emotions can be derived from the basic ones [35, 128]. The advantage of such approach is the use of natural language and intuitive terms for emotions. On the other side, the drawback is that the number of basic emotions is very small in comparison with the range of various emotions experienced by humans. Additionally, the language used to describe emotions can be ambiguous and subjective.

Affective Dimensional Model (ADM)

On the contrary to the categorical approach, a psychotherapist Michael Graham [47] describes all emotions as existing on a continuum of intensity. According to his proposal, for instance, fear might range from mild concern to terror. Therefore, the ADM (also known as the continuous dimension model) treats emotions as fundamentally similar, differing only in terms of one of more dimensions. Most often, the first two dimensions uncovered by factor analysis are valence and arousal [119]. Valence can be defined as how positive or negative the experience feels, while arousal represents the intensity of the feeling [52]. According to this model, emotions are therefore identified based on their position in the emotion space. The two-dimensional valence-arousal (VA) space is represented in Figure 2.3. The arousal and valence in the suggested model can be associated to certain music features. For instance, the arousal can be related to changes in tempo, pitch, loudness, timbre, while valence can be connected to the scale used (e.g., major or minor) or harmony (e.g. consonant or dissonant) [46].

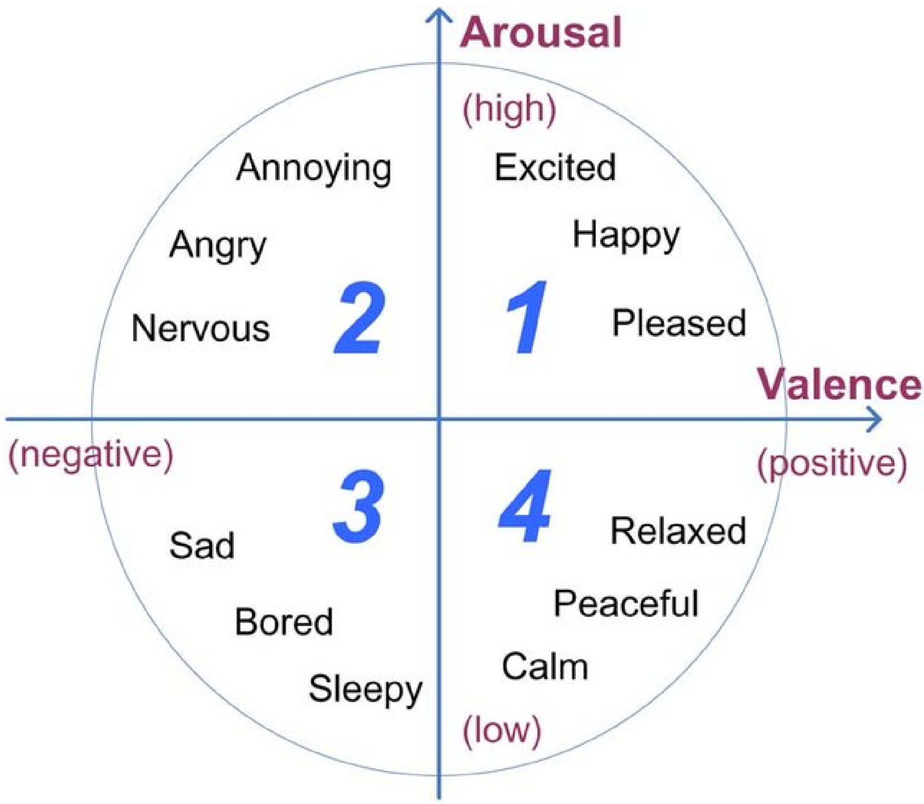

Figure 2.3: Thayer's arousal-valence emotion space [143]

The dimensional approach is often used because of its simple representation using a limited number of dimensions. On the other hand, it is less intuitive for subjects than the categorical approach as it can blur psychological distinctions (e.g. anger and fear could be placed close to each other in the VA space, but have very different impact on the organism). Adding more dimensions to the model could make the annotation process more complex and put additional cognitive load on the subjects.

2.1.3. Emotions and music

The relationship between music and emotions has been a subject of research in various disciplines, such as philosophy, musicology, and sociology. For instance, the chapter [134] provides a psychological explanation for the link between musical activity and emotional states. Meyer's book "Emotion and meaning in music" [97] provides ideas on how musical expectations are created, maintained, confirmed, or disrupted. According to a study by Juslin and Laukka, majority of people report that their primary motivation to listen to music is the emotional effect it has on them [68]. It has also been suggested that musical training is not required to perceive emotions in music [64]. In psychological studies, emotions are often classified in three categories: the emotion expressed by the performer, the emotion perceived by the listener, and the evoked emotion induced by music.

Even though the most subjective one, the induced emotion also turns out to be the most interesting to explore in research. Namely, researchers are usually interested in cases where emotions in music are perceived similarly by different listeners, which is referred to as *listener agreement* (where the music seems to express a particular emotion with a certain level or agreement among listeners). Music is also often used to "manipulate" the emotions of listeners in many areas of society (e.g., in health, sport, or advertising) [68]. Additionally, the results of research in this field can be applied to various technical innovations, such as software that provides recognition of musical emotion or automatic music synthesis.

Inducing basic emotions for physiological data collection in an experiment requires certain guidelines and standard protocols. Music is among five common elicitation techniques, along with audio visual, imagery, memory recall, and the situational procedure [125]. There are different ways in which a musical event may evoke emotions [67]. For instance, some emotions may be aroused by structural characteristics of the music, such as variations in lyrics, melody, tempo, and more [77], while others are related to personal associations and memories and not based on musical factors: what Davies [28] refers to as the "Darling, they are playing our tune" phenomenon. In fact, research shows that listeners often use music as a reminder of valued past events [127] and that specific pieces of music may be associated with particular time periods of an individual's life (e.g., [121]). Therefore, the relationships among music features and emotions are only probabilistic (i.e., uncertain) and are therefore best thought of as correlational, as described by the Juslin's Lens Model [65].

2.1.4. Emotion evaluation

Recognizing and evaluating emotions is not such a straight-forward task; it is usually based on three kids of evidence: *self-report*, *behavior*, and *physiological reaction*. The most common and simple approach to measure emotional responses to music is through a first-person perspective or self-report – either *verbal* (e.g., adjective checklist, quantitative ratings, questionnaire, free description) or *non-verbal* (moving a slider, pressing a bar, drawing a picture) [68]. The questionnaires can vary according to the emotional model that is used. 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 true conscious and unconscious experiences about the stimuli.

In their early work, Salovey and Mayer [94] defined Emotional Intelligence (EI) as the "the ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions". They considered that the high EI is correlated with the individual's emotional abilities and skills related to appraising and regulating emotions in the self and others. Accordingly, it was argued that individuals high in EI could accurately perceive certain emotions in themselves and others and also regulate emotions in themselves and others in order to achieve a range of adaptive outcomes or emotional states [103]. Some of the tools used for self-report emotion evaluation are described in details in Subsection 3.2.2. Another kind of evidence used to infer emotions involves various physiological measures of emotion, details of which are presented in Subsection 3.1.1.

Based on the most previous research on expression, perception, and induction of emotions, in can be considered that most of it has neglected the social context of musical emotion [67]. Even though several studies on emotion in music have revealed limits on reliability of music conveying certain emotions, in some situations the context could add more information to the study, thus helping the analysis of the induced emotions. Nevertheless, such assumption is not certain for now, since previous research has only measured listeners' de-contextualized responses to music in laboratory settings [68]. The lack of social context particularly affects the studies on inducted emotions, as it neglects several critical issues that could help understanding music and emotions, such as listeners' motivation for listening to music and their uses of music in different daily contexts. Therefore, it would be beneficial to consider the functions of musical emotions in their particular context [68].

2.2. Music features

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. Results of several studies over the past century suggest that changes in music attributes are correlated with changes in emotional interpretation [134].

2.2.1. Music Information Retrieval (MIR)

Music Information Retrieval (MIR) is a multidisciplinary field born in the last century, with the aim to describe music and extract information about it. MIR connects fields such as musicology, informatics, signal processing, psychoacoustics, machine learning, and more. It was first mentioned as a term in 1966 by Kassler [70], and nowadays it finds applications in music recommendation systems, automatic transcription, automatic composition, classification or recognition of music genres, artists, instrumentation etc. The goal of MIR is to describe music with features, for means of better understanding, describing, and categorizing music. There are numerous features of music that have been reported to be eliciting emotions. These features include tempo, mode, harmony, tonality, pitch, micro-intonation, contour, interval, rhythm, sound level, timbre, timing, articulation, accents on specific notes, tone attacks and decays, vibrato, and more. The same feature can be used in a similar manner in more than just one emotional expression (e.g., if considering a Discrete Emotional Model, fast tempo can be used to induce both anger and happiness) [68]. Therefore, "each cue is neither necessary nor sufficient, but the larger the number of cues used, the more reliable the communication" ([66], p. 430).

The features can be roughly divided into low-level features (LLF), mid-level features (MLF), and high-level features (HLF). The low-level features are characterized by being objective and directly computable, usually by extracting from the signal or the spectrum. Examples of LLFs are frequency, sound pressure level, spectral centroid, and more. The mid-level features introduce some semantics and musicological notions, such as pitch, chord, timbre, beat, or tempo. Lastly, the high-level features are the least objective and can be very ambiguous as they often refer to human's interpretation of music. They are also the most meaningful and intuitive for describing. Some examples of HLF are mood, emotion, style, and other [82]. Therefore, the semantic richness in features is followed by the decrease of objectiveness.

2.2.2. Tension in music

Tension experience is the basis for music emotion [131], but it is also subjective and challenging to describe precisely. In his seminal work, the musicologist Leonard Meyer [97] wrote that understanding and enjoyment of music depend upon the perception of and response to attributes such as tension and repose, instability and stability, and ambiguity and clarity. He proposed that expectations have the main role in experiencing emotions while listening to music. Namely, some parts of a music piece can create strong expectations for continuation, giving the sense of tension and instability, while other points in a music piece would fulfill expectations and be perceived as "closed" and completed.

Therefore, the term tension in music represents a sense of unrest, instability, excitement, or anticipation of a listener. It can also be described as balance between expectation of something familiar and surprise or curiosity about what is next. The "exchange" of building and releasing the tension is what makes the music "move forward" and unfold in time. Meyer [97] discussed that arousal through interruption of expectations has little value; to have any aesthetic meaning, the arousal or tension must be followed by a satisfying resolution of the tension.

Music can convey tension and release through changes in various musical layers: harmony, melody, rhythm, timbre, structure, dynamics, and more. In general, tension can be brought by less expected elements, changes or "violated rules" in these layers. For instance, in every scale there is a hierarchy of importance among the scale tones: some are more stable, significant of "final sounding" than others, which makes the listeners experience varying amounts of tension and resolution [82]. An example of "musical expectancy formation" is the bigram table of chord transition probabilities extracted from a corpus of Bach chorales [118], showing, for instance, that after a dominant seventh chord, the most likely chord to follow is the tonic, and that a supertonic is nine times more likely to follow a tonic than a tonic following a supertonic [134].

The tension in music largely depends on tonal features and their relative importance in the musical piece but can be enhanced by expressive features (such as dynamics) [81]. Tension in the harmonic layer can be expressed with a complex chord or certain harmonic role of a chord, such as dominant or cadential 6-4. Examples of tension in melodic layer could be a large leap in the melody, a change of a melodic range, certain (often dissonant) intervals or particular notes in a scale. The rhythmic tension can be conveyed through complexity of rhythm or effects such as anticipation of a note, repetition, accentuation, syncopation, acceleration, off-beat notes. The growing tension in the layer of music dynamics is often expressed with crescendo[1], but also a sudden change of dynamics - even silence used in expressive way - can contribute to a surprise and therefore the tension [33].

Nevertheless, the perceived tension largely depends on the musical background of the listener, because their expectations are built with their education. Listeners can experience the same music in drastically different ways due to differences in their background, culture, and preferences. From a listener's point of view, the amount of tension can be correlated with the amount of attention they need to put into listening. In fact, according to Berlyne's theory [12], listeners tend to prefer music that gives them an optimum level of arousal (e.g., perceived complexity). If the arousal potential is too high or too low, listeners will reject the music. Berlyne modelled this hypothesis on relations between listeners' preferences and arousal in the form of an inverted U-shaped curve [68], represented in Figure 2.4.

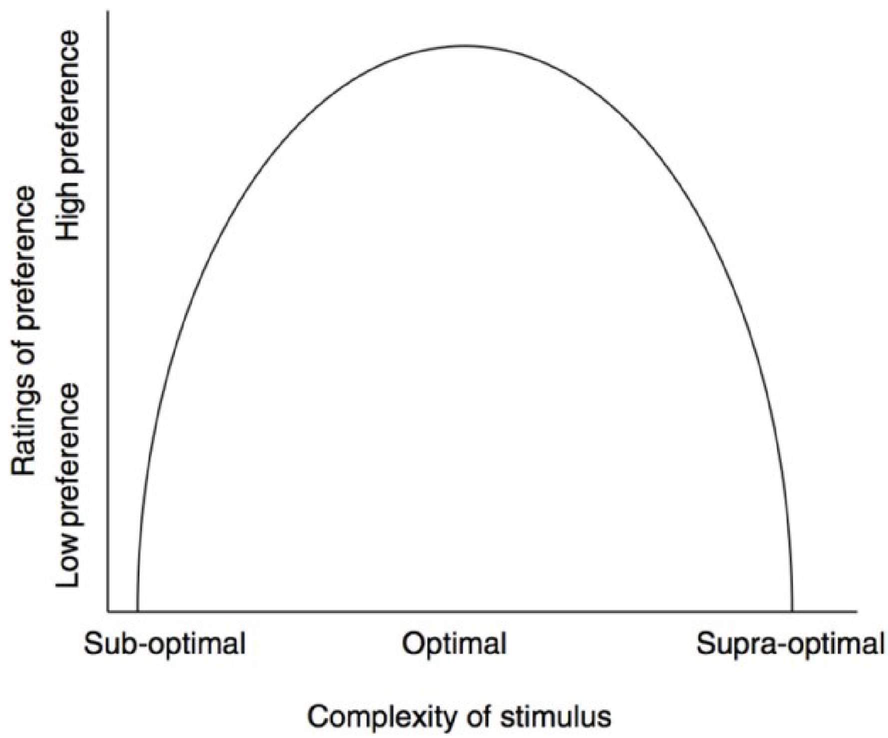

Figure 2.4: The inverted-U relationship between complexity and preference, as theorized by Berlyne. Taken from [30].

[1]a gradual increase in loudness or intensity

2.3. Physiological signals and sensors

As mentioned in Subsection 2.1.1, affective computing is a multidisciplinary study that revolves around computer science, psychology, cognition, and physiology, aiming to integrate human affects and emotions with artificial intelligent systems [130]. Currently, there are a lot of ways to build an such systems using various techniques and algorithms. However, it is considered that the physiological approach towards emotion recognition has become a better alternative to facial expressions, gestures, and vocal traits [54]. The summary of the signals used in physiological approach follows.

2.3.1. Electroencephalogram (EEG)

The electroencephalography (EEG) is a method to record electrical activity of the human brain using electrodes placed on their scalp. The mentioned electrical activity generated by the brain underneath reflects how many neurons in the brain network communicate with each other via electrical impulses and in what way [132]. In particular, the electrical function of the brain is measured with a electroencephalogram (EEG), which represents a difference in potentials from two points of brain function recorded in time domain [132]. The electroencephalography is mainly a non-invasive method that involves placing electrodes along the scalp, called *extracranial EEG*. Since the voltage is recorded from the scalp, the EEG signals are rather weak and affected by various biological and external artifacts [123]. Therefore, optimizing the quality of recorded EEG data is crucial for the analysis [39]. Another type of EEG, *intracranial EEG*, provides an EEG recording directly from the brain, using surgically implanted, therefore invasive electrodes. Consequently, the intracranial EEG can provide more detailed and targeted information on brain activity but is less often used due to the invasive setup [132]. 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 [71].

The EEG was discovered in the 19th century: a British physician Richard Caton is credited with giving the first contribution to the field by presenting his findings about electrical phenomena of the cerebral hemispheres of animals. Back in 1875, Caton observed variations in currents measured by a galvanometer, from unipolar electrodes placed on the surface of both brain hemispheres of animals. The first EEG recordings were done on paper [51]. However, the first EEGs from humans were recorded in 1924 by Hans Berger, a German professor of neurology and psychiatry who used to investigate electrical activity of the brain. Berger was the one who came up with the term "electroencephalogram" and defined the wave patterns, including *alpha* and *beta* waves [51].

The development of electroencephalography contributed greatly to the knowledge about the brain and its functions, as well as to exploration of various changes in psychological and physiological behavior [71]. Nowadays, EEG is mainly used for clinical purposes, which includes observing and interpreting abnormal EEG signal patterns, in order to support a clinical diagnosis of epilepsy or other conditions such as sleep disorders, coma, Alzheimer's disease, and more [124]. Additionally, the EEG used to be the main method of diagnosis for tumours, stroke, and other focal brain disorders, but with development of high-resolution anatomical imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT), the use of EEG in those fields has decreased [124]. Other than for clinical purposes, EEG is also used in the fields of neuroscience, cognitive psychology, psychiatry, neurolinguistics, and psycho-physiological research [39, 124]. Applications of EEG are mainly focused either on event-related analysis or on its spectral content. The event-related analysis explores potential fluctuations in time interval around the stimulus onset, while the spectral analysis focuses on neural oscillations (popularly called "brain waves") that can be observed in the frequency domain of an EEG signal.

EEG: biological background

A human brain is a complex organ that is constantly active, absorbs information, re-wires existing data, and integrates everything into a consistent experience of oneself and of the surrounding. It is the main organ of the human central nervous system (CNS) and it can be divided into brainstem, limbic system (also known as "the emotional brain"), the cerebellum (also known as "little brain") and the cerebrum (or cortex). The cerebral cortex is further divided into four basic sections or lobes: Occipital, Temporal, Parietal, and Frontal. The Occipital lobe is located in the rearmost portion of the skull and is responsible for the visual processing, spatial orientation, and motion perception. The Temporal lobe is associated with processing sensory input using visual memories, language, and emotional association, as well as the comprehension of written and spoken language. The Parietal lobe is responsible for integrating information coming from external sources as well as internal sensory feedback from skeletal muscles, head, eyes, etc. From these sources, Parietal cortex creates a coherent representation of how the body relates to the environment and vice-versa. Lastly, the Frontal lobe is where most of conscious thoughts and decisions are made, and is associated with attention, short-term memory, planning, and motivation.

The previously described activity of the brain refers to neurons in brain areas generating diverse "firing" patterns and thus creating postsynaptic potential. The postsynaptic potential of a single neuron is obviously too small to be detected; however, if it happens in synchrony for hundreds of thousands of similarly oriented neurons, they generate an electric field that is rapidly propagated through the brain tissue and the skull. Within milliseconds, the generated electric field can be measured from the scalp. Nevertheless, since the scalp is relatively far from the source of the postsynaptic potentials, it is not possible to precisely locate the spatial portion of the brain that created the potentials (since the measured potential on the scalp is the average response of several spatial areas activated by a particular stimulus [71]). Additionally, since the measured electrical signals are of very low amplitude, the recorded data needs to be sent to an amplifier.

EEG: Recording protocol

For standardization reasons, electrode locations and names are specified by the international 10-20 electrode system, established by the General Assembly of the International Federation in Paris in 1949 [60]. The locations of the electrodes are chosen based on standard landmarks of the skull: the nasion, inion, and the left and right preauricular points. The anterior-posterior plane from the nasion to the inion is divided into 5 separate areas: Fronto Polar, Frontal, Central, Parietal, and Occipital areas. Therefore, the names of the electrodes placed in those areas start with the marks *Fp*, *F*, *C*, *P*, and *O*, respectively [60]. On the left and right sides, there are also Temporal areas, labeled by the mark *T*. The described landmarks and brain areas are represented in Figure 2.5.

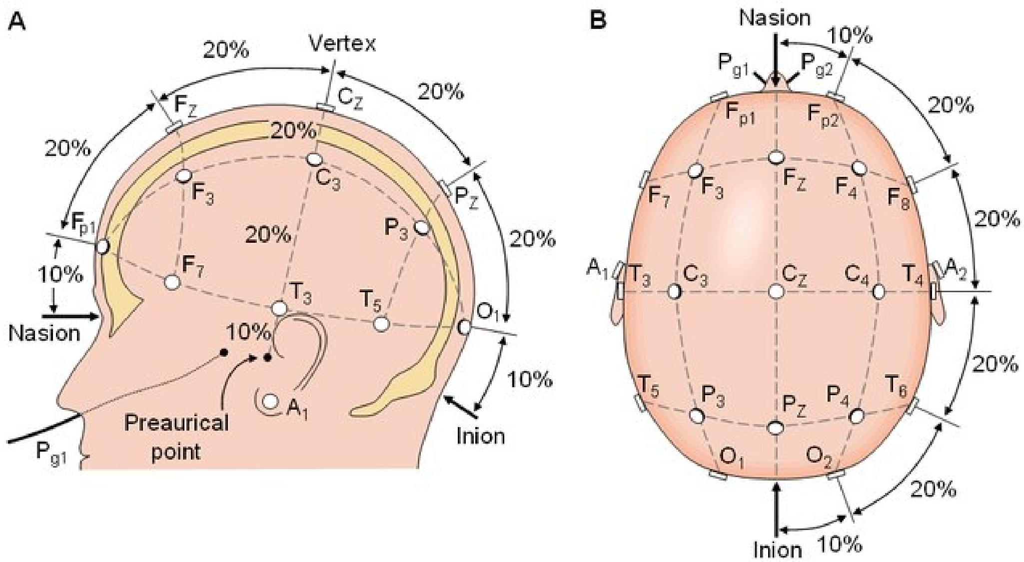

Figure 2.5: EEG electrodes names and locations from two sides: 10-20 setup, [4]

To distinguish between left and right homologous regions, a numbering system of the electrodes was also established. The electrodes placed on the right hemisphere are labeled with even numbers (e.g., F4, O2), while those placed on the left hemisphere are labeled with odd numbers (e.g., C3, T4). Thus, each electrode placement site has a letter that identifies the area of the brain, and a number that identifies the exact position. The electrodes on the on the central nasion-inion line are labeled with "z" (zero) (e.g., Fpz, Cz, Oz) and are often used as "ground" or "reference" points. As for the use of additional electrodes, the American Electroencephalographic Society adopted another guideline, described in details in the latest IFCN standards [102] and represented in Figure 2.6.

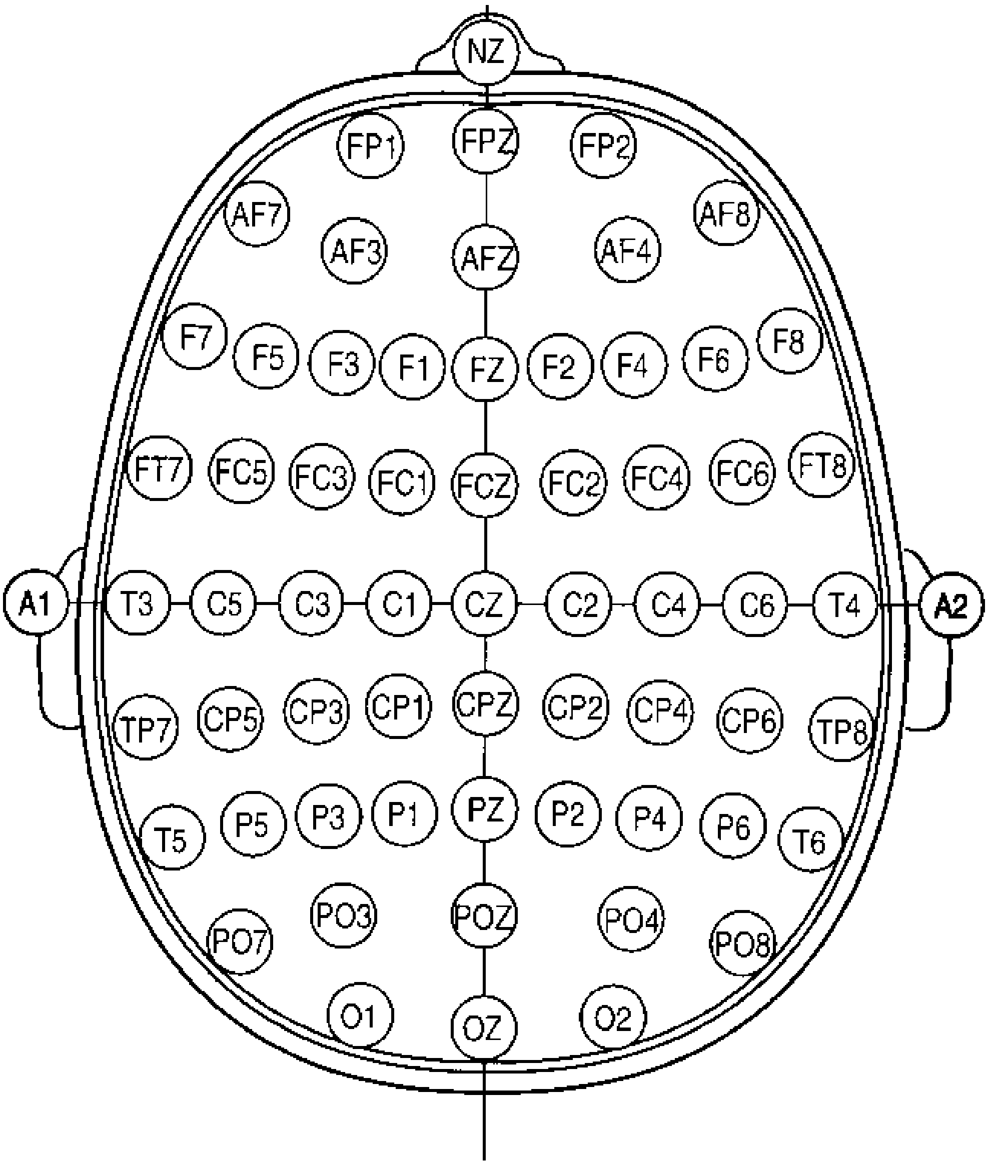

Figure 2.6: EEG electrodes names and locations according to the 10-20 setup, [102]

During an EEG recording procedure, electrodes are placed on the subject's scalp to record the electrical activity. If the EEG system includes a large number of electrodes, they are usually embedded into an EEG cap or a net for easier use. To increase the conductivity between the electrodes and the subject's scalp, a water-soluble electrolytic gel or paste is applied into a hole in each electrode - usually with a syringe. Each electrode is attached to an individual wire and connected to one of the two inputs of a differential amplifier. The subjects are advised to arrive with clean and dry hair, to remove any ponytails, braids, hair clips etc, and to not put sprays, oils, or creams in their hair beforehand [39].

The advantage of such EEG system with wet electrodes is that it has a better signal-to-noise ratio than a system with dry electrodes, hence it is easier to design and less expensive. On the other hand, the disadvantages are that they cause a sense of discomfort to the subject, and that the setting up procedure is more complicated and time consuming. As raw EEG signals are rather weak, they are usually contaminated by various types of external artifacts such as eye blinks, muscular movement, heart rate, etc. Moreover, the recording equipment is sensitive and easily picks interference from the external sources. The artifacts coming from eye blinks or eye movement are typically characterized by low-frequency signal (< 4Hz) with high amplitude, and therefore they represent one of the main problems in EEG analysis [123]; thus, they need to be identified and removed from the raw EEG signal. The main methods used to remove artifacts are Independent Component Analysis (ICA) [63, 90], Principal Component Analysis (PCA), regression based methods, high-pass filtering (e.g., Butterworth, above 2Hz) and adaptive filtering [55]. Other than artifacts coming from eye activity, any type of muscular movement or muscular tension (e.g., clenching the teeth) adds high-amplitude noise to the signal.

EEG: Features

The relevant features in electroencephalogram can be event-related (therefore, in time domain) or focused on the spectral content (therefore, in frequency domain). The event-related analysis refers to a certain stimulus and focuses on the time interval around the stimulus onset. In particular, event-related potential (ERP) is a measured brain response that is a direct result of a specific external stimulus. Those stimuli need to be clearly defined in time and to follow a resting state (baseline) in order to be observable. Since the EEG reflects thousands of simultaneously ongoing brain processes, the response to a single stimulus or event of interest is not usually visible in the EEG recording of a single trial. Thus, to see the brain's response to a stimulus, it is important that the same stimuli are repeated many times in the experiment, so that averaging across many trials reduces the randomness in the brain activity and leaves the relevant waveform of the ERP.

The spectral analysis works with neural oscillations or "brain waves" that can be observed in the frequency content of an EEG signal. Human EEG power spectrum is divided into five frequency bands: delta (0-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-50 Hz). Each frequency band is considered to be related to specific functions. The frequency bands are represented in Figure 2.7.

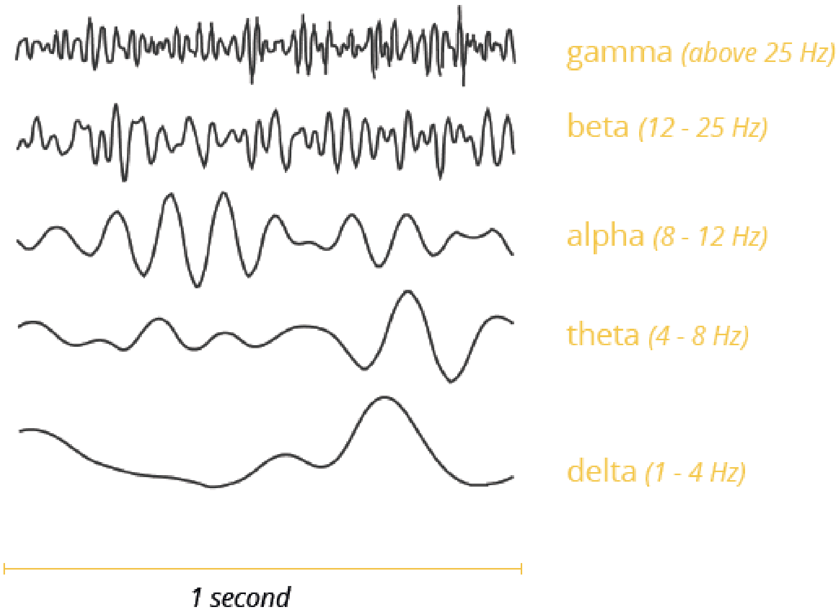

Figure 2.7: EEG bands, taken from [104]

Spectral analysis is one of the most used methods to analyze EEG data, which implies the decomposition of the EEG into its frequency components. Starting from the Fourier transform of the signal (which gives the amplitude and phase values of the signal at a specific frequency), it is then common to compute the squared magnitude and thus obtain an estimate of the power spectral density, expressed in squared microvolts per Hertz ($\mu V^2/Hz$). One of the features that could be obtained from power spectral density is the average band power for any of the bands represented in Figure 2.7, which describes with a single number the contribution of a given frequency band to the overall power of the signal in a certain time interval.

2.3.2. Electrocardiogram (ECG)

Electrocardiography is the process of recording of the human heart's electrical activity through sensors attached to the skin. The electrocardiogram (also called ECG or EKG) produces a graph of voltage across time of the electrical activations that lead to the contraction of the heart muscles. The ECG recording protocol involves attaching a small number of electrodes on a subject's body (usually on the chest near the heart) and connecting them by wires to a data acquisition machine [54, 85]. These electrodes measure the electrical activation of the heart during each cardiac cycle. A common source of noise in ECG signals are motion artefacts resulting from sensor displacement due to participant movement [135]. Traditionally, "ECG" usually refers to a 12-lead ECG that is acquired while the subject is lying down, however, many other devices can record ECG, such as various portable devices or even some models of smartwatches. While it is widely used in medical and clinical fields to investigate and monitor symptoms of possible health problems, ECG is being used more and more by researchers exploring physiological arousal (often combined with other biosensor methods). Common features of the ECG signal are heart rate (HR), inter-beat interval (IBI) and heart rate variability (HRV), while it is considered that HR reflects emotional activity [73].

ECG: biological background

The human nervous system is a very complex structure and can be divided into two major regions: the central nervous system (CNS) and the peripheral nervous systems (PNS). The CNS includes the brain and spinal cord, while the PNS is named so because it is on the periphery and it consists of ganglia and nerves outside the brain and spinal cord [14]. The representation of this structure can be seen in Figure 2.8. The PNS is then divided into the somatic nervous system and the autonomic nervous system.

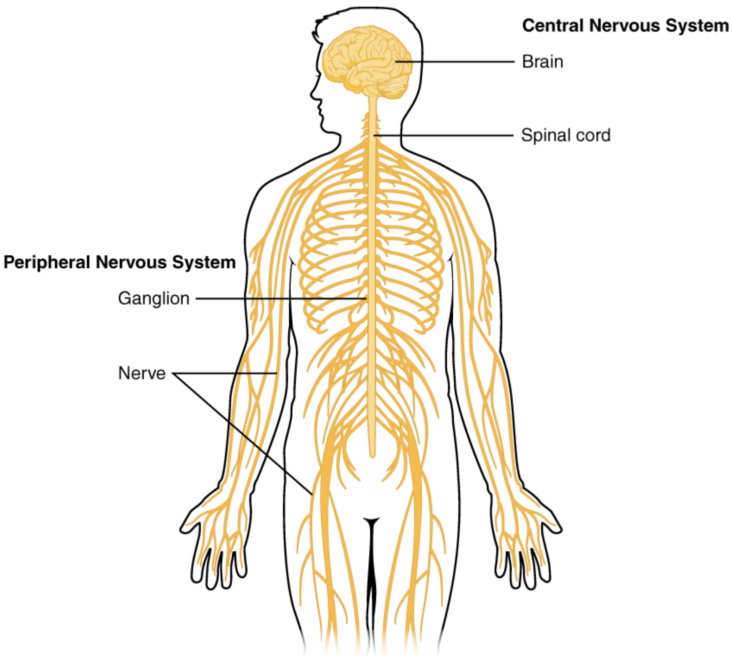

Figure 2.8: Central and Peripheral Nervous System [14]

Both the brain and heart are connected via the autonomic nervous system (ANS) and their communication affect one's perception, emotion, intuition, and general health. The ANS regulates many bodily functions, such as the heart rate, respiratory rate, digestion, pupillary response, and more. The ANS is subdivided into two distinct components: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The more detailed structure of the nervous system is represented in Figure 2.9.

The SNS is often considered the "fight or flight" or "quick response" system that dominates during stressful states, while the PNS is referred to as "rest and digest" or "relaxed response" system that dominates in relaxing states. Therefore, these two systems often have "opposite" actions of activating and inhibiting a physiological response. The oscillations of a healthy heart are complex and constantly changing, while the SNS and the PNS provide complementary stimulation for the cardiovascular system to adjust to sudden changes and maintain the regulatory balance or homeostasis. Increased sympathetic drive is associated with increases in heart rate, blood pressure and sweating, which can be referred to as "autonomic arousal". Upon excitation, the cardioaccelerator releases the neurotransmitter norepinephrine and causes the increase of the heart rate, which occurs throughout the SNS. On the other hand, the PNS is responsible for the decrease in the heart rate, which happens when the cardioinhibitory centers release the neurotransmitter acetylcholine [54].

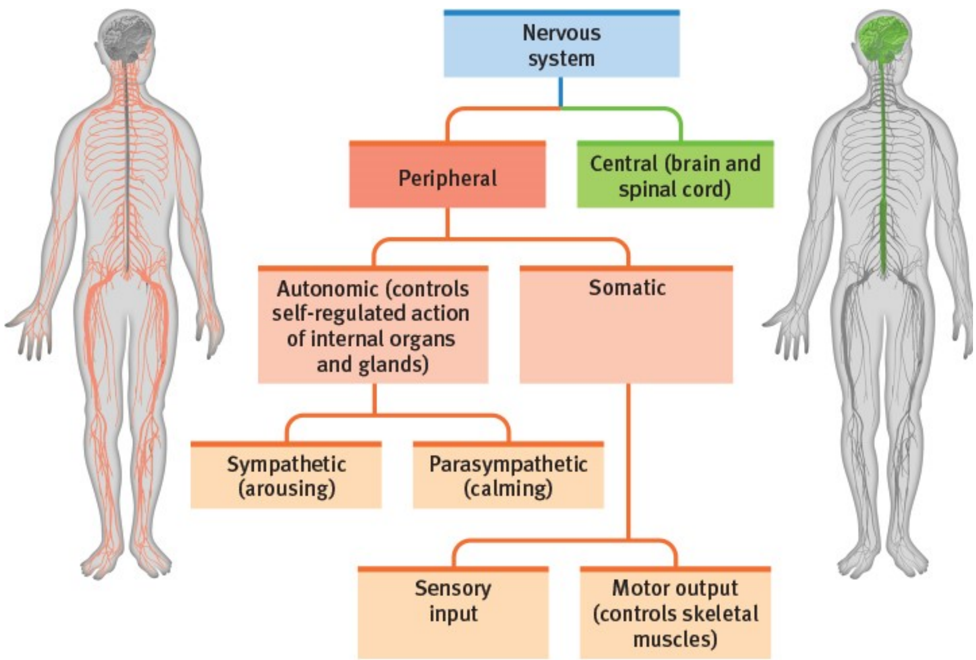

Figure 2.9: General Nervous System structure [32]

ECG: Features

A normal ECG signal should have three segmented waves in a single cycle, which is represented in Figure 2.10. The cycle consists of the P wave, which represents depolarization of the atria; the QRS-complex, which represents depolarization of the ventricles; and the T wave, which represents repolarization of the ventricles [83, 85]. The most notable feature of the ECG is the QRS-complex, which represents the electrical activation that leads to the ventricles contracting and expelling blood from the heart muscle. The R-peak is strongly present in the signal and is used for extracting heart beats. The ECG generally provides a good signal/noise ratio, and the R-peak that is of interest generally has a large amplitude compared to the surrounding data points and a small amplitude variation through time. Starting from detected R-peaks, some of the common ECG features used in research are heart rate (HR) and heart rate variability (HRV).

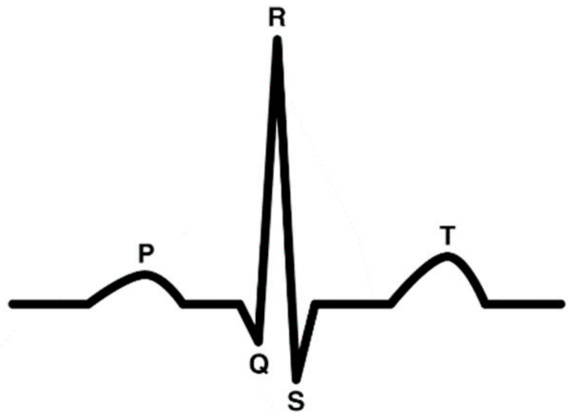

Figure 2.10: Normal ECG waveform, where P-wave, QRS-complex and T-wave represent the contraction/depolarization of atria, contraction/depolarization of ventricles and repolarization of ventricles, respectively [57].

Heart rate (HR) is measured in beats per minute, and represents the average number of heart beats in a given time period. Common measures expressing the HR found in the literature are the beats per minute (BPM) and the mean inter-beat interval (IBI). For extraction of the instantaneous heart rate (BPM), the peak detection precision is not crucial, due to the BPM being an aggregate measure, calculated as the average beat-beat interval across the entire analyzed signal or a longer segment.

Heart rate variability (HRV) is a measure that represents the variation of the heart rate. It refers to how much an individual's heart rate signal changes through time and is measured by the variation in the beat-to-beat interval. Overall, HRV can be considered an indicator of physiological stress or arousal, with increased arousal associated with a low HRV, and a decreased arousal associated with high HRV [96, 100]. Research suggests that a person with higher HRV while resting can provide more appropriate emotional responses compared to those that have low HRV at rest. It is important to note that, even though it is largely used in research of emotional states, the HRV cannot be a moment-by-moment indicator of emotion. Namely, the data collected over a time interval cannot accurately relay information on response to a stimulus, but can rather help understanding the total reaction to a signal in the taken time interval.

Contrary to HR, when extracting HRV measures, the R-peak positions are crucial, since the measures are designed to capture the slight natural variation between peak-peak intervals in the heart rate signal. Thus, even a few incorrectly detected peaks can introduce large measurement errors [135]. The measures that express HRV can be divided into time-domain and frequency-domain ones. Time-domain methods are based on beat-to-beat or R-R intervals. Frequency domain methods assign bands of frequency and then count the number of R-R intervals that match each band.

The measures in time-domain that are commonly used for HRV are:

- MAD (Median absolute deviation of RR intervals)
- SDRR (Standard deviation of RR intervals)
- SDSD (Standard deviation of successive difference between heart beats intervals)
- RMSSD (Root mean square of successive RR interval differences),

out of which SDSD and RMSSD are the most often used [135]. RMSSD is computed as presented in Equation 2.1.

$$RMSSD = \sqrt{\frac{1}{n-2}\sum_{i=0}^{n-2}(RR_i - RR_{i+1})^2}, \tag{2.1}$$

When HRV is expressed in frequency domain, the following measures are usually considered:

- Power of the low frequency band (LF, 0.04–0.15 Hz), which is related to short-term blood pressure variation [13],
- Power of the high frequency band (HF, 0.16–0.5Hz), which is a reflection of breathing rate [100],
- The ratio between power in HF and LF bands.

These measures are computed from the PSD (Power Spectral Density), which is most commonly estimated using the Welch's method [142]. High-frequency (HF) activity has been linked to PSNS and found to decrease under conditions of acute time pressure and emotional strain, possibly related to focused attention and motor inhibition. While relatively short recordings of 5 min or less can effectively capture high-frequency HRV, long-term recordings of at least 24 hours are needed to reliably assess lower-frequency components [110].

2.3.3. Electrodermal activity (EDA)

Electrodermal activity (EDA) refers to changes in electrical conductance measured at the skin surface, usually corresponding to changes elicited by the ANS [72]. The skin conductance can be quantified by applying an electrical potential between two points of skin contact and measuring the resulting current flow between them [17]. For the majority of people, when they experience emotional activation, increased cognitive workload or physical exertion, their brain sends signals to the cells in skin to increase the level of sweating. Even though the increase of sweat on the skin surface may not be noticeable, electrical conductance can increase significantly as the pores of the skin begin to fill [45]. A common factor that elicits EDA responses is subjective salience, a concept closely related to motivational importance. EDA is also a useful indicator of attention, and it is widely recognized that attention-grabbing stimuli and attention-demanding tasks evoke increased EDA responses [24]. EDA can also be referred to as skin conductance (SC) or galvanic skin response (GSR).

The skin conductance can be further divided into two components: the phasic skin conductance (PSC) and the tonic skin conductance (TSC). The tonic part corresponds to slow shifts and therefore varies slowly over time. It is considered to be the raw level of skin conductance when there are no external stimuli, thus also called Skin Conductance Level (SCL). On the contrary, the phasic part measurements are usually associated with discrete external stimuli, therefore with rapid changes or transient events. The changes in the phasic component are shown as increases or peaks in skin conductance and are generally referred to as Skin Conductance Responses (SCR) or Galvanic Skin Responses (GSR). Although associated to fast external events, a phasic response can occur 1 to 5 seconds after the stimulus [16]. An example of a typical SCR following a stimulus is represented in Figure 2.11, where the reader can notice a latency before the onset of the response, as well as before reaching the peak amplitude.

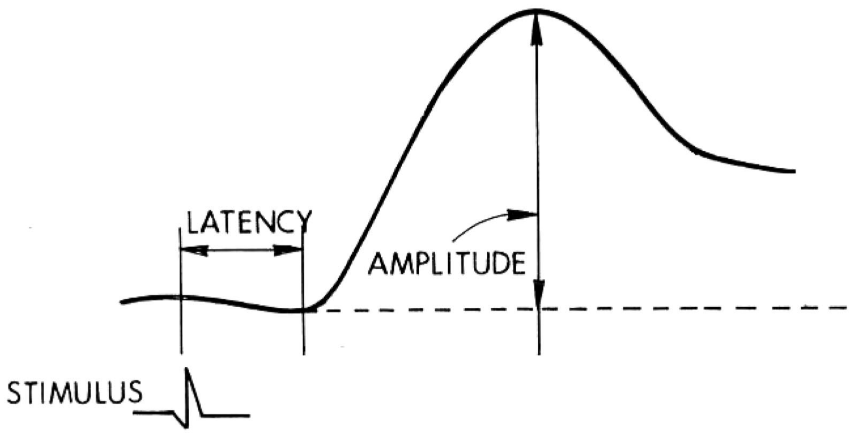

Figure 2.11: An example of a typical SCR, following a stimulus onset. Taken from [29]

The phasic component is therefore sensitive to specific arousing stimulus events, and those SCRs are called event-related SCRs or ER-SCRs. The event-related analysis refers to the physiological changes that occur in response to a certain event and is usually done by time-segmenting the signal in epochs. An epoch is a short chunk of the signal (usually below 10 seconds) around the stimulus. The event-related features can be rate changes, peak characteristics, and phase characteristics. On the other hand, is no specific events are considered in the signal, such SCRs are called non-specific SCRs, and the analysis is based on longer time intervals. The described systematic division is represented in Figure 2.12.

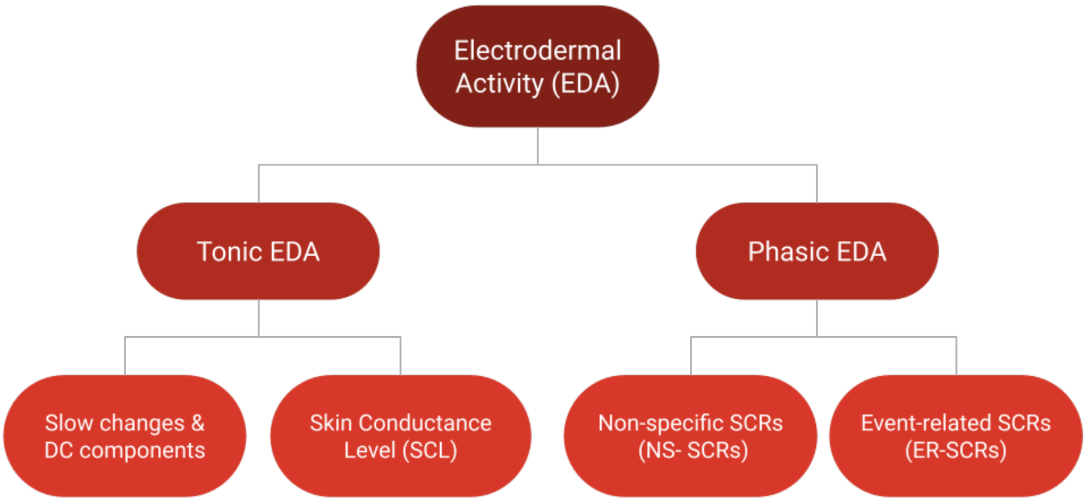

Figure 2.12: EDA division [17]

EDA: biological background

The skin is the largest organ of a human body and the main interface between an organism and the environment. Since it is very complex in its functions, it is also densely innervated. When sweat glands are innervated, a change in skin conductance can be measured at the surface. Similar as for the cardiovascular arousal (increase in heart rate and blood pressure), the increased sympathetic drive causes increase in sweating [24]. The activity of sweat glands is triggered by sudomotor fibers, each of which innervates a skin area of about 1.28cm$\hat{2}$. When multiple sudomotor fibers fire at the same time, it is identified as a "nerve burst" in the integrated nerve record. A sudomotor nerve burst corresponds to an observable skin conductance response (SCR), visible as modulation of the conductance of an applied current. Therefore, the SCR amplitude is considered as an index of sympathetic activity [11].

EDA: Recording protocol

The EDA sensor measures electrical conductance (the inverse of resistance) of the skin, by passing a small amount of direct current between two electrodes that are in contact with the skin. The signal is usually measured in microsiemens (μS). According to [17], the amplitudes of phasic SCRs can range from the threshold (0.01–0.05 μS) to a maximum value of around 2-3 μS. In experiments with highly aversive and fearful stimuli, the maximum value can increase to around 8 μS, although rarely. As for SCLs, [138] mention a range of 1-40 μS, but averages are more often between 2-16 μS. Based on developed recommendations [45], the EDA signal acquisition should be done when subjects are seated and asked not to move. Since the sweat glands are at highest density in palmar and plantar regions, it is considered better to record EDA from the palmar surface or fingers of the non-dominant hand. If the EDA is recorded from the wrist, the phasic signals could be with lower SCR magnitude [24].

A common procedure is to acquire EDA in conditions that can then be subtracted from each other, in such a way that the resultant measure is a relative difference within the individual, and not based on the absolute values. It is recommended to record the baseline signal representing the resting state, that is several minutes long, in order to give enough time for the subject to relax and the skin conductivity to decrease. The subtraction procedure serves as a form of normalization for the individual's EDA data [45].

It is also important to distinguish between normalization and standardization. While normalization refers to data transformations in order to correct them and make them fit for parametric statistical analysis, it does not help with inter-subjects comparisons.

On the other hand, standardization refers to correcting the data so that they can be directly and meaningfully compared between subjects. Therefore, standardization facilitates individual-difference comparisons where all SCRs have been transformed, relative to their physiological responsiveness, and thus to factor out issues like thickness of skin, etc. There are several proposed methods for standardization, while there is no universally agreed one, considering there are pros and cons to every procedure. The examples of some EDA standardization methods are presented below:

1. **SCL data: Range-Corrected Scores**
   This method computes the minimum SCL during a baseline or rest period and a maximum SCL during the most arousing period. The resulting SCL for a subject can then be computed as a proportion of their individual maximum range, using the formula

$$\frac{SCL - SCL_{\min}}{SCL - SCL_{\max}} \tag{2.2}$$

2. **SCR data: Proportion of Maximal Response**
   The SCR data can be assumed to have a minimum equal to zero and a maximum corresponding to the result of a startle stimulus, such as surprising audio stimuli, hand clap, balloon pop, etc. In such a way, each individual SCR can then be standardized by dividing it by the subject's maximum SCR.

3. **Transformations into standard values**
   Some studies recommend transforming SCRs into Z-scores [10], which requires the mean and standard deviation to be used instead of a hypothetical maximum. In particular, each raw SCR, a mean SCR value and standard deviation of SCRs, are used to compute the Z-score which is normally distributed, has an average of 0 and a standard deviation of 1 [17]

.

2.3.4. Respiratory activity (RSP)

The respiratory activity 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, represented in BPM (breaths per minute). It is typically measured by counting how many times the chest rises, through a sensor attached to a human body. While heart rate and pulse rate are involuntary parameters, respiratory rate can be both voluntary and involuntary. The differentiation between voluntary and automatic (metabolic) breathing is that automatic breathing requires no attention to maintain, whereas voluntary breathing involves a given amount of focus (e.g. meditation) [129]. A healthy adult in a resting state can have a respiratory rate between 12-15 BPM: the average length of an inhalation in a quiet state is around two seconds, while an exhalation is around three seconds. This respiratory rhythm can change with some medical conditions [1]. The respiratory rate can be affected by excitatory nervous system arousal; it has also been reported that factors such as stress, crying, sleeping, agitation and age have a significant influence on the respiratory rate.

There are various methods to monitor and measure the respiratory rate; some of the novel techniques allow its estimation from the electrocardiogram, photoplethysmogram, or accelerometry signals [7, 22, 79], and thus can be mounted in wearable sensors. Additionally, a video recording of chest and abdomen movement can be recorded to analyze the breathing activity. The respiratory activity is widely used to monitor the physiology of acutely-ill hospital patients, by using regular measures to identify changes in physiology and other vital signs. It is also a good tool for short-term analysis of autonomic nervous system (ANS) issues.

RSP: Biological background

The respiratory system of humans consists of the respiratory tract, which is divided into an upper and a lower respiratory tract. The upper tract includes the nose, nasal cavities, sinuses, pharynx and the part of the larynx above the vocal folds, while the lower tract includes the lower part of the larynx, the trachea, bronchi, bronchioles and the alveoli. The primary purpose of the respiratory system is the equalization of the partial pressures of the respiratory gases in the alveolar air with those in the pulmonary capillary blood, by simple diffusion across a thin membrane which forms the walls of the pulmonary alveoli.

The inhalation at rest mainly happens with the contraction of the diaphragm. When the diapraghm contracts, the sheet of muscle that separates the thoracic cavity from the abdominal cavity flattens (moves downwards), pushing the abdominal organs downwards. That movement increases the volume of the thoracic cavity. The automatic rhythmical breathing in and out, can be interrupted by forms of very forceful exhalation (coughing, sneezing), by the expression of a wide range of emotions (laughing, sighing, crying out in pain, exasperated intakes of breath) and by such voluntary acts as speech, singing, whistling and playing wind instruments. All of these actions rely on the muscles within the respiratory system, and their effects on the movement of air in and out of the lungs.

RSP: Features

The features used in RSP analysis can be calculated from the respiration rate and the breath-to-breath interval (BBI). The BBI is calculated by noting a time stamp at the peak of each breath pulse and measuring the time interval between these stamps. The BBI is used as input for calculation of respiratory rate variability features. The respiratory rate variability (RRV) or breath rate variability (BRV) represents variations in respiratory rhythm. The RRV features can be grouped into time-domain features, frequency-domain features, and nonlinear parameter calculation of RRV.

1. **Time domain measures of RRV**

   Some of the statistical measures relying on time-domain signal are standard deviation of the BBI (SDBB), root mean square value of successive differences of the BBI (RMSSD), and standard deviation of SDBB (SDSD).

   Root mean square of successive differences shows the square root of the mean of the square of the successive differences between adjacent BBs.

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(BB_{\mathrm{j+1}} - BB_{\mathrm{j+1}})^2}, \tag{2.3}$$

   Standard deviation of B-B intervals (SDBB) is a measure of changes in breath rate due to cycle longer than 5 min.

$$SDABB = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(meanIBI_i - \overline{meanIBI})}, \tag{2.4}$$

   where M is a total number of segments.

   Standard deviation of SDBB (SDSD) is a standard deviation of successive differences between adjacent BBs.

2. **Frequency-domain measures of RRV**

   Frequency-domain methods show the number of BB counts in each frequency band. Bands are defined as high frequency (HF) band (0.15–0.4 Hz), low frequency (LF) band (0.04–0.15 Hz), and very low frequency (VLF) band ($<$0.04 Hz). The most common feature is the power distribution across each of these bands, calculated using algorithms based on FFT or Welch estimation [142]. The power distribution across LF and VLF bands reflects the sympathetic modulation and parasympathetic tone, while the power in HF band is a pure measure of parasympathetic activity. Another common features is the ratio between LF and HF bands (rLFHF).

3. **Nonlinear analysis**
Approximate entropy (ApEn) is a measure that quantifies the unpredictability in a signal, by estimating the likelihood that similar observation will not be followed by additional similar conditions. Lower ApEn values are related to more regular signal, while larger ApEn values are related to more irregular and thus less predictable signals. ApEn is defined as:

$$ApEn(m, r, L) = \phi^m(r) - \phi^{m+1}(r), \tag{2.5}$$

where r is the tolerance, m is the embedded dimension, L is the signal length, and $\phi^m(r)$ is the normalized counting number of different vectors, calculated as:

$$\phi^m(r) = \frac{1}{L-m+1} \sum_{i=1}^{L-m+1} \frac{N^m(i)}{L-m+1}, \tag{2.6}$$

Sample entropy (SampEn) is a modification of the ApEn that attempts to assess the complexity of physiological signals. The unconditional probability of randomly selecting two sequences of length m from a signal that have a distance less than r using the relative frequency methods is $C_i^m = \frac{n_i^m}{N-m}$. The averaged probability is then given by:

$$\phi^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} C_i^m, \tag{2.7}$$

2.4. Statistical and spectral measures

This section describes the background for relevant statistical and spectral measures used in this work.

2.4.1. Power spectral density (PSD)

Power spectral density (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. Using the PSD measure, frequency components having stronger or weaker power levels in the given frequency range can be identified. The PSD applies to signals existing over the whole signal length, or over a large enough time period, in which case it refers to the spectral energy distribution per unit of time. Windowed power spectrum breaks down the data into consecutive (overlapping or not) segments or windows; the entire PSD is then computed for each segment.

The PSD estimation techniques can be using parametric or non-parametric methods, and can be based on time-domain or frequency-domain analysis. The spectral density is usually estimated using Fourier transform methods (such as the Welch method [142]), but other techniques such as the maximum entropy method can also be used. The Welch's period gram [142] is commonly used to estimate the PSD and it is based on discrete Fourier transform (DFT). The N-point DFT of a random variable X (n) is given by:

$$DFT_x(f) = \sum_{n=0}^{N-1} X(n)e^{i2\pi fn}, \tag{2.8}$$

Incorporating a weighted windowing function $w(n)$ to the input series, the data near the edges of the time series are given less weight compared to the data near the center. In such a way modified period gram is given by:

$$P_M(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} X(n)w(n)e^{-i2\pi fn}\right|^2, \tag{2.9}$$

where $U = \frac{1}{M}\sum_{n=0}^{M-1} w(n)^2$. Finally, the averaged PSD is computed using all segments. PSD by Welch period gram is given by:

$$P_w(f) = \frac{1}{N}\sum_{i=0}^{N-1} P_{M,i}(f), \tag{2.10}$$

where $P_{M,i}(f)$ is the i^{th} modified periodogram from the data series.

2.4.2. Correlation

In statistics, correlation is defined as any statistical relationship, whether causal or not, between two random variables [25]. Correlations are useful measures because they can indicate a predictive relationship. The most familiar measure of dependence between two quantities is Pearson's correlation coefficient, commonly called simply "the correlation coefficient". The correlation coefficient is a value that indicates the strength of the relationship between variables. It can take any value ranging from -1 to 1, interpreted as following:

- **-1: Perfect negative correlation.** The variables tend to move in opposite directions (i.e., when one variable increases, the other variable decreases).
- **0: No correlation.** The variables do not have a relationship with each other.
- **1: Perfect positive correlation.** The variables tend to move in the same direction (i.e., when one variable increases, the other variable also increases).

The correlation coefficient that indicates the strength of the relationship between two variables can be found using the following formula:

$$r_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2(y_i - \overline{y})^2}}, \tag{2.11}$$

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. A correlation matrix is a table of rows and columns that shows the extent of correlation between variables. All the numbers in the cells of a correlation matrix represent correlation coefficient values of the column and row variables. The diagonal values (always equal to 1) show that the correlation between a variable and itself is a perfect correlation. The correlation matrix is a symmetrical matrix. In terms of usage, a correlation matrix mainly serves to quantitatively summarize large data to identify correlation patterns.

3 | State of the art

This chapter provides an overview of various studies done in the field and discusses the characteristics of provided options.

3.1. State of the art on experimental protocols

Several research methods have been used to explore the relation between music and psycho-physiological responses of humans. A review of 251 studies by Eerola et al. [34] represents a valuable summary of research approaches, developed methodologies, and experimental protocols. The document includes an overview of the equipment used in the field, the number and profile of recruited subjects, the type of stimuli used to elicit emotions, the number of stimuli repetitions, the length of the experiment, and more.

3.1.1. Methodologies used to record physiological responses

There are dozen of methods that can be used to track and analyze human emotional states: some of them use self-report measures, some others use the physical signals such as facial expression [37], speech, posture and more, while others measure the internal physiological signals [54]. Considering the number of available options, selecting measures and devices to use is a challenging task. Among methodologies that record internal physiological signals, the most commonly used are:

- Electroencephalogram (EEG)
- Functional Magnetic Resonance Imaging (fMRI)
- Functional near-infrared spectroscopy (fNIRS)
- Electrocardiogram (ECG)
- Blood volume pressure (BVP)
- Electromyogram (EMG)
- Electrooculogram (EOG)

- Electrodermal activity (EDA)

  Skin conductance (SC)

  Skin temperature (ST)

- Respiratory activity (RSP)

As for the number of sensors used in a study, a division can be made to unimodal and multimodal approaches. Many studies focused on multimodal approach, by combining different physiological signals from biosensors such as ECG, EEG, EMG, EDA, or RSP. Even though the multimodal approach generally showed better performance, the clear advantage of the unimodal approach is the simpler data collection and lower data processing time [54, 106]. The used devices can be clinical, academic, or commercial; many wearable devices have lately been developed to monitor biological signals of an individual during their daily life [71]. According to [68], physiological measures should be used in connection with other measures, such as self-report.

Electroencephalogram (EEG)

As far as devices for EEG signal acquisition are concerned, they can be classified into those used in clinical or laboratory setting and those used for commercial purposes. The clinical EEG devices are usually equipped with more electrodes but are also less comfortable to wear and more complicated to mount. On the contrary, the commercial devices can often be in the form of a headband that are visually more appealing and easier to wear, but that often have a very small number of electrodes. An example of such a headband with 4 electrodes is given in Figure 3.1.

Figure 3.1: A wearable headband device for EEG acquisition, model Muse 2, https://choosemuse.com/

The IFCN standards for digital recording of clinical EEG [102] and the EEG recording protocol for cognitive and affective human neuroscience research [39] represent a valuable summary of requirements for a proper EEG recording. According to the IFCN standards, the amplification and channel-acquisition for EEG must be available for at least 24 channels, and preferably 32 channels. In their work, Fiedler et al. [43] explored the use of 256-channel EEG cap, emphasizing that such high-density electroencephalography (HD-EEG) also requires skilled staff and extensive preparation. The IFCN standards also recommend the minimum digital sampling rate for acquisition and storage of the EEG data to be 200 Hz; however, some models (e.g. eego™mylab by ANT Neuro [144]) claim to have a sampling rate up to 16 kHz. The high temporal resolution is a known advantage of EEG signals. However, being very receptive to noise and interferences, they are also characterized by low signal-to-noise ratio (SNR) and are therefore challenging for data processing and analysis. The data can vary significantly between individuals and even for the same person over time. Hence, EEG is often paired with PET scans and fMRI scans to help study the affective picture processes in the brain.

It has been shown in research that most of the EEG channels represent redundant information [4]. Namely, brain imaging and EEG studies have suggested that listeners' responses to music involve sub-cortical and cortical regions of the brain that are known from earlier research to be involved in emotional reactions [68, 120]. For that reason, the EEG data is often analysed only for specific regions of interest, by grouping electrodes according to their position on the scalp. For instance, in their study on tension based on nested structures in music [131], Sun et al. analyze the EEG data in four regions of interest: left anterior (F1, F3, F5, FC1, FC3), right anterior (F2, F4, F6, FC2, FC4), left posterior (P1, P3, P5, CP1, CP3), and right posterior (P2, P4, P6, CP2, CP4). Maity [89] extracted EEG data for only frontal electrodes (F3, F4, F7, F8, Fp1, Fp2, Fz), claiming that earlier works have shown frontal electrodes to be the most significant for studying simple auditory musical stimuli. Schmidt [120] used only 4 scalp locations (F3, F4, P3, P4) and reported greater relative left frontal EEG activity related to joy and happy musical excerpts, and greater relative right frontal EEG activity to fear and sad musical excerpts. In their study on music listening effect on Parkinson's disease patients, Maggioni et al. [88] use only 8 electrodes (Fp2, C4, T4, O2, Fp1, C3, T3, O1).

According to several mentioned studies [88, 89], the frequency bands of interest for a listening experiment are alpha and theta bands, corresponding to ranges of [4,8] Hz and [8,13] Hz, respectively. The frequency analysis is mainly based on computing the Fast Fourier Transformation (FFT) [120] or Power Spectral Density (PSD) [43, 88], usually using a moving window with 0.5-2 seconds length.

Functional near-infrared spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) is an optical brain monitoring technique which uses near-infrared spectroscopy for the purpose of functional neuroimaging [40]. fNIRS uses near-infrared light to measure brain activity in order to estimate cortical hemodynamic activity that occurs as a response to neural activity. Alongside EEG, fNIRS is one of the most common non-invasive neuroimaging techniques which can be used in portable contexts. The signal is often compared with the BOLD signal measured by fMRI.

Electrodermal activity (EDA)

An overview by Fowles [45] summarizes recommendations related to using EDA in experimental protocols. According to Braithwaite [17], the sample rate of EDA can be set quite low for long-term ambulatory measurements or experiments that do not require a high level of temporal precision. However, if an experimenter wants to run an event-related analysis, then the accuracy of 1 millisecond is required. It is thus recommended to set the acquisition rate to a minimum of 2000 samples per second (2KHz). Generally, 500Hz - 2000Hz sample rates are sufficient for laboratory studies. For instance, the Empatica E4 [78] wristbands that were used in the initial project described in Section 1.3 have a sampling rate of 4Hz, which is not enough for a temporal analysis, according to the described resources.

As described in Subsection 2.3.3, the EDA signal can be decomposed into its tonic and phasic components; the process of decomposition can be done in multiple ways. The most simple methods are using a median of highpass filter; however, there are several model-based methods such as [2, 48] that are shown to have a greater validity compared to heuristic and ad-hoc approaches. Several studies [3, 49] use the cvxEDA model [48], which is based on convex optimization and robust to noise coming from overlapping SCRs. According to [49], EDA features such as the maximum value, the mean or the variance seemed more significant for estimating arousal than valence, if taking into account the valence-arousal emotion model presented in Subsection 2.1.2.

Since the background tonic SCL is constantly changing for an individual and can be very different between them, some researchers made a conclusion that the actual SCL level is not that informative on its own nor that easy to derive. The protocol usually requires recording a baseline, which is considered to be the average tonic level of a subject during their rest state and in absence of any specific external stimuli. The standard practice is to record a baseline signal for 5-15 minutes of rest prior to any tasks. During the baseline period, considering that the subject is in resting state, the EDA signal level should be smooth and slowly drop in amplitude. The baseline value is computed by averaging over the lowest values of in such way recorded signal.

3.1.2. Music stimuli used in state of the art

To measure specific patterns of responses to stimuli in experiments, researchers have used various stimuli like cognitive tasks, motor imaginary movement, visual or audial stimuli. When it comes to the music stimuli, the music genres used in the field are mainly classical and commercial music. The nature of music stimuli can be roughly categorized into natural music (used in [75])and computerized music [34]. The latter is easier to describe, control and measure, but also more limited in the aspect of musical expression. The set of music stimuli usually included short excerpts of few seconds, but there could also ne whole music pieces that are several minutes long [34].

Exploring the effects of music harmony, Koelsch [74, 75] has focused on analyzing the violated music expectations. In [75], he analyzed sequences of chords, where some chords were more or less expected, according to the rules of classical music harmony. In [74], he transformed the tonic resolution of existing classical pieces by shifting them a semi-tone above (which represents the most distant harmonic relationship according to the circle of fifths (cite)), thus violating the expectation. The results showed that the transposed chords evoked electric brain activity in range 180-350ms with negative polarity.

In their study [131], Sun, Feng, and Yang studied the behavioral and physiological responses to music tension in the context of complex nested structures. Then, the study on emotion recognition [15] used one-minute-long audio tracks to explore the effect of five different genres. One of the most common databases for decoding affective states is DEAP (Koelstra, 2011), which consists of 40-minutes-long EEG and other biosensors recordings of 32 subjects while they are watching music videos.

3.2. Overview of studies based on self-report methods

3.2.1. Musicality and music preferences assessment

Musical skills and expertise can vary greatly among individuals, starting from their repertoire of musical behaviors to the level of skill they display in any musical behavior. The types of musical behaviors can be ranging from performance on an instrument and listening expertise, over the ability to employ music in functional settings, to communicating about music. The assumption behind adding this aspect to the study is that the individuals' habits, skills, and preferences related to music could influence their emotional and cognitive experience while they listen to certain music stimuli.

Musical Sophistication Index (Gold-MSI)

Musicality is generally difficult to define as it is not necessarily correlated with formal musical education. The Goldsmiths Musical Sophistication Index (Gold-MSI) [101] was developed for that purpose. Being musically "sophisticated" in this sense means being more likely responsive to music. The questionnaire is also considered sensitive to differences among 'non-musicians'.

The self-report inventory also allows the scoring of a General Musical Sophistication factor that incorporates aspects from five sub-scales:

- Active Engagement (AE)
- Perceptual Abilities (PA)
- Musical Training (MT)
- Singing Abilities (SA)
- Emotions (EM)
- General Musical Sophistication (GM).

Music preferences

Acknowledging the individual's interests in and attitudes to different musical genres and activities is an important aspect of this research. Music preferences could be measured at different levels of abstraction, aiming to classify by artists, genre, instrumentation, etc. Some studies also explore the connection of music preference with personality [20, 114].

Assessment of music preference is usually performed via self-report questionnaires that use Likert scale [62] or in the form of music excerpts to listen. Some examples of the first group are Short Test of Music Preferences (STOMP) [113], built with Likert scale on 14 genres and mapped on 4 music-preference dimensions, then Revised Short Test of Music Preferences (STOMP-R) extended to 23 genres, and Music Preference Questionnaire (MPQ) [126] that evaluates preferences across 11 genres, and maps them to 4 dimensions (Rock, Elite, Urban and Pop/Dance). On the other hand, a questionnaire based on listening to music excerpts representing certain genres and then reporting the opinions are MUSIC (standing for Mellow, Urban, Sophisticated, Intense, and Campestral music) [115, 116] which counts music excerpts from mainly unknown pieces, representing 26 genres. A study by Ang et al. [6] was based on comparing preferences for ethnic and non-ethnic music among Malaysian students using 30 music excerpts. A study by Ferrer et al. [42] mentions the artist-based musical preferences (AMP) assessment as a "more ecologically valid instrument" to collect musical preferences.

Assessing one's music preferences brings various difficulties and biases. For instance, a difficulty on the side of the subject could be the lack of knowledge on certain genres or music in general, or certain lexical limitation (not being able to describe their preferences with words or numbers precisely and distinctively). When the study is based on listening to certain music excerpts, the subjects could be biased by the popularity of the piece, or their personal memories, and therefore judge for the entire genre based on their opinion on that piece. Additionally, their reactions and preferences could depend on a certain situation or a period in their lives, and can be changed through time, which is why some studies tend to repeat the questionnaire after several weeks. An obvious large bias present in the field of music is created by cultural differences, since individuals coming from different cultures are exposed to different scales, instruments, ornamentation, and more. Lastly, a bias exists on the experiment's side, as it is very challenging to find a proper representative piece of a certain genre, or even objectively classify genres into wide enough and distinctive enough categories.

3.2.2. Emotional competencies assessment

Aiming to develop Emotional Intelligence (EI) measures with good psychometric properties, the researchers had a difficulty of constructing questions that could be objectively scored. Namely, compared to cognitive ability measures that have objectively right or wrong answers, items designed to measure emotional ability rely on expert judgment to define correct answers, which is problematic for multiple reasons [117].

Based on whether the EI measure is a test off maximal performance or a self-report questionnaire, a classification into ability EI and trait EI measures, respectively, was proposed by Petrides and Furnham in 2000 [109]. Following this distinction, the ability EI tests measure an individual's theoretical understanding of emotions and emotional functioning, while the trait EI questionnaires measure typical behaviors in emotion-relevant situations, as well as self-rated abilities. O'Connor [103] provides an exhaustive review of the Emotional Intelligence measures mostly used in research, considering also factors such as the test length, number of facets measured and if tests are freely available. Summaries of some tests relevant for this study follow.

The most popular measure of ability EI is the Mayer, Salovey, Caruso Emotional Intelligence Test (MSCEIT) [93] that measures ability dimensions of perceiving emotions, facilitating thought, understanding emotions, and managing emotions. However, MSCEIT is a commercialized test that is expensive to use, relatively long (141 items) and thus time consuming to complete.

One of the main commercial trait EI instrument is the Bar-On Emotional Quotient Inventory (EQ-i) [8, 9], which comprises 133 items and is also expensive to use. The Schutte Self-Report Emotional Intelligence Scale (SSREI, SREIT, or SEIS) [61, 122]) is one of the most widely used self-report EI measures based on the earlier ability model by Salovey and Mayer [94]). It comprises 33 items and therefore, compared to Bar-On EQ-i represents an easier and shorter measure, that is freely available.

3.2.3. Emotional music scale

According to Juslin [69], the instruments commonly used for self-report are:

- Likert scales
- Adjective checklist
- Visual analogue scales
- Self-report instruments
- Diary study,

while some of the standardized mood/emotion scales are:

- Differential Emotion Scale (DES) [58]
- Positive and Negative Affect Schedule (PANAS) [141]
- Affect Intensity Measure (AIM) [80]

- Activation–Deactivation Adjective Check List [133]
- Geneva Emotional Music Scales (GEMS) [145]
- Profile of Mood States (POMS) [95]
- UWIST Mood Adjective Checklist [92]

In the DES, participants are requested to rate 30 emotion words, while the PANAS is based on 60 words that characterize emotion or feeling (e.g., inspired, calm, distressed) on a 5-point scale. The POMS is another commonly used instrument for assessing mood including 65 adjectives, again to be rated on a 5-point scale. The adjective checklists, such as the Activation–Deactivation Adjective Check List consists of sets of emotions of interest, presented to the participants in a scrambled order. Lastly, the AIM aims to quantify the intensity of an individual's emotional experience.

Additionally, Subsection 2.1.2 presents several models of measuring emotions.

Geneva Emotional Music Scales (GEMS)

Geneva Emotional Music Scales (GEMS) [145] is a Discrete Emotional model aimed at assessing emotions induced by music. The GEMS is considered to be the first instrument specifically designed to capture the richness of musically evoked emotions. It is based on multiple studies that included a wide range of music and listener samples. The model comprises nine categories of musical emotions: wonder, transcendence, tenderness, nostalgia, peacefulness, energy, joyful activation, tension, and sadness. This domain-specific model accounts for ratings of music-evoked emotions more powerfully than multi-purpose scales that are based on non-musical areas of emotion research. In addition, we also showed that the experience of the musical emotions tends to activate distinct emotive brain sites.

The original GEMS (GEMS-45) that is frequently used in studies on music and emotion has 9 scales and 45 emotion labels. However, a short and ultra-short version have been developed: GEMS-25 and GEMS-9, respectively. The GEMS-25 is based on the same adjective items pool and data-analytic procedures as the GEMS-45, although obviously providing less detailed results. The GEMS-9 presents only the primary nine factors along with a few emotion adjectives describing each factor. Each of the factors is supposed to be rated with a Likert scale of 5 levels. This ultra-short version is made for data collection in short studies. The use the GEMS for academic research purposes in a university environment is free, but any use for commercial purposes is prohibited.

4 | Methodology

This chapter describes the steps taken in this study, starting from the organization of the listening experiment, to data acquisition and data processing. Relevant choices, concerning sensors, music stimuli, questionnaires and more, are also described in detail.

4.1. Subjects of the experiment

A total number of 10 subjects whose age ranged from 22 to 36, out of which 7 males and 3 females, were recruited for the study from a variety of settings in a metropolitan area of Milan, Italy. All subjects were university students of different nationalities and cultural backgrounds. The subject's age ranged from 22 to 36, the average age being 26.1, with standard deviation 4.5. Upon invitation, a simple sorting process was conducted, mainly based on the applicant's basic health status, gender, and time availability. The gender factor was considered not only with the goal to obtain gender balance, but also due to certain limitations related to the size of the available EEG caps. As for the health status, the participation in the study was allowed to the subjects who do not suffer from any hearing difficulties, who are not carries of a pacemaker, nor those who show pathological inheritance concerning cardiovascular diseases, respiratory diseases, or neural system disorders.

The invitation contained general information about the study, the explanation of the protocol, and the Informed Consent document[1] that was approved by the research Ethical Committee of Politecnico di Milano. The Informed Consent describes the way in which the data is acquired, stored, and later processed. Prior to taking part in the study, all subjects were required to read and agree to what is described in the document, after which each of them was assigned a two-hours-long session, according to their availability. All 10 subjects gave their consent to take part in the acquisition session and did not interrupt the session until it was finished.

[1]https://drive.google.com/file/d/1I4dKPL1g0sg3R5JH-HseiQqHvHiLXODy/view

4.2. Self-report questionnaires

The questionnaires aimed to acquire data related to subjects' lifestyle, musical expertise, music preference and emotional competencies, as well as the emotions elicited while the subject listen to the music pieces during the data acquisition session. The questionnaires were added into the study as an additional element to analyze and categorize the results and explore conditional correlations. The assumptions behind this choice were that the subjects' preferred music genre, emotional competencies, and lifestyle habits related to music could be correlated with their reactions and thus their physiological responses to certain music stimuli. The questionnaires used in this work can be divided into two categories: those to be filled out before the acquisition session, aimed to profile the subjects and to collect general information about them, and those to be filled out during the session, collecting responses about elicited emotions during specific listened music pieces.

4.2.1. Subjects' background and preferences

The questionnaire that was supposed to be filled out before the session was assembled as an online Google Form, consisting of five sections:

I. Informed consent

II. Basic data collection

III. Questionnaire on Musical Sophistication (Gold-MSI) [101]

IV. Questionnaire on Music Preference (STOMP-R) [113]

V. Questionnaire on Emotional Competencies (SREIT) [122]

The subjects were asked to fill the form before coming to the session. Additional reason for filling out the form beforehand was to put more emphasis on reading the Informed consent and to get the subjects more familiar with the task requirements. The part aimed at collecting basic data consisted of questions about the subjects' age, gender, nationality, occupation, and similar. The three questionnaires chosen from the literature, in particular Gold-MSI, STOMP-R, and SREIT, were mainly constructed by a number of sentences to be rated on a Likert scale. These questionnaires are in depth described and discussed in Section 3.2.

4.2.2. Assessment of elicited emotions

The questionnaire to be filled during the data acquisition session, after listening to each choral piece was based on GEMS-9 (Geneva Emotional Music Scales) [145]. GEMS-9 is a self-report questionnaire that aims to assess the subject's emotions elicited while listening to a specific music piece. The original questionnaire was enriched with two general questions that were considered of value for further analysis and categorization:

1. How well do you know this music piece?
2. How much do you like this music piece?

Both questions were to be answered using the same Likert scale as for the original GEMS-9 questionnaire. In such a way adjusted document is presented in Appendix A.

4.3. Music stimuli

The initial set of music stimuli that was chosen for the experiment contained recordings of four pieces of polyphonic choral music, performed in a live concert in spring of 2019 by the "Discanto Vocal Ensemble", directed by Giorgio Brenna, in Milan, Italy. Then, following some examples in the literature [75], and in order to have a part of music stimuli that is represented and controlled in a simpler and more direct fashion, the initial set was enriched with 12 short sequences of chords played on a MIDI keyboard. The whole set of stimuli can be found online[2].

4.3.1. Chord sequences

The part of the music stimuli that included chord sequences was recorded using a MIDI keyboard and the digital audio workstation REAPER (Cockos, Inc) [23], and then rendered into audio files using a free "Soft piano" plugin by Spitfire Audio [56]. In total there were 12 chord sequences, each of which consisted of 5 consecutive chords. Six sequences had the initial tonic chord of C major, while the other six started from C minor chord. All sequences represented different chord progressions used in Western music, some with more complexity and tension (therefore "less expected"), and some more natural to a non-educated listener (therefore "more expected"). In particular, the 5th chord had the most variations, providing expected or unexpected resolutions. Some examples are presented in Figures 4.1 and 4.2.

[2] `https://1drv.ms/u/s!AvcHs1ydRWzggZ1Zp8YGjxTauNInXg?e=dvGmmd`

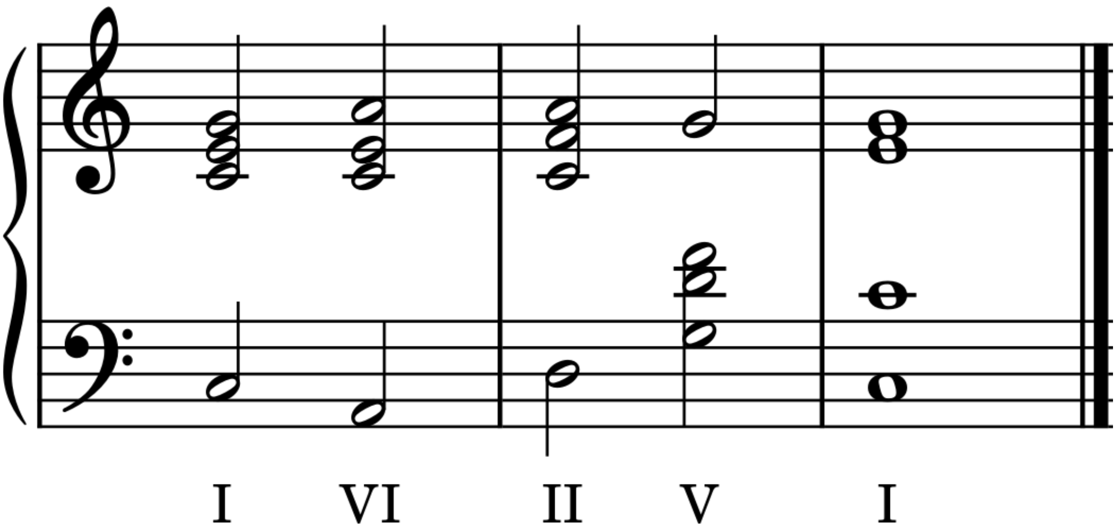

Figure 4.1: An example of a chord progression with an expected resolution from the dominant chord to the tonic chord

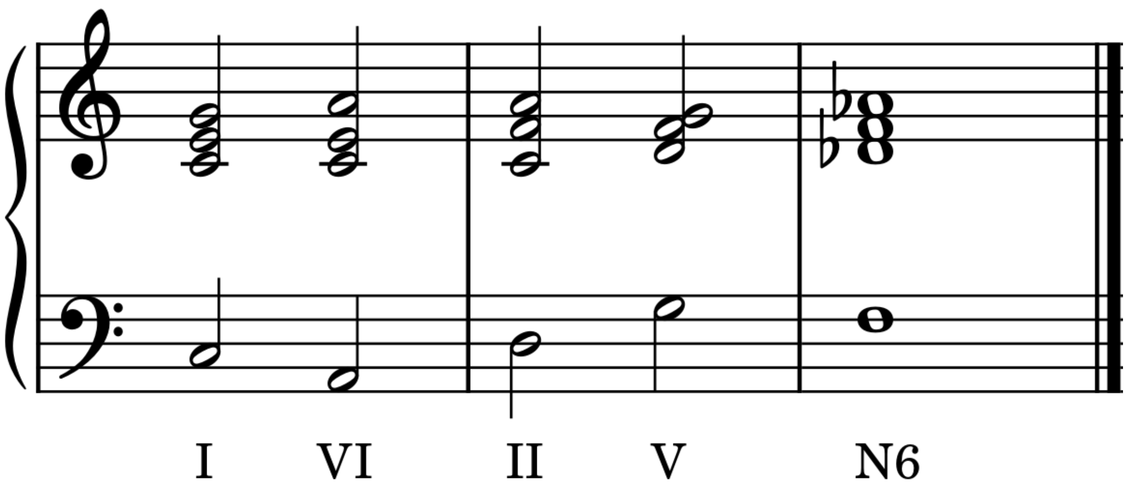

Figure 4.2: An example of a chord progression with a non-expected resolution from the dominant chord to the Neapolitan 6th chord

Each chord was set to last one beat, with exception of the very last chord of each sequence that took two beats. With BPM[3] set to 120, each beat lasted two seconds, and therefore each chord sequence lasted 12 seconds, as represented in Figure 4.3. The sequences were played one after another, and in order to minimize the bias and the mutual influence in the inter-subjects statistics, the order of sequences was randomized for each subject using a random sequence generator. Then, to give the subject some rest between the stimuli, six seconds of silence were added between each chord sequence. Following the described order, the 12 sequences were then merged into a single audio file for each subject, summing up to 10 different audio files. The total length of the resulting audio files was 3 minutes and 44 seconds.

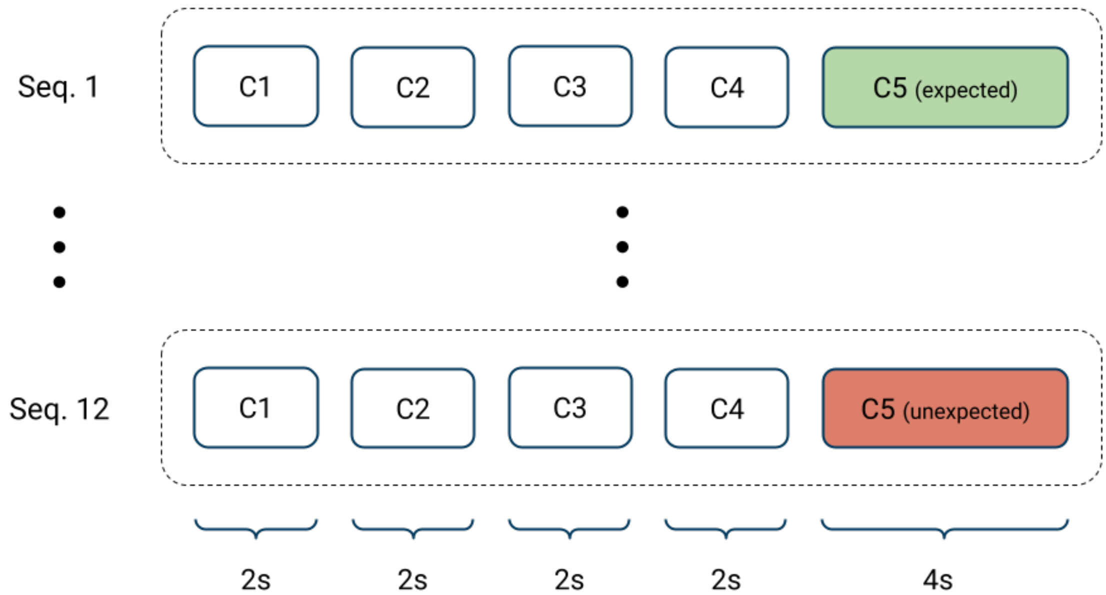

Figure 4.3: The representation of the structure of the 12 chord sequences. Each sequence consists of five chords, the last of which lasts longer than the rest and can be harmonically expected or unexpected.

[3]beats per minute

4.3.2. Choir recordings

This part of the music stimuli includes four pieces recorded during the live concert of the polyphonic choir "Discanto Vocal Ensemble", directed by Giorgio Brena in 2019. All four pieces can be classified into the genre of Western choral "sacred music" and thus represent a novelty in the field of researching human responses to music stimuli using physiological signals. The details about the four music pieces used in the experiment are represented in Table 4.1.

Piece title	**Composer**	**Length [min]**
"Abide with me"	William Henry Monk	1:52
"Intellige Clamorem Meum"	Alessandro Scarlatti	3:07
"Ubi Caritas"	Maurice Duruflé	2:00
"Ubi Caritas"	Ola Gjeilo	2:48

Table 4.1: The list of the choral pieces used in the experiment

For each of the pieces, the different aspects of music tension (further explained in Subsection 2.2.2) through time were manually annotated by music experts, including the choir director himself. The annotation process was done taking into account the corresponding music scores, but also the live performances, thus the layer of musical expression. The tension was represented with numbers ranging from 1 to 5, standing for the lowest and highest levels of tension, respectively. In such a way represented tension was then noted down in a tabular way, with a column for time latency and another column for the tension level. An example of such annotated harmonic tension through time for the piece "Ubi Caritas" by M. Duruflé is given in Table 4.2, while a corresponding graphical representation for the same piece is shown in Figure 4.4.

Latency [s]	Harmonic tension [1-5]
0	1
7.908	2
10.914	1
18.474	2
35.736	1
39.078	3
47.49	2
48.576	3
51.516	2
54.228	1
56.316	3
61.542	2
69.954	1
90.354	2
105.84	1
112.08	0

Table 4.2: The annotations of the harmonic tension for the piece "Ubi Caritas" by M. Duruflé.

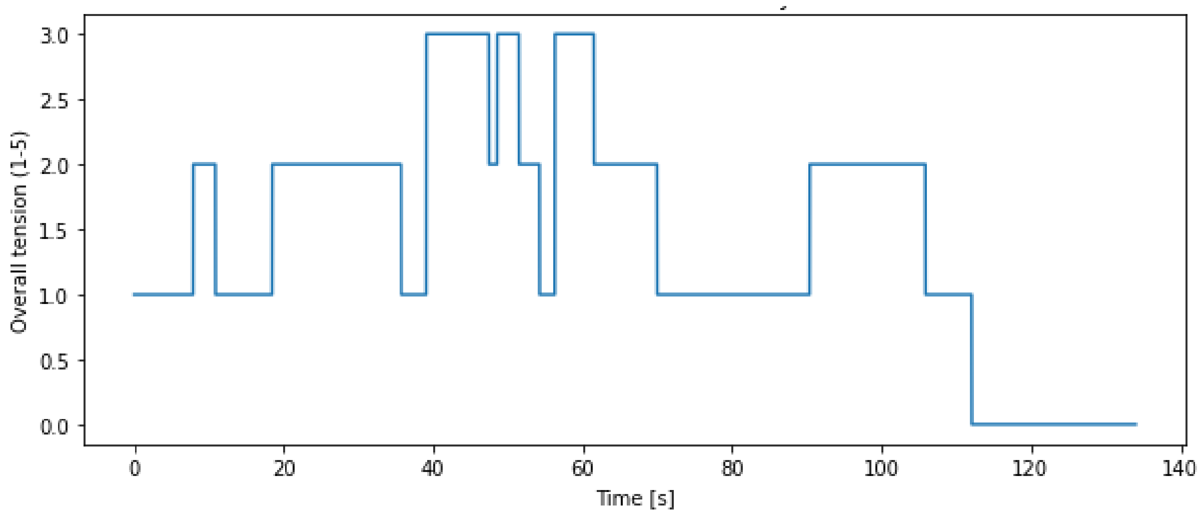

Figure 4.4: The plot corresponding to the annotations of the harmonic tension for the piece "Ubi Caritas" by M. Duruflé.

4.4. Technical equipment

The physiological signals chosen for the continuous acquisition were the following: 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.

Electroencephalogram (EEG)

The EEG signals were measured using a digital electroencephalograph system developed by Micromed S.p.A. The EEG cap with 61 electrodes was placed on the subject's head according to the international 10-20 EEG placement system [102], as presented in Figure 4.5. Each electrode was connected with wires to the acquisition device. The signals were recorded with a sampling rate of 1024 Hz. In order to improve the conductivity between the electrodes and the scalp, a water-soluble electrolytic gel was applied on subjects' scalp before the data acquisition. The subjects were previously informed about the procedure, were advised to keep their hair clean and dry, without ponytails, braids, hair clips etc, as well as to avoid using any sprays, oils, or gels in their hair beforehand. They were also explained that the gel could be removed from their hair using warm water at the end of the acquisition session. The gel was applied using a syringe, via a hole in each electrode.

Figure 4.5: An example of an EEG cap with 61 electrodes and a strap [5]

Electrodermal activity (EDA)

The electrodermal activity was recorded using the ProComp Infiniti System, model SA7500 (Thought Technology Limited, Montreal, QC, Canada), which is a multi-modality device for real-time biofeedback and data acquisition. This system was used for acquiring not only EDA, but also ECG and RSP signals. The sampling rate used for the EDA signals was 2048 Hz. The recording protocol required wrapping two clamps, each coupled with a sensor, around the medial phalanx of the index and middle finger of the non-dominant hand, as presented in Figure 4.6. The clamps were wrapped in a way that both ensures stable conductivity and is comfortable for the subject. The subject was asked to avoid engaging the hand's muscles, to minimize interferences from physical movement.

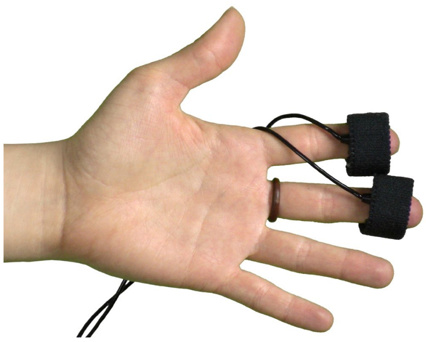

Figure 4.6: The placement of two clamps, each with EDA sensors, around fingers [87]

Electrocardiogram (ECG)

The electrocardiogram setup included three adhesive and disposable electrodes, placed on the subjects' thorax to record the cardiac activity. The electrodes were placed as presented in Figure 4.7 and were connected to the ProComp Infiniti System described above. The sampling rate of the recorded ECG signal was 256 Hz.

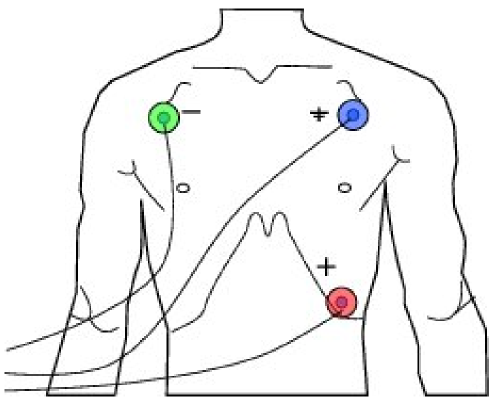

Figure 4.7: The placement of thee adhesive ECG electrodes on the thorax [44]

Respiratory activity (RSP)

The respiratory activity was recorded using a belt with a sensor, wrapped around the subject's thorax, just under the sternum. The sensor could track the movement of the thorax while the subject breathes in and out and was connected to the previously described ProComp Infiniti System. The samples were taken with a frequency of 256 Hz.

Figure 4.8: The placement of a respiratory belt around the thorax [140]

4.5. Experimental protocol

The acquisition sessions were conducted in a light and quiet room inside Building 21 of Politecnico di Milano, Italy, in December 2021. The BrainLab of Politecnico di Milano provided not only the room and equipment, but also educational support for this study. Two trial acquisition sessions were organized in order to build up the protocol and adapt it to the needs and limitations concerning time, space, and personnel. Within the official experiment, there were usually two sessions scheduled per day (one from 10AM to 12PM and the other from 2PM to 4PM), taking into account the time required for setting up the equipment before each session, as well as the cleaning after it. Before their acquisition session, the subjects were asked to fill out the provided online questionnaire and were given additional instructions related to the acquisition session. Upon arrival, the subjects were verbally reminded about the duration and tasks involved in the experiment, as well as the possibility to withdraw their participation at any given moment. The subjects were then asked to give a formal written consent of participation in the study and the involved data collection by signing the printed informed consent form.

The approximate duration of each acquisition session was 1 hour 45 minutes, including:

- Setting up the equipment (45 min)
- Acquiring the data (45 min)
- Cleaning (15 min)

The data acquisition part included continuous recording of the listed physiological signals and compiling the adjusted GEMS-9 questionnaires after each listened piece. During the session, the subject was sitting in the room together with the experimenter, while wearing the mentioned equipment and the AKG headphones [50], model K92.

4.5.1. Equipment setup

Before the subject arrived, the room was prepared for the session, which included organizing the space where the subject will be seated, the space for the experimenter, and the equipment. A movable desk with disposable electrodes, syringes, conductive gel, measure tape, dry wipes, and similar objects was prepared. Then, all the sensors were set up and connected to the PC on the experimenter's desk, with two screens to monitor several signals continuously. Next, the software for data acquisition were set up, entering general data about the subject and the session. The software used for acquisition of EDA, ECG, and RSP was BioGraph Infiniti (Thought Technology Limited), while the software used for EEG was Micromed System Plus Evolution. Lastly, a notes insert was prepared, in order to keep track of possible disruptions or other details during the session.

After the subject arrived, they were welcomed and seated in a comfortable chair, at a comfortable distance from the table. The disposable electrodes for ECG were attached to the subject's body, the two clamps for EDA were wrapped around their two fingers, and the respiratory belt was wrapped around the subject's thorax. At that point, all the incoming signals were checked, adjusting the equipment accordingly if necessary. As far as the EEG is concerned, first it was needed to measure the subject's head circumference and the nasion-inion distance with a measuring tape. According to the measures, one of the two EEG caps (one smaller, the other larger) was chosen and then set up on the subject's head, in such a way that the cap 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, it could be proceeded with filling the holes in the electrodes with the conductive gel. The gel was applied doing a circular motion with the syringe to move the hair and thus get closer to the skin, and then slowly getting the syringe out of the hole.

When all the holes were filled with gel in a systematic order, the signals coming from all channels were checked on the monitor. If a certain channel showed that there was not enough conductivity between the corresponding electrode and the scalp (e.g. if the signal of the channel drifted or was very noisy), the electrode was adjusted by adding more gel into the corresponding hole or moving the hair, or adjusting the whole EEG cap. The Figure 4.9 represents the setup of equipment on a subject's body: the reader can notice the respiratory belt around the thorax, as well as the wires of the ECG electrodes attached under the shirt.

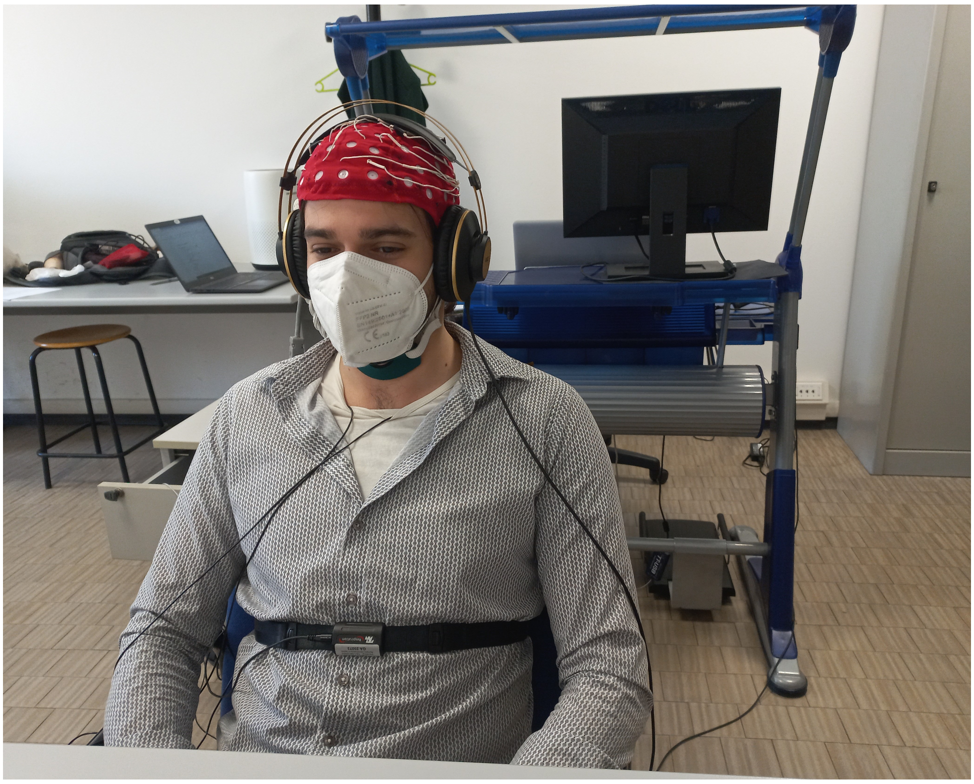

Figure 4.9: A subject with set-up equipment during an acquisition session

When the equipment was finally set up, as represented in Figure 4.10, the subject was informed about the task in details, and advised to stay relaxed and keep any physical movement at minimum, especially while listening to the music stimuli. A short trial listening session was then performed, in order to adjust the volume of the music stimuli playback according to the subject's preferences. After confirming with the subject that they are ready, the data acquisition part could start.

4.5.2. Data acquisition and monitoring

Before starting the task, it was needed to record the baseline signals, corresponding to the subject's resting state. The EEG recording protocol required acquisition of one-minute-long signal for the baseline, for two cases of the subject having eyes open and closed, while the EDA recording protocol needs a longer baseline signal in order to capture the slow changes while the subject reaches a resting state. Therefore, it was decided to record one-minute-long EEG baseline signal while the subject keeps their eyes closed, and then 5 minutes of both EEG and EDA baseline signals while the subject keeps their eyes open. The described process was performed with the help of a chronometer and an external button connected to the apparatus, to mark the beginnings of certain time intervals. The subjects were asked to keep their eyes open during the task.

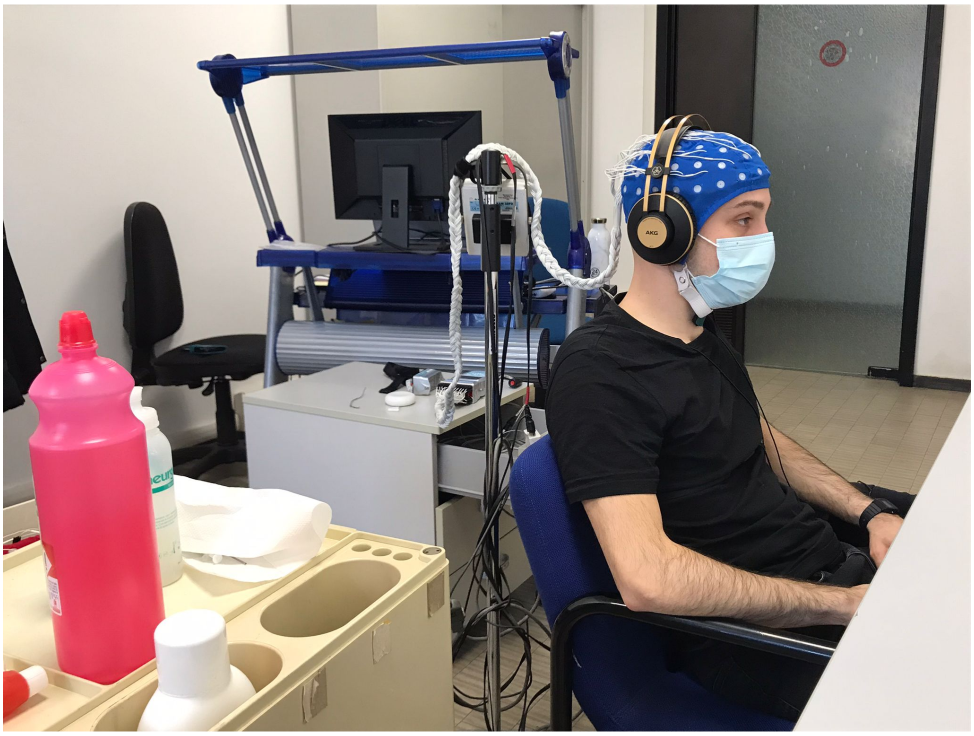

Figure 4.10: A subject with set-up equipment during an acquisition session

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 PC keyboard button that played the music stimuli from a PC. Both during the acquisition of the baseline data and the task-related data, the signals were continuously monitored during the session on the screens. The notes were kept on any disruptions in the experimental protocol.

4.5.3. Cleaning

When the task was finished, the EEG cap was carefully taken off of the subject's head and the gel was partially removed from their hair using dry napkins. After taking off the EEG cap, the rest of the equipment was taken off, including the disposable electrodes and the respiratory belt. The subject was then asked for feedback about the whole process; with that, the session was concluded, and the subject could leave the room.

After the subject has left, the acquired signals were again checked and saved. All the used equipment was dissembled and packed back where they were taken from, unless there is another session upcoming right after. The disposable electrodes and used napkins were thrown away. Lastly, the reusable part of the equipment needed to be cleaned: particularly the EEG cap and the syringes. The EEG cap was first soaked in water, and then the remains of the gel were taken out from each electrode hole with the help of toothpicks, ear picks, napkins, toothbrush, and similar objects. The whole process of cleaning the EEG cap could take around an hour, after which it was left to dry before the next use. The Figure 4.11 represents the whole above-described experimental protocol.

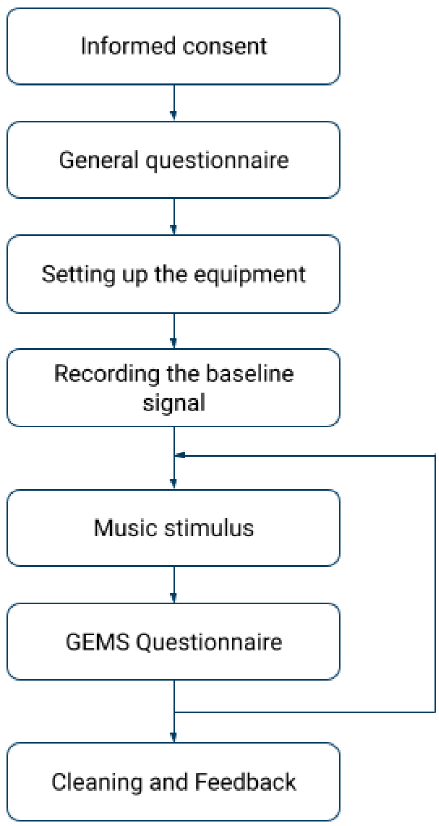

Figure 4.11: The diagram representing the proposed experimental protocol

4.6. Data processing

The raw EEG data processing was performed using EEGLAB [31] in MATLAB R2021b (the Mathwords, Inc.), while the processing of the acquired EDA, ECG, and RSP signals was performed in Python [137]. The created pipeline that includes eliciting emotions using the music stimuli, acquiring, importing, and preprocessing the data, as well as feature extraction and data analysis is represented in Figure 4.12.

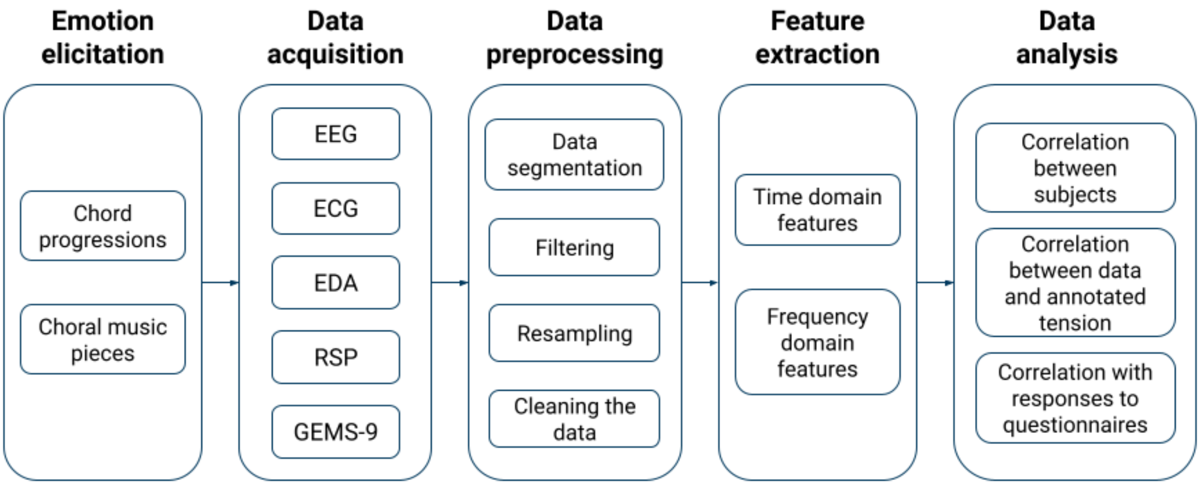

Figure 4.12: The diagram representing the data processing pipeline

4.6.1. EEG

At first, the raw *.TRC* files were imported and read into MATLAB and then transformed into *.SET* files. Among the 65 channels of the imported data, four of them (*elA32*, *elB31*, *elB32*, and *MKR*) were not EEG channels, so they were directly removed. Then, the information about button-induced events was read from the files' header, which allowed indicating the beginning of each audio track. Then, the imported files were segmented into tracks, according to the imported events' latency. Since the subjects' physical movement and other external interferences were increased in the parts between the listening tasks, segmenting into tracks before signal processing allowed disposing the parts with large artifacts. The first 16 channels of the raw EEG data corresponding to a part of a listened music track is represented in Figure 4.13.

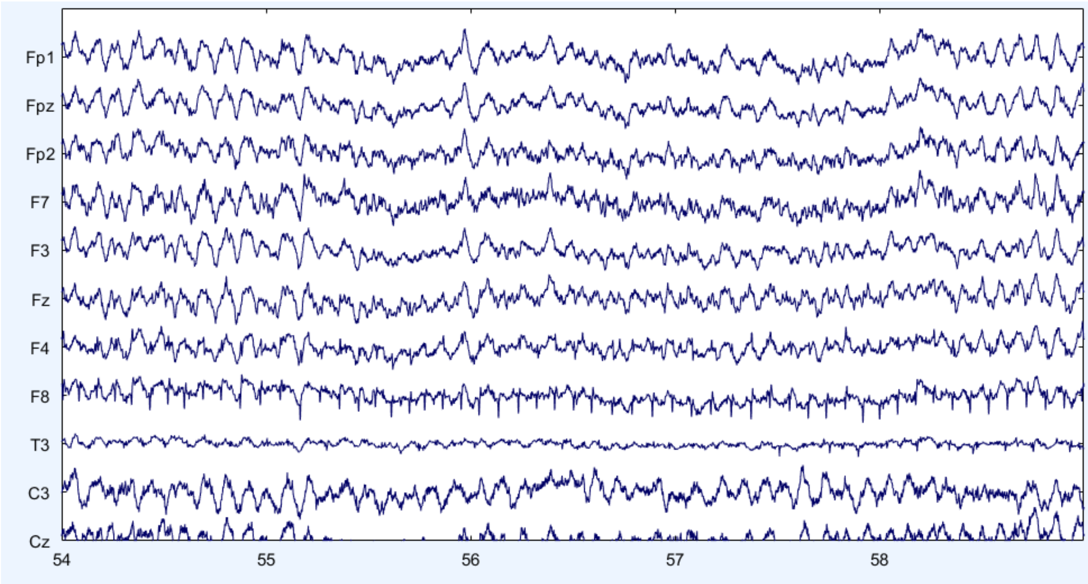

Figure 4.13: The first 16 channels of raw EEG data in the EEGLAB environment [31]

EEG: Data preprocessing

After having the EEG data organized according to the audio tracks, the signal preprocessing was performed. The data were initially filtered with the basic finite impulse response (FIR) filter, set at 1Hz as the lower edge and 45Hz as the higher edge of the frequency pass band. The reasoning behind the higher edge cutoff frequency is that there is a line-noise present at 50Hz, while the frequency range that concerns this task does not require high frequencies. On the other hand, the lower edge was set to reduce the noise present at low frequencies, while not affecting the spectrum of interest. After filtering, the downsampling of the data was performed, reducing the sampling frequency from 1024Hz to 256Hz, in order to reduce the memory load and required computational power.

Next step in the preprocessing pipeline was cleaning the data, by inspecting the channels that could be considered not useful, due to high frequency noise or bad conductivity of the corresponding electrodes. The detection of such bad channels was done using Clean Rawdata plugin [76, 99]. This procedure often detected those channels whose corresponding electrodes are physically positioned near the reference electrode *Cz*, due to their lower voltage amplitude. However, since those channels are not by default noisy or artifactual, after manual inspection and confirmation of channels *Pz*, *P1*, *P2*, *Cpz*, *Cp1*, and *Cp2*, they were discarded from the set of channels to be removed. Then, the channels that are finally selected as bad were removed from the data, after which spherical interpolation [41, 108] of the missing channels was performed, which led to having 61 channels in the data for each subject. The information about detected and removed channels was kept in the fields of the EEG data structure. Since the task required having continuous data through time, no additional removal of temporal data portions was performed.

After removing the bad channels from the data, the artifacts embedded in the data (muscular movement, eye blinks, or eye movements) were corrected or subtracted, using an independent component analysis (ICA) algorithm [90]. The ICA is used to separate and estimate signal sources, which also allows removal of artifactual components. In the very end of the preprocessing pipeline, the re-referencing of the channels was performed, using the method that computes the common average reference (CAR) [84].

EEG: Feature extraction

After pre-processing, the next step was to extract meaningful features from the cleaned EEG data. The event-related potential (ERP) described in Subsection 2.3.1 was not used in the end, since there was a low number of music stimuli repetitions to obtain meaningful results.

The analysis in frequency domain was performed for segments of 1 second or 256 samples for all EEG recordings. For each segment of 1 second, standardization was performed, by subtracting the mean value of the samples in the segment and dividing by the standard deviation of these samples. Such operation was performed for each channel. Then, power spectral density (PSD) was calculated for each of the described standardized segments, using the Welch estimation method [142], again for each channel. The mean power of the alpha band (8-13 Hz) and the theta band (4-8 Hz) was calculated for each segment and for each channel. The resulting matrices for average power in alpha and theta bands are of dimension $N_{samples}$ x $N_{channels}$, where $N_{samples}$ corresponds to the number of segments, which is in this case equal to the length of the data in seconds, while $N_{channels}$ stands for the number of channels. Since the locations of the electrodes are known and are same for each subject, the channels have been grouped into five regions: Frontal, Central, Occipital, Parietal, and Temporal, and additionally divided into left and ride sides of the scalp. The described process was performed for each of the subjects, for each of the music tracks and baseline files, and for both alpha and theta bands.

4.6.2. EDA

The raw data for EDA, ECG, and RSP were organized in *.txt* files, where several first rows contained basic information about the session and the subject, while the rest of the rows contained data points for the time latency and the amplitude values of the three signals, each sample in a new row. After importing and reading the *.txt* files, as well as importing the events, the data were segmented into tracks and saved as JSON (JavaScript Object Notation) files. The raw EDA data of one of the separated tracks is represented in Figure 4.14.

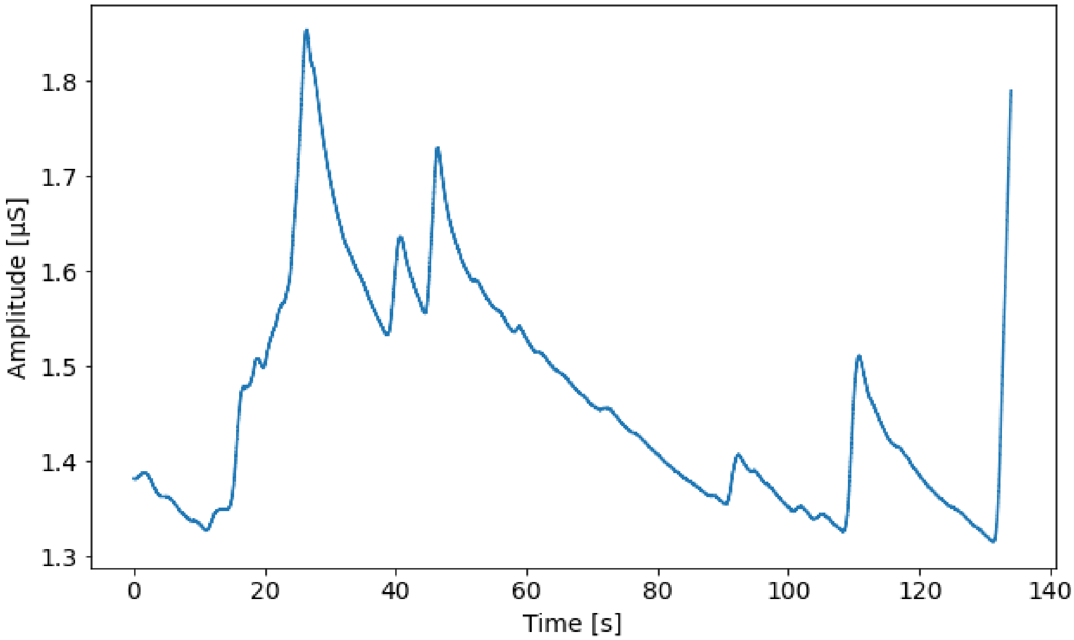

Figure 4.14: EDA raw data for subject S08

EDA: Preprocessing

The processing of EDA data was done in Python, using Python library NeuroKit2 [91]. First, the data were downsampled from 2048 Hz to 1024 Hz. Then, the cleaning of the data was performed, which included filtering the signal with a Butterworth low-pass filter of the 4th order and then smoothing it as described in [21]. The comparison between the raw and the cleaned signal is represented in Figure 4.15.

As explained in Subsection 2.3.3, the EDA consists of the tonic and the phasic part, so considering the signal as a whole would not give adequate measures of SCL and SCR. Therefore, the decomposition into tonic and phasic components needed to be done, which was performed using a function *eda_phasic* from NeuroKit2. The function provides several models to extract the components, while the model chosen for this study is cvxEDA [48], based on convex optimization and robust to noise coming from overlapping SCRs. The result is shown in Figure 4.16.

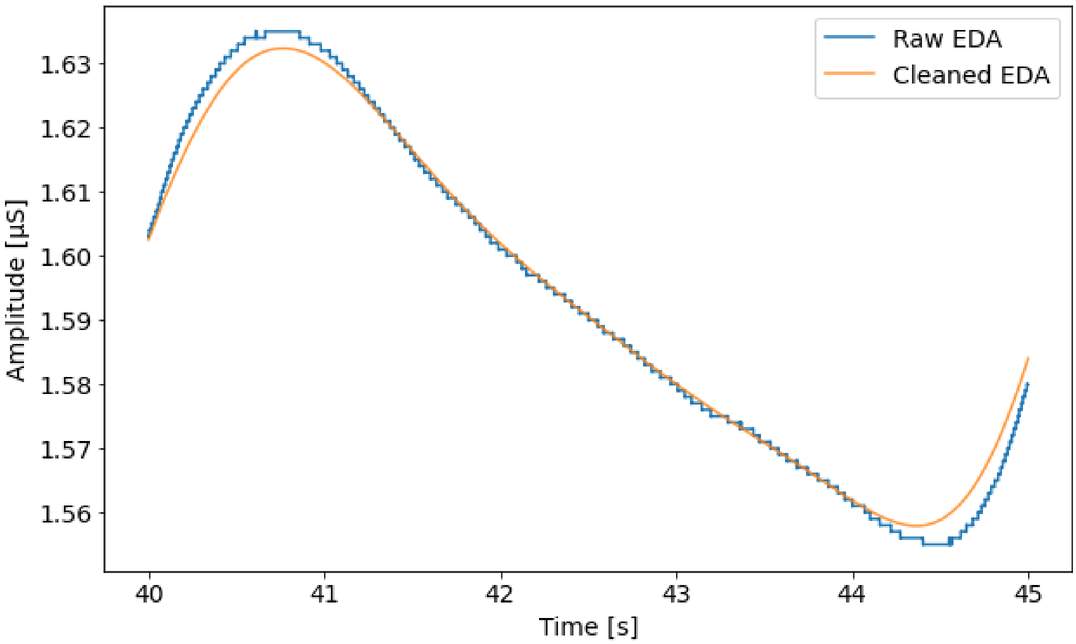

Figure 4.15: Raw and cleaned EDA data for subject S08

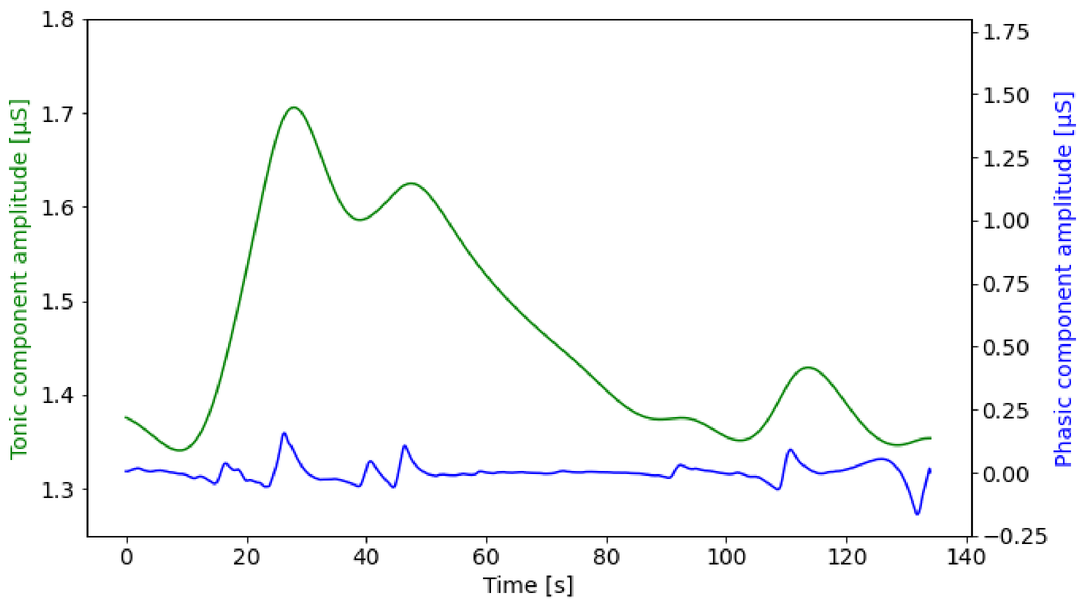

Figure 4.16: EDA tonic and phasic components for "Abide with me", for the subject S08.

EDA: Feature extraction

Other than extracting the SCL (tonic component) and SCR (phasic component) from the EDA data, the library Neurokit [91] helps extracting features such as SCR onsets, SCR peaks, SCR amplitude, or SCR rise time.

4.6.3. ECG

For the processing of the ECG data, a Python library HeartPy [135, 136] was used. As described in Subsection 2.3.2, the defining feature of the ECG is the QRS-complex, out of which the R-peak is strongly present in the signal and it is used for heart beat extraction.

ECG: Preprocessing

Since the ECG data were acquired by the same device as the EDA, the number of samples written in a second is equal to that of the EDA (2048 Hz), even though the practical sampling frequency of the ECG is 256 Hz. Therefore, as the first step, the downsampling was performed, lowering the sampling frequency from 2048 Hz to 256 Hz. In such a way downsampled, but still raw ECG signal is presented in Figure 4.17.

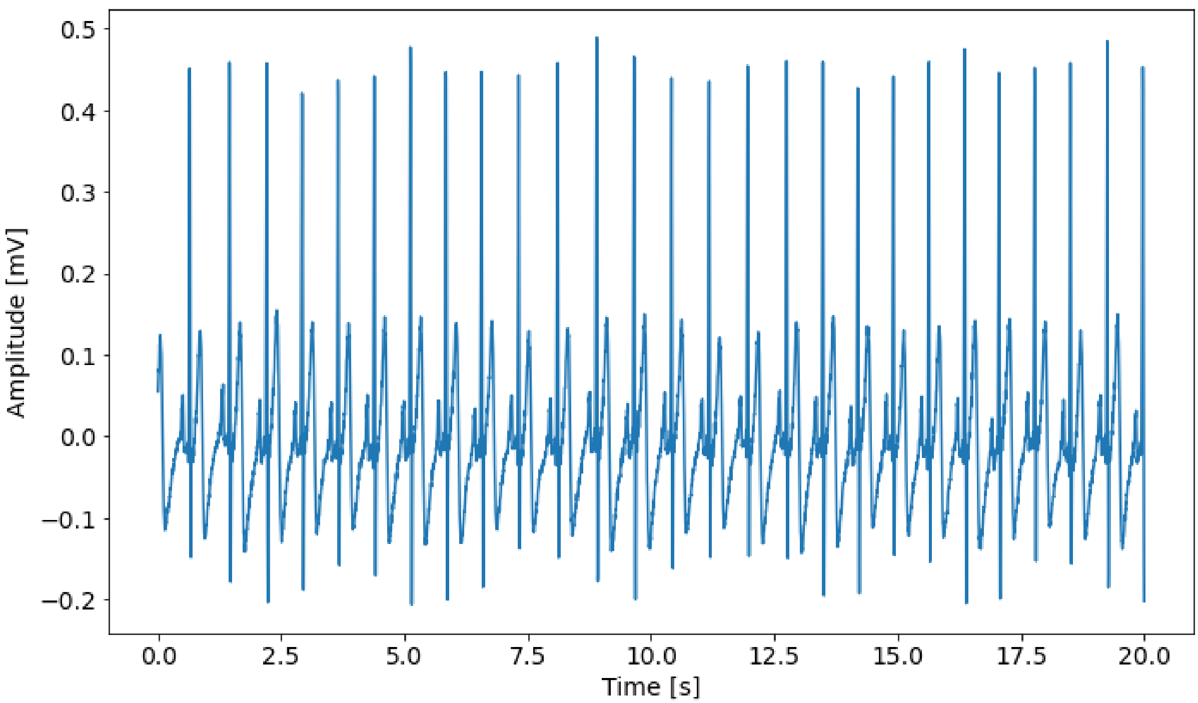

Figure 4.17: Raw ECG signal for a part of "Ubi Caritas" by Gjeilo, for the subject S08.

The preprocessing of the ECG data starts with preparing the raw signal for the R-peak detection. This involves filtering the signal with a bandpass filter with [3, 45] Hz as cut-off frequencies. Then, a peak detection function from the Python SciPy library [139] was used to locate peaks in the signal. In such a way located R-peaks are shown in Figure 4.18.

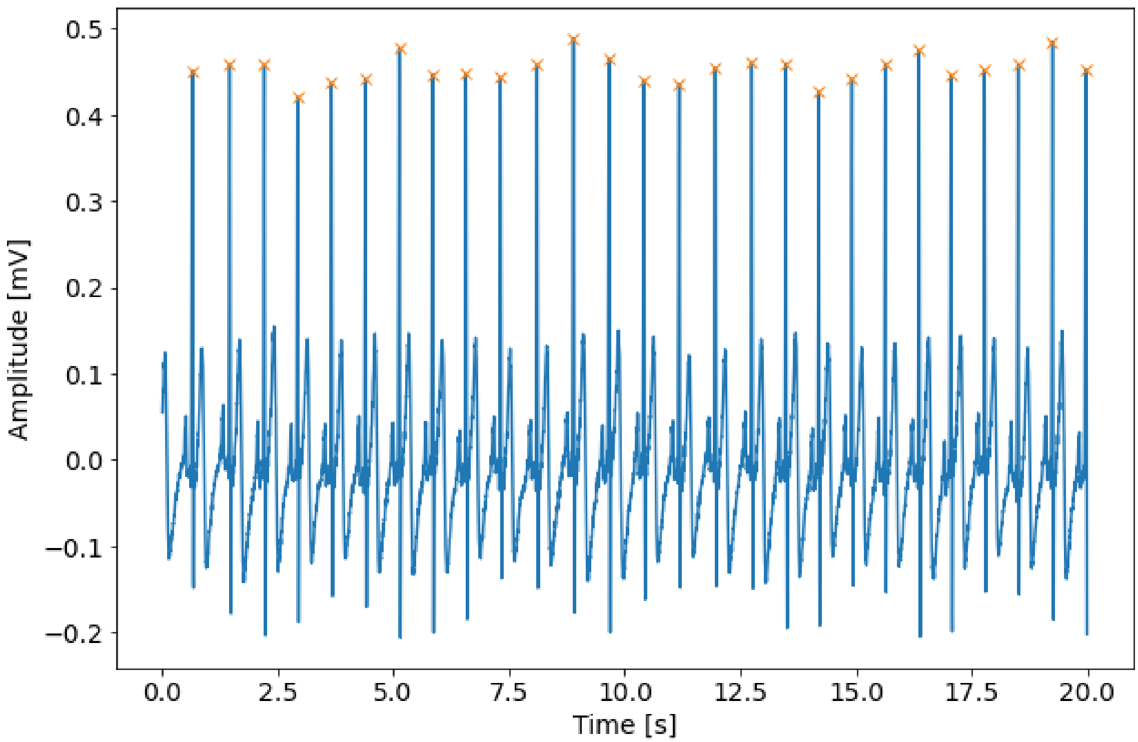

Figure 4.18: Raw ECG with R-peaks for a part of "Ubi Caritas" by Gjeilo, subject S08

ECG: Feature extraction

The peak detection algorithm can incorrectly detect peaks if the signal is very noisy. Therefore, after the R-peaks are detected, they are tested and possibly rejected based on a threshold value for the RR-intervals (that is, intervals between two R-peaks) in the analyzed section. The thresholds are determined as $RR_{\text{mean}} \pm 30\%$ of RR_{mean}, with minimum value of 300. If the RR-interval exceeds one of the thresholds, it is ignored. Only accepted peaks in the segment are used for computing the ECG measures [135, 136].

The processing of ECG data usually requires using longer time windows (e.g., more than 40 seconds) in order to have enough material to compute features based on averaging R-peaks positions. Therefore, the ECG signal is not the most appropriate for continuous tracking of emotional response to listened music. Using the library HeartPy, several time-domain features were computed across the whole length of music tracks:

- BPM (Beats per minute)
- IBI (Interbeat interval)
- SDNN (Standard deviation of the NN (R-R) intervals)
- SDSD (Standard deviation of successive RR interval differences)
- RMSSD (Root mean square of successive differences),

out of which SDNN, SDSD, and RMSSD are the measures of HRV [54].

4.6.4. RSP

The processing of RSP signals was done using a Python library NeuroKit2 [91], same as for EDA signals. After organizing the RSP data, the first step of preprocessing was downsampling from 2048 to 256 Hz, for the same reason as described in Subsection 4.6.3. The raw respiratory activity signal is represented in Figure 4.19.

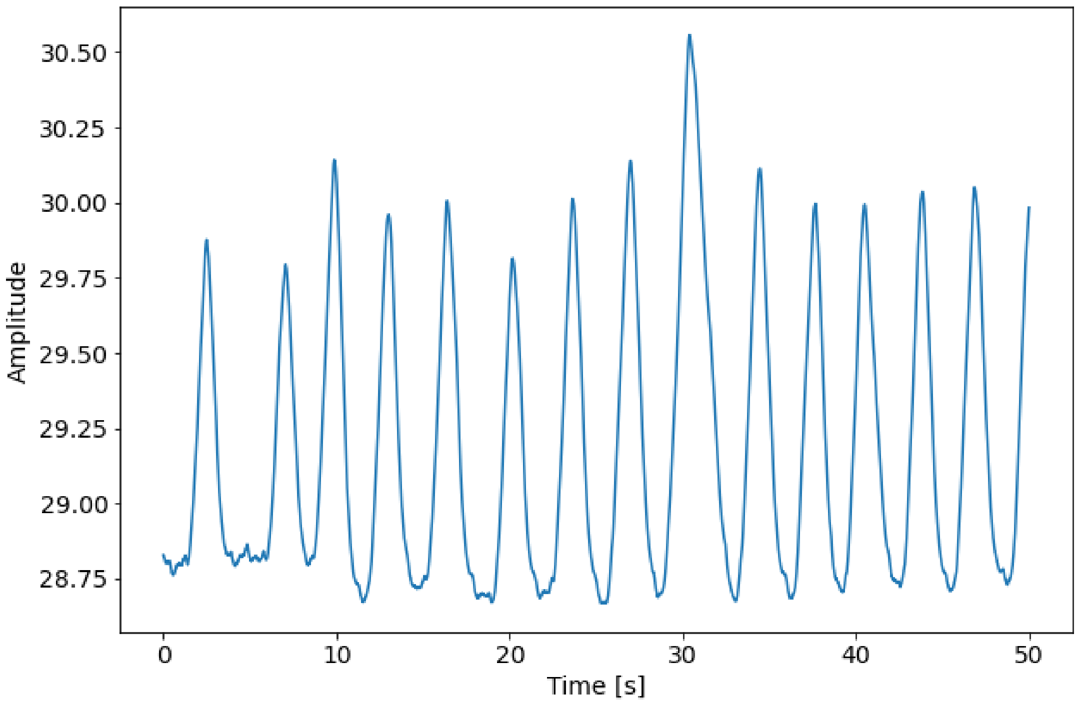

Figure 4.19: Raw RSP signal for "Ubi Caritas" by Gjeilo for the subject S05.

Next, it was necessary to clean the raw signal using the function *rsp_clean()* and extract the inhalation peaks of the signal using *rsp_peaks()* [91]. Then, one of the first features to extract for RSP signals is the respiratory rate, representing the number of breaths an individual takes per minute and measured in breats per minute (BPM).

The respiratory rate was computed using a function *rsp_ rate*, and a graph for one of the music pieces and for one of the subjects is presented in Figure 4.20

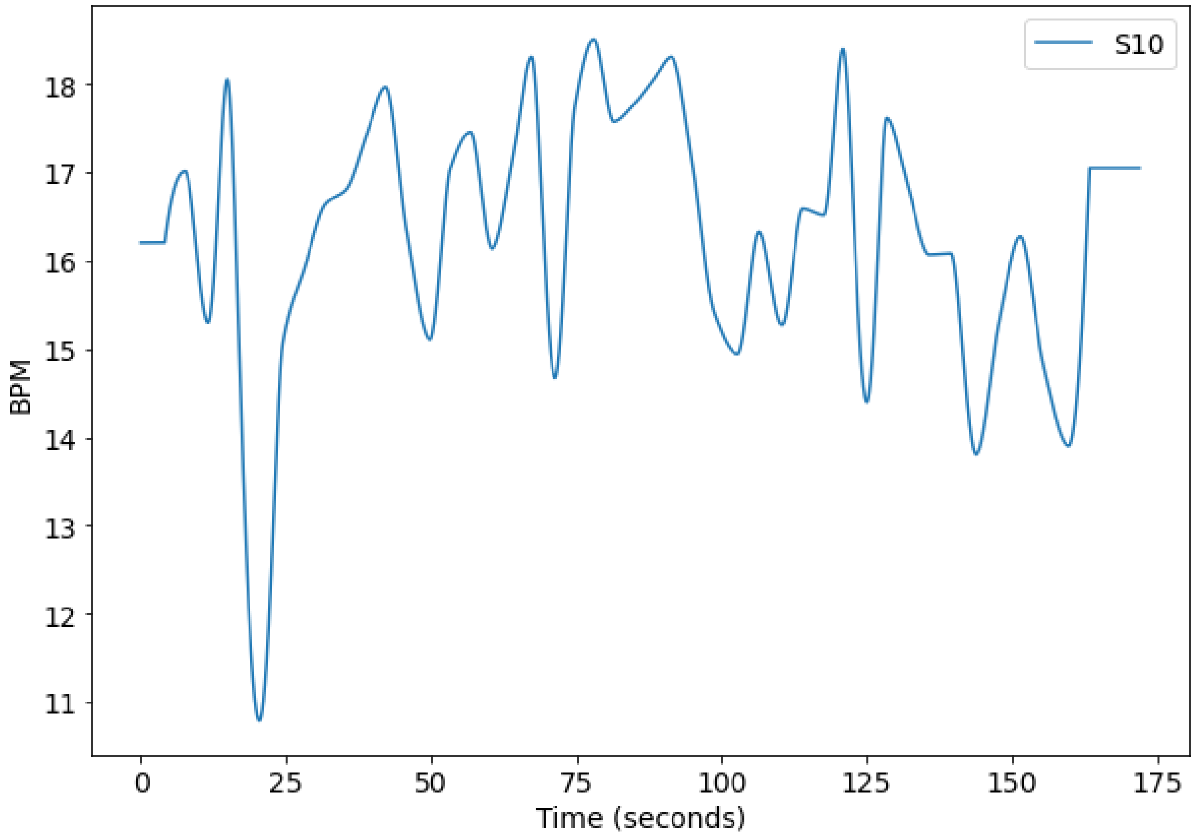

Figure 4.20: RSP rate signal for "Ubi Caritas" by Gjeilo, for the subject S10.

After having extracted the respiratory rate signal and the peaks, a Respiratory Rate Variability (RRV) can be computed, using a function *rsp_ rrv()* [91]. This function gives as outputs several RRV measures including time domain, frequency domain, and nonlinear features, some of which are listed below:

- Time-domain features

    RMSSD (root-mean-squared standard deviation)

    SDBB (standard deviation of the breath-to-breath intervals)

- Frequency-domain features

    Power of the LF band

    Power of the HF band

    Ratio between LF and HF bands

4.6.5. Questionnaires

Gold-MSI

The scoring system of a General Musical Sophistication factor incorporates aspects from five sub-scales:

- Active Engagement
- Perceptual Abilities
- Musical Training
- Singing Abilities
- Emotions
- General Musical Sophistication.

The summed and average values for each of the sub-scales were extracted for each subject.

STOMP-R

Following the scoring instructions from the literature, a possible scoring of STOMP-R (Revised Short Test of Music Preferences) for four dimensions implies classifying the music preferences into the following categories: Reflective & Complex, Intense & Rebellious, Upbeat & Conventional, and Energetic & Rhythmic. According to the MUSIC model, the classification into five groups is also possible: Mellow, Urban, Sophisticated, Intense, and Campestral music.

SREIT

The SREIT (Schutte Self-Report Emotional Intelligence Scale) questionnaire is comprised of 33 items based on a Likert scale. Thus, the total SREIT score for each participant was derived by summing up the item responses. For a larger number of subjects, factor analysis could be performed, grouping the responses into several categories.

5 | Analysis

This chapter presents the steps taken in analysis; considering the direction of the study, it is mainly focused on EEG signals, due to their high temporal resolution.

5.1. EEG

The analysis of EEG signals was mainly performed in the frequency domain. This section presents and discusses the correlation between various features related to music and EEG.

5.1.1. Power spectral density

The power spectral density for EEG signals was estimated using the Welch method, for each of the recorded files and for each EEG channel. The estimated PSDs of the EEG signals corresponding to two baseline states - with subjects' eyes closed (EC) and open eyes (EO) - averaged across time and for all subjects are represented in Figure 5.1.

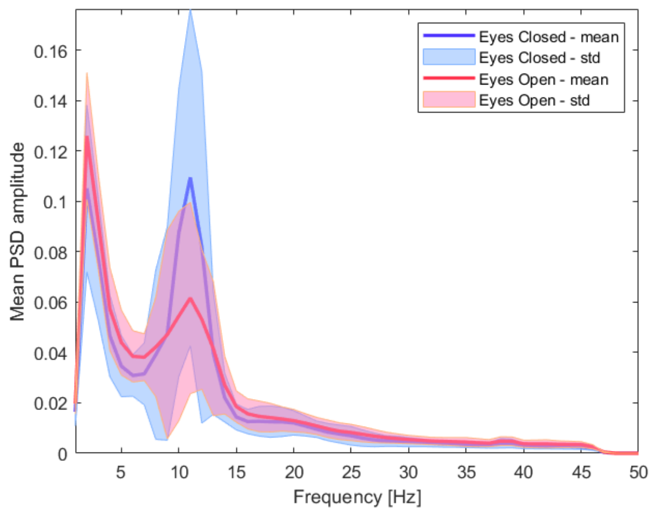

Figure 5.1: A comparison between estimated PSDs of the EEG signals corresponding to two baseline states. The blue and red colors correspond to the baseline states with subjects' eyes closed (EC) and open (EO), respectively.

It can be observed that the averaged estimated PSD corresponding to the EC-baseline is significantly stronger than EO-baseline in the frequency area around 10 Hz, which corresponds to the alpha frequency band (8-13 Hz), as defined in Section 2.3.1 (Figure 2.7). Increased power for frequencies below 2 Hz is visible for both baseline conditions, with EO-baseline being slightly higher.

The same comparison was then performed between the estimated PSD of EO-baseline and each of the music tracks, again averaged across time and for all the subjects. The corresponding plots are shown in Figure 5.2.

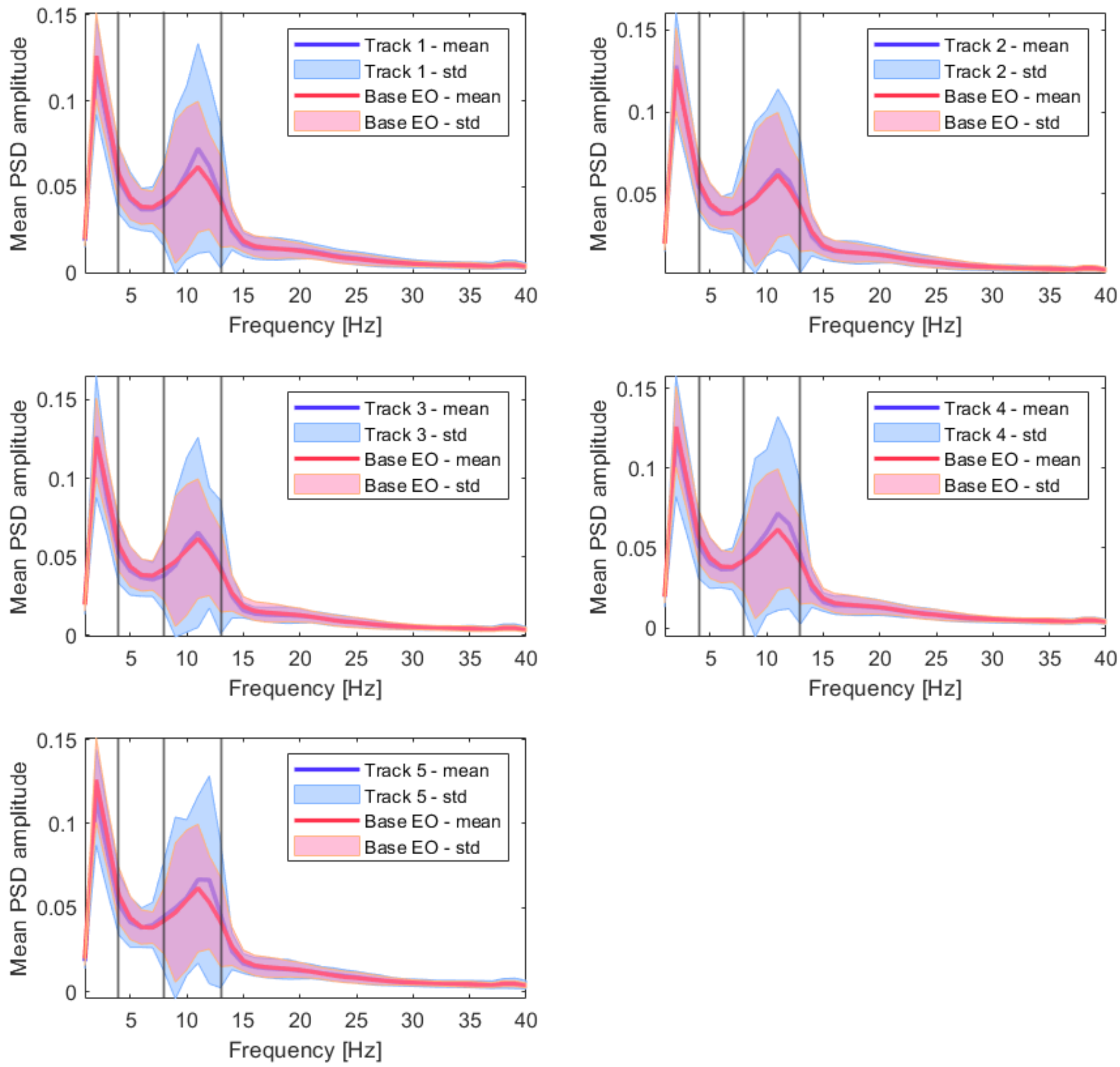

Figure 5.2: A comparison between estimated PSDs of the EEG signals corresponding to the baseline-EO state and the five tracks. The black lines mark the boundaries of theta and alpha frequency range.

Again, it can be observed that the average alpha activity for the EEG signals corresponding to the music tracks is stronger compared to those corresponding to the EO-baseline signals. However, in the case of music tracks versus the EO-baseline, the difference is not as prominent as in Figure 5.1.

Extracting the theta and alpha temporal activity

The PSD was then estimated for segments of 1 second, using a moving window with no overlap. Normalization was performed for each segment and each EEG channel, subtracting the mean value of the samples and dividing by their standard deviation. Then, the integration over theta (4-8 Hz) and alpha (8-13 Hz) frequency bands was performed, again for each segment and each EEG channel. The PSDs in theta and alpha bands over time for the subject S02 are shown in Figure 5.3.

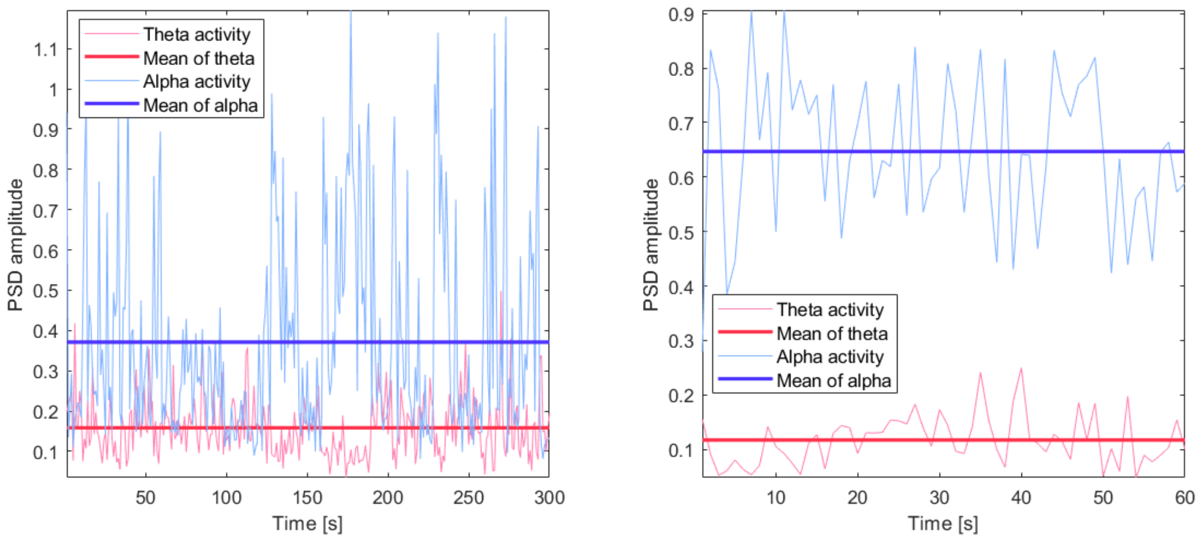

Figure 5.3: EEG baseline EC and EO, alpha and theta, subject S02

It can again be noticed that the alpha activity is stronger for the baseline EC state, as expected. However, not all the subjects have such high amplitude of alpha power. For analysis and demonstration purposes, the mean value and the standard deviation were computed for each subject and for each music track. The resulting values are represented in Figure 5.4. Observing the presented values, a pattern can be noticed among certain subjects; namely, subjects S02, S07, and S10 have higher values of both mean and standard deviation in the alpha band, for both tracks.

Subject	"Abide with me"				"Intellige Clamorem Meum"			
	Mean θ	Mean α	STD θ	STD α	Mean θ	Mean α	STD θ	STD α
S01	0.25	0.18	0.15	0.14	0.23	0.19	0.13	0.13
S02	0.12	0.48	0.09	0.31	0.13	0.47	0.09	0.30
S03	0.21	0.16	0.13	0.10	0.22	0.18	0.14	0.12
S04	0.15	0.47	0.10	0.29	0.15	0.55	0.10	0.28
S05	0.24	0.26	0.14	0.15	0.23	0.26	0.14	0.15
S06	0.22	0.18	0.13	0.14	0.24	0.16	0.13	0.10
S07	0.17	0.21	0.10	0.13	0.17	0.24	0.10	0.14
S08	0.14	0.25	0.09	0.17	0.14	0.20	0.09	0.14
S09	0.16	0.25	0.10	0.16	0.10	0.15	0.08	0.13
S10	0.42	0.49	0.25	0.24	0.36	0.53	0.25	0.24

Figure 5.4: Values of mean and standard deviation for the 10 subjects, for "Abide with me" and "Intellige Clamorem Meum". The darker colors correspond to larger values.

5.1.2. Inter-subjects analysis

The inter-subjects analysis refers to comparing the EEG signals among subjects and exploring the correlation between their features.

Standardization

Due to numerous physiological and cognitive differences across subjects, it is not possible to directly compare with each other the EEG signals corresponding to different subjects listening to the same stimuli. For that reason, we use the mentioned baseline files that were recorded in subject's resting state while they kept their eyes open because it is more similar condition to the one of the tasks. After having computed the mean power for alpha and theta bands, the power corresponding to the music tracks was standardized with respect to the power of the baseline of the same subject. The described standardization was performed by subtracting the mean value of the baseline and then dividing by the standard deviation of the baseline. This procedure was performed for each channel, for each music track, and for each subject. The estimated PSD across time for subjects S01, S02, and S10, for the case of the original PSD, the PSD with subtracted baseline mean, and the one divided by the standard deviation, are presented in Figure 5.5, corresponding to alpha activity and Figure 5.6, corresponding to theta activity.

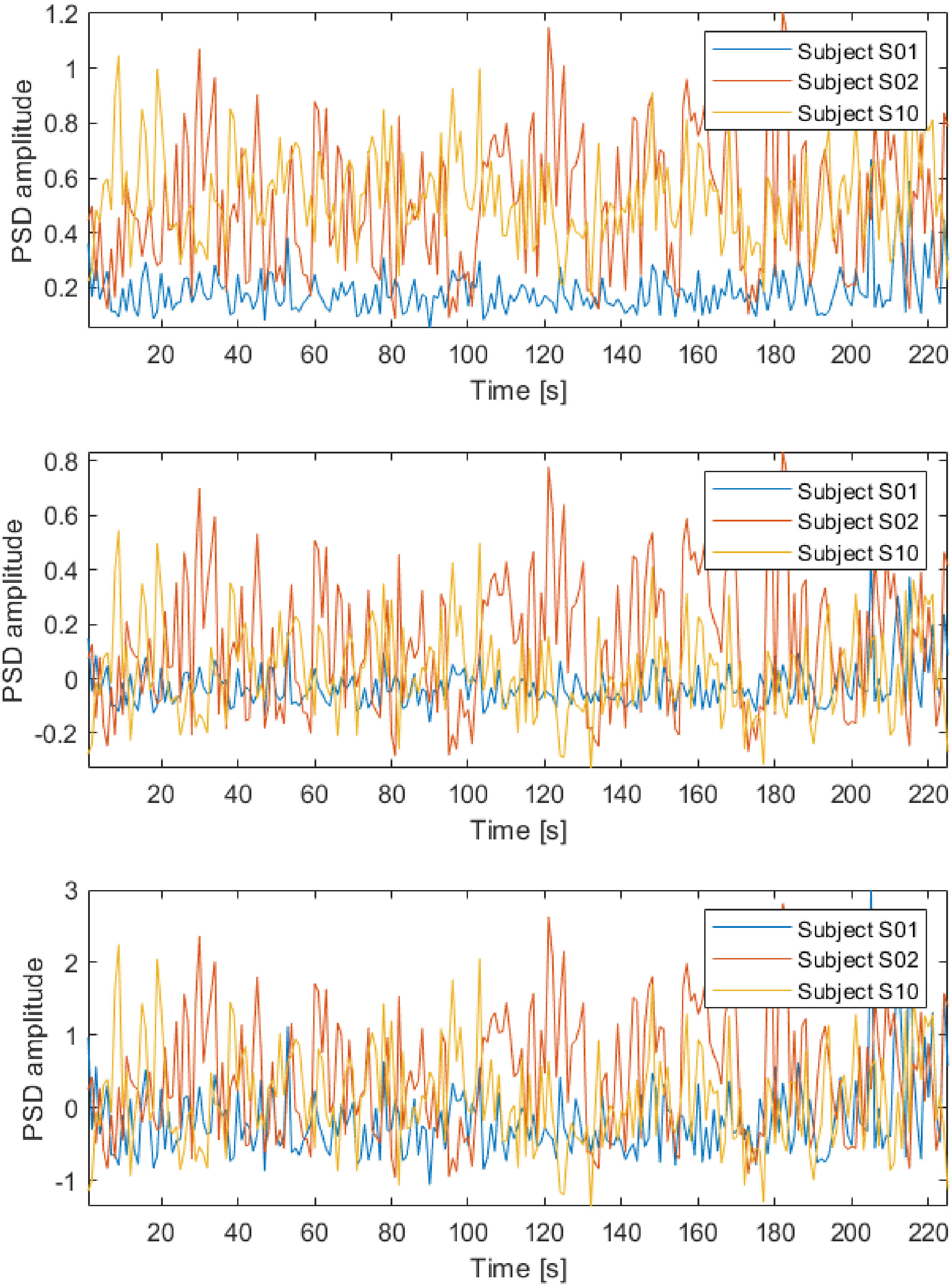

Figure 5.5: Standardization steps of the PSD of the alpha band for subjects S01, S02, and S10, for track 1. (a) Without normalization, (b) Subtracting the mean of the baseline-EO, (c) Dividing by the standard deviation of the baseline-EO.

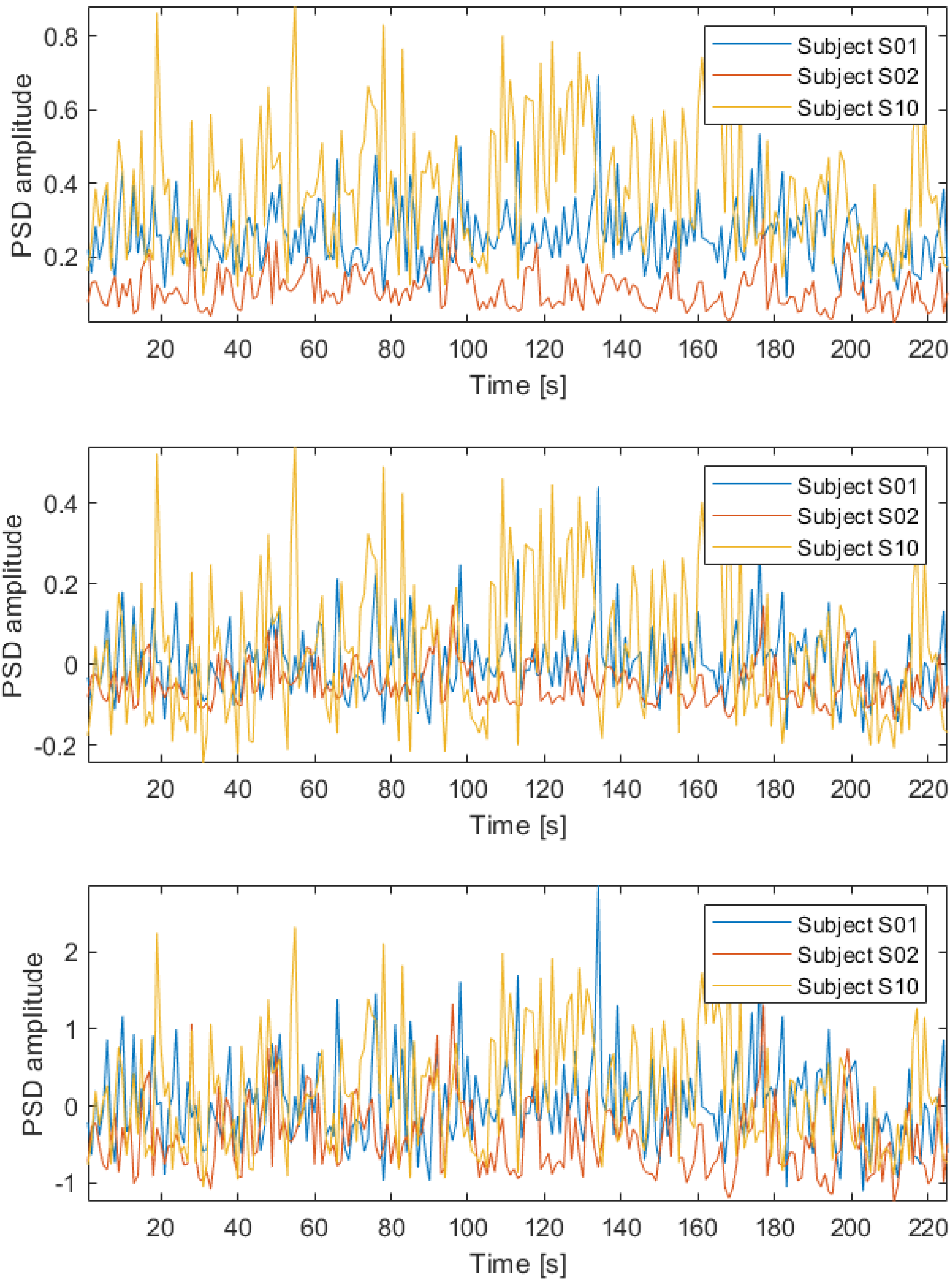

Figure 5.6: Standardization steps of the PSD of the theta band for subjects S01, S02, and S10, for track 1. (a) Without normalization, (b) Subtracting the mean of the baseline-EO, (c) Dividing by the standard deviation of the baseline-EO.

Correlation between subjects

The arrays containing values of estimated PSD for segments of 1 second were compared among all 10 subjects. The resulting 10x10 correlation matrix measuring the correlation between the PSDs of the alpha band of all subjects for tracks “Intellige Clamorem Meum” and “Ubi Caritas” are shown in Figures 5.7 and 5.8, respectively.

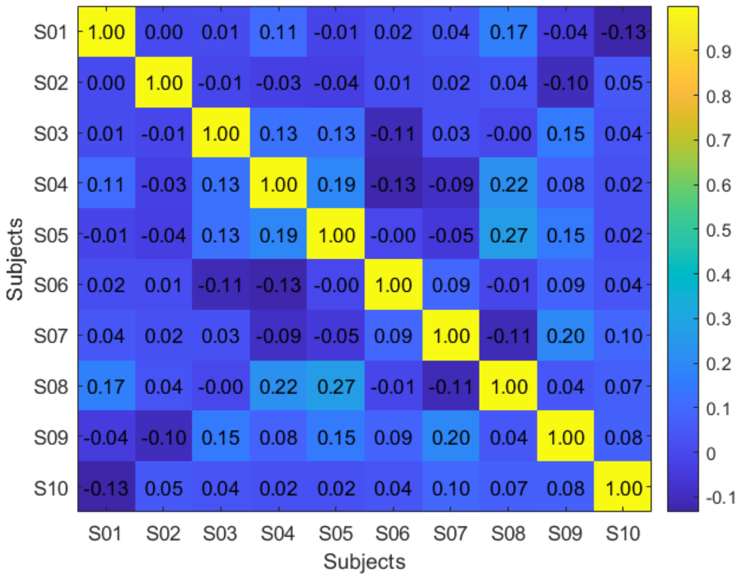

Figure 5.7: Correlation matrix between all subjects for "Intellige Clamorem Meum"

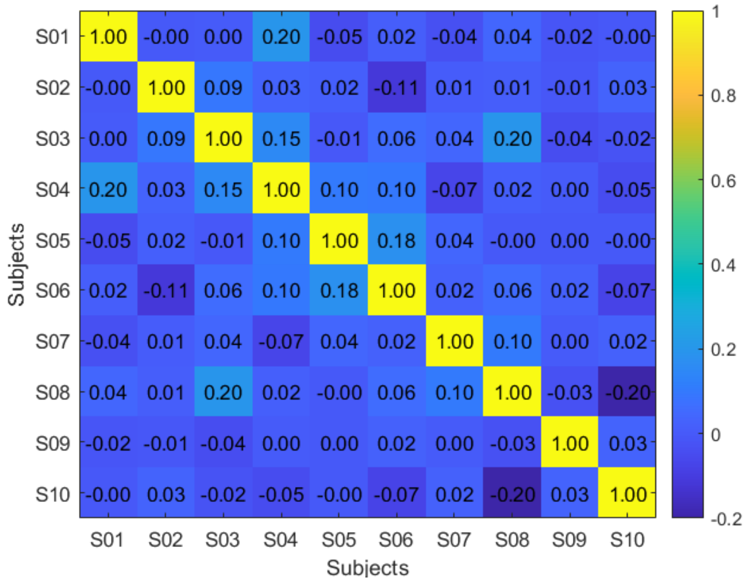

Figure 5.8: Correlation matrix between all subjects for "Ubi Caritas"

As it can be seen, the non-diagonal values of the matrix, corresponding to correlation coefficients between each of the subject pairs, are very low. It can thus be concluded that there is no meaningful correlation when considering the whole temporal domain of the PSD corresponding to the music tracks.

Division into brain regions

Since the locations of the electrodes are known and are the same for all subjects, the channels have been grouped into five regions: Frontal, Central, Occipital, Parietal, and Temporal. The described process was performed for each subject, for each music track and baseline file, and for both alpha and theta bands. Figures 5.9 and 5.10 represent the correlation between the extracted alpha activity of all subjects, for "Ubi Caritas" by Gjeilo, taking into account only those EEG channels that belong to the frontal and temporal regions of the brain, respectively.

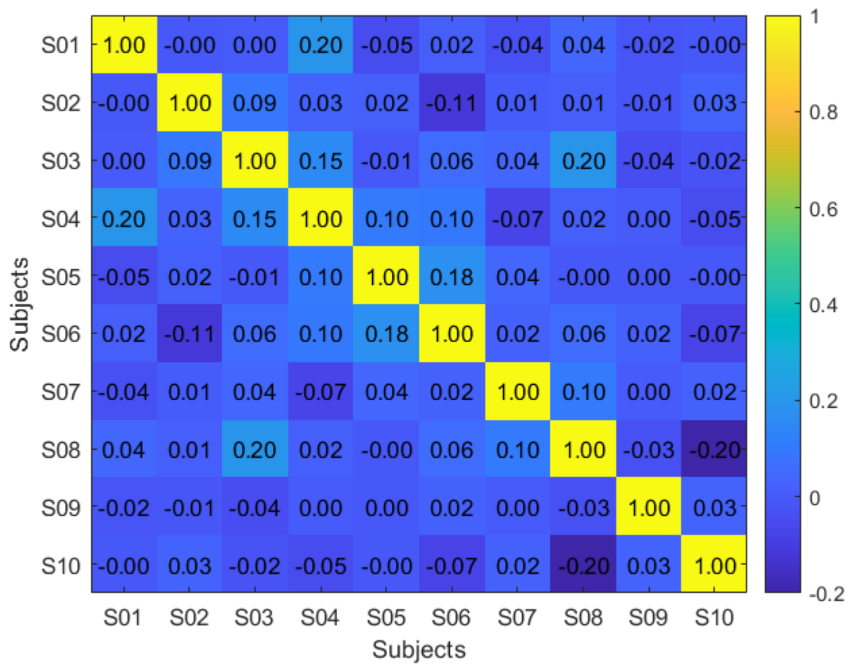

Figure 5.9: Correlation matrix of extracted alpha activity for all subjects for "Ubi Caritas" for the channels belonging to the frontal region

Observing both matrices, it can be noticed that the values of correlation coefficients are still very low; therefore, no meaningful correlation can be extracted between the PSD of the alpha band throughout the whole temporal domain. Considering that EEG signal is highly sensitive to noise and therefore contains a lot of randomness, no significant result could be obtained observing the whole time domain.

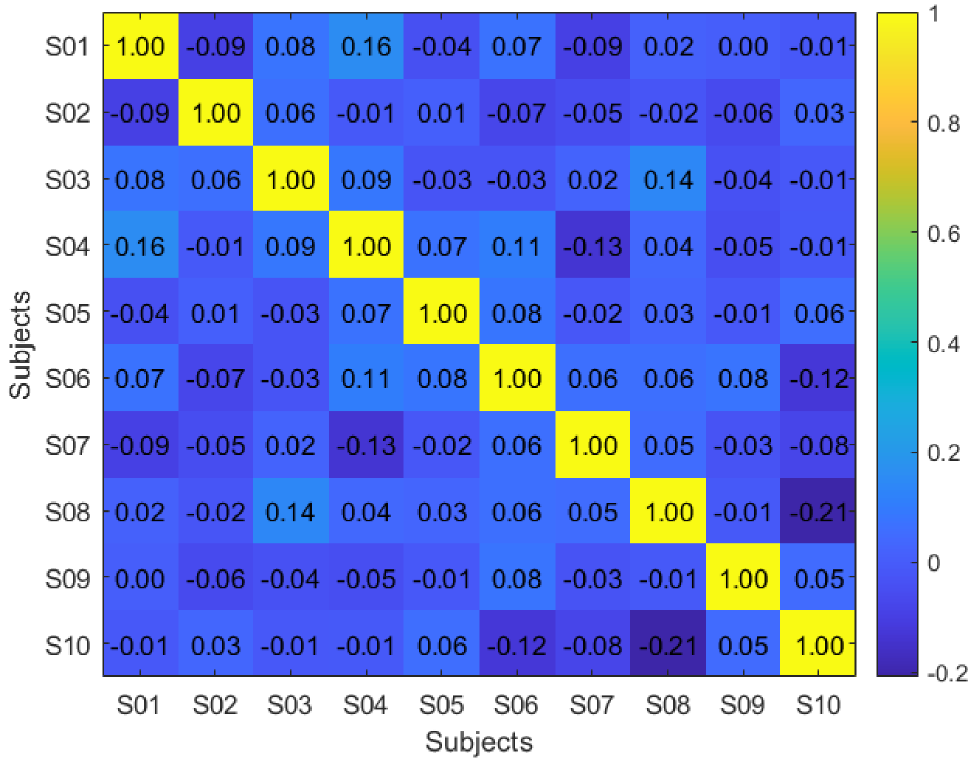

Figure 5.10: Correlation matrix between all subjects for "Ubi Caritas" for the channels belonging to the temporal region

Then, a choice was made to explore correlation between the subjects not for the whole temporal domain, but for shorter segments, using a moving window to split the entire tracks into shorter sections. In this case, taking a window that is 5 seconds long, the correlation matrix has more positive values, as shown in Figure 5.11. Therefore, such analysis in shorter segments could bring more interesting results.

However, in that case, the analysis would include working with high-dimensional matrices, which are difficult to visually represent, as well as to process numerically. Thus, another way of representing the correlation corresponding to temporal segments had to be found.

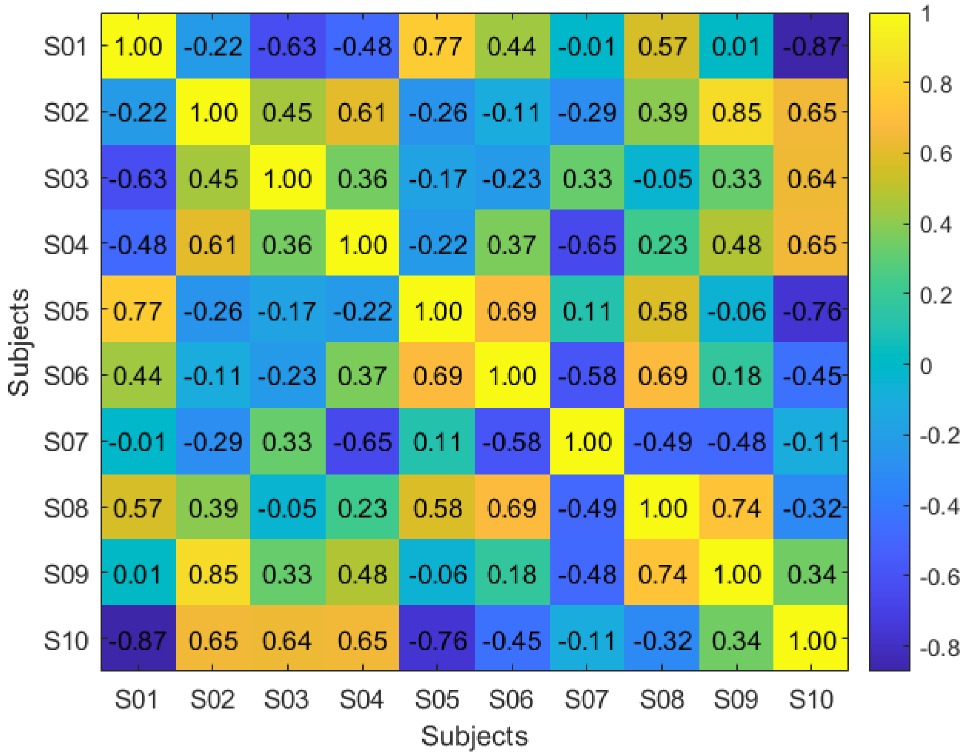

Figure 5.11: Correlation matrix between all subjects for "Intellige Clamorem Meum" in a window corresponding to the range between 130 and 135s of the track

Correlation between the temporal segments of interest

Considering that a correlation matrix is symmetrical - that is, the values above and under the main diagonal are identical - a measure to represent could be taken by averaging or summing those values and therefore having only one number representing the correlation matrix corresponding to a certain window of interest. Some of the possible measures to represent the upper-diagonal values as one number are to transform the matrix into a 1-D array and to compute its statistical measures, such as the mean, average, or standard deviation. Another interesting measure could be the number of correlation coefficients in the array that are greater than a certain value. Following this reasoning, Figure 5.12 shows the average values of the correlation matrices across the whole temporal domain of "Abide with me", while Figure 5.13 represents the number of correlation coefficients that larger than 0.5, for the same track. The window length used in this case is 3 seconds.

The segments that are interesting to consider are the windows where the correlation reaches certain value; thus, the segments where there is certain similarity between EEG signals of all subjects. The next step would be exploring whether the found similarity corresponds to a certain musical event in the considered track.

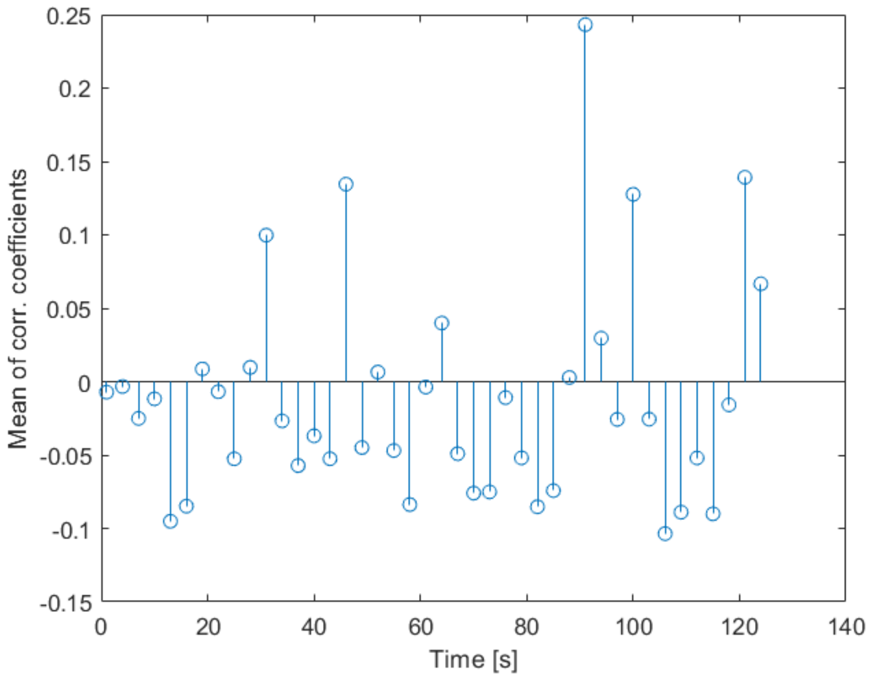

Figure 5.12: The average values of the upper-diagonal cells of the correlation matrices across the whole temporal domain of "Abide with me"

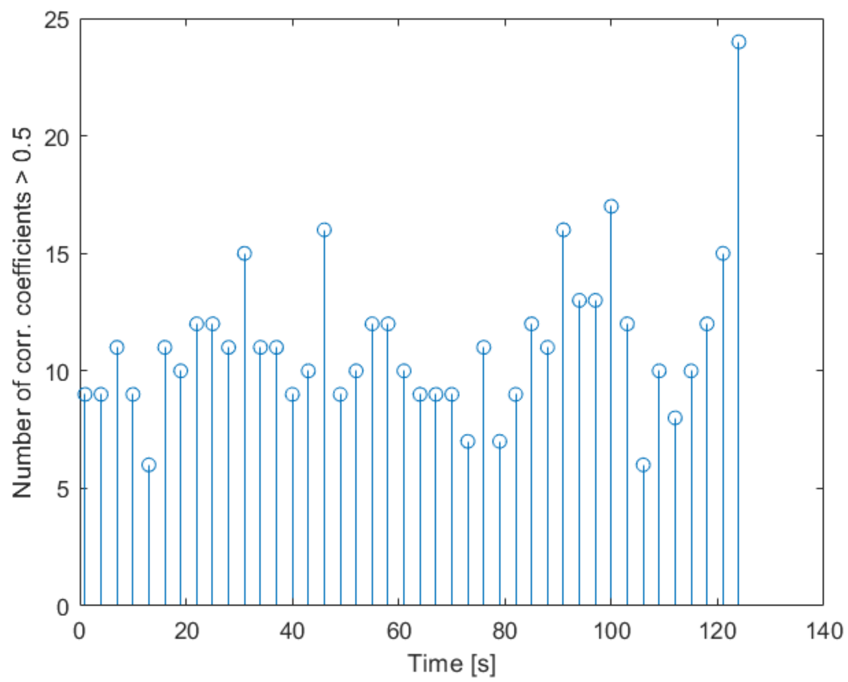

Figure 5.13: The number of corr. coefficients that are larger than 0.5 across the whole temporal domain of "Abide with me"

Both figures have several peaks in common, the highest of which correspond to the windows ranging from 31 to 34, 46 to 49, and 91 to 94 seconds of the track.

Therefore, the correlation matrices corresponding to these segments are shown in Figures 5.14, 5.15, and 5.16, respectively.

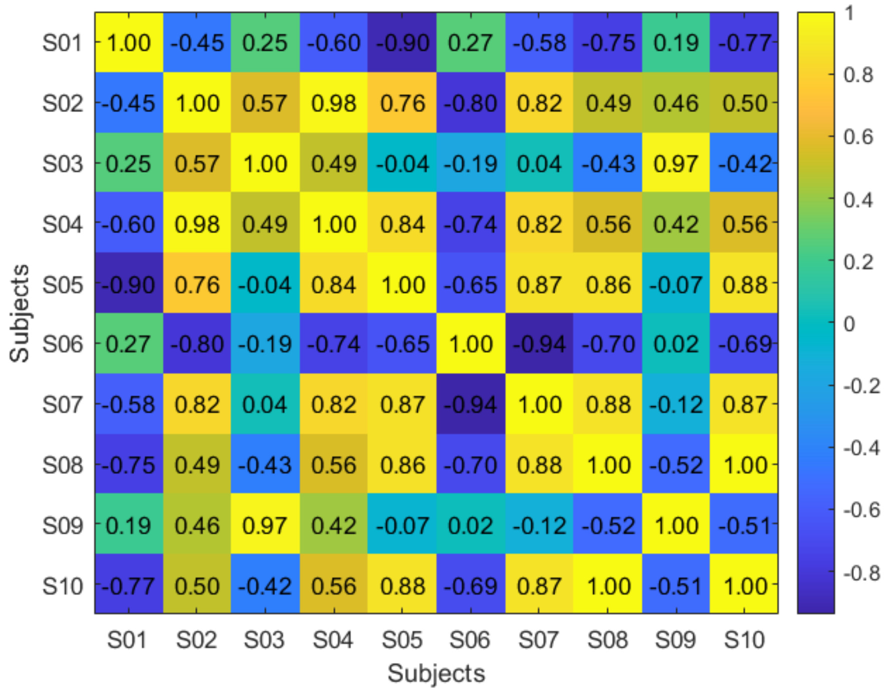

Figure 5.14: Correlation matrix between all subjects for "Abide with me" in a window ranging from 31 to 34 seconds of the track

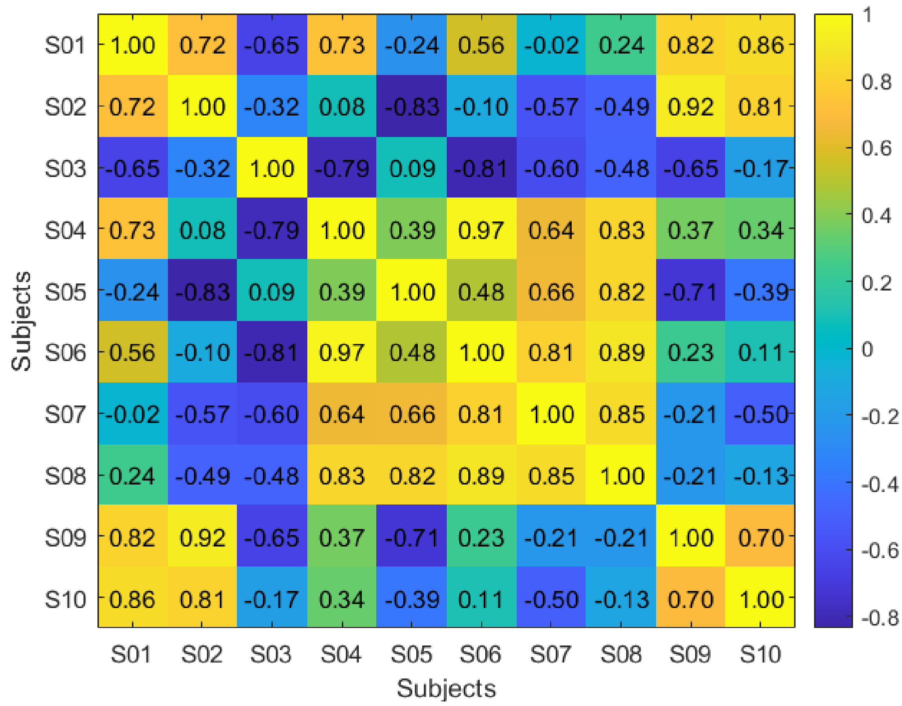

Figure 5.15: Correlation matrix between all subjects for "Abide with me" in a window ranging from 46 to 49 seconds of the track

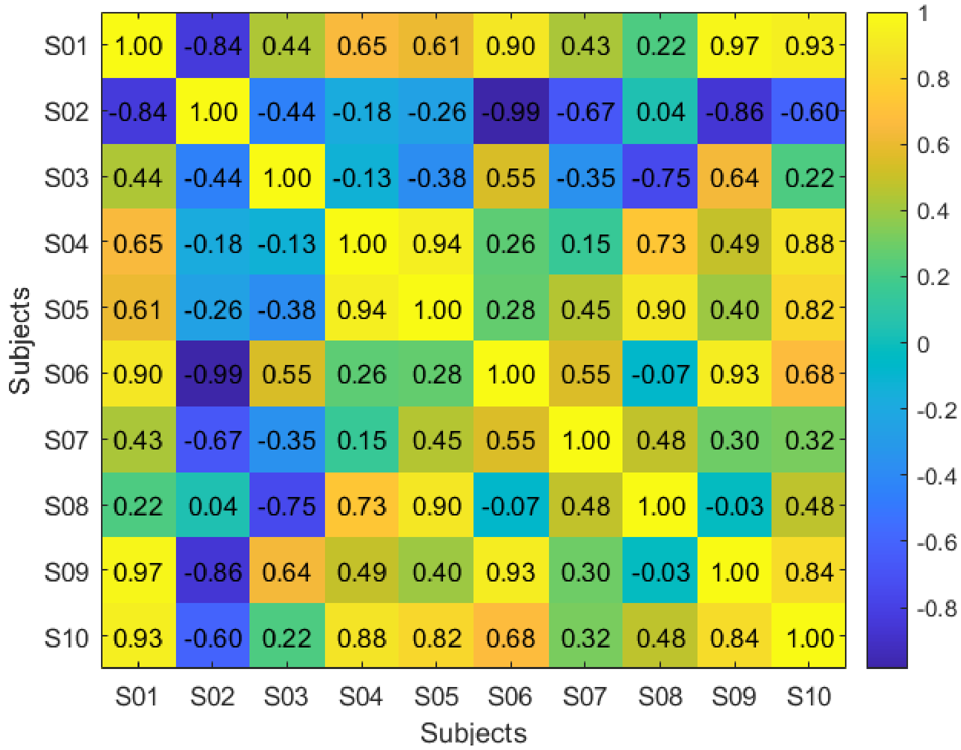

Figure 5.16: Correlation matrix between all subjects for "Abide with me" in a window ranging from 91 to 94 seconds of the track

It can be noticed that the number of positive correlation coefficients is significantly larger than in previously presented correlation matrices, which means that computing the correlation between EEG signals of subjects gives more informative results when shorter segments are considered. Now, it is interesting to compare the time position of these segments with the annotated tension of "Abide with me". The tension levels annotated for "Abide with me" (described in detail in Subsection 4.3.2) are presented in Table 5.1. Observing the Table, it can be noticed that the highest levels of annotated tension, marked with color, corresponds to the temporal segments ranging from 32.27 to 36.7 seconds, and from 88.22 to 91.5 seconds. Therefore, in this case, it can be stated that the largest values of correlation correspond to the highest levels of annotated tension.

Latency [s]	Harmonic tension [1-5]
0	1
5.35	0
7.70	2
8.86	2
10.66	0
15.34	0
17.05	2
19.26	2
20.03	1
22.70	1
23.47	2
26.73	1
30.74	1
32.27	3
35.83	3
36.70	2
39.87	1
50.64	0
59.48	0
62.01	2
63.58	2
64.97	0
71.70	0
73.16	2
74.86	2
77.09	1
78.08	1
81.32	2
85.65	1
88.22	1
90.32	3
91.50	2
94.36	1
105.94	0

Table 5.1: The annotations of the harmonic tension for the piece "Abide with me"

5.1.3. Correlation with music tension

As each music track is assigned numerically annotated tension, another direction to explore the similarity between the subjects' EEG signals is to start from the time intervals that correspond to the certain level of tension and compute the correlation between subjects for those intervals. Figures 5.17 and 5.18 represent the correlation matrices corresponding to the segments of highest level of tension in the track "Abide with me".

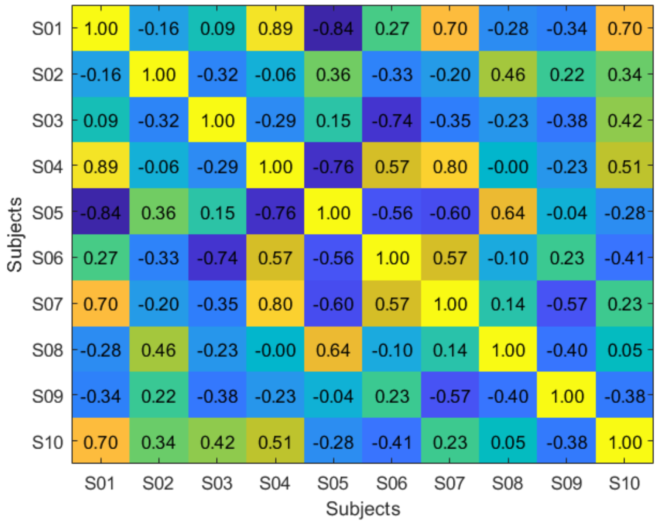

Figure 5.17: Correlation matrix between all subjects for "Abide with me" for all channels, in the first segment of the highest annotated tension (interval 33-36 seconds).

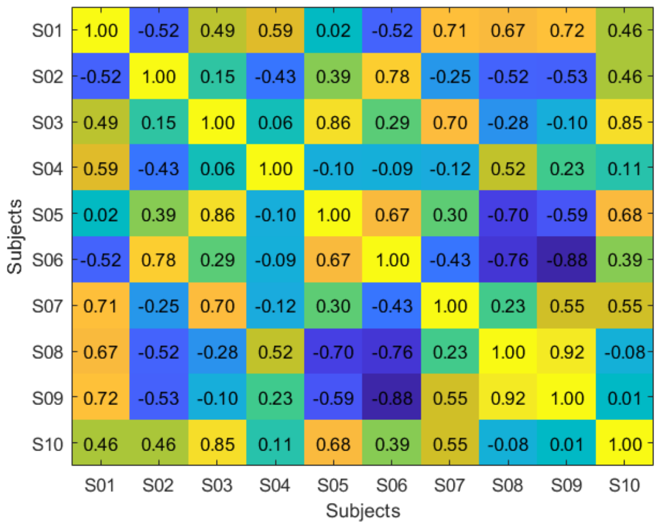

Figure 5.18: Correlation matrix between all subjects for "Abide with me" for all channels, in the second segment of the highest annotated tension (interval 88-91 seconds).

5.2. EDA

Upon visual evaluation of the acquired EDA signals, those corresponding to subject S09 were excluded. As it can be noticed in Figure 5.19, the signal is corrupted, the reason for which could be the low conductivity between the sensor and the skin during the data acquisition. Therefore, the EDA signals corresponding to the subject S09 were excluded from the study.

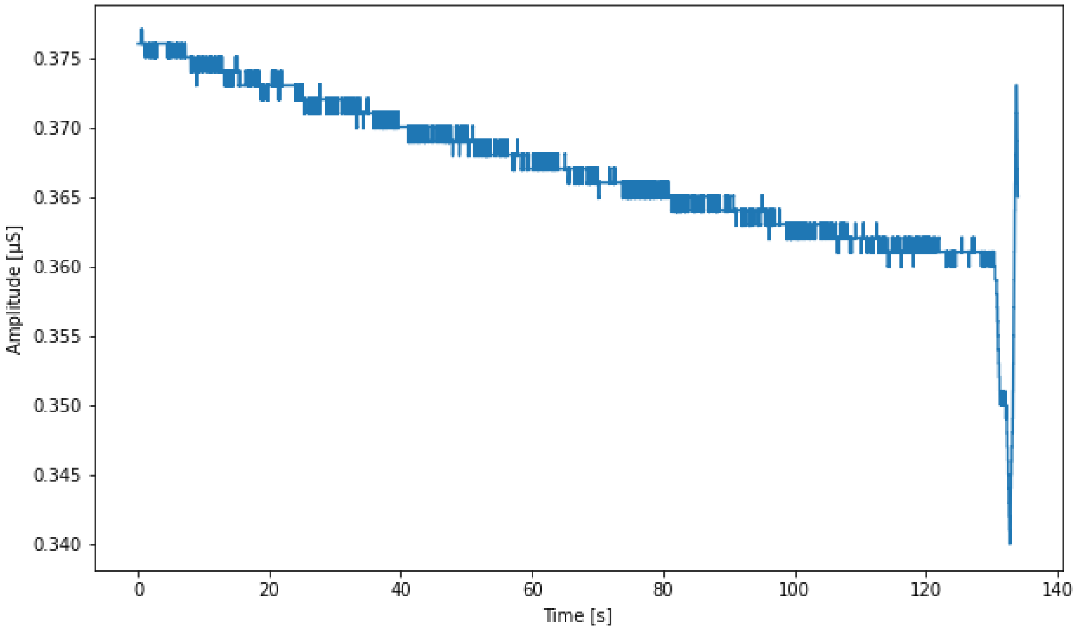

Figure 5.19: Bad EDA signal for subject S09

Since EDA signals are considered "slow", they might not be able to catch responses during continuous or overlapping stimuli. Several features described in Chapter 4 were extracted, including the SCR onsets, SCR peaks, SCR amplitude, and SCR rise time. Due to not having observed significant results in preliminary visual inspection of the signals, and due to the time limitations, it was not proceeded further with the analysis of EDA signals.

5.3. ECG

According to the reasons provided in Subsection 2.3.2, the analysis of ECG signals was mainly based on extracting long-term features, taking into account the whole temporal domain of a signal.

Peak detection

In order to extract HR and HRV features, the R-peaks have to be detected, as described in 2.3.2. An example of a successful peak detection is represented in Figure 5.20.

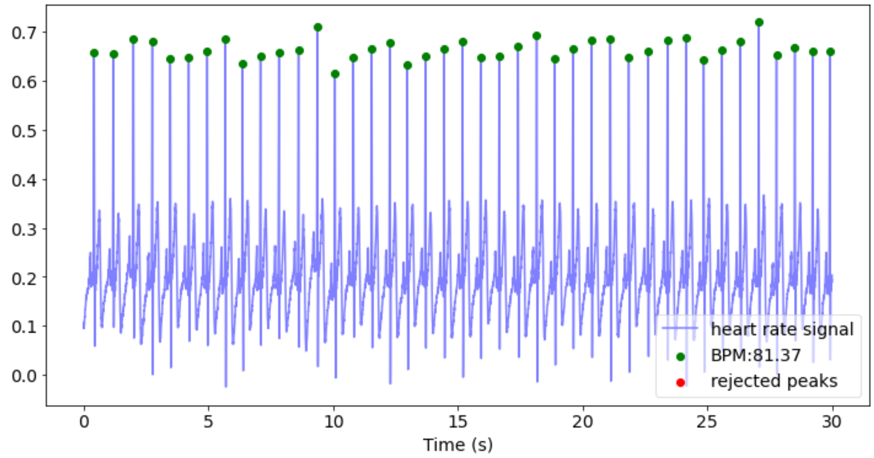

Figure 5.20: Successful peak detection for "Intellige Clamorem Meum", subject S08.

The peak detection was unsuccessful for signals corresponding to some subjects or some tracks due to noisy signal, example of which are shown in Figure 5.21. According to this, the ECG data corresponding to subjects S07 and S10 were excluded from the analysis.

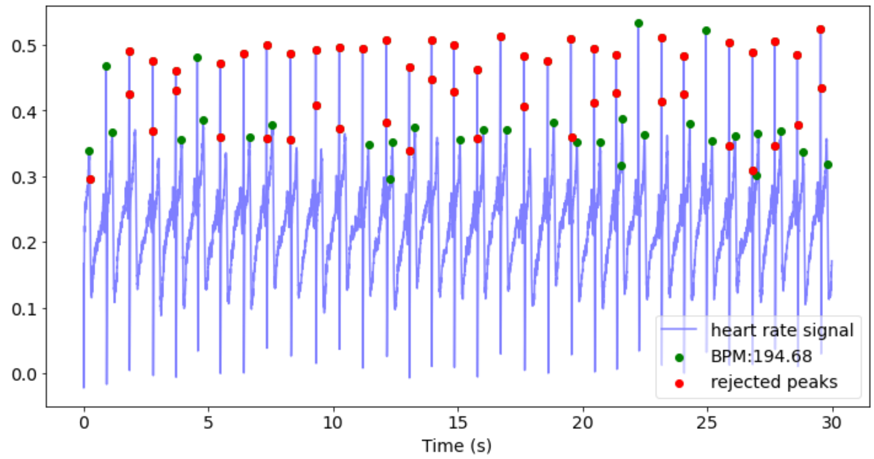

Figure 5.21: Unsuccessful peak detection for "Intellige Clamorem Meum", subject S10.

Measures of HR and HRV

The relevant features for the heart rate and the heart rate variability were extracted for each song and each subject. An example of the features for the track "Abide with me" is shown in Figure 5.22.

Subject	BPM	IBI	SDNN	SDSD	RMSSD
S01	63.22	949.05	70.43	49.04	96.06
S02	77.18	777.38	32.74	14.74	24.6
S03	84.45	710.52	40.89	19.22	29.27
S04	95.25	629.9	22.96	6.77	10.44
S05	69.53	862.98	61.12	17.12	31.91
S06	55.8	1075.22	71.08	44.13	86.09
S08	82.31	728.96	33.72	14.12	23.3
S09	71.71	836.69	43.76	24.74	42.18

Figure 5.22: Heart rate (HR) and heart rate variability (HRV) features extracted for the track "Abide with me"

5.4. RSP

The RRV (respiratory rate variability) features were extracted for each track and each subject. An example of the extracted features for the track "Abide with me" is given in Figure 5.23.

Subject	"Abide with me"						
	RMSSD	SDBB	SDSD	CVBB	CVSD	MadBB	LF
S01	283.46	219.72	287.54	0.074	0.095	118.72	0.005
S02	660.44	442.50	670.52	0.144	0.214	237.45	0.010
S03	766.78	734.99	782.68	0.175	0.183	897.67	0.000
S04	1,082.07	865.25	1,100.93	0.242	0.302	610.99	0.002
S05	624.79	441.50	634.73	0.141	0.200	318.53	0.003
S06	219.19	185.29	222.26	0.056	0.066	144.79	0.001
S07	762.46	739.08	774.23	0.177	0.183	364.86	0.001
S08	325.31	259.94	330.50	0.078	0.098	191.12	0.001
S09	263.24	191.17	268.06	0.053	0.073	121.62	0.001
S10	210.81	198.71	213.70	0.073	0.078	124.52	0.003

Figure 5.23: The extracted RRV features for "Abide with me".

5.5. Questionnaires

5.5.1. Gold-MSI

Responses of all 10 subjects to the Gold-MSI questionnaire are presented in Figure 5.24, where measures AE, PA, MT, SA, EM, and GM stand for Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, Emotions, and General Musical Sophistication, respectively.

Subject	AE_mean	PA_mean	MT_mean	SA_mean	EM_mean	GM_mean
S01	4.11	5.22	4.00	5.71	5.83	5.06
S02	5.78	6.78	6.29	7.00	5.67	6.44
S03	5.56	5.78	5.71	5.29	6.50	5.67
S04	4.33	6.00	4.29	6.14	6.17	5.22
S05	5.33	6.22	5.43	5.43	6.00	5.28
S06	4.00	5.33	2.00	2.71	4.50	2.94
S07	4.11	5.89	3.86	3.57	6.50	4.17
S08	4.22	4.89	4.14	3.43	5.67	4.28
S09	5.22	4.33	2.43	5.00	4.17	4.50
S10	5.33	5.67	4.86	3.29	6.67	4.89

Figure 5.24: The responses to Gold-MSI questionnaire. The darker colors correspond to larger values.

Correlation with EEG features

Figures 5.25 and 5.26 represent the correlation between Gold-MSI responses and EEG features - precisely, mean and standard deviation for alpha and theta bands - for each of the choral tracks, for all 10 subjects. Figure 5.27 represents the same correlations averaged across all choral tracks, for all 10 subjects.

Subject	"Abide with me"				"Intellige Clamorem Meum"			
	Mean θ	Mean α	STD θ	STD α	Mean θ	Mean α	STD θ	STD α
S01	0.25	0.18	0.15	0.14	0.23	0.19	0.13	0.13
S02	0.12	0.48	0.09	0.31	0.13	0.47	0.09	0.30
S03	0.21	0.16	0.13	0.10	0.22	0.18	0.14	0.12
S04	0.15	0.47	0.10	0.29	0.15	0.55	0.10	0.28
S05	0.24	0.26	0.14	0.15	0.23	0.26	0.14	0.15
S06	0.22	0.18	0.13	0.14	0.24	0.16	0.13	0.10
S07	0.17	0.21	0.10	0.13	0.17	0.24	0.10	0.14
S08	0.14	0.25	0.09	0.17	0.14	0.20	0.09	0.14
S09	0.16	0.25	0.10	0.16	0.10	0.15	0.08	0.13
S10	0.42	0.49	0.25	0.24	0.36	0.53	0.25	0.24

Corr. with								
AE	0.15	0.39	0.19	0.29	0.08	0.27	0.21	0.37
PA	- 0.02	0.49	0.01	0.49	0.17	0.59	0.15	0.59
MT	0.07	0.43	0.12	0.36	0.15	0.48	0.20	0.50
SA	- 0.39	0.33	- 0.32	0.44	- 0.40	0.32	- 0.35	0.50
EM	0.37	0.28	0.39	0.13	0.46	0.48	0.46	0.31
GM	- 0.09	0.48	- 0.01	0.48	- 0.10	0.48	- 0.01	0.59

Figure 5.25: Correlation between Gold-MSI and EEG features, tracks 2 and 3

Subject	"Ubi Caritas" by Duruflé				"Ubi Caritas" by Gjeilo			
	Mean θ	Mean α	STD θ	STD α	Mean θ	Mean α	STD θ	STD α
S01	0.24	0.17	0.15	0.10	0.26	0.17	0.14	0.10
S02	0.11	0.51	0.08	0.32	0.11	0.56	0.09	0.33
S03	0.23	0.15	0.14	0.09	0.22	0.15	0.14	0.10
S04	0.14	0.54	0.09	0.30	0.17	0.50	0.12	0.30
S05	0.26	0.29	0.17	0.17	0.26	0.30	0.15	0.17
S06	0.21	0.14	0.12	0.10	0.23	0.18	0.13	0.14
S07	0.14	0.36	0.09	0.23	0.16	0.26	0.10	0.15
S08	0.15	0.24	0.09	0.15	0.15	0.25	0.09	0.16
S09	0.14	0.25	0.10	0.17	0.14	0.21	0.09	0.17
S10	0.43	0.53	0.26	0.23	0.44	0.52	0.26	0.23

Corr. with								
AE	0.22	0.29	0.29	0.26	0.11	0.38	0.23	0.34
PA	0.02	0.54	0.06	0.57	0.01	0.60	0.12	0.54
MT	0.14	0.41	0.19	0.39	0.08	0.48	0.20	0.37
SA	- 0.35	0.28	- 0.28	0.40	- 0.38	0.34	- 0.27	0.44
EM	0.38	0.41	0.37	0.27	0.38	0.34	0.45	0.12
GM	- 0.04	0.44	0.03	0.46	- 0.09	0.50	0.04	0.47

Figure 5.26: Correlation between Gold-MSI and EEG features, tracks 4 and 5

Subject	All tracks averaged			
	Mean θ	Mean α	STD θ	STD α
S01	0.24	0.18	0.14	0.12
S02	0.12	0.50	0.09	0.31
S03	0.22	0.16	0.13	0.10
S04	0.15	0.52	0.10	0.29
S05	0.25	0.28	0.15	0.16
S06	0.23	0.16	0.13	0.12
S07	0.16	0.26	0.10	0.16
S08	0.14	0.23	0.09	0.15
S09	0.13	0.21	0.09	0.16
S10	0.41	0.52	0.25	0.24

Corr. with	AE	0.14	0.34	0.23	0.33
	PA	0.04	0.57	0.09	0.56
	MT	0.11	0.46	0.18	0.42
	SA	- 0.38	0.32	- 0.31	0.46
	EM	0.40	0.39	0.42	0.21
	GM	- 0.08	0.48	0.01	0.51

Figure 5.27: Correlation between Gold-MSI and EEG features

It can be seen from the figures that relatively significant correlation exists between the feature Perceptual Abilities and the alpha activity of the subjects (values of 0.57 and 0.56 for the mean and the standard deviation of the alpha activity, respectively), as well as between the General Musical Sophistication factor and the alpha activity (values of 0.48 and 0.51 for the mean and the standard deviation of the alpha activity, respectively).

Correlation with ECG features

Figure 5.28 represents the correlation between Gold-MSI responses and ECG features, averaged across all choral tracks, for 8 subjects.

Subject	All tracks averaged				
	BPM	IBI	SDNN	SDSD	RMSSD
S01	66.91	899.2	69.14	40.76	79.12
S02	84.04	731.6	54.85	19.76	118.2
S03	82.41	728.8	35.53	14.73	23.44
S04	93.36	642.8	27.93	8.353	12.61
S05	68.92	871.1	60.92	22.05	37.19
S06	57.87	1038	64.47	37.77	71.52
S08	81.4	737.1	31.28	14.1	24.46
S09	71.94	834	48.95	26.32	44.85

Corr. with	AE	0.29	- 0.34	- 0.05	- 0.36	0.18
	PA	0.36	- 0.30	0.05	- 0.32	0.31
	MT	0.56	- 0.59	- 0.19	- 0.52	0.10
	SA	0.55	- 0.56	- 0.03	- 0.29	0.26
	EM	0.57	- 0.59	- 0.36	- 0.52	- 0.28
	GM	0.63	- 0.67	- 0.17	- 0.46	0.19

Figure 5.28: Correlation between Gold-MSI and ECG features

When it comes to correlation between Gold-MSI responses and the ECG features, positive correlation can be observed between the BPM (beats per minute) and the Gold-MSI features MT, SA, EM, and GM, with values of 0.56, 0.55, 0.57, and 0.63, respectively. Additionally, negative correlation is present between IBI (Interbeat interval) and the same Gold-MSI features. Since the average values of IBI are also negatively correlated with average values of BPM (with a correlation coefficient equal to -0.988), this result makes sense. Not that strong negative correlation exists also between SDSD (standard deviation of the differences between successive NN (or RR) intervals) and the Gold-MSI features MT, EM, and GM.

Correlation with RRV features

Figure 5.29 represents the correlation between Gold-MSI responses and RRV features, averaged across all choral tracks.

Subject	All tracks averaged						
	RMSSD	SDBB	SDSD	CVBB	CVSD	MadBB	LF
S01	240.2	194.5	243.0	0.07	0.08	141.2	0.004
S02	692.3	507.1	701.3	0.16	0.22	346.0	0.005
S03	503.3	441.0	510.5	0.12	0.14	391.6	0.004
S04	1,399.9	1,152.0	1,418.2	0.31	0.38	681.9	0.002
S05	463.7	345.7	469.1	0.11	0.15	293.9	0.005
S06	221.8	188.8	224.4	0.06	0.07	162.9	0.002
S07	666.4	564.2	675.7	0.14	0.16	347.5	0.001
S08	396.6	320.8	401.7	0.10	0.12	235.3	0.001
S09	401.8	279.3	407.0	0.08	0.11	204.9	0.003
S10	239.2	269.0	241.6	0.09	0.08	221.5	0.002

Corr. with	**AE**	- 0.04	- 0.07	- 0.04	0.00	0.04	0.08	0.62
	PA	0.44	0.44	0.44	0.49	0.49	0.50	0.34
	MT	0.24	0.24	0.24	0.32	0.32	0.36	0.55
	SA	0.51	0.43	0.51	0.50	0.58	0.46	0.69
	EM	0.28	0.36	0.28	0.38	0.28	0.42	0.02
	GM	0.36	0.33	0.36	0.41	0.44	0.41	0.68

Figure 5.29: Correlation between Gold-MSI and RRV features

The last figure shows that the largest values of correlation coefficients are between power spectral density of the low frequency band (LF) and the Gold-MSI responses, in particular, the AE, MT, SA, and GM, with values 0.62, 0.55, 0.69, and 0.68, respectively.

5.5.2. SREIT

Responses of all 10 subjects to the SREIT (Schutte Self-Report Emotional Intelligence Scale) questionnaire, described in 3.2.2, are shown in Figure 5.30.

Subject	SREIT
S01	129
S02	126
S03	102
S04	113
S05	131
S06	106
S07	132
S08	115
S09	123
S10	149

Figure 5.30: SREIT results

Correlation with EEG features

Figure 5.31 represents the correlation between SREIT values and EEG features.

Subject	All tracks averaged			
	Mean θ	Mean α	STD θ	STD α
S01	0.24	0.18	0.14	0.12
S02	0.12	0.50	0.09	0.31
S03	0.22	0.16	0.13	0.10
S04	0.15	0.52	0.10	0.29
S05	0.25	0.28	0.15	0.16
S06	0.23	0.16	0.13	0.12
S07	0.16	0.26	0.10	0.16
S08	0.14	0.23	0.09	0.15
S09	0.13	0.21	0.09	0.16
S10	0.41	0.52	0.25	0.24

Corr. with	SREIT				
		0.52	0.45	0.57	0.31

Figure 5.31: Correlation between SREIT and EEG features

The correlation coefficients between SREIT responses and the EEG features show that a positive correlation exists between the SREIT values and the EEG theta activity - specifically, the mean and the standard deviation, with values of 0.52 and 0.57, respectively.

Correlation with ECG features

Figure 5.32 represents the correlation between SREIT values and ECG features.

Subject	All tracks averaged				
	BPM	IBI	SDNN	SDSD	RMSSD
S01	66.91	899.2	69.14	40.76	79.12
S02	84.04	731.6	54.85	19.76	118.2
S03	82.41	728.8	35.53	14.73	23.44
S04	93.36	642.8	27.93	8.353	12.61
S05	68.92	871.1	60.92	22.05	37.19
S06	57.87	1038	64.47	37.77	71.52
S08	81.4	737.1	31.28	14.1	24.46
S09	71.94	834	48.95	26.32	44.85

Corr. with	SREIT	- 0.16	0.09	0.51	0.26	0.41

Figure 5.32: Correlation between SREIT and ECG features

In the last figure, it can be noticed that only SDNN is relatively strongly positively correlated with the SREIT values, the correlation coefficient being equal to 0.51.

Correlation with RRV features

Figure 5.33 represents the correlation between SREIT responses and the RRV (respiratory rate variability) features - averaged across all choral tracks.

Subject	All tracks averaged						
	RMSSD	SDBB	SDSD	CVBB	CVSD	MadBB	LF
S01	240.2	194.5	243.0	0.07	0.08	141.2	0.004
S02	692.3	507.1	701.3	0.16	0.22	346.0	0.005
S03	503.3	441.0	510.5	0.12	0.14	391.6	0.004
S04	1,399.9	1,152.0	1,418.2	0.31	0.38	681.9	0.002
S05	463.7	345.7	469.1	0.11	0.15	293.9	0.005
S06	221.8	188.8	224.4	0.06	0.07	162.9	0.002
S07	666.4	564.2	675.7	0.14	0.16	347.5	0.001
S08	396.6	320.8	401.7	0.10	0.12	235.3	0.001
S09	401.8	279.3	407.0	0.08	0.11	204.9	0.003
S10	239.2	269.0	241.6	0.09	0.08	221.5	0.002

Corr. with	SREIT	- 0.23	- 0.21	- 0.23	- 0.16	- 0.20	- 0.27	0.05

Figure 5.33: Correlation between SREIT and RRV features

According to the last figure, no significant correlation can be noticed between any of the RRV features and the SREIT values.

5.5.3. STOMP-R

The elaboration of responses to STOMP-R (Revised Short Test of Music Preferences) questionnaire, in the form of MUSIC model (standing for Mellow, Urban, Sophisticated, Intense, and Campestral music) is shown in Figure 5.34. Due to not having a standardized evaluation model, and due to the limited number of subjects, it was not proceeded with the analysis including the responses to STOMP-R questionnaire.

Subject	M	U	S	I	C
S01	8	18	28	16	18
S02	17	16	24	25	25
S03	15	17	32	24	24
S04	15	12	24	17	18
S05	15	12	26	19	17
S06	16	11	33	26	24
S07	11	10	23	27	20
S08	13	14	27	25	18
S09	18	14	26	20	24
S10	10	24	23	13	13

Figure 5.34: The elaboration of responses to STOMP-R according to the MUSIC model. The darker colors correspond to larger values.

6 | Conclusions and future work

This chapter contains the discussion and conclusions of this research, as well as suggestions for future developments.

6.1. Summary of the study

The goal of this study was to explore the correlation between music features and induced emotions using various physiological signals in a listening experiment, supported by self-report questionnaires. The listening experiment was designed and conducted with 10 recruited subjects, in laboratories of Politecnico di Milano (Brain Lab) in December 2021. Detailed research has previously been done on the state of the art to understand what methods are commonly used and the most fitting for the nature of the study, which was focused on harmonic tension in music pieces. The chosen physiological signals to acquire were electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), and respiratory activity (RSP). Additionally, the subjects provided responses to several self-report questionnaires, aiming to collect data about their lifestyle, musical expertise and preferences, and emotional competencies, as well we the emotions they feel while listening to the music stimuli. After acquiring the signals for all 10 subjects, the data was organized and preprocessed according to the literature. The analysis of various features was performed, exploring the correlation between the manually annotated tension in music and the responses captured by the recorded physiological data. The correlation between subjects' responses was also analyzed, as well as the differences between statistical measures of each subject. The pipelines for signal preprocessing and feature extraction for each of the physiological signals was proposed, with suggestions for improvement.

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, considering the limitations on the number of subjects, 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. 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.

6.2. Contributions

The main contributions of this work are the created experimental protocol, the acquired datasets containing data for four physiological signals for 10 subjects, as well as their responses to self-report questionnaires, then the proposed pipeline for data processing and analysis, and the preliminary results. The experimental protocol described in Chapter 4 contains detailed instructions for the experimenter that could be used in the continuation of this study or for one exploring a similar topic.

The whole dataset can be found using the following hyperlinks:

- EEG signals: `https://1drv.ms/u/s!AvcHs1ydRWzggZ4t3R9FFhLezMROPg?e=yJB7j2`
- EDA, RSP, and ECG signals: `https://1drv.ms/u/s!AvcHs1ydRWzggZ4bsbStszR3FrAMSw?e=FK9Y80`

6.2.1. Subjects' feedback

All 10 subjects provided some feedback on the procedure and the task they had to perform. They reported that the experiment was clear, well explained, and well organized, and that the equipment setup was not too uncomfortable. Related to the task, some subjects reported that the part with chord sequences was long, not musically engaging, and therefore tiring to listen to, especially not being allowed to move. On the other hand, they noted that the choral pieces were very similar to each other and suggested that a study with more diverse examples could be more interesting. The subjects liked filling out the self-report questionnaires and showed interest for a deeper understanding of their theoretical background, their use in this study, and the obtained results.

6.3. Criticism and suggestions for improvement

6.3.1. Comments on project management

As the study involved several departments and universities, putting together the whole project was challenging from the organizational point of view. Since numerous aspects depended on the University personnel, for instance acquiring the needed technical equipment, receiving the approval from the ethical committee, or organizing the acquisition sessions, the whole process took longer than expected. The planning should therefore be more realistic, taking into consideration possible delays, the human factor, and the whole project's complexity. Additionally, if the student performs such an experiment for the first time, it is advised that a mentor is physically present at least during the first acquisition sessions, to educate, monitor, and support them.

6.3.2. Comments on the experimental protocol

As this study aimed to conduct preliminary assessment with limited resources, only 10 subjects were recruited. This represents a drawback in analyzing data and drawing conclusions, especially when working with such noisy signals as EEG, and when aiming to extract relevant statistics based on inter-subjects analysis. For the possible future work, according to the state of the art, at least 30 subjects would be sufficient to obtain significant results. Additionally, the selected group of students could be more diverse, in terms of age, nationality, but also occupation (only university students participated in the data collection process).

The experimental setup, even if not using invasive equipment, was complex, and therefore it took a lot of time to acquire data for each participant, including setting up the equipment and cleaning it afterwards. On average, it took three to four hours of the experimenter's work per subject. It also took time and assistance of a mentor to get educated and gain practice using the equipment and properly setting up the whole experimental protocol.

6.3.3. Comments on music stimuli

Since this study started as a continuation of the "Sound Resonance Project", described in details in Section 1.3, the set of music stimuli consisting of polyphonic choral music pieces was predefined. The advantage of having choral pieces recorded in a live performance is the novelty that it brings to the field, since the music used in the state of the art mainly belongs to genres of classical instrumental or popular music. On the other hand, the drawback of this dataset is that the pieces are similar to each other and that the annotations of music tension are not easily and objectively measurable. 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, for instance drastic changes in dynamics, rhythm, and similar. Additionally, having more diverse pieces would be useful in terms of exploring the variety of induced emotions.

According to the state of the art, the stimuli should be repeated a few times (especially the short stimuli), in order to increase the signal-to-noise ration. Considering the analysis related to specific temporal events, extracting features based on Event Related Potential (ERP) of EEG could provide interesting results, which in this case was not applicable due to low number of repetitions. Another prerequisite for ERP analysis is having a resting state (baseline) right before the stimuli, which in the case of continuous music pieces or chords progressions is not possible. More exploration on the topic could be done.

6.4. Future work

Regarding the continuation of this study, it would be interesting to explore more features and connections between features for all collected data. The EEG signals could be used in combination with fMRI or fNIRS to obtain information on spatial distribution of elicited responses. After acquiring data for more subjects, possibly even for different music stimuli, some machine learning methods could be used, such as supervised learning. Classification methods could be additionally used, classifying patterns of acquired signals into one of predefined classes, such as various elicited emotions.

With these improvement at place - on a longer timescale - the development of more advanced affective computing systems would become feasible. These systems could be used in healthcare, therapeutic fields, entertainment, marketing, education and other fields where it is desired to monitor or affect the emotional state and responses of individuals.

Bibliography

[1] Vital signs 101. https://www.hopkinsmedicine.org/. Last visited on 02/07/2022.

[2] Acq*Knowledge*. https://www.biopac.com/knowledge-base/phasic-eda-issue/. Last visited on 02/07/2022.

[3] D. M. Alexander, C. Trengove, P. Johnston, T. Cooper, J. August, and E. Gordon. Separating individual skin conductance responses in a short interstimulus-interval paradigm. *Journal of neuroscience methods*, 146(1):116–123, 2005.

[4] M. Alomari, E. Awada, and O. Younis. Subject-independent EEG-based discrimination between imagined and executed, right and left fists movements. *European Journal of Scientific Research*, 118(3):364–373, 2014.

[5] B. Andrea and M. Enrico. OT Bioelettronica SRL. https://www.otbioelettronica.it/en/, 2007. Last visited on 20/06/2022.

[6] M. K. Ang and M. Yeoh. Music preferences of malaysian students and kbsm curriculum implications. *Pertanika Journal of Social Sciences and Humanities*, 10:43–51, 2001.

[7] R. Bailón, L. Sörnmo, P. Laguna, et al. Ecg-derived respiratory frequency estimation. *Advanced methods and tools for ECG data analysis*, 1, 2006.

[8] R. Bar-On. *BarOn emotional quotient inventory*. Multi-health systems, 1997.

[9] R. Bar-On. The bar-on emotional quotient inventory (eq-i): Rationale, description and summary of psychometric properties. 2004.

[10] G. Ben-Shakhar. Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22(3):292–299, 1985.

[11] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010.

[12] D. E. Berlyne. Aesthetics and psychobiology. *Journal of Aesthetics and Art Criticism*, 31(4), 1973.

[13] L. Bernardi, S. Leuzzi, A. Radaelli, C. Passino, J. A. Johnston, and P. Sleight. Low-frequency spontaneous fluctuations of rr interval and blood pressure in conscious humans: a baroreceptor or central phenomenon? *Clinical Science*, 87(6):649–654, 1994.

[14] J. G. Betts, K. A. Young, J. A. Wise, E. Johnson, B. Poe, D. H. Kruse, O. Korol, J. E. Johnson, M. Womble, and P. DeSaix. *Anatomy and physiology*. OpenStax, 2013.

[15] A. M. Bhatti, M. Majid, S. M. Anwar, and B. Khan. Human emotion recognition and analysis in response to audio music using brain signals. *Computers in Human Behavior*, 65:267–275, 2016.

[16] W. Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[17] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1):1017–1034, 2013.

[18] P. Broca. Anatomie comparée des circonvolutions cérébrales. le grand lobe limbique et la scissure limbique dans la série des mammifères. *Rev Anthrop*, 1:385–498, 1978.

[19] M. Cabanac. What is emotion? *Behavioural processes*, 60(2):69–83, 2002.

[20] E. Carlson, P. Saari, B. Burger, and P. Toiviainen. Personality and musical preference using social-tagging in excerpt-selection. *Psychomusicology: Music, Mind, and Brain*, 27(3):203, 2017.

[21] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred, et al. BioSPPy: Biosignal processing in Python, 2015–. URL `https://github.com/PIA-Group/BioSPPy/`.

[22] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. Pimentel, A. E. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton. Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE reviews in biomedical engineering*, 11:2–20, 2017.

[23] I. Cockos. REAPER (Rapid Environment for Audio Production, Engineering, and Recording). `https://www.reaper.fm/`, 2006. Last visited on 20/06/2022.

[24] H. D. Critchley. Electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142, 2002.

[25] F. E. Croxton and D. J. Cowden. Applied general statistics. 1939.

[26] I. Daly, A. Malik, F. Hwang, E. Roesch, J. Weaver, A. Kirke, D. Williams, E. Miranda, and S. J. Nasuto. Neural correlates of emotional responses to music: an eeg study. *Neuroscience letters*, 573:52–57, 2014.

[27] A. R. Damasio. Emotion in the perspective of an integrated nervous system. *Brain research reviews*, 26(2-3):83–86, 1998.

[28] J. B. Davies. The psychology of music. *Journal of Aesthetics and Art Criticism*, 37 (3), 1979.

[29] M. E. Dawson, A. M. Schell, and C. G. Courtney. The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics*, 4(2):111, 2011.

[30] T. deClercq. Corpus studies of harmony in popular music: A response to gauvin. *Empirical Musicology Review*, 10(3):239–244, 2015.

[31] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[32] C. N. DeWall and D. G. Myers. *Psychology in everyday life*. Worth, 4 edition, 2016. ISBN 9781319013738.

[33] M. E. What is tension and release in music? `https://www.schoolofcomposition.com/what-is-tension-and-release-in-music/`, 2022. Last visited on 18/06/2022.

[34] T. Eerola and J. K. Vuoskoski. A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, 30 (3):307–340, 2012.

[35] P. Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[36] P. Ekman and D. Cordaro. What is meant by calling emotions basic. *Emotion review*, 3(4):364–370, 2011.

[37] P. Ekman, W. V. Freisen, and S. Ancoli. Facial signs of emotional experience. *Journal of personality and social psychology*, 39(6):1125, 1980.

[38] P. E. Ekman and R. J. Davidson. *The nature of emotion: Fundamental questions*. Oxford University Press, 1994.

[39] J. L. Farrens, A. M. Simmons, S. J. Luck, and E. S. Kappenman. Electroencephalogram (EEG) recording protocol for cognitive and affective human neuroscience research. 2020.

[40] M. Ferrari and V. Quaresima. A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. *Neuroimage*, 63(2):921–935, 2012.

[41] T. C. Ferree. Spherical splines and average referencing in scalp electroencephalography. *Brain topography*, 19(1):43–52, 2006.

[42] R. Ferrer, T. Eerola, and J. K. Vuoskoski. Enhancing genre-based measures of music preference by user-defined liking and social tags. *Psychology of Music*, 41(4): 499–518, 2013.

[43] P. Fiedler, C. Fonseca, E. Supriyanto, F. Zanow, and J. Haueisen. A high-density 256-channel cap for dry electroencephalography. *Human brain mapping*, 43(4):1295–1308, 2022.

[44] A. Fortin-Côté, N. Beaudin-Gagnon, A. Campeau-Lecours, S. Tremblay, and P. L. Jackson. Affective computing out-of-the-lab: The cost of low cost. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 4137–4142. IEEE, 2019.

[45] D. C. Fowles, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken, and P. H. Venables. Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3):232–239, 1981.

[46] A. Gabrielsson and E. Lindström. The influence of musical structure on emotional expression. 2001.

[47] M. C. Graham, L. Priddy, and S. Graham. *Facts of Life: ten issues of contentment.* Outskirts Press, 2014.

[48] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804, 2015.

[49] A. Greco, G. Valenza, L. Citi, and E. P. Scilingo. Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, 17(3): 716–725, 2016.

[50] R. Görike and E. Pless. AKG Acoustics (Akustische und Kino-Geräte Gesellschaft m.b.H.). https://www.akg.com/, 1947. Last visited on 23/06/2022.

[51] L. F. Haas. Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 2003.

[52] E. Harmon-Jones, C. Harmon-Jones, and E. Summerell. On the importance of both dimensional and discrete models of emotion. *Behavioral sciences*, 7(4):66, 2017.

[53] P. Hartzbech. iMotions. https://imotions.com/, 2007. Last visited on 24/06/2022.

[54] M. A. Hasnul, N. A. A. Aziz, S. Alelyani, M. Mohana, and A. A. Aziz. Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors*, 21(15):5015, 2021.

[55] P. He, G. Wilson, and C. Russell. Removal of ocular artifacts from electroencephalogram by adaptive filtering. *Medical and biological engineering and computing*, 42(3):407–412, 2004.

[56] C. Henson and P. Thomson. Spitfire Audio Holdings LTD. https://www.spitfireaudio.com/, 2007. Last visited on 20/06/2022.

[57] A. Isin and S. Ozdalili. Cardiac arrhythmia detection using deep learning. *Procedia computer science*, 120:268–275, 2017.

[58] C. E. Izard, D. Z. Libero, P. Putnam, and O. M. Haynes. Stability of emotion experiences and their relations to traits of personality. *Journal of personality and social psychology*, 64(5):847, 1993.

[59] W. James. What is an emotion? mind, os-ix, 188-205, 1884.

[60] H. H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, 10:370–375, 1958.

[61] C. S. Jonker and C. Vosloo. The psychometric properties of the Schutte Emotional Intelligence Scale: empirical research. *SA Journal of Industrial Psychology*, 34(2): 21–30, 2008.

[62] A. Joshi, S. Kale, S. Chandel, and D. K. Pal. Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4):396, 2015.

[63] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10):1745–1758, 2000.

[64] P. N. Juslin. Emotional communication in music performance: A functionalist perspective and some data. *Music perception*, 14(4):383–418, 1997.

[65] P. N. Juslin. Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human perception and performance*, 26(6):1797, 2000.

[66] P. N. Juslin. A brunswikian approach to emotional communication in music performance. 2001.

[67] P. N. Juslin. Communicating emotion in music performance: A review and a theoretical framework. 2001.

[68] P. N. Juslin and P. 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.

[69] P. N. Juslin and J. Sloboda. *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, 2011.

[70] M. Kassler. Toward musical information retrieval. *Perspectives of New Music*, pages 59–67, 1966.

[71] B. Kaur, D. Singh, and P. P. Roy. A novel framework of EEG-based user identification by analyzing music-listening behavior. *Multimedia tools and applications*, 76 (24):25581–25602, 2017.

[72] Z. Kekecs, A. Szekely, and K. Varga. Alterations in electrodermal activity and cardiac parasympathetic tone during hypnosis. *Psychophysiology*, 53(2):268–277, 2016.

[73] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12): 2067–2083, 2008.

[74] S. Koelsch and J. Mulder. Electric brain responses to inappropriate harmonies during listening to expressive music. *Clinical Neurophysiology*, 113(6):862–869, 2002.

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

[76] C. A. Kothe and S. Makeig. BCILAB: a platform for brain–computer interface development. *Journal of neural engineering*, 10(5):056014, 2013.

[77] C. L. Krumhansl. Music: A link between cognition and emotion. *Current directions in psychological science*, 11(2):45–50, 2002.

[78] M. Lai, S. Tognetti, M. Garbarino, and R. Picard. Empatica Srl. `https://www.empatica.com/`, 2013. Last visited on 27/06/2022.

[79] S. Lapi, F. Lavorini, G. Borgioli, M. Calzolai, L. Masotti, M. Pistolesi, and G. A. Fontana. Respiratory rate assessments using a dual-accelerometer device. *Respiratory physiology & neurobiology*, 191:60–66, 2014.

[80] R. J. Larsen, E. Diener, and R. A. Emmons. Affect intensity and reactions to daily life events. *Journal of personality and social psychology*, 51(4):803, 1986.

[81] M. Lehne. *Emotional experiences of tension and suspense: psychological mechanisms and neural correlates*. PhD thesis, 2014.

[82] D. J. Levitin. *This is your brain on music: The science of a human obsession*. Penguin, 2006.

[83] L. S. Lilly. *Pathophysiology of heart disease: a collaborative project of medical students and faculty*. Lippincott Williams & Wilkins, 2012.

[84] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, and D. R. Kipke. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of neurophysiology*, 101(3):1679–1689, 2009.

[85] A. Luthra. *ECG made easy*. Jaypee Brothers Medical Publishers, 2019.

[86] P. D. MacLean. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalography and clinical neurophysiology*, 4(4):407–418, 1952.

[87] A. Macy. BIOPAC Systems Inc. `https://www.biopac.com/`, 1985. Last visited on 20/06/2022.

[88] E. Maggioni, F. Arienti, S. Minella, F. Mameli, L. Borellini, M. Nigro, F. Cogiamanian, A. M. Bianchi, S. Cerutti, S. Barbieri, et al. Effective connectivity during rest and music listening: An EEG study on Parkinson's disease. *Frontiers in aging neuroscience*, 13:208, 2021.

[89] A. K. Maity, R. Pratihar, A. Mitra, S. Dey, V. Agrawal, S. Sanyal, A. Banerjee, R. Sengupta, and D. Ghosh. Multifractal detrended fluctuation analysis of alpha

and theta EEG rhythms with musical stimuli. *Chaos, Solitons & Fractals*, 81:52–67, 2015.

[90] S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, 94(20):10979–10984, 1997.

[91] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods*, 53(4):1689–1696, 2021.

[92] G. Matthews, D. M. Jones, and A. G. Chamberlain. Refining the measurement of mood: The uwist mood adjective checklist. *British journal of psychology*, 81(1): 17–42, 1990.

[93] J. D. Mayer and P. Salovey. *Mayer-Salovery-Caruso emotional intelligence test*. Multi-Health Systems Incorporated Toronto, 2007.

[94] J. D. Mayer, M. DiPaolo, and P. Salovey. Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of personality assessment*, 54(3-4):772–781, 1990.

[95] D. M. McNair, M. Lorr, L. F. Droppleman, et al. Manual profile of mood states. 1971.

[96] B. Mehler, B. Reimer, and Y. Wang. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In *Driving Assesment Conference*, volume 6. University of Iowa, 2011.

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

[98] J. Mitchell. "both sides, now" / words and music by joni mitchell, 1967.

[99] M. Miyakoshi, A. Delorme, and S. Makeig. Clean rawdata EEGLAB plug-in, 2013–. URL `https://github.com/sccn/clean_rawdata/`.

[100] N. Montano, A. Porta, C. Cogliati, G. Costantino, E. Tobaldini, K. R. Casali, and F. Iellamo. Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. *Neuroscience & Biobehavioral Reviews*, 33(2):71–80, 2009.

[101] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart. The musicality of non-

musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642, 2014.

[102] M. R. Nuwer, G. Comi, R. Emerson, A. Fuglsang-Frederiksen, J.-M. Guérit, H. Hinrichs, A. Ikeda, F. J. C. Luccas, and P. Rappelsburger. IFCN (International Federation of Clinical Neurophysiology) standards for digital recording of clinical EEG. *Electroencephalography and clinical Neurophysiology*, 106(3):259–261, 1998.

[103] P. J. O'Connor, A. Hill, M. Kaya, and B. Martin. The measurement of emotional intelligence: A critical review of the literature and recommendations for researchers and practitioners. *Frontiers in psychology*, 10:1116, 2019.

[104] U. of Zaragoza. Bitbrain Technologies. `https://www.bitbrain.com/blog/ai-eeg-data-processing`, 2010. Last visited on 24/06/2022.

[105] J. Panksepp. *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press, 2004.

[106] M. Pantic, G. Caridakis, E. André, J. Kim, K. Karpouzis, and S. Kollias. Multimodal emotion recognition from low-level cues. In *Emotion-Oriented Systems*, pages 115–132. Springer, 2011.

[107] J. W. Papez. A proposed mechanism of emotion. 1937. *The Journal of neuropsychiatry and clinical neurosciences*, 7(1):103–112, 1995.

[108] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, 72(2):184–187, 1989.

[109] K. V. Petrides and A. Furnham. On the dimensional structure of emotional intelligence. *Personality and individual differences*, 29(2):313–320, 2000.

[110] T. Pham, Z. J. Lau, S. A. Chen, and D. Makowski. Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors*, 21(12):3998, 2021.

[111] R. W. Picard. *Affective computing*. MIT press, 2000.

[112] R. Plutchik. Emotions in the practice of psychotherapy-clinical implications of affect theories. 2000.

[113] P. J. Rentfrow and S. 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.

[114] P. J. Rentfrow, L. R. Goldberg, and D. J. Levitin. The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6): 1139, 2011.

[115] P. J. Rentfrow, L. R. Goldberg, and D. J. Levitin. The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6): 1139, 2011.

[116] P. J. Rentfrow, L. R. Goldberg, D. J. Stillwell, M. Kosinski, S. D. Gosling, and D. J. Levitin. The song remains the same: A replication and extension of the music model. *Music perception*, 30(2):161–185, 2012.

[117] R. D. Roberts, M. Zeidner, and G. Matthews. Does emotional intelligence meet traditional standards for an intelligence? some new data and conclusions. *Emotion*, 1(3):196, 2001.

[118] M. Rohrmeier and I. Cross. Statistical properties of tonal harmony in bach's chorales. In *Proceedings of the 10th international conference on music perception and cognition*, volume 6, pages 123–1319. Hokkaido University Sapporo, Japan, 2008.

[119] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

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

[121] M. D. Schulkind, L. K. Hennis, and D. C. Rubin. Music, emotion, and autobiographical memory: They're playing your song. *Memory & Cognition*, 27(6):948–955, 1999.

[122] N. S. Schutte, J. M. Malouff, L. E. Hall, D. J. Haggerty, J. T. Cooper, C. J. Golden, and L. Dornheim. Development and validation of a measure of emotional intelligence. *Personality and individual differences*, 25(2):167–177, 1998.

[123] M. Shahbakhti, V. Khalili, and G. Kamaee. Removal of blink from EEG by empirical mode decomposition (emd). In *The 5th 2012 Biomedical Engineering International Conference*, pages 1–5. IEEE, 2012.

[124] K. Shimoji, A. Nader, and W. Hamann. *Chronic Pain Management in General and Hospital Practice*. Springer, 2021.

[125] E. Siedlecka and T. F. Denson. Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review*, 11(1):87–97, 2019.

[126] P. Sikkema. Jongeren 890–990: Een generatie waar om gevochten wordt [Youth 890–990: A generation that is being fought for]. *Amsterdam: Interview-NSS*, 1999.

[127] J. A. Sloboda and S. A. O'neill. Emotions in everyday listening to music. *Music and emotion: Theory and research*, 8:415–429, 2001.

[128] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui. Mped: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7: 12177–12191, 2019.

[129] R. Soni and M. Muniyandi. Breath rate variability: a novel measure to study the meditation effects. *International journal of yoga*, 12(1):45, 2019.

[130] M. Strauss, C. Reynolds, S. Hughes, K. Park, G. McDarby, R. Picard, J. Tao, and T. Tan. Affective computing: A review, 2005.

[131] L. Sun, C. Feng, and Y. Yang. Tension experience induced by nested structures in music. *Frontiers in Human Neuroscience*, 14:210, 2020.

[132] W. O. Tatum IV. *Handbook of EEG interpretation*. Springer Publishing Company, 2021.

[133] R. E. Thayer. Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological reports*, 58(2):607–614, 1986.

[134] W. F. Thompson, L. Quinto, et al. Music and emotion: Psychological considerations. *The aesthetic mind: Philosophy and psychology*, pages 357–375, 2011.

[135] P. van Gent, H. Farah, N. Nes, and B. van Arem. Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data. In *Proceedings of the 6th HUMANIST Conference*, pages 173–178, 2018.

[136] P. Van Gent, H. Farah, N. Van Nes, and B. Van Arem. Heartpy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research part F: traffic psychology and behaviour*, 66:368–378, 2019.

[137] G. vanRossum. Python reference manual. *Department of Computer Science [CS]*, (R 9525), 1995.

[138] P. H. Venables and M. J. Christie. Electrodermal activity. *Techniques in psychophysiology*, 54(3), 1980.

[139] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

[140] R. Warnke. Meditech Electronic GmbH. `https://www.meditech.de/en_US/`, 1996. Last visited on 20/06/2022.

[141] D. Watson and L. A. Clark. The panas-x: Manual for the positive and negative affect schedule-expanded form. 1994.

[142] P. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[143] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen. A regression approach to music emotion recognition. *IEEE Transactions on audio, speech, and language processing*, 16(2):448–457, 2008.

[144] F. Zanow. ANT Neuro. `https://www.ant-neuro.com/`, 1997. Last visited on 24/06/2022.

[145] M. Zentner, D. Grandjean, and K. R. Scherer. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8(4):494, 2008.

A | Appendix: Adjusted GEMS-9 questionnaire

Subject: ________________ Date: ____________________

Geneva Emotional Music Scale (GEMS-9)

Instructions

When providing your ratings, please describe how the music piece you listen to makes you *feel* (e.g., this music piece makes me *feel sad*). Do not describe the music (e.g., this music piece is sad) or what the music may be expressing (e.g. this music piece expresses sadness). Keep in mind that a piece of music can be sad or can sound sad without making you feel sad. Please rate the intensity with which you *felt* each of the following feelings on a scale ranging from 1 (*not at all*) to 5 (*very much*).

1	2	3	4	5
Not at all	Somewhat	Moderately	Quite a lot	Very Much

1. **Music piece no. 1**

   a. To what extent does the music piece make you feel the following emotions?

1.	**Wonder** Filled with wonder, Dazzled, Moved	1	2	3	4	5
2.	**Transcendence** Fascinated, Overwhelmed, Feelings of transcendence and spirituality	1	2	3	4	5
3.	**Power** Strong, Triumphant, Energetic	1	2	3	4	5
4.	**Tenderness** Tender, Affectionate, In love	1	2	3	4	5
5.	**Nostalgia** Nostalgic, Dreamy, Melancholic	1	2	3	4	5
6.	**Peacefulness** Serene, Calm, Soothed	1	2	3	4	5
7.	**Joyful Activation** Joyful, Amused, Bouncy	1	2	3	4	5
8.	**Sadness** Sad, Sorrowful	1	2	3	4	5
9.	**Tension** Tense, Agitated, Nervous	1	2	3	4	5

   b. How well do you know this music piece? 1 2 3 4 5

   c. How much do you like this music piece? 1 2 3 4 5

B | Appendix: STOMP-R questionnaire

STOMP-Revised

Please indicate your basic preference for each of the following genres using the scale provided.

1	2	3	4	5	6	7
Dislike Strongly	Dislike Moderately	Dislike a Little	Neither like nor dislike	Like a Little	Like Moderately	Like Strongly

1. _____ Alternative
2. _____ Bluegrass
3. _____ Blues
4. _____ Classical
5. _____ Country
6. _____ Dance/Electronica
7. _____ Folk
8. _____ Funk
9. _____ Gospel
10. _____ Heavy Metal
11. _____ International/Foreign
12. _____ Jazz
13. _____ New Age
14. _____ Oldies
15. _____ Opera
16. _____ Pop
17. _____ Punk
18. _____ Rap/hip-hop
19. _____ Reggae
20. _____ Religious
21. _____ Rock
22. _____ Soul/R&B
23. _____ Soundtracks/theme song

Music preference dimensions scoring:

Reflective & Complex: 2, 3, 4, 7, 11, 12, 13, 15

Intense & Rebellious: 1, 10, 17, 21

Upbeat & Conventional: 5, 9, 14, 16, 20, 23

Energetic & Rhythmic: 6, 8, 18, 19, 22

C | Appendix: SREIT questionnaire

The Schutte Self Report Emotional Intelligence Test (SSEIT)

Instructions: Indicate the extent to which each item applies to you using the following scale:

1 = strongly disagree
2 = disagree
3 = neither disagree nor agree
4 = agree
5 = strongly agree

1. I know when to speak about my personal problems to others
2. When I am faced with obstacles, I remember times I faced similar obstacles and overcame them
3. I expect that I will do well on most things I try
4. Other people find it easy to confide in me
5. I find it hard to understand the non-verbal messages of other people*
6. Some of the major events of my life have led me to re-evaluate what is important and not important
7. When my mood changes, I see new possibilities
8. Emotions are one of the things that make my life worth living
9. I am aware of my emotions as I experience them
10. I expect good things to happen
11. I like to share my emotions with others
12. When I experience a positive emotion, I know how to make it last
13. I arrange events others enjoy
14. I seek out activities that make me happy
15. I am aware of the non-verbal messages I send to others
16. I present myself in a way that makes a good impression on others
17. When I am in a positive mood, solving problems is easy for me
18. By looking at their facial expressions, I recognize the emotions people are experiencing
19. I know why my emotions change
20. When I am in a positive mood, I am able to come up with new ideas
21. I have control over my emotions
22. I easily recognize my emotions as I experience them
23. I motivate myself by imagining a good outcome to tasks I take on
24. I compliment others when they have done something well
25. I am aware of the non-verbal messages other people send
26. When another person tells me about an important event in his or her life, I almost feel as though I have experienced this event myself
27. When I feel a change in emotions, I tend to come up with new ideas
28. When I am faced with a challenge, I give up because I believe I will fail*
29. I know what other people are feeling just by looking at them
30. I help other people feel better when they are down
31. I use good moods to help myself keep trying in the face of obstacles
32. I can tell how people are feeling by listening to the tone of their voice

33. It is difficult for me to understand why people feel the way they do*

List of Figures

List of Tables

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, while not knowing life at all [98].