Modeling ANS Responses to Acoustic Features for Personalized Music Therapy Interventions in Mental Health Care Using Machine Learning

UNIVERSITY OF TURKU
Department of Computing
Master of Science (Tech) Thesis
Biomedical Engineering and Health Technology
May 2026
Niko Itänen

Supervisors:
Katri Karhinoja
Jonas Sandelin

UNIVERSITY OF TURKU
Department of Computing

Niko Itänen: Modeling ANS Responses to Acoustic Features for Personalized Music Therapy Interventions in Mental Health Care Using Machine Learning

Master of Science (Tech) Thesis, 58 p.
Biomedical Engineering and Health Technology
May 2026

The purpose of this study is to model the effects of music's acoustic features on autonomic nervous system (ANS) responses and thereby contribute to the personalization of receptive music therapy interventions. Although the physiological responses related to music properties are known, identifying individual differences among complex acoustic and physiological signals is challenging.
The study conducted an empirical experiment in which physiological signals were collected from 13 participants, measuring heart and respiratory activity as well as skin conductance. Each participant attended a total of three measurement sessions on different days, during which individual physiological changes were measured while listening to music. Correlations between these features were examined at both the group and individual levels using statistical methods, a population-level mixed-effects model, and an individual-level Ridge regression model.
The statistical methods of the study showed that acoustic features related to the brightness and energy distribution of music drive physiological responses, such as heart rate and respiration, at the group level. Individual-level modeling was used to examine the practical predictive power of the machine learning model using the R^2 metric in the validation. The model's performance proved to be negative, indicating that predictions cannot be made using the model based on the current sample size. Individual-level permutation tests revealed statistical significance for genuine individual-specific signals in cardiac responses.
The results suggest that there is a genuine connection between music and physiological reactions, but that there are very strong individual variation in these reactions. For this reason, utilizing generalizable machine learning models cannot explain the reactions of all individuals. In future research, the development of personalized models will require significantly larger datasets. Since the study used low-level acoustic features to interpret physiological reactions, including higher-level musical features - such as harmonic and rhythmic structure and the emotional nature of the music - into the analysis could improve the model's interpretive power in the future.

Keywords: music therapy, autonomic nervous system, machine learning, acoustic features, physiological signals, personalized modeling

TURUN YLIOPISTO
Tietotekniikan laitos

NIKO ITÄNEN: Modeling ANS Responses to Acoustic Features for Personalized Music Therapy Interventions in Mental Health Care Using Machine Learning

Diplomityö, 58 s.
Lääketieteellinen tekniikka ja terveysteknologia
Toukokuu 2026

Tämän tutkimuksen tarkoituksena on mitata musiikin akustisten piirteiden vaikutusta autonomisen hermoston (ANS) vasteisiin ja pyrkiä siten edistämään reseptiivisen musiikkiterapian interventioiden personointia. Vaikka musiikin luomia fysiologisia vasteita tunnetaan, yksilöllisten erojen tunnistaminen monimutkaisten akustisten ja fysiologisten signaalien välillä on haasteellista.
Tutkimuksessa suoritettiin empiirinen koe, jossa kerättiin 13 koehenkilön fysiologisia signaaleita, jotka mittasivat sydämen ja hengityksen toimintaa sekä ihon sähkönjohtavuutta. Jokaiselle osallistujalle suoritettiin kolme erillisestä mittauskertaa eri päivinä, missä yksilön fysiologista muutosta mitattiin musiikkia kuunnellessa. Piirteiden välisiä yhteyksiä tutkittiin sekä ryhmä- että yksilötasolla tilastollisin menetelmin, populaatiotason sekamallin (mixed-effects) sekä yksilökohtaisen Ridge-regressiomallin avulla.
Tutkimuksen tilastolliset menetelmät osoittivat musiikin energiaan ja kirkkauteen liittyvien akustisten piirteiden ohjaavan fysiologisia vasteita, kuten sykettä ja hengitystä populaatiotasolla. Yksilötason mallinnuksella tutkittiin koneoppimismallin suorituskykyä R^2 mittarin avulla sekä syvennyttiin personoinnin tarpeen validointiin. Mallin suorituskyky osoittautui negatiiviseksi, minkä perusteella nykyisen aineiston pohjalta tarkkoja ennusteita ei voida mallin avulla suorittaa. Yksilötason permutaatiotestit todistivat tilastollista merkittävyyttä sydämen toiminnan vasteiden aidoille yksilöllisille signaalille.
Tulokset viittaavat siihen, että musiikin ja fysiologisten vasteiden välillä on havaittavissa aitoja yhteyksiä, mutta vasteissa esiintyy erittäin vahvaa yksilöllistä vaihtelua. Tästä syystä yleistävät ryhmätason koneoppimismallit eivät kykene selittämään reaktioita kaikkien yksilöiden kohdalla. Jatkotutkimuksissa personoitujen mallien kehittäminen vaatii huomattavasti laajempaa ainestoa. Koska tässä tutkimuksessa hyödynnettiin fysiologisten vasteiden tulkinnassa matalan tason akustisia piirteitä, saattaisi korkeamman tason musiikin piirteiden - kuten harmonisen ja rytmisen rakenteen sekä musiikin emotionaalisen luonteen - hyödyntäminen analyysissä parantaa mallin tulkintakykyä tulevaisuudessa.

Asiasanat: musiikkiterapia, autonominen hermosto, koneoppiminen, akustiset piirteet, fysiologiset signaalit, personoiva mallintaminen

Contents

List of Figures

List of Tables

1 Introduction

Mental health disorders are a common group of illnesses that cause a high amount of disabilities and deaths globally every year. According to the *World Health Organization* (WHO), there are more than a billion people who suffer from some sort of mental health disorders. This has evoked many countries to make large investments for their healthcare systems in this particular field [1]. Even though the relative incidence of these disorders is not estimated to increase globally, population growth and societal pressure are expected to increase the total number of mental health patients in the following years. Despite the fact that significant investments are constantly being made in the sector, current resources related to the treatment needs cannot cover care for all patients, which highlights the need to develop scalable digital interventions and machine learning methods that can improve the accessibility and efficiency of care by personalizing treatment for each individual patient. [1], [2]

Current treatment methods for mental health disorders consist primarily of biological, psychological, and lifestyle-based approaches. Biological treatment contains interventions such as pharmacological and neurostimulation approaches [3, pp. 129–154]. Psychological treatment encompasses various therapies that focus on the patient's understanding of personal emotion regulation and environmental influences [3, pp. 155 – 170]. Life-based approaches are typically connected to psychosocial interventions, but their core idea is the reorganization of personal lifestyle to improve individual quality of life [3, pp. 171 – 178]. Even though current methods provide a solid foundation, their limitations lie in the need for in-person interactions, which require a significant amount of professional resources—where digital health care solutions could offer more accessibility and tailored support. [4]

Although multiple different methods and approaches exist to provide care for patients, access to treatment remains one of the greatest challenges in the field of psychiatry. For instance, an individual with impaired mental condition may suffer from several different conditions simultaneously, while others may face social anxiety or stigma that makes traditional face-to-face therapy difficult to arrange. Furthermore, for others, severe depression can significantly reduce a patient's motivation to engage in traditional therapy sessions. [4]

Since digital devices are now available to a large portion of the global population, psychosocial and lifestyle-based approaches can be made more accessible without requiring the patient to exert excessive effort for in-person interaction. In this context, artificial intelligence and digital interventions are essential not only for improving the accessibility but also for personalizing treatment on a patient-by-patient basis. By utilizing machine learning methods, healthcare professionals can be supported in tailoring the most effective treatment options for each individual, which is a critical factor in improving overall treatment outcomes.

This research focuses on music-based therapy, which has been found to produce positive outcomes in patients suffering from impaired mental conditions. [5] However, as with many other therapeutic approaches, a significant limitation of music-based therapies is that they affect individuals differently, which means these methods are required to be tailored even more closely to individual needs. [6] In music therapy, an individualized perspective is an essential factor of the therapeutic process, as personal reactions to the emotional tone, dynamics, genre, and other musical aspects can vary significantly [7]. To address this challenge, the technical section of this study examines machine learning methods for the purpose of personalizing interventions. This is achieved by investigating how different acoustic features of music affect the autonomic nervous system (ANS), which is closely linked to the regulation of emotional life. Modeling these physiological responses provides a data-driven foundation for developing more effective, personalized music therapy interventions.

In the preparation of this thesis, various artificial intelligence models have been utilized as tools to support the different stages of the research and writing process. The Gemini 3 Flash and Gemini 3.1 Pro models, developed by Google, were used for outlining the work's structure, clarifying scientific concepts, supporting proofreading, and automatically compiling LaTeX tables. In addition, the Claude Haiku 4.5 model was used to support the design and programming of the technical analysis pipeline.

The collection of empirical research data, development of the analytical methods, and the final interpretations were carried out independently, based on the scientific literature utilized in the work, maintaining the originality and contribution of the study. The literature used was primarily accessed through Google Scholar, the University of Turku digital library (Volter), PubMed and ScienceDirect.

This thesis consists of a total of six chapters. Chapter 2 introduces music therapy, describes the characteristics of music elements, and finally explains how different statistical and modeling methods are applied to personalize hierarchically complex data for therapeutic purposes. Chapter 3 provides a straightforward overview of the empirical experiment, the data obtained, and the methods used to analyze the collected data. In Chapter 4, results from the analysis stages are presented, including the direct statistical relationships between features, as well as the performance of the used modeling methods. The obtained results are discussed and compared to the existing literature in Chapter 5, where the study's limitations in the experimental and modeling stages are also identified. Finally, Chapter 6, discusses the final conclusions, assesses the significance of the study for the topic domain, and considers how current limitations could be addressed in future research.

1.1 Research Questions

The purpose of this thesis is to answer a total of three research questions:

1. Which acoustic features of music are most strongly associated with autonomic nervous system responses?
2. Do individuals' physiological responses to the acoustic features of music differ significantly from one another?
3. To what extent can individual physiological responses be predicted based on the acoustic features of music using machine learning models?

The first question is addressed through statistical analysis based on data collected from volunteer participants in the empirical experiment. The second question is addressed using the same data analysis, but applied at a different level of data examination that accounts for subject-wise influences instead of population-level. Research questions one and two also draw on previously published articles that address different acoustic features and their relationship to ANS. The third question is addressed in the model evaluation process, where model performance metrics are used to determine how well both the population-level mixed-effects and the subject-specific Ridge regression models perform with the collected data.

2 Background

This chapter presents the theoretical foundation of methods for analyzing the physiological effects of music, based on previous research on the topic. Section 2.1 discusses music therapy as a clinical intervention, as well as its current and future requirements. It also defines the acoustic properties of music. Section 2.2 examines the neurobiological processing of music perception and its measurable effects on the autonomic nervous system. Finally, Section 2.3 presents how machine learning can be used to model individual physiological responses and explore the challenges and solutions related to processing hierarchically complex data.

2.1 Music therapy and acoustic characteristics

Music is a cultural universal that is generally linked to the processing, reinforcement, and expression of emotional experiences and is used as a tool for interpersonal communication and socialization [8, pp. 3–14], [9] [10]. To narrow the topic for objective analysis in this study, music is defined as organized sound patterns that include properties such as rhythm, pitch variation, repetition, and temporal order [8, pp. 31–48] [11] [12, pp. 25–30]. The elements of music are generally constructed intentionally to influence the listener's emotional states, where various instrumental and structural properties have been shown to be linked to relaxation, stress levels, self-awareness and cognitive engagement [8, pp. 427–448] [12, pp. 224–281] [13] [14]. However, these effects are not tied to the general population level, as various factors such as music styles, genres and cultural traditions may lead to variations in individual responses. [15] These connections between musical properties and their effects on human physiology and psychology form the basis for the use of music as a therapeutic intervention.

2.1.1 Music therapy

Music therapy is a clinical method based on high-evidence research in which different elements of music are used for both rehabilitation and treatment in various psychological, cognitive, and even physical conditions [12, pp. 224–281] [16]. In clinical practice, music therapy is typically used as an adjunctive method for mental health disorders—such as major depressive disorder, anxiety disorders, schizophrenia, and post-traumatic stress disorder—alongside already established psychotherapy and pharmacological interventions [17] [18]. It is particularly effective for individuals who are not always easily reached through verbal or cognitive methods alone, especially in areas of emotional expression, physiological regulation, and nonverbal communication [16] [19, pp. 9–13].

Outcomes of music therapy have shown a reduction in anxiety symptoms and stress levels, improvements in mood and emotional regulation, and enhancements in cognitive processes [12, pp. 505–568] [18]. However, differences in the type and duration of interventions, as well as variation at the population level, highlight the need for a better understanding of the mechanisms of how music exerts its effects [12, pp. 505–568] [19, pp. 35–52].

Music therapy can be applied in two different forms: as an active form, such as playing an instrument, singing, and self-expression, and as a receptive form based on structured music listening [16] [19, pp. 9–13]. Although active music therapy methods are widely used in clinical settings, the receptive form of music therapy primarily focuses on analyzing physiological reactions to auditory stimuli. Receptive therapy methods apply listening approaches, such as rhythmic auditory stimulation (RAS) and physioacoustic treatment (PAT). RAS is based on the playback of repetitive isochronous pulses generated by a metronome to trigger the brain's motor network and enhance natural reactions to rhythmic and steady sounds [20]. The optimal rhythmic cue for RAS is selected and matched to individual preferences, with the rhythm being gradually adjusted over time. [21] PAT, in turn, uses various low-frequency sound waves to primarily treat physiological challenges [22] [23]. Based on previous studies, both RAS and PAT have shown positive outcomes in the treatment of anxiety and depression [24] [25] [26] [27]. Both RAS and PAT focus on specific musical elements that have been shown to induce physiological changes. However, these methods focus specifically on isolated elements, such as musical rhythm or resonance, which is limited compared to natural music listening that includes a broader range of elements. To better account for the complexity of music and high variation in individual preferences, a deeper analysis of musical elements in a broader context is required to expand these current interventions and enhance music therapy personalization.

Because music therapy is methodologically diverse in terms of treatment design, duration, and other patient-specific factors, there are still many conflicting conclusions regarding its effectiveness as a therapy method. [5] [7]

A key limitation in music therapy is the highly individualized form of treatment, where individual responses to musical features such as tempo, rhythm, timbre, and harmonic structure can vary significantly [7]. While there may be certain factors between types of music that are generally considered to provide relaxation for people, it cannot be said that a single specific set of features in music would provide relaxation for all people [5] [28].

Addressing these challenges requires a central focus on the individual associations between musical features and physiological responses. The use of machine learning in analyzing the effects of music can expand the potential for developing more personalized music therapy methods. Previous studies examining machine learning-based personalization have shown that treatment effectiveness depends on the patient's individual physiological state and baseline conditions. These studies highlight the current direction in music therapy toward data-driven tailoring approaches. [6]

2.1.2 Acoustic features

Music is a broad combination of different but organized overlapping sounds that can be described as pressure fluctuations that propagate as acoustic waves [29, pp. 4–7] [30, pp. 28–42] [31]. These acoustic waves can be divided into two different domains: the time domain and the frequency domain. The frequency domain describes sounds as a combination of frequency components and signal amplitude, which determine how pitch, loudness, harmonic structure, and timbre are perceived. The time domain, in turn, describes sound signals as continuous pressure variations that the human brain interprets as temporal features such as onsets, transients, and tempo. [30, pp. 43–55] [32] [33, pp. 19–25] [34, pp. 19–41]

Frequency and pitch

Frequency is a fundamental physical property of periodic waves, determining the number of oscillation cycles per time unit, measured in hertz (Hz). In the context of music, absolute frequency relates to the perceived pitch of a sound, meaning that higher-pitched sounds have higher frequencies, and vice versa [33, pp. 19–25]. Frequencies such as 261.63 Hz (C4), 329.81 Hz (E4), and 440 Hz (A4) are absolute frequencies that produce completely different perceived pitches [30, pp. 80–95]. However, the human brain does not perceive sound frequencies linearly, but rather on a logarithmic scale. This is reflected in the structure of octaves, where frequency is doubled (e.g. 220 Hz, 440 Hz, and 880 Hz), but perceived as belonging to the same pitch class because of their harmonic ratio [34, pp. 19–41] [35]. In modern Western music theory, the logarithmic scale is formed by dividing an octave into 12 equally spaced semitones, ensuring stable perception across the entire audible spectrum (See Figure 2.1). [30, pp. 80–95]

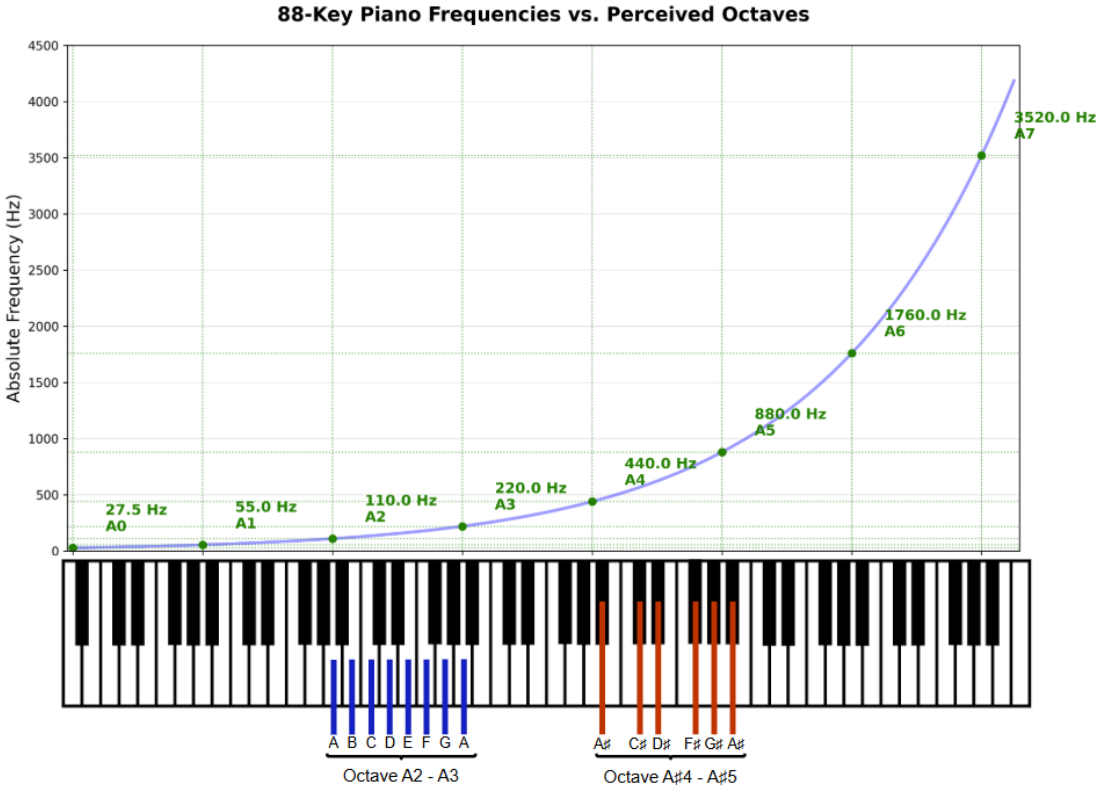

Figure 2.1: An 88-key piano keys, with the absolute frequencies of A notes plotted on a linear scale. The piano illustration modified from [36], and the plot generated by author in Python.

In audio signal processing, representations in the frequency domain are typically obtained using the Fourier transform, which decomposes the signal into its constituent frequency components. From the extracted frequency components, acoustic features describing the brightness and harmony of music can be calculated, such as the spectral centroid, spectral bandwidth, spectral rolloff, as well as Mel-frequency cepstral coefficients (MFCCs), which represent the short-term power spectrum of the sound. In addition, features representing the harmonic and tonal content of music, such as the tonal centroid (Tonnetz), can be extracted to identify relationships between musical harmony and physiological responses. [32] [37] [38]

Amplitude and loudness

In sound signals, amplitude determines the magnitude of pressure fluctuations and the energy carried by the wave. The perceptual experience of sound intensity is called loudness. Similar to the perception of pitch, the human brain processes sound intensity logarithmically, which is expressed using the decibel (dB) scale [30, pp. 16–27]. While amplitude determines the amount of energy transmitted into the auditory system, the perceived loudness is also influenced by other factors, such as frequency and temporal features [31].

In signal processing, energy is typically expressed as root mean square energy (RMSE), which captures the average energy of the signal over a specific time interval, providing a clearer presentation of total energy distribution, rather than focusing on a single amplitude points [37].

Temporal structure

In a musical context, time determines the order of sounds through factors such as tempo, transients and onsets, which together form the dynamic structure of music [39].

Transients are fast, non-stationary events where the amplitude and spectrum of a signal change rapidly and typically last only milliseconds. They are typically linked with events such as drum beats, descending strokes of piano keys, and consonants in vocals [31] [40] [41]. Onsets typically refer to precise points in time when high spectral energy sound events, such as transient, begin. The auditory system recognizes these onsets and uses them as reference markers that are essential for musical timing and rhythm [31]. In signal processing, transients are strongly related to the zero-crossing rate, but in the frequency-domain, the spectral centroid and bandwidth can be used to identify transients because they exhibit strong frequency peaks [32] [41].

Tempo refers to the rate at which various events occur in music and is measured in beats per minute (BPM). It is derived from the periodicity of onsets and is one of the most perceptually central temporal features of music, having a significant influence on physiological arousal and the regulation of the ANS [42] [43]. The brain perceives tempo of music by estimating the duration between continuous onsets [39].

All of these acoustic features form a physical input which the human body can process. The following section describes how the body processes music and how this can be measured into analyzable data using various measurement methods.

2.2 Neurobiological mechanisms and physiological responses

The perception of music is a multi-level process in which physical properties of sounds are transformed into various responses within the body. At the lowest level, music perception involves the recognition and processing of sound waves, where these are first encoded in the auditory system and further transformed into perceptual representations [44].

In addition to recognizing direct acoustic features, the human brain also processes music to evoke emotions, memories, and context-dependent interpretations. This neural processing is clinically significant in the context of music therapy, as brain activation related to music interpretation and emotion processing largely occurs in shared neural pathways [8, pp. 427–448] [45, pp. 926–970] [46]. Interaction with these shared neural networks modulates the ANS by regulating the responses of various organs to stimuli, enabling music to be perceived and experienced at the whole body level. [42] [47]

2.2.1 Auditory pathway

The processing of music begins in the cochlea and brainstem, from where it proceeds to the thalamus, and ultimately reaches the auditory cortex. (See Figure 2.2) Along this pathway, auditory information is converted into more complex perceptual representations for further processing related to meaning and emotion. [33] [45, pp. 583–586]

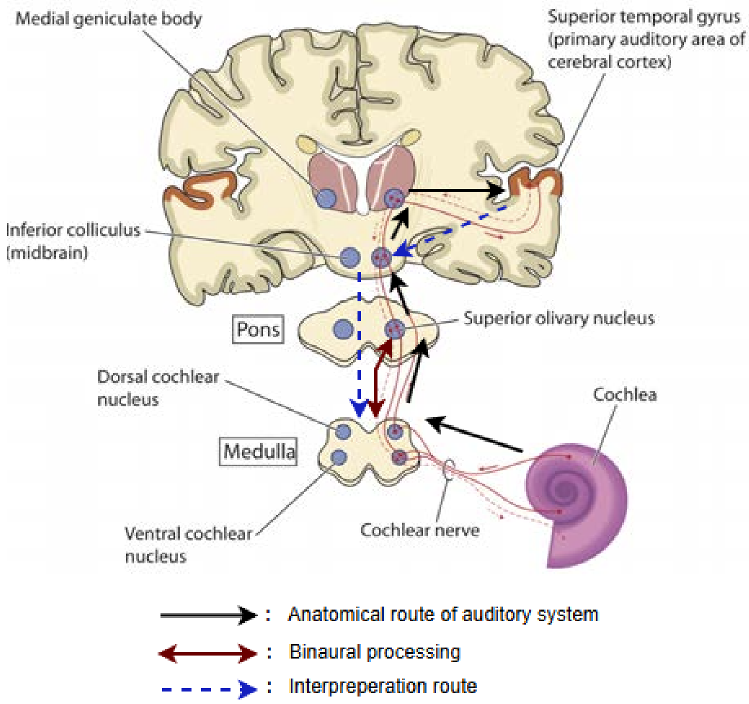

Figure 2.2: The main brain regions and neural pathways involved in sound processing. Modified from [48]

The auditory signals are first converted into mechanical vibrations in the cochlea. This signal is transmitted via the vestibulocochlear nerve to the inferior colliculus in the midbrain [33, pp. 233–235] [49]. The inferior colliculus participates in anticipation and salience determination, while also playing a critical role in encoding the frequency and intensity of auditory signals [44]. It also indirectly influences emotions and regulation of the ANS through brainstem circuits involved in stress and arousal responses. Finally, the signals are relayed through the medial geniculate nucleus of the thalamus to the auditory cortex [33, pp. 233–235] [45, pp. 560–604] [50].

The auditory cortex can be divided into three main regions: the primary auditory cortex, the secondary auditory cortex, and the tertiary auditory cortex [18]. The primary auditory cortex receives auditory information from the medial geniculate nucleus and performs initial processing of sound frequency differences using the brain's tonotopic mapping [33, pp. 223–249] [45, pp. 602–609]. In the belt region, surrounding the primary auditory cortex, much more complex sound analysis is performed, including sound localization as well as the storage and recall of various auditory patterns, which is a key factor in the perception of timbre, rhythm, and spatial features of music. [12, pp. 224–281] Finally, the parabelt area surrounding the belt region, functions as the brain's recognition mechanism, where the high-level interpretation of auditory information is processed [33, pp. 223–249] [45, pp. 602–609] .

2.2.2 Limbic activity

Although auditory pathways strongly influence how music is perceived, they do not process the subjective emotional and contextual processes to the extent of the limbic and frontal brain regions [28]. The limbic region is the center of emotional and motivational processing, including variety of different structures, such as the amygdala, hippocampus, and insula [28] [45, pp. 926–970]. Another significant brain region is the prefrontal cortex, which modulates limbic activity and is involved in emotional awareness, evaluation, and control. [45, pp. 337–361]

Within this network, the amygdala evaluates salient stimuli, associated with arousal and stress reactions. Because auditory cortex is connected directly to the amygdala, the processing of music does not necessarily depend on cognitive interpretation [28]. Activation of the amygdala stimulates the ANS, particularly through sympathetic nervous system (SNS), leading to physiological changes, such as increased heart rate (HR), respiration, and sweating. [42] Additionally, the hippocampus links the context of a stimulus with memory, which explains why context-dependent musical experiences— such as those associated with personal memories— can activate the amygdala and indirectly stimulate the SNS. Finally, the insula acts as a link between the ANS and emotional processing, integrating information about ANS into conscious awareness and emotional control. [28] [45, pp. 926–970]

Regarding the regulation of these responses, the prefrontal cortex influences the amygdala through top-down regulation, shaping the expectations and interpretations of auditory stimuli. It evaluates emotions and can indirectly lead to a reduction in heart rate and respiration during stressful situations. Ultimately, the hypothalamus acts as a mediator, linking the limbic and cortical emotional processing to physiological responses, making it possible to measure emotional reactions to various stimuli by examining physiological changes in the body. [45, pp. 337–361] [47]

Conditions such as anxiety, depression, and psychosis typically share certain neurological features related to disturbances in brain network regulation, imbalances in neurotransmitter systems, and disruptions in the integration of sensory and cognitive functions [51] [52]. Therapeutic effects of music are processed through the same brain regions that have been shown to influence mental health disorders, such as the limbic and paralimbic systems, the dopaminergic reward system, and neuroplasticity related to memory integration and learning [17] [18] [53] [54] [55]. This overlap between the neural processing of mental illnesses and music processing suggests that music therapy can modulate these interconnected body systems, forming the foundation for music-based clinical applications. [46].

Simultaneously, the ANS reacts to musical stimuli by regulating physiological arousal, which is closely linked to central nervous system activity. [47] [28] [45, pp. 926–970]. The direct connection between limbic processing, hypothalamic regulation, and the ANS provides the basis for measuring responses to musical stimuli through physiological signals. [45, pp. 926–970]

2.2.3 ANS responses and biosignal acquisition

The autonomic nervous system is responsible for regulating the body's internal homeostasis, the cardiovascular system, and stress responses by operating through two distinct divisions: the sympathetic nervous system and the parasympathetic nervous system. [56] These two divisions typically innervate specific organs and function in opposite ways to maintain a dynamic balance between arousal and relaxation. [45, pp. 795–804] [57] Activation of the ANS drives various responses in the body related to arousal regulation, in which the SNS drives the "fight or flight" response by increasing the physiological activation to prepare the body for various situations. This increased activation is associated with the release of noradrenaline from sympathetic nerve endings, which increases the rate of cardiac and respiratory activity [45, pp. 795–804], [56]. The PNS, in turn, takes responsibility for the "rest-and-digest" state by promoting relaxation and recovery. [57] Parasympathetic control of the body depends largely on the activity of the vagus nerve, which releases acetylcholine to inhibit SNS activation in organs that are innervated by both ANS divisions, such as the heart and lungs, thereby reducing physiological arousal. [45, pp. 795–804] [58]

Through the direct connection between the limbic system and the ANS via the amygdala, the emotional interpretation of musical elements actively influences the balance between the SNS and PNS. For example, the tempo and rhythm of music have a powerful effect on ANS activity, where simple rhythms, small changes in dynamics, and slow tempos promote PNS activation, leading to a decrease in heart rate and an increase in heart rate variability (HRV) [47], [59]. In contrast, faster tempos, higher pitches and stronger energy distribution – such as strong transients and high RMSE – can increase ANS arousal and thereby trigger the SNS activation [42]. Furthermore, respiration can involuntarily synchronize with the tempo of the music, while cardiac and respiratory modulation are influenced by this rhythmic synchronization. [47] Even though cardiac and respiratory functions are highly regulated by the balance between SNS and PNS, electrodermal activity (EDA) serves as a distinct indicator that can reveal experiential responses to emotional arousal and excitement through its unique interaction with sympathetic control. Since EDA is regulated exclusively by sympathetic activation, it reveals arousal reactions through the activity of skin sweat glands only in relation to sympathetic responses without being influenced by parasympathetic activation. [60] [61] [59]

In order to utilize these various physiological responses for analysis, these biological processes must be recorded and transformed into digital signals. In this context, biosignal acquisition is focused on extracting measurable features from cardiovascular, respiratory, and electrodermal activity that reflect the regulation of the subject's autonomic nervous system.

Measuring cardiac activity reveals insight into the vagal tone, where heart rate and heart rate variability (HRV) can serve as indicators of emotional regulation. Heart rate variability is largely driven by respiratory sinus arrhythmia, which reflects the bidirectional interaction between heart rate and respiration [58] [62]. Since cardiac regulation is not only a peripheral reaction of the body but is closely related to central nervous system activity, it has been shown that HRV reflects the regulation of the prefrontal cortex and the insula, providing a connection to the measurement of limbic system activity. [63], [64]

To capture the cardiac activity, electrocardiography (ECG) is a method used to monitor electrical changes caused by depolarization and repolarization in the heart muscle via electrodes attached to the skin [65]. The recorded signal typically consists of a repetitive waveform, called the P-QRS-T complex. In simple heart rate monitoring, which aims to produce a clear QRS-complex for R-wave detection, the heart's electrical activity is measured using a single lead — typically Lead II in a three-lead system — while any additional leads serve as components to improve signal quality in more challenging measurement purposes. [66] [67]

The interval between successive R-peaks can be analyzed based on temporal distance in order to calculate metrics such as heart rate and heart rate variability, as well as their time-domain metrics such as the standard deviation of interbeat intervals (SDNN), the root mean square of successive interval differences (RMSSD), and the ratio of heart rate differences greater than 50 ms (pNN50). [68]

Additionally, photoplethysmography (PPG) is a method for measuring cardiac activity, in which optical data on changes in blood volume is collected from microvascular tissues, such as the fingertip. An advantage of PPG is its ability to provide data on arterial pressure fluctuations. These fluctuations can be evaluated through systolic rise time by measuring how quickly blood flows through a vessel. This rise time is a sensitive ANS reaction, which strongly correlates with blood pressure levels associated with the emotional valence. [69] [70]

Respiration activity is a significant factor in the measurement of ANS activation due to its bidirectional relationship with cardiac regulation. Respiratory rate and the intervals between breaths reflect emotional states, where a high breathing rate and shorter intervals are typically associated with SNS activation and vice versa. [71], [72]

A respiration transducer is a device that uses a displacement transducer based on the principle of a strain gauge. It measures changes in the chest or abdominal area through the mechanical stretch. When the subject breathes, the expansion of the chest area causes linear stretch, which is converted into digital data. With this device, respiratory rate, its variability, relative changes in lung volume, and the duration of inhalation and exhalation can be recorded. [73], [74] An electrodermal activity sensor is a device that measures the skin's electrical conductivity from the fingertips or palm. EDA data is typically analyzed in two main components: skin conductance level (SCL) and skin conductance responses (SCR). SCL acts as a tonic component, representing the slowly changing baseline level of skin conductance, whereas SCR acts as a phasic component representing rapid temporary spikes in response to SNS activation. [75]

According to previous studies, measuring emotional states based on the ANS responses has revealed promising results, particularly when utilizing cardiac and respiratory activity, where the balance between SNS and PNS activation can serve as a predictable variable in music-related tasks [76], [77]. However, these physiological responses are highly individual, which is why they must be analyzed on an individual basis rather than assuming a single physiological response pattern occurs in all people. This subject-specific variability of physiological features highlights the importance of personalization in data-driven analysis or machine learning models. [6], [76]

2.3 Machine learning for personalization

The main purpose of personalization is to achieve better results in terms of specificity and relevance for individual users, patients, or specific groups, depending on the context. [78] In the medical context, machine learning offers significant potential for this by identifying patterns in clinical single or multimodal data, enabling more precise individualized treatment approaches and profiling aimed at improving clinical outcomes. [2] [79] As mentioned in the Chapter 2.1.1 music therapy depends heavily on the patients' personal preferences. While music therapy is strongly linked to high-evidence studies, traditional clinical heuristic or rule-based personalization solutions can be limited by several factors, such as the type of illness, used clinical methods, the type of played music, and other ongoing treatments. An effective machine learning model can enable higher scalability in personalization, allowing for evaluation in certain areas beyond traditional personalization solutions, for example, where a deeper understanding of musical elements provides value. [2], [6] [80]

2.3.1 Feature extraction and data hierarchy

In order for a machine learning model to understand the meaning of music structure, audio-data is required to be extracted into different acoustic features [81], based on the physics which was discussed in Chapter 2.1.2.

Music and audio analysis tools, such as the Python library called Librosa, can be used to analyze and extract musical features based on frequency and time domains, and to generate representations such as spectrograms, tempo, and harmonic-percussive components [38]. This audio-based information can be represented in several hierarchical levels: the lowest level (raw spectral and temporal variation), the middle level (interconnected music properties like rhythm or dynamic structure), and the highest level (meaningful features such as rhythmic regularity, musical dynamics, or elements that affect mood). [82] [83] Now that acoustic features can be used as input data for a model, the corresponding target features must be defined based on physiological features discussed in Chapter 2.2.3.

In music therapy approaches, physiological responses to various musical elements can vary significantly across different patients. Factors such as cultural dynamics, changes in personal experiences, and environmental influences can affect how a patient experiences music over time [6]. Because the topic is highly subjective and physiological data is not independent, traditional machine learning models cannot be used to generalize effects across the entire population [84]. Since many statistical models are based on the assumption that observations are independent and identically distributed, this assumption cannot be applied to physiological data, as observations within subjects are correlated with each other and there are real baseline differences between subjects [84]. Treating the data as independent would artificially inflate the sample size and lead to overoptimistic estimates of the model's performance [85]. Processing small and interdependent datasets requires different approaches in the model architecture and evaluation strategy. With a small amount of data, the model is at high risk of overfitting, in which case it may learn more from noise or these subject-specific artifacts. This leads to unreliable statistical test results, such as biased p-values and R^2 values. [84]

This hierarchical data can be structured into three levels: observations, sessions, and subjects. (See Figure 2.3) In physiological measurements, multiple records taken from the same subject strongly correlate with one another due to a shared physiological baseline, mood, and sensitivity to stimuli [84]. For this reason, the data must be analyzed hierarchically. A single observation represents a physiological measurement recorded while a stimulus was presented. Sessions contain session-specific observations, which are grouped under subjects, reflecting the individual variation in physiological characteristics. [86, pp. 237–250]

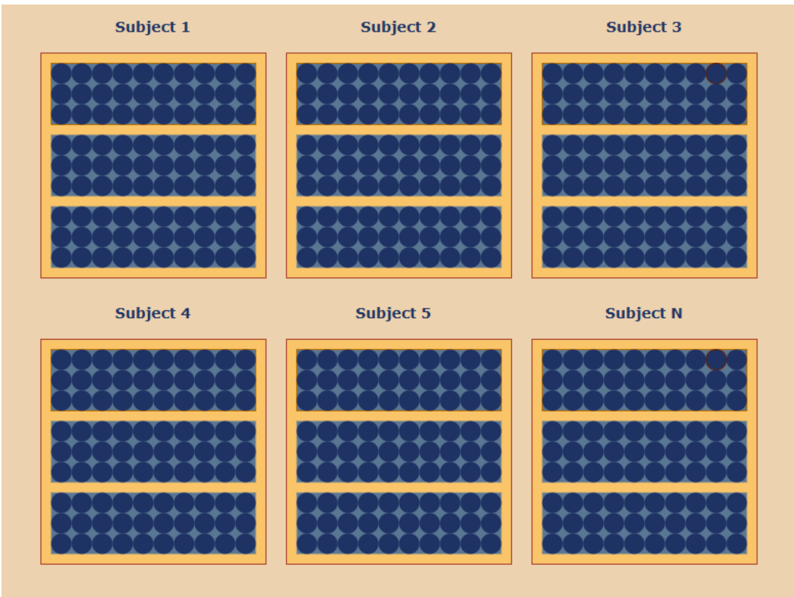

Figure 2.3: A three-level data hierarchy: the effects of subjects (yellow), sessions (light blue), and observations (dark blue) on their inter-dependencies.

In order for the machine learning model to identify the best possible response for an individual person, the model must be adapted. Theoretically, by updating the population-level model with subject-specific data and using their intercepts as random effects, the personalized model can be more flexible and produce better results for a single subject [85]. Personalizing the model design using these subject-specific intercept values is an excellent solution for creating a more personalized analysis based on a population-level model design [85]. Since the data hierarchy has three levels, the evaluation of the models must be made at two level: the population level and the subject-specific level, where both levels provide their own insights into the relationships between acoustic features and physiological responses. The population level evaluates how relationships between acoustic features and physiological responses occur across subjects, providing a baseline understanding of patterns identified in the data. Subject-level evaluation focuses on internal relationships by evaluating models separately for each subject, where the goal is to identify patterns existing within a single subject, supporting the feasibility of the personalized model approach [85]. Together, these methods provide an understanding of both shared and personalized effects from the data. This requires combining knowledge of acoustic features with measurable physiological responses, which forms the basis for modeling individual reactions in receptive music therapy. [76]

2.3.2 Statistical analysis and predictive modeling

Pearson's correlation matrix is a method used to reveal which variables in the dataset correlate with one another. Correlation reveals positive numerical connections between features, whose magnitude is calculated using the formula of:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

where cov is the covariance, σ_X is the standard deviation of feature X, and σ_Y

is the standard deviation of feature Y.

Correlation does not reveal a true cause-and-effect relationship, meaning that even though the numbers indicate a mutual relationship, the actual effect of the relationship is not necessarily due specifically to this relationship. [2] [87, pp. 59–62]

When calculating correlation values for numerous different feature pairs, the probability of obtaining statistically significant results due to chance alone increases significantly. To account for this, the false discovery rate (FDR) correction can be applied as a verification method, in which all rejected null hypotheses are examined based on the observed correlation coefficients so that false discoveries are not included in the statistically significant results. The FDR correction can be used to tighten the requirements so that stronger evidence is obtained to support the statistically found results. [88] [89, pp. 553–573]

If features in the data are strongly correlated, the data can be reduced using principal component analysis (PCA), in which correlated features are combined into new components, called principal components (PC). PCA reduces the overall complexity of the data and preserves meaningful information through eigenvectors. It is used to reduce multicollinearity, enabling a more stable model and preventing overfitting by finding the direction in which the variance of data is maximized. [87] [90, pp. 205–220]

PCA requires normalized data to calculate covariance-matrix, which is used to determine the variance of the eigenvector direction, as shown by the mathematical formula:

$$\Sigma = \frac{1}{n} X^T X$$

$$\Sigma v = \lambda v$$

where v is an eigenvector and λ is the variance in the direction of the eigenvector. [90, pp. 205–220]

A mixed-effects model can be used to account for repeated measurements across different subjects. In this approach, predictor features are treated as fixed effects, reflecting their shared influence on the responses of the target feature at the population level [86, pp. 237–250]. According to research about mixed-effects models applied to the physiological data, subject-specific model adaptation is crucial, as it prevents the effects of actual stimuli from being masked by large variations in baseline levels across subjects. [85] For this reason, the mixed-effects model is a powerful choice compared to standard linear regression models. The mixed-effects model uses a randomly determined intercept to capture the variation in physiological signals across individuals, allowing for their individual baseline level. [84] Mathematically, the mixed-effects model is defined as follows:

$$Y_{ij} = \beta X_{ij} + u_j + \epsilon_{ij}$$

where Y_{ij} represents the target variable for observation i and subject j, X_{ij} represents the vector of selected predictor features, β represents the coefficients of fixed-effects, u_j is the subject-specific random intercept, and ϵ_{ij} is the residual error term. As a model estimator, restricted maximum likelihood is used to estimate the variance effect of the random intercept. [85]

In subject-level modeling, where the dimensionality of the features, small sample size, and strong correlation between physiological features increase the risk of multicollinearity and overfitting in traditional modeling approaches, ridge regression can be a suitable solution to this challenge. The ridge regression aims to minimize the sum of squared residuals and includes an L2 regularization term that penalizes large coefficient values, leading to more stable model estimates. [91] [89, pp. 237–250] This approach is examined separately for each subject revealing possible relationships at the individual level without interference caused by inter-subject variability.

The model performance evaluation with subject-level physiological data must be performed using K-fold cross-validation, utilizing individually partitioned data across sessions. In the cross-validation, all subsets are processed such that each fold is used once for both training the model and testing its performance. Each of these subsets is iterated as many times so that every session has been included in both processes. [89, pp. 197–225] [92] The final evaluation metrics, such as R^2, are computed as the average of all cross-validation rounds.

While cross-validation can be used to estimate the model's predictive accuracy, evaluating the true statistical significance of its performance can be challenging with limited and interdependent datasets. In such cases, permutation testing can be used to empirically evaluate its performance. In permutation testing, the target features are randomly shuffled and the model is run numerous times to break any true relationships with the predictors, after which the null distribution of the evaluation metric is computed. Finally, the model's actual performance is compared against the computed distribution, which provides a p-value that confirms whether the signal is genuine and based on statistically significant responses rather than random noise. [89, pp. 464–469] [93]

3 Data and Methods

3.1 Data collection

A total of 13 volunteer subjects (10 females and 3 males) participated in the study, with each subject completing three separate measurement sessions, resulting in a total of 39 experimental recordings. The participants were aged 18–40 (mean age = 24.3, SD = 3.5) and had no history of serious cardiovascular or neurological diseases. Recruitment of participants was conducted through various channels, such as student association mailing lists, as well as direct invitations. The first participant in the experiment served as a pilot measurement to verify the setup.

All test subjects participated in three sessions where five distinct music genres were played in a random order. The playback order of music genres and individual songs was randomized using a custom Python script. The raw data collected during these sessions consisted of continuous multi-channel biosignals and timestamps of events during the music listening task. For this study's analysis, these raw biosignals were segmented using timestamps related to the experimental events and compiled into a final feature matrix, containing a total of 1178 discrete observations, comprising 21 physiological and 12 acoustic features.

All music used in this experiment was loaded from the Free Music Archive (FMA). This study uses a medium-sized subset of the dataset, which contains a total of 25,000 30-second audio clips from different unbalanced genres, as well as database metadata including basic song information such as name, artist, genre, and possible parent genre. [94]

Data collection was conducted in accordance with the Declaration of Helsinki [95]. All data were anonymized, and no personal data were shared for any purposes. All test subjects and their entries were pseudonymized with a nominal variable, and the final dataset is published under a CC-0 license. During the experiment, all subjects were allowed to discontinue their participation at any time. Only non-invasive methods were used, and the volume of the musical stimuli was adjusted in accordance with general hearing recommendations.

3.1.1 Music stimuli

Five music genres were selected as musical stimuli based on their general musical properties: classical music, jazz, hip-hop, electronic music (aimed at techno music), and experimental music (aimed at noise music). Since all the songs were selected randomly and genres can vary widely, a selection algorithm was needed in the song randomization program to ensure that selected songs remained as similar as possible. The program dividing songs into the song pools used metadata of FMA to determine which genre group each song belongs to. Since not all the songs actually matched their genre labels, an audio analysis algorithm in the program aimed to evaluate all songs based on whether they contained the most typical acoustic features of each genre. The selection criteria were based on the theoretical foundation of the typical properties of each genre.

For the selection, classical music was defined to contain songs with a slower tempo and a wide spectral centroid. For selecting songs under jazz, a wide range of tempos and spectral centroid values was defined. For hip-hop, both the minimum limit of the tempo and the spectral centroid were set very low. Since the FMA dataset contained only a genre called electronic, which included numerous subgenres, techno music was re-extracted from the dataset using parameters with a tempo range of 120 to 150 BPM and spectral centroid range at very high frequencies. Finally, experimental music was re-extracted to include noise and unpleasant music in order to examine the effects of general musical chaos, where the spectral centroid and RMSE level were set very high. The conceptual basis for genre selection has been presented in Table 3.1, while Table 3.2 defines the selected acoustic threshold values used in the selection algorithm.

Table 3.1: Selected music genres, their expected physiological effects, and corresponding selection criteria.

Genre	**Arousal**	**Selection criteria**	**Reference**
Classical	↓	Low tempo, wide centroid	[42] [96] [97] [98]
Jazz	↑	Broad tempo and spectral ranges	[97] [99] [100]
Hip-hop	↑↑	Low tempo and low centroid thresholds	[101]
Electronic	↑↑↑	High tempo and high-frequency thresholds	[98] [102]
Experimental	↑↑↑↑	Extreme centroid and RMSE thresholds	[103] [100]

Table 3.2: Defined acoustic features for the selection algorithm.

Genre	**Tempo**	**Spectral centroid**	**Mean RMSE**	**Spectral rolloff**
Classical	70 – 110 BPM	900 – 2100 Hz	0.02 – 0.1	2000 – 4500 Hz
Jazz	75 – 170 BPM	1200 – 2600 Hz	0.04 – 0.16	2800 – 5200 Hz
Hip-hop	65 – 110 BPM	850 – 2300 Hz	0.09 – 0.22	2200 – 4800 Hz
Electronic	120 – 150 BPM	2000 – 4500 Hz	0.3 – 0.8	4200 – 8500 Hz
Experimental	70 – 200 BPM	2000 – 4500 Hz	0.22 – 0.65	5500 – 11000 Hz

3.1.2 Measurement setup

The experiment sessions were conducted in a laboratory environment where environmental factors, such as lighting, temperature, and the test subject's position were balanced. The lights were kept on, the window blinds were closed, and the subjects lay on a bed with instructions to remain as still as possible. The total duration of the session was an average of 28 minutes per test subject.

Each measurement session followed a predetermined procedure. The session began with five minutes of silence in order to establish a baseline for the experiment.

After the baseline phase, randomly selected music genres were played in five-minute listening trials. To minimize the effects of different music genres, the listening trials were separated by a two-minute period of silence. Finally, the session ended with a four-minute recovery period, during which the recovery of the ANS following the last stimulus was recorded.

Biosignal data were collected at a sampling rate of 1000 Hz using a multi-channel BioNomadix Wireless acquisition system and AcqKnowledge software manufactured by BIOPAC Systems. The setup used PPGED-R and RSPEC-R amplifiers and their corresponding receivers to record physiological responses from the subjects. The EDA signal was measured from the subjects' palms using standard Ag/AgCl ECG electrodes, while the PPG signal was measured from the middle finger. Respiratory activity was recorded using a respiration transducer belt positioned around the subjects' diaphragm area. For the 3-lead ECG, electrodes were placed below the left clavicle (red, positive), below the right clavicle (white, negative), and around the lower left ribcage (black, ground). Before the data acquisition, the PPGED-R receiver was calibrated in the AcqKnowledge software.

3.1.3 Feature extraction

The time segmentation of all session phases was labeled automatically with the Python program. A music randomization program was developed to log all changes occurring during music playback including timestamps for the start and end of the baseline, each 30-second track, breaks, and the recovery phase. All phases during the measurement were defined with their own block IDs. This time segmentation was used to synchronize physiological changes with the events in the experiment during the data processing pipeline. The continuous biosignal data collected during the music listening phases were divided into 30-second-long, non-overlapping windows corresponding to the playback of each track. A second Python module was developed to process all biosignals and the acoustic features of the songs to compile the final feature matrix. In this module, all collected biosignal data within these time segments were first preprocessed according to the specific requirements of each signal and finally extracted as features suitable for further analysis.

To synchronize the physiological data with the experiment, the logged block IDs and timestamps were used to align the biosignals with the specific experiment phases. Once the data were loaded and merged, all physiological data for each 30-second track segment were preprocessed using the NeuroKit2 library [104]. NeuroKit2 handled signal filtering, artifact removal, normalization, and resampling. After the biosignals were preprocessed and their physiological features extracted, each played song was processed using the Librosa library to extract its acoustic features [38]. Finally, the extracted physiological data and acoustic features were combined into the feature matrix for further analysis.

All the extracted features from the raw signals and audio files were compiled into a single feature matrix. For this study, the acoustic features were intended to be used as predictor features, while the physiological ANS changes were used as target features. Finally, this matrix included information on physiological changes such as cardiac modulation, respiratory activity, and electrodermal activity. Since audio-based machine learning models often rely on the time-domain features—such as energy distribution, zero-crossing rate, and tempo—these features have been included in the matrix. Features such as spectral centroid, bandwidth, and roll-off, as well as MFCCs from 1 to 5 dimensions, and tonnetz were included as well [37]. A more detailed description of all the features is shown in Table 3.3.

3.2 Analysis of data

This study conducted a deeper analysis of data by using Python and its data analysis libraries, such as Pandas, NumPy, Matplotlib, SciPy, and statsmodels. In order

Table 3.3: Columns of feature matrix and their descriptions

Data Column	Description of column
genre	Genre of music played [1]
Acoustic features	
averageTempo	Average tempo of played tracks in single block [1]
averageMeanRMSE	Average mean value of RMSE (dB) during single block [1]
averageMeanZCR	Average mean value of zero crossing rate during single block [1]
averageMeanSpectralCentroid	Average mean value of spectral centroid during single block [1]
averageMeanSpectralBandwidth	Average mean value of spectral bandwidth during single block [1]
averageMeanSpectralRolloff	Average mean value of spectral rolloff during single block [1]
averageMFCCxMean	Average value of MFCC coefficient x during single block [1] [2]
averageMeanTonnetz	Average mean value of tonal centroid during single block [1]
Physiological features	
meanHR	Average heart rate during the block
meanHRV	Average heart rate variability
HRVSDNN	Standard deviation of NN intervals
HRVRMSSD	Root mean square of heart rate variability
HRVpNN50	Percentage of successive interbeat interval differences greater than 50 ms
breathsPerMinute	Average breathing rate during the block
IBI	Average inter-breath interval
IBISD	Standard deviation of inter-breath intervals
meanInhaleDuration	Average duration of inhalation phases
meanExhaleDuration	Average duration of exhalation phases
meanRespiratoryAmplitude	Average amplitude of the respiration signal
SCL	Average skin conductance level during the block
SCRCountMinute	Number of skin conductance responses per minute
SCRMeanAmplitude	Average amplitude of skin conductance responses
meanPulseRate	Average pulse rate per minute
meanPRV	Average pulse rate variability
PRVSDNN	Standard deviation of pulse-to-pulse intervals
PRVRMSSD	Root mean square between pulse intervals
pulseMeanAmplitude	Average amplitude of pulse
pulseAmplitudeSD	Standard deviation of pulse amplitude
systolicRiseTime	Average time from pulse onset to systolic peak

[1] Value is NaN for blocks where no music was played (baseline, break, recovery).
[2] In fact, the average MFCC values for the matrix are all in separated columns.

to successfully understand the data, it is crucial to identify any patterns and trends the MT-dataset may reveal, which can used to build a connection between its features. Analytical methods can be used both to identify statistical relationships and thereby to provide initial information for more effective modeling strategies. This study uses analytical tools such as Pearson's correlation analysis, regression analyses, and statistical tests to answer for the research question whether there is genuine relationship between ANS and acoustic features.

3.2.1 Processing and preparation

Initially, the features in the matrix were examined using descriptive functions and visualizations to identify invalid values, outliers, and scale differences. This exploratory analysis led to the processing pipeline of four different tasks: handling invalid and outlier values as well as performing baseline correction, polynomial detrending, and Z-score normalization.

Invalid values and outliers were handled in the feature matrix, as it contained missing values in acoustic features during periods when music was not played. Rows containing acoustic features were separated into a distinct new subset, excluding baseline data, which was required for baseline correction. Outliers were identified through descriptive statistics (See Section 4.1) and imputed with the subject's median value for that feature, if the feature exceeded a manually selected threshold value based on physiological literature.

Baseline correction was performed by subtracting the session-specific baseline value from all experimental blocks. This resulted in adjusted values that reflect physiological changes during the experimental trials rather than absolute baseline levels. [105] [106]

Polynomial detrending was applied because the EDA values were strongly influenced by the slowly changing baseline level of the tonic component. During the

28-minute experiment, the subject's body likely began to relax, which was reflected in a continuous natural decline of the baseline level. [107]

Finally, Z-score normalization was performed to account for significant scale differences between physiological and acoustic features, such as heart rate values (60–100) and MFCC values (-250–40). Normalizing these values is essential, particularly for regression models and PCA, to ensure that features representing large orders of magnitude do not excessively affect the model weights. [87, pp. 48–59] [108]

3.2.2 Statistical relationship analysis

Statistical relationship analysis was performed in two distinct stages: first at the group-level to investigate whether direct linear relationships between the acoustic features of music and physiological responses exist, and second at the subject-specific level, using each subject's own data to determine how these relationships appear individually. For these purposes, Pearson's correlation analysis was selected, as it is highly suitable for determining relationships between continuous numerical values. All correlation values were first calculated for the group-level data to reveal potential general relationships, such as whether tempo is associated with the activation of heart regulation across all subjects.

A separate analysis was then performed for each subject based on their own data to determine whether subjective associations between these features exist. For example, a single subject might react strongly to differences in musical energy, even if the group-level analysis does not reveal that relationship as significant.

In each of these analyses, statistical significance (p-values) was calculated to verify whether the observed relationships were statistically genuine rather than based on chance or random noise. Furthermore, the false discovery rate (FDR) correction was applied to the individual analyses to ensure that the large number of calculated Pearson's correlations did not lead to random false positives.

3.3 Modeling approach

The main objective of the modeling stage is to investigate the detection of relationships between the acoustic features of musical stimuli and their impact on emotion-related ANS responses. Unlike typical machine learning applications, the primary goal is not effective generalization at the population level, but rather to analyze how these meaningful relationships vary across individuals. Instead, the goal is to identify potential relationships, estimate the magnitudes, and ultimately evaluate the method itself, rather than draw conclusions about the generalization of the models' effects.

As the collected data consists of physiological data which is known to be sensitive to individual differences, a single, generalizing model may fail to capture the variation among individuals. Therefore, the modeling method is required to be designed to account for both within-subject dependencies and between-subject variations.

The modeling strategy is divided into two perspectives that can validate each other, but also reveal potential conflicts between them. First, population level modeling is used to estimate the overall relationship as baseline information. Second, subject-level analysis of internal connections within the subjects is conducted to determine the feasibility of a personalized modeling approach. Combining these modeling approaches provides insight into whether the effects are significant across the general population and whether genuine patterns exist within individuals.

3.3.1 Feature selection and dimensionality reduction

Since the dataset contains a large number of diverse features that do not necessarily reveal how the body's physiological responses are influenced by acoustic features, feature selection was used to reduce the dimensionality of the model data significantly. In this model construction, feature selection was based on the information from Pearson's correlation matrix (See Section 4.2.1).

Based on these correlations, a new subset of acoustic features was selected, the first two MFCCs, spectral bandwidth, spectral rolloff, and RMSE. These features primarily represent the timbre and spectral features of the musical stimuli. Interestingly, RMSE, which describes the energy distribution caused by musical transients and kicks, correlated well for certain subjects. As the target features, variables such as inter-breath interval, systolic rise time, mean pulse rate variability, respiration rate, and mean heart rate were selected.

Since many of the acoustic features in the feature matrix were strongly correlated with one another, PCA was used alongside manual feature selection, which allowed the models to capture the full acoustic variance of the dataset while effectively removing redundancy in the data. The use of both of these methods in this study provides insight into how well unsupervised design could perform in this type of modeling approach.

3.3.2 Population-level model

Since the multilevel hierarchical data structure of the MT-dataset contains numerous observations from different measurements and subjects, it presents various challenges that must be taken into account when designing a population-level model. Applying a traditional linear regression model to this dataset would be inappropriate, as such models assume that data points are independent of one another. Because physiological signal baselines vary significantly among individuals, treating all observations as independent would cause a systematic bias that masks the actual physiological modulation caused by the acoustic features.

To resolve this, a linear mixed-effects model was applied to the dataset, designed directly to adapt separate stimuli's effects from individual variance. The acoustic predictors, which were manually selected in Section 3.3.1, were used as fixed effects to estimate their overall impact on the physiological target variable across the entire population. At the same time, individual subjects were used to determine subject-specific random intercepts. This is critical for allowing each subject to retain their own physiological baseline within the model and ensuring that the estimates are not distorted by inter-subject physiological variability.

Because the collected data in this study is limited in size and lacks the statistical power to support complex random slope structures, only random intercepts were included in the model. The model was ultimately fitted using the restricted maximum likelihood estimator, which provides reliable variance estimates for random effects in small sample sizes.

3.3.3 Subject-level model

In addition to the population-level analysis, another model was developed to evaluate potential relationships at the individual level. The primary objective of this approach was to examine each subject independently, thereby evaluate the results separately from inter-subject variability. For this task, a subject-specific ridge regression model was developed, as it is better suited for small sample sizes and helps to reduce potential overfitting when the predictor features are strongly correlated. [91]

The subject-level modeling followed a straightforward pipeline to evaluate model stability and validity. First, all manually selected features were used in the model performance analysis, after which the performance was analyzed separately using PCA components as predictors.

The model's predictive accuracy was evaluated using five-fold cross-validation, in which the data for each subject was shuffled and split into five separate subsets, where a total of 80 % of the data were used for training and 20% was used for testing in each evaluation round. Evaluation rounds were performed until each subset had been used exactly once for a testing. Finally, the model's performance was computed to produce its R^2 evaluation metric, which was averaged based on these cross-validation rounds to form the final performance statistics. [89, pp. 197–225] [92]

Finally, permutation testing was performed 1000 times to ensure that the observed relationships were not due to random noise in the small dataset. From this process the statistical significance was calculated for different target features by utilizing the results of all permutation rounds. The same permutation test was performed for both the manually selected predictors and the PCA component predictors, which allowed the verification of whether music's acoustic features had a statistically significant impact on the subjects' physiological responses. [89, pp. 464–469]

4 Results

The results section first presents the information obtained from the exploratory and statistical analyses to evaluate the quality of data and statistical relationships directly related to the raw numbers. After that, the performance of the machine learning models has been analyzed at both the population and individual levels.

4.1 Exploratory analysis

The exploratory analysis revealed that the feature matrix contains a total of 156 missing values on acoustic features during the baseline, break, and recovery periods, as no music was played during these periods. However, these periods, except the baseline, were not used in this study, while the analysis focused on periods, where music was played, and baseline was used in the baseline correction.

The physiological data however revealed total of 14 missing rows in the respiration activity features and 3 missing rows in the EDA peak amplitudes. This happened because the data processing pipeline for these extraction functions prevented data creation if a sufficient number of peak or trough values were not detected during music playback.

As shown in Table 4.1, the data clearly contains outlier values that had to be taken into account, such as negative and extremely high or low values for several features. Furthermore, since the feature matrix contains physiological data that varies significantly among different subjects, and since the study focuses on physiological changes rather than on the exact feature values between subjects, data standardization was necessary.

Table 4.1: Descriptive statistics of the physiological features

Feature	N	Mean	STD	Min	25%	50%	75%	Max
meanHR	1177	69.21	11.14	47.66	61.73	67.26	74.05	112.35
meanHRV	1177	892.30	130.57	534.27	813.53	897.05	980.28	1261.65
HRVSDNN	1177	58.48	36.27	9.67	32.07	45.59	81.70	233.57
HRVRMSSD	1177	59.29	45.38	5.27	26.92	41.96	83.29	324.00
HRVpNN50	1177	27.16	24.85	0.00	3.57	19.35	48.39	86.67
breathsPerMinute	1163	14.86	4.35	4.88	11.53	15.03	17.78	40.51
IBI	1163	4.44	1.50	1.48	3.37	3.99	5.20	12.30
IBISD	1163	0.57	0.55	0.00	0.23	0.40	0.74	4.59
meanInhaleDuration	1163	2.17	0.98	0.59	1.52	1.91	2.53	10.78
meanExhaleDuration	1163	2.28	0.89	0.70	1.72	2.07	2.59	8.91
meanRespiratoryAmplitude	1163	2.66	2.20	0.01	1.01	1.91	3.67	11.76
SCL	1177	9.12	10.89	-16.78	2.21	4.30	14.78	49.54
SCRCountMinute	1177	23.45	23.02	0.40	7.95	15.89	29.84	118.95
SCRMeanAmplitude	1174	0.599	1.540	0.000	0.007	0.011	0.652	35.551
meanPulseRate	1177	70.22	11.36	49.97	62.65	68.05	74.93	112.39
meanPRV	1177	0.89	0.13	0.53	0.81	0.90	0.97	1.24
PRVSDNN	1177	0.09	0.10	0.00	0.04	0.06	0.11	1.51
PRVRMSSD	1177	0.11	0.14	0.00	0.03	0.06	0.13	2.16
pulseMeanAmplitude	1177	0.67	0.49	0.01	0.18	0.66	1.00	2.01
pulseAmplitudeSD	1177	0.09	0.08	0.00	0.03	0.07	0.14	0.68
systolicRiseTime	1177	0.40	1.16	0.17	0.27	0.33	0.43	38.83

4.2 Statistical analysis

The results of the statistical data analysis are presented at two different levels using the corresponding correlation statistics. The general relationships across subjects have been presented at the population level with a Pearson's correlation matrix, and individual differences by examining the correlations at the individual level.

4.2.1 Population-level correlations

The statistical analysis of this study reveals potential direct relationships between physiological and acoustic features. The Pearson's correlation matrix in Figure 4.1 indicates that the highest correlation coefficients ($r \approx 0.12 - 0.15$) are observed in associations with respiration activation, primarily regarding acoustic features such as mean values of MFCC1, spectral bandwidth, spectral rolloff, spectral centroid, RMSE, and zero-crossing rate, many of which are statistically significant ($p < 0.05$). In addition to respiration activation, EDA and some PPG values show correlation as well.

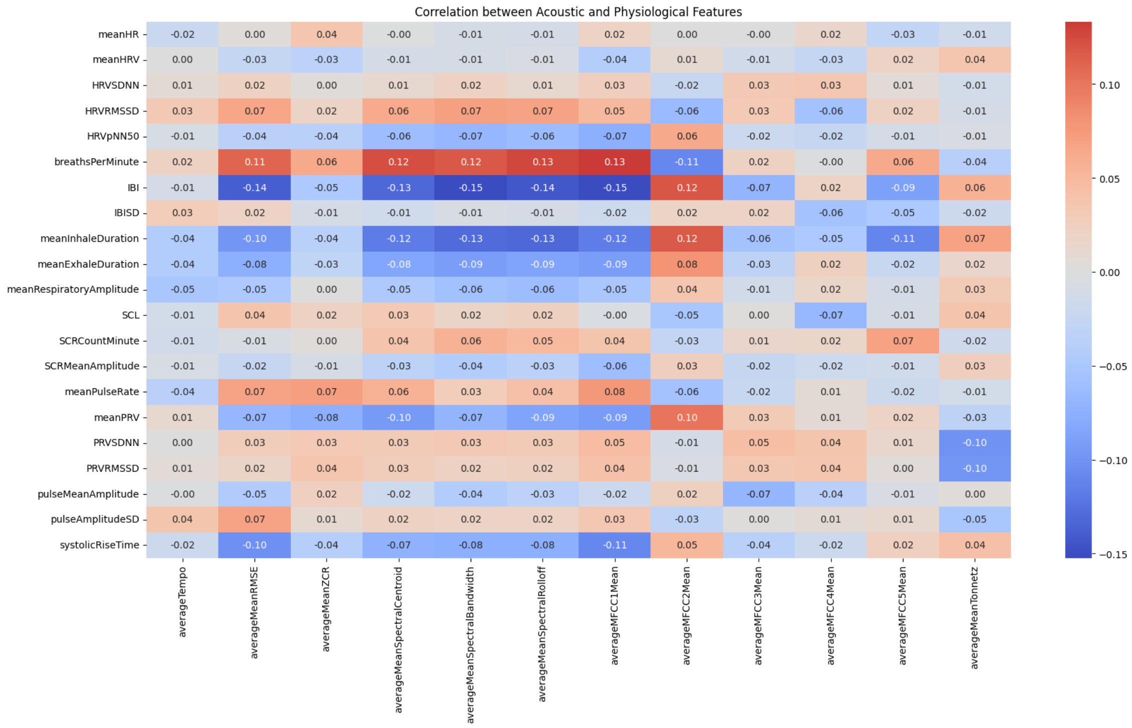

Figure 4.1: Pearson's correlation matrix of the numerical values in the feature matrix

At the population level, the three strongest correlation pairs with different target features (shown in Figure 4.2) were the breathing interval, which correlated most strongly with MFCC1 ($r \approx 0.152$, $p \approx 0.000001$), breathing rate with MFCC1 ($r \approx 0.133$, $p \approx 0.00002$), and inhaling duration with spectral rolloff ($r \approx 0.133$, $p \approx 0.00002$).

4.2.2 Subject-specific responses

In the subject-specific correlation analysis, stronger correlation coefficients were observed for several acoustic features, and many of the features that were dominant in

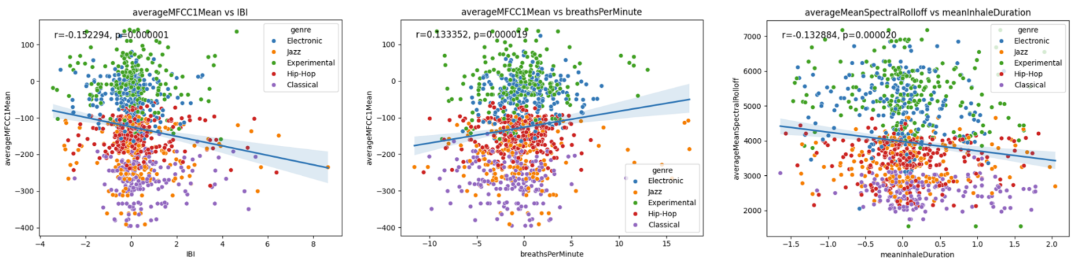

Figure 4.2: Scatter plots of the three strongest correlations between acoustic and physiological features.

the population-level analysis appeared to be dominant at the individual level as well. A new finding in the subject-specific responses, compared to the population-level analysis, was that heart rate was also a significant physiological feature showing correlations in certain subjects.

After applying the FDR correction to all subjects, a total of 35 significant correlations out of 97 passed the FDR correction (See Table 4.2). The FDR correction in this analysis strongly supports the assumption that there are robust correlations passing a strict testing process even at the statistical level, meaning that ANS responses are not random noise, but rather, real observable relationships between features.

When comparing certain subjects in the dataset, it was observed that subject 5, 9, and 13 had the highest correlation values, and their individual strongest correlations were all close to $r \approx 0.46 - 0.5$, with FDR-corrected statistical significance. Interestingly, however, all of these subjects with the highest correlations had different physiological features that reacted to a specific acoustic feature. For example, subject 5 had the highest correlation pair between breathing interval and MFCC1 ($r = -0.502,\ p\_corrected = 0.0009$), subject 13 between the tonic EDA component and spectral bandwidth ($r = -0.488,\ p\_corrected = 0.0010$), and subject 9 between heart rate and MFCC1 ($r = 0.466,\ p\_corrected = 0.0019$).

Table 4.2: Number of subjects with statistically significant correlations after FDR correction between acoustic features and physiological signals.

Physiological	MFCC1	MFCC2	MFCC3	RMSE	SpecBW	SpecCent	SpecRoll	Tempo
IBI	3	2	0	2	2	2	2	0
SCL	1	1	1	2	1	1	1	0
meanHR	1	3	0	2	3	2	2	1

4.3 Machine learning performance

4.3.1 Mixed-effects model

As explained in Section 3.3.1, ANS responses to various acoustic features were examined using a total of five different selected feature pairs that showed the highest correlation based on both population- and subject-specific correlation analyses.

The results obtained from the model are summarized below in Table 4.3, showing how strongly given predictor features influence different target features when all subjects have been analyzed based on their corresponding intercept values. As shown in the population-level correlation analysis, many acoustic features indicated weak population-level correlation with a statistical significance. However, when subject-specific differences are taken into account in the mixed-effects model, their correlation potential increases, while statistical significance remains.

The strengths of the observed relationships were confirmed for spectral centroid mean values. In particular, the effect magnitudes for respiration features, such as breathing interval and breathing rate, increased from population-level correlation of $r = -0.13$ and $r = 0.12$ to mixed-effects coefficients of $\beta = -0.3329$ ($p = 0.0048$) and $\beta = 0.3585$ ($p = 0.0025$), respectively.

In addition to manual feature selection, PCA was used to reduce the dimensionality of the acoustic features. The PCA was able to explain 95% of the variance in the acoustic features using eight principal components, where loadings (See Figure 4.3) of the first principal component primarily included features such as spectral rolloff, spectral centroid, spectral bandwidth, MFCC1, and RMSE. These features

Table 4.3: Mixed-effects model coefficients and p-values for physiological targets

Target feature	**Predictor**	**Coefficient**	**p-value**	
IBI	averageMFCC1Mean	0.0018	0.9816	
	averageMFCC2Mean	-0.0879	0.2199	
	averageMeanSpectralRolloff	0.2497	0.0250	*
	averageMeanSpectralCentroid	-0.3329	0.0048	*
	averageMeanRMSE	-0.0903	0.1417	
systolicRiseTime	averageMFCC1Mean	-0.1159	0.1425	
	averageMFCC2Mean	0.0466	0.5172	
	averageMeanSpectralRolloff	0.0059	0.9579	
	averageMeanSpectralCentroid	0.0203	0.8637	
	averageMeanRMSE	0.0177	0.7739	
meanPRV	averageMFCC1Mean	-0.1898	0.0166	*
	averageMFCC2Mean	0.1280	0.0757	
	averageMeanSpectralRolloff	0.1803	0.1075	
	averageMeanSpectralCentroid	-0.0346	0.7703	
	averageMeanRMSE	0.0436	0.4806	
breathsPerMinute	averageMFCC1Mean	-0.0140	0.8600	
	averageMFCC2Mean	0.1075	0.1364	
	averageMeanSpectralRolloff	-0.2559	0.0225	*
	averageMeanSpectralCentroid	0.3585	0.0025	*
	averageMeanRMSE	0.0985	0.1113	
meanHR	averageMFCC1Mean	0.1849	0.0205	*
	averageMFCC2Mean	-0.0928	0.2009	
	averageMeanSpectralRolloff	-0.1985	0.0783	
	averageMeanSpectralCentroid	0.0827	0.4885	
	averageMeanRMSE	-0.0750	0.2277	

* p-value < 0.05, statistical significance revealed

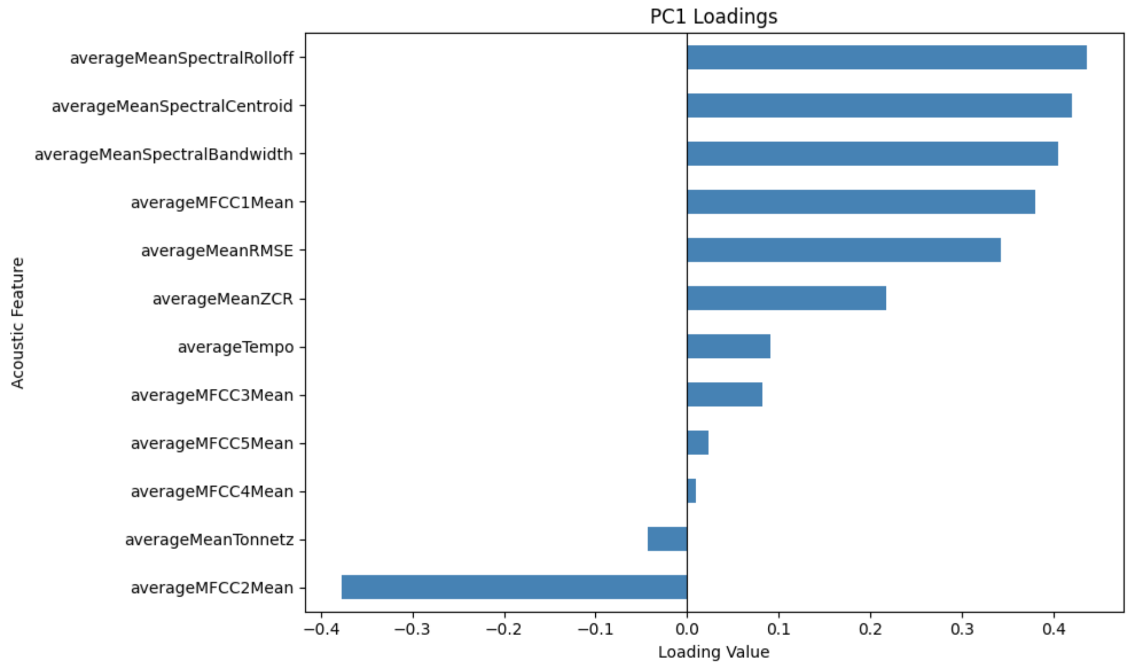

Figure 4.3: Feature loadings of the first principal component (PC1) for acoustic features.

are related to the brightness or energy distribution of the music. An interesting observation regarding PC1 is that MFCC2 and tonnetz have opposite loading values, which is due to their inverse relationship with the other spectral features within the principal component.

The results of the mixed-effects model using various principal components (See Table 4.4) show that the model was able to predict statistically significant signal using PC1 for nearly all target features, even though the model coefficients were significantly lower than in most of the computations based on manually selected features. Furthermore, components PC2 through PC8 did not show any statistical significance, confirming that ANS responses are primarily influenced by the combined low-dimensional representations of musical energy and brightness rather than by other acoustic features.

Table 4.4: Mixed-effects model coefficients and p-values using PCA components as predictors.

Target Feature	**Predictor**	**Coefficient**	**p-value**	
IBI	PC1	-0.0285	0.0410	*
	PC2–PC8	(Not significant)	> 0.05	
systolicRiseTime	PC1	-0.0443	0.0015	*
	PC2–PC8	(Not significant)	> 0.05	
meanPRV	PC1	-0.0410	0.0033	*
	PC2–PC8	(Not significant)	> 0.05	
breathsPerMinute	PC1	0.0267	0.0566	
	PC2–PC8	(Not significant)	> 0.05	
meanHR	PC1	0.0285	0.0429	*
	PC2–PC8	(Not significant)	> 0.05	

* p-value < 0.05, statistical significance revealed

4.3.2 Ridge regression model

The subject-specific model was developed in this study to provide evaluation metrics to determine whether individual ANS responses could be predicted. Figure 4.4 illustrates how the model can capture observable patterns in the three best-performing subjects, indicating that signals are detectable in certain sensitive individuals. However, as the raw cross-validated R^2 values remained below zero for most target features across the entire dataset, it confirms that under these conditions, the model is unable to produce genuine predictions of target features at the subject-level.

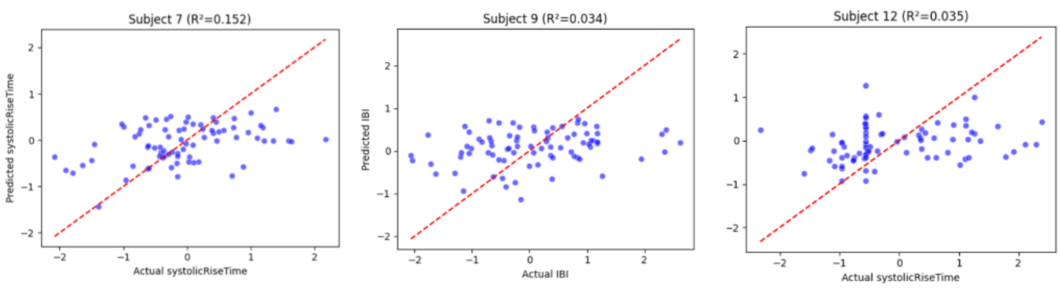

Figure 4.4: Predicted vs. actual values for the three best performing subjects in the subject-level Ridge regression model.

The model's evaluation capabilities were tested using a 1000-iteration permutation test to determine if any real signal could be captured in the data. The results of the permutation test for manually selected features (See Table 4.5) revealed that the model detected a statistically significant signal, particularly in cardiac-related features such as systolic rise time, pulse rate variability, and heart rate, suggesting that a genuine signal can be found in the data.

Table 4.5: Results of the 1000-iteration permutation test for all target features based on the subject-level ridge regression model.

Target feature	**Real R^2**	**p-value**	
IBI	-0.077	0.1179	
systolicRiseTime	-0.054	0.0090	*
meanPRV	-0.066	0.0480	*
breathsPerMinute	-0.087	0.2627	
meanHR	-0.055	0.0150	*

* p-value $<$ 0.05, statistical significance revealed

To further validate this, a permutation test was performed on the same ridge regression model using the principal components as predictors (Table 4.6). The repeated test confirmed exactly the same pattern, where cardio-based responses exhibit a statistically significant signal, whereas respiratory features do not.

Although the pattern remained, the overall R^2 values decreased compared to manual feature selection due to the PCA. This is expected, as PCA compresses the predictive variance into a lower dimensionality. However, the significance of the signal is still preserved in the permutation with PCA.

Finally, individual differences were analyzed in the ridge regression model to illustrate how variations in individual ANS responses affected the data variation. As shown in Figure 4.5, the variation among subjects differs clearly across various acoustic features. For instance, spectral rolloff strongly increases the heart rate of subject 7 ($\beta = 0.437$), but at the same time, it decreases the heart rate of subject 13 with a very strong negative correlation ($\beta = -0.674$).

Table 4.6: Results of the 1000-iteration permutation test for all target features based on the subject-level ridge regression model using principal components as predictors.

Target Feature	**Real** R^2	**p-value**	
IBI	-0.1512	0.3596	
systolicRiseTime	-0.1138	0.0260	*
meanPRV	-0.1040	0.0090	*
breathsPerMinute	-0.1255	0.0849	
meanHR	-0.1124	0.0450	*

* p-value $<$ 0.05, statistical significance revealed

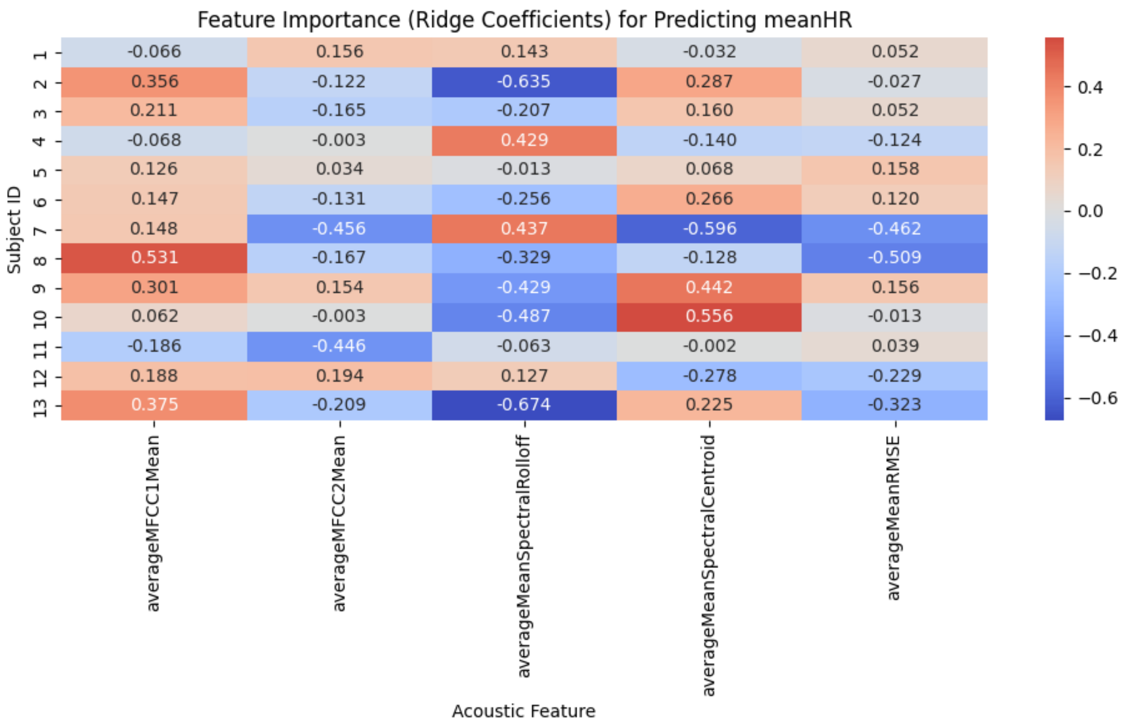

Figure 4.5: Feature importance for predicting target features with manually selected predictors at subject level

4.3.3 Comparison of results

All of these results can now be interpreted as a single entity in which they support one another. Statistical analysis revealed that the strongest correlations between the features were weak, with a maximum of $r \approx 0.15$, whereas the mixed-effects model succeeded in raising the coefficient values to a maximum of $\beta = 0.3585$ while keeping the statistical significance within acceptable limits, $p = 0.0025$. This improvement in the mixed-effects model demonstrated that, to better identify these patterns, it is necessary to account for individual differences among the subjects.

In subject-specific ridge regression, the group-level mean R^2 remained below zero, while certain individuals produced extremely low but positive predictive accuracies, at most $R^2 = 0.152$. Permutation testing validated that the signal is detectable, but it lacks true statistical power across all target features.

5 Discussion

This discussion chapter addresses all three research questions (see Section 1.1) based on the obtained results. The relationship between acoustic features and ANS responses is examined in the context of previous studies, the impact of individual responses and their variability is discussed, and the performance of the developed machine learning approaches is examined in greater depth. Finally, the limitations of this study are reviewed by exploring how experimental conditions, data quality issues, and the methodological challenges of modeling hierarchical physiological data have influenced the results.

The correlation results (see Chapter 4.2) revealed weak but statistically significant correlations between ANS reactions and acoustic features, primarily regarding how respiration and cardiac activity relate to acoustic features associated with the energy and brightness of the music. This relationship is scientifically reasonable, as the harmonic structure of music has been shown to influence these ANS functions [42] [59]. Although some studies suggest that tempo is closely linked with respiration and heart regulation, the statistical analysis of this study does not unequivocally support this pattern. [37] [42] [47]

At the population level, the mixed-effects model used in this study suggests a similar pattern in the relationships between acoustic and physiological features, but with improved coefficient values. When the model accounts for individual baseline levels, stronger correlations can be obtained, suggesting that the strength of the relationship can be better identified using machine learning methods. Statistically, these findings indicate that certain acoustic features related to the musical energy and brightness either activate or inhibit the ANS with a mild influence. Although the correlation values are not particularly high, they indicate that there are dependencies that a machine learning approach could learn to recognize, which is significant in answering the research question of whether personalization can be used to better identify patterns of individual variation.

In the mixed-effects model, PCA applied to acoustic features revealed that the first principal component had a statistically significant correlation with four different physiological target variables: inter-breath interval, systolic rise time, pulse rate variation, and heart rate. As indicated by the loadings of PC1, the principal component contained variance in acoustic features, such as spectral rolloff, spectral centroid, spectral bandwidth, the first and second MFCCs, RMSE, zero-crossing rate and tempo of music. Together both the statistical analysis and the results of the mixed-effects model suggest that factors related to higher energy and spectral brightness may be associated with increased arousal in the ANS and activation of the SNS, patterns on which the literature has focused. [38] [71] [77]

The subject-specific analysis was intended to help answer the question of whether individual responses could be observed with regard to specific acoustic features. The subject-specific statistical analysis revealed highly significant variation in correlations among the study's participants, which was explored even deeper with machine learning methods. As the results in Section 4.2.2 presented, certain subjects stood out from the group with high sensitivity to various acoustic features after FDR correction. These values revealed that the strength of the correlations and the target features appeared to vary heavily among different subjects, which reinforces the need for subject-specific personalization. All the strongest correlations at the individual level were linked to different feature pairs, but according to the PCA, the acoustic features still belonged to the same group of musical nature. This most likely explains why a generalized model design can be challenging.

Individual weights based on the estimates from the ridge regression model were separated, showing that the large differences in data diversity affected different target features differently, but followed the same pattern. Together, these two methods suggest that there are significant individual differences on these factors, where a single person may exhibit increased ANS regulation while others may not. [98]

Features related to cardiac activity or respiratory activity also showed large differences between subjects, which confirms the inter-subject variability. This sensitivity is directly reflected in the subject-level results, where the model failed to achieve positive predictive performance from the available data.

Although the analysis of feature importance in the subject-specific analysis confirmed the importance of individual differences, the performance of the utilized ridge regression model with subject-specific data remained the greatest challenge in this study. As the results in Section 4.3.2 show, the model's cross-validated R^2 values were primarily negative and nearly zero for all subjects, indicating poor predictive accuracy.

However, it is crucial to distinguish between practical predictive power and the existence of a true statistical signal when addressing the third research question, whether machine learning is capable of predicting physiological reactions. The two permutation tests applied in this study represent a significant accomplishment, as they made it possible to detect a statistically significant signal in the data despite low R^2 values. In the permutation testing, the significance of respiration activity as a target feature was no longer maintained, but the significance of cardio activity remained within acceptable limits. This means that the regulation of cardiac activity has a measurable signal in the data, even though its strength cannot be predicted accurately.

The second permutation test was performed using PCA components as predictors, revealing a similar pattern, but the model's performance decreased slightly due to the variance lost in the PCA. Although the R^2 values obtained using PCA decreased compared to manual feature selection, the significance of the associations between the same target features was preserved. This confirms that while the ridge regression model cannot predict precise results with limited amount of data, the significant variance in the data is preserved in these strongly correlated acoustic features, and it can be presented in a more efficient form using PCA. [87, pp. 48–51] [91]

5.1 Limitations

5.1.1 Experimental limitations

Although the results from this study have obtained several findings, they must be interpreted critically, taking into account the possible limitations that may have influenced the results.

First, it is important to consider the limitations that may be associated with laboratory measurements. Even though environmental factors such as stable lighting, posture, and external disturbances were controlled, activities related to other studies being conducted simultaneously in the laboratory may have disturbed the test subject, possibly leading to minor measurement artifacts in some cases. Listening to music in a laboratory setting is also less familiar to participants than in real-world situations, which may affect an individual's ability to concentrate. All measured participants were instructed to remain still in a controlled environment, which removes many contextual factors related to music listening, such as movement, social interaction, and potential task-related listening. All of these effects may reflect in that the generalization cannot be fully applied to natural music listening conditions.

Second, the selection of music used as stimuli can be considered ambiguous, as all the music utilized was sourced from the Free Music Archive dataset and filtered based on specific criteria to represent different genres. Although the filtering worked well according to participant's experience, the music did not account for the subjective or emotional perspectives of the listeners. Unfamiliar, emotionally neutral music may weaken physiological responses compared to music that the participants themselves recognize as meaningful or preferable. However, the use of new, unfamiliar stimuli serves an essential methodological purpose in this study, as it allows for the direct physiological reactivity to acoustic features being isolated from personal memories or strong existing emotional associations. Although this limits direct generalization to clinical music therapy, these limitations do not significantly undermine the results regarding the physiological baseline effects of acoustic properties.

The temporal structure of the experiment also introduces additional limitations. The music used in the experiment consisted of short, 30-second-long music clips with immediate transitions between songs, which may not have allowed sufficient time for physiological reactions to fully develop or stabilize. Because autonomic responses, particularly those related to cardiac or respiratory activity, respond to external stimuli with a delay—sometimes requiring up to a minute for metrics like HRV to stabilize—the true effect of a specific song may not be revealed, which can lead to potential overlap in reactions during transitions between songs. [47], [46] This temporal instability between rapidly changing acoustic features and slower physiological recovery is a major limiting factor that causes the poor predictive power observed in the analytical results.

Finally, even though significant limitations on physiological signals and their analysis have been identified, there is an important limitation that has not been discussed yet. The ANS responses do not always indicate a single specific causal relationship depending on the source of arousal or relaxation. In this study, these two ANS reactions, parasympathetic and sympathetic activity, have been considered to be key factors in determining whether the ANS triggers a direct stress reaction based on the arousal, but it was not taken into account whether the arousal is caused by positive valence increase or stress reaction, as both may reflect similar ANS reactions. [28], [75] Without a clear subjective measurement of how musical stimuli affected the subject's mood, it is challenging to identify the causal nature of the ANS response. The lack of subjective measurement in each experiment may be a weakness of this study, and addressing this could resolve this problem better in future research.

5.1.2 Modeling limitations

Various limitations have a significant impact on the modeling performance and its ability to make interpretations. Machine learning is highly dependent on the amount of data available, which in this study is already limited itself, but also contains challenging features that are not easy to analyze statistically. Physiological data rarely reveals generalizable patterns, as all biological systems have their own unique characteristics. [2] Using acoustic features as predictors also opens up a large feature space where minor elements can easily be hidden and cannot be represented in numerical form.

All the physiological signals used in this study include significant variability across three different levels of the data hierarchy which was discussed in Section 2.3.1. At the observation level, physiological signals are highly sensitive to noise and artifacts caused by different factors such as measurement device connections and movement artifacts, which remove some information already during the data preprocessing stage. At the session level, SCL values showed large magnitude differences (0.001 to 23.5 μS) even within the same subject, as skin sweat gland activity could vary between sessions at different measurement times. Although detrending was performed for the SCL data, it does not remove all the variation between sessions and may, in some cases, even remove meaningful information from the measurement.

In this study, the representation of music was primarily extracted from the lowest level of analysis, with limited addition from the middle-level feature extraction. While the use of high-level representation would allow for a deeper analysis of music, the main focus of this study was to retain these lower two levels to keep the data complexity manageable. However, many musical details may not have been represented well without the high-level analysis. [83] Including these features could potentially explain more about physiological responses to features such as harmonic complexity, melodic motion, and rhythmic structure.

The PCA method was used in the modeling stage to reduce dimensional complexity, which revealed that the acoustic features could be compressed into eight principal components while retaining the 95% of their variance. However, PCA is an unsupervised method that maximizes the total variance of acoustic features, but may at the same time discard certain subtle effects specific to individual acoustic triggers, that are required for accurate predictions. This loss of predictive variance likely reduces the accuracy of the Ridge regression model compared to the use of manually selected features. Although some significant relationships with physiological responses were captured, the current feature space is too narrow to capture the full range of acoustic features from music that could drive physiological responses. This opens up opportunities for future studies to investigate whether high-level musical features are possible to reveal stronger and more interpretable relationships between acoustic features and physiological responses.

The used modeling methods are particularly limited by the overall amount of data. Although the dataset contains multiple observations per subject, the number of independent units in a population-level analysis remains limited. The statistical power of the model to detect correlations with $r = 0.15$ was calculated to be around

8% with a dataset of 13 subjects, whereas the standard threshold for detecting a genuine effect requires around 80% statistical power. The required number of subjects and measurements is possible to calculate, which indicates that the total number of subjects with three measurements would need to increase to up 350 subjects.

The observed population-level effects need to be interpreted as indicative results rather than definitive conclusions, because the sample size of 13 subjects does not support the random intercepts structure in the mixed-effects model well enough. The mixed-effects model used in this study indicates variation in subject-specific components, which are most likely not estimated accurately, resulting in overly optimistic p-values.

The poor statistical power crucially limits the ability of mixed-effects model to reliably identify patterns in a noisy signal, which increases the uncertainty of the estimated coefficients. In the subject-wise analysis, this poor statistical power is an even greater limitation. Since generalizable information based on the mixed-effects model is already difficult to achieve, the subsets used to identify individual patterns are even weaker, which explains the poor predictive accuracy in the subject-specific model analysis.

The goal of the subject-level modeling approach in this study was to evaluate whether individual relationships among physiological and acoustic features could be identified. However, while permutation tests revealed genuine individual-specific signals supporting personalization in the modeling, the obtained cross-validated R^2 values were mostly negative for most subjects across all physiological features. The reason for the poor performance is most likely caused by two major factors. First, the number of observations for each subject was only around 80 music blocks, which, according to the statistical power analysis, is nowhere near the theoretical minimum requirement for detecting a weak correlation effect. Second, as physiological variability has been discussed several times, the patterns identified in one subject do not correlate with other subjects, which is why the personalization was intended from the beginning, but it also explains the reason why more individual data per subject must be collected. Based on statistical power analysis, the required number of subject-wise observations could rise up to 150 to 200 observations, which necessitates a significantly expanded measurement protocol compared to the current three-session study protocol.

As with all machine learning models, the quantity, quality and relevance of the data are significant factors in terms of model performance. In the field of medicine, data must always be collected and used with care due to ethical and privacy considerations. [2] Machine learning models do not understand true causal relationships, while their main task is to evaluate as well as predict numbers and their relationships to ground-truth data. The results given by the model must be interpreted carefully, and in potentially critical applications, the final assessment of the patient must always be made by a healthcare professional. [80] As in statistical analysis, unexplained correlations are always possible, even if they do not actually reveal a positive effect on the task. This can pose significant challenges, particularly in tasks that attempt to answer real-world subjective questions, such as what kind of music actually reduces patient's stress level.

6 Conclusions

The contribution of this study lies in expanding the research on machine learning methods for music and emotional regulation by providing a methodological framework that allows computational audio features to be combined with multimodal physiological data. The study demonstrated how a synchronized acquisition, preprocessing, and feature extraction pipeline can be used to investigate relationships between specific acoustic features and physiological signals.

The obtained results indicate that acoustic features related to the energy and spectral brightness of music are associated with the arousal regulation of the autonomic nervous system, particularly cardiac and respiration activity, at the population level. Furthermore, the study confirmed that there are real individual differences in the ANS responses of different subjects, which highlights the importance of personalization in music therapy interventions based on emotion regulation. The predictive power of the used machine learning approaches was insufficient for making predictions at the individual level, as the performance evaluation metrics were nearly zero. However, the permutation tests confirmed that a signal was statistically identified in the collected data, although its statistical power was extremely low.

The findings of this study introduce several future research directions, where different levels of opportunities could produce interest in this domain. In the next stage of this research domain, improving the statistical power is essential to demonstrate more statistically significant connections between music, physiology and psychology.

To address this, improved experiment methods are needed, including the recruitment of a larger and more diverse group of participants to obtain more reliable estimates in both population-level and subject-level analyses. In addition to increasing the number of participants in the study protocol, the scale of data collection must be expanded to obtain more individual data from each subject, which will further improve the potential for personalized approaches.

If sufficient statistical power with the current scale of feature dimensions can be achieved, the future opportunities could be opened for more complex data extraction of music data. In such a case, higher-level acoustic features could be utilized to better identify individual preferences in music therapy by focusing on properties such as harmonic complexity, rhythmic structure and the emotional content of music.

The ultimate goal of this study was to explore the possibilities of integrating the collected data into more practical applications that could be used to implement personalized music-based therapeutic interventions. If a reliable connection could be demonstrated at the individual level between musical features and their effects on emotional regulation, it could open up possibilities for adaptive therapy systems that use music as a regulating method for arousal, stress, or various cognitive states in real-time applications. The potential applications related to the topic of this study could be, for example, personalized music interventions in therapeutic methods, stress management tools, and wellness applications that enhance individual emotional feedback. However, achieving this first requires the development of methods based on the earlier mentioned approaches, which take into account both improved data collection and the consideration of individual differences.

References

[1] World Health Organization, *Over a billion people living with mental health conditions – services require urgent scale-up*, en. Accessed: Dec. 9, 2025. [Online]. Available: https://www.who.int/news/item/02-09-2025-over-a-billion-people-living-with-mental-health-conditions-services-require-urgent-scale-up.

[2] O. Oyebode, J. Fowles, D. Steeves, and R. Orji, "Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness", *International Journal of Human–Computer Interaction*, vol. 39, no. 9, pp. 1938–1962, May 2023, ISSN: 1044-7318. DOI: 10.1080/10447318.2022.2089085.

[3] R. McKnight, J. Price, and J. Geddes, *Psychiatry*. Oxford University Press, May 2019, ISBN: 978-0-19-875400-8. DOI: 10.1093/oso/9780198754008.001.0001.

[4] U. Schnyder, "Future perspectives in psychotherapy", en, *European Archives of Psychiatry and Clinical Neuroscience*, vol. 259, no. 2, pp. 123–128, Nov. 2009, ISSN: 1433-8491. DOI: 10.1007/s00406-009-0051-z.

[5] S. K. de l'Etoile, "The effectiveness of music therapy in group psychotherapy for adults with mental illness", *The Arts in Psychotherapy*, vol. 29, no. 2, pp. 69–78, Apr. 2002, ISSN: 0197-4556. DOI: 10.1016/S0197-4556(02)00139-9.

[6] A. Raglio et al., "Machine learning techniques to predict the effectiveness of music therapy: A randomized controlled trial", *Computer Methods and Programs in Biomedicine*, vol. 185, p. 105 160, Mar. 2020, ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2019.105160.

[7] S. L. Robb, D. S. Burns, and J. S. Carpenter, "Reporting Guidelines for Music-based Interventions", *Journal of health psychology*, vol. 16, no. 2, pp. 342–352, Mar. 2011, ISSN: 1359-1053. DOI: 10.1177/1359105310374781.

[8] N. L. Wallin, B. Merker, S. Brown, and S. R. Brown, *The origins of music* (A Bradford book), eng. Cambridge (Mass.): MIT Press, 2000, ISBN: 978-0-262-23206-7.

[9] S. A. Mehr et al., "Universality and diversity in human song", *Science*, vol. 366, no. 6468, eaax0868, Nov. 2019, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.aax0868.

[10] S. Mithen, I. Morley, A. Wray, M. Tallerman, and C. Gamble, "The Singing Neanderthals: The Origins of Music, Language, Mind and Body, by Steven Mithen. London: Weidenfeld & Nicholson, 2005. ISBN 0-297-64317-7 hardback £20 & US$25.2; ix+374 pp.", en, *Cambridge Archaeological Journal*, vol. 16, no. 1, pp. 97–112, Feb. 2006, ISSN: 1474-0540, 0959-7743. DOI: 10.1017/S0959774306000060.

[11] W. F. Thompson, N. J. Bullot, and E. H. Margulis, "The psychological basis of music appreciation: Structure, self, source.", eng, *Psychological Review*, vol. 130, no. 1, pp. 260–284, Jan. 2023, ISSN: 1939-1471. DOI: 10.1037/rev0000364.

[12] W. F. Thompson, A. Lamont, R. Parncutt, and F. A. Russo, *Music in the social and behavioral sciences: an encyclopedia*, eng, 1st ed. Los Ange-

les: SAGE Reference, 2014, vol. 2, ISBN: 978-1-4522-8303-6. DOI: 10.4135/9781452283012.

[13] Suomalainen Lääkäriseura Duodecim, *Musiikkiterapia*, 2019. Accessed: Dec. 9, 2025. [Online]. Available: https://www.kaypahoito.fi/nix03414.

[14] G. R. Watkins, "Music Therapy: Proposed Physiological Mechanisms and Clinical Implications", en-US, *Clinical Nurse Specialist*, vol. 11, no. 2, p. 43, Mar. 1997, ISSN: 0887-6274.

[15] I. Cross, "Music and Science: Three Views", *Revue belge de Musicologie / Belgisch Tijdschrift voor Muziekwetenschap*, vol. 52, pp. 207–214, 1998, ISSN: 0771-6788. DOI: 10.2307/3686926.

[16] P. Saukko, "Musiikkiterapian vaikutus kehitysvammaisten lasten, nuorten ja aikuisten toimintakykyyn. Kirjallisuuskatsaus", fi, Kansaneläkelaitos (Kela), Sosiaali- ja terveysturvan raportteja 18, 2019.

[17] X. Wang et al., "Music therapy for depression: A narrative review", en, *Brain-X*, vol. 2, no. 3, e72, 2024, ISSN: 2835-3153. DOI: 10.1002/brx2.72.

[18] J. W. O'Kelly, "Music Therapy and Neuroscience: Opportunities and Challenges", en, *Voices: A World Forum for Music Therapy*, vol. 16, no. 2, Apr. 2016, ISSN: 1504-1611. DOI: 10.15845/voices.v16i2.872.

[19] J. Edwards, Ed., *The Oxford handbook of music therapy* (Oxford library of psychology), en, First published in paperback. Oxford: Oxford University Press, 2017, ISBN: 978-0-19-963975-5 978-0-19-881714-7.

[20] Y. N. Wang et al., "Effects of movement training based on rhythmic auditory stimulation in cognitive impairment: A meta-analysis of randomized controlled clinical trial", English, *Frontiers in Neuroscience*, vol. 18, Apr. 2024, ISSN: 1662-453X. DOI: 10.3389/fnins.2024.1360935.

[21] S. Gonzalez-Hoelling, C. Bertran-Noguer, G. Reig-Garcia, and R. Suñer-Soler, "Effects of a Music-Based Rhythmic Auditory Stimulation on Gait and Balance in Subacute Stroke", *International Journal of Environmental Research and Public Health*, vol. 18, no. 4, p. 2032, Feb. 2021, ISSN: 1661-7827. DOI: 10.3390/ijerph18042032.

[22] T. Norlander, C. Sandholm, and O. Anfelt, "The Physioacoustic Method and the Creative Process", EN, *Perceptual and Motor Skills*, vol. 86, no. 3, pp. 1091–1096, Jun. 1998, Publisher: SAGE Publications Inc, ISSN: 0031-5125. DOI: 10.2466/pms.1998.86.3.1091.

[23] A. J. van Os, L. Aziz, D. Schalkwijk, J. M. Schols, and R. A. de Bie, "Effectiveness of Physio Acoustic Sound (PAS) therapy in demented nursing home residents with nocturnal restlessness: Study protocol for a randomized controlled trial", *Trials*, vol. 13, p. 34, Apr. 2012, ISSN: 1745-6215. DOI: 10.1186/1745-6215-13-34.

[24] Z. Yang et al., "Music tempo modulates emotional states as revealed through EEG insights", *Scientific Reports*, vol. 15, p. 8276, Mar. 2025, ISSN: 2045-2322. DOI: 10.1038/s41598-025-92679-1.

[25] Y. Zhao, H. Xu, and J. Fu, "Integrating rhythmic auditory stimulation in intelligent rehabilitation technologies for enhanced post-stroke recovery", *Frontiers in Bioengineering and Biotechnology*, vol. 13, p. 1 649 011, Aug. 2025, ISSN: 2296-4185. DOI: 10.3389/fbioe.2025.1649011.

[26] M. KARKKAINEN and J. MITSUI, "The effects of sound based vibration treatment on the human mind and body The Physioacoustic Method(Research Report,The 21st Symposium on Life Information Science)", *Journal of International Society of Life Information Science*, vol. 24, pp. 155–164, May 2019. DOI: 10.18936/islis.24.1_155.

[27] F. C. S. Frickmann, R. D. Urman, K. Siercks, G. Burgermeister, M. M. Luedi, and F. E. Lersch, "The Effect of Perioperative Auditory Stimulation with Music on Procedural Pain: A Narrative Review", en, *Current Pain and Headache Reports*, vol. 27, no. 8, pp. 217–226, Aug. 2023, ISSN: 1534-3081. DOI: 10.1007/s11916-023-01138-x.

[28] S. Koelsch, "Brain correlates of music-evoked emotions", en, *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 170–180, Mar. 2014, ISSN: 1471-0048. DOI: 10.1038/nrn3666.

[29] Charles Brown and Tobias Riede, *Comparative Bioacoustics: An Overview*. Sharjah, UNITED ARAB EMIRATES: Bentham Science Publishers, 2017, ISBN: 978-1-68108-317-9. Accessed: Jan. 6, 2026. [Online]. Available: http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=4801106.

[30] P. Photinos, *The Physics of Sound Waves (Second Edition): Music, Instruments, and Sound Equipment*. Bristol, UNITED KINGDOM: Institute of Physics Publishing, 2021, ISBN: 978-0-7503-3539-3. Accessed: Jan. 6, 2026. [Online]. Available: http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=31252805.

[31] F. E. Toole, E. A. G. Shaw, G. A. Daigle, and M. R. Stinson, "The Physical Nature of Sound", en, in *Encyclopedia of Acoustics*, John Wiley & Sons, 1997.

[32] J. O. Smith III, *Spectral Audio Signal Processing*. Stanford University: Center for Computer Research in Music and Acoustics (CCRMA), 2011, ISBN: 978-0-9745607-3-1. Accessed: Apr. 12, 2026. [Online]. Available: https://ccrma.stanford.edu/~jos/sasp/.

[33] J. O. Pickles, *Introduction to the Physiology of Hearing*. Bradford, UNITED KINGDOM: BRILL, 2012, ISBN: 978-1-78052-167-1. Accessed: Dec. 15, 2025.

[Online]. Available: http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=992501.

[34] A. D. Patel, *Music, language, and the brain*, eng, 1st ed. New York ; Oxford University Press, 2008, ISBN: 978-0-19-989017-0.

[35] W. Thompson, “Intervals and Scales”, *The Psychology of Music*, pp. 107–140, Dec. 2013, ISSN: 9780123814609. DOI: 10.1016/B978-0-12-381460-9.00004-3.

[36] Session Town, *Musical Notes Explained Step-By-Step [And the Piano Notes]*, en. Accessed: Apr. 23, 2026. [Online]. Available: https://www.sessiontown.com/en/courses/easy-music-theory/musical-notes.

[37] H. Bahuleyan, *Music Genre Classification using Machine Learning Techniques*, arXiv:1804.01149 [cs], Apr. 2018. DOI: 10.48550/arXiv.1804.01149.

[38] B. McFee et al., “Librosa: Audio and Music Signal Analysis in Python”, en, *SciPy 2015*, Jun. 2015. DOI: 10.25080/Majora-7b98e3ed-003.

[39] C. L. Krumhansl, “Rhythm and pitch in music cognition”, *Psychological Bulletin*, vol. 126, no. 1, pp. 159–179, 2000, ISSN: 1939-1455. DOI: 10.1037/0033-2909.126.1.159.

[40] T. Painter and A. Spanias, “Perceptual coding of digital audio”, *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000, ISSN: 1558-2256. DOI: 10.1109/5.842996.

[41] H. Thornburg, “Detection and modeling of transient audio signals with prior information”, en, Ph.D. dissertation, Stanford University, 2005.

[42] M. C. Jacob Rodrigues, O. Postolache, and F. Cercas, “The Influence of Stress Noise and Music Stimulation on the Autonomous Nervous System”, *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–19, 2023, ISSN: 1557-9662. DOI: 10.1109/TIM.2023.3279881.

[43] N. F. Bernardi et al., "Increase in Synchronization of Autonomic Rhythms between Individuals When Listening to Music", *Frontiers in Physiology*, vol. 8, p. 785, Oct. 2017, ISSN: 1664-042X. DOI: 10.3389/fphys.2017.00785.

[44] A. Angulo-Perkins and L. Concha, "Music Perception: Information Flow Within the Human Auditory Cortices", en, in *Neurobiology of Interval Timing*, H. Merchant and V. de Lafuente, Eds., New York, NY: Springer, 2014, pp. 293–303, ISBN: 978-1-4939-1782-2. DOI: 10.1007/978-1-4939-1782-2_15.

[45] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*. Burlington, UNITED STATES: Jones & Bartlett Learning, LLC, 2025, ISBN: 978-1-284-28688-5. Accessed: Dec. 15, 2025. [Online]. Available: http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=31929812.

[46] S. Koelsch, "Investigating Emotion with Music", en, *Annals of the New York Academy of Sciences*, vol. 1060, no. 1, pp. 412–418, 2005, ISSN: 1749-6632. DOI: 10.1196/annals.1360.034.

[47] R. J. Ellis and J. F. Thayer, "Music and Autonomic Nervous System (Dys)Function", *Music Perception: An Interdisciplinary Journal*, vol. 27, no. 4, pp. 317–326, 2010, ISSN: 0730-7829. DOI: 10.1525/mp.2010.27.4.317.

[48] S. Beginnings, *Know your Neurobiology: The Auditory System*, en-GB, Jan. 2022. Accessed: Apr. 13, 2026. [Online]. Available: https://sensorybeginnings.com/blog/know-your-neurobiology-the-auditory-system/.

[49] M. E. Driscoll and P. Tadi, "Neuroanatomy, Inferior Colliculus", eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Dec. 17, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK554468/.

[50] F. C. F. Müller-Ribeiro, A. K. Goodchild, S. McMullan, M. A. P. Fontes, and R. A. L. Dampney, "Coordinated autonomic and respiratory responses evoked

by alerting stimuli: Role of the midbrain colliculi", *Respiratory Physiology & Neurobiology*, Brain and breathing, vol. 226, pp. 87–93, Jun. 2016, ISSN: 1569-9048. DOI: 10.1016/j.resp.2015.10.012.

[51] P. J. Uhlhaas, "Neural dynamics in mental disorders", *World Psychiatry*, vol. 14, no. 2, pp. 116–118, Jun. 2015, ISSN: 1723-8617. DOI: 10.1002/wps.20203.

[52] P. J. Uhlhaas and W. Singer, "Neuronal Dynamics and Neuropsychiatric Disorders: Toward a Translational Paradigm for Dysfunctional Large-Scale Networks", *Neuron*, vol. 75, no. 6, pp. 963–980, Sep. 2012, ISSN: 0896-6273. DOI: 10.1016/j.neuron.2012.09.004.

[53] L. M. Jenkins et al., "Individuals with more severe depression fail to sustain nucleus accumbens activity to preferred music over time", *Psychiatry Research: Neuroimaging*, vol. 275, pp. 21–27, May 2018, ISSN: 0925-4927. DOI: 10.1016/j.pscychresns.2018.03.002.

[54] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion", *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11 818–11 823, Sep. 2001. DOI: 10.1073/pnas.191355898.

[55] S. Koelsch, "A Neuroscientific Perspective on Music Therapy", en, *Annals of the New York Academy of Sciences*, vol. 1169, no. 1, pp. 374–384, 2009, ISSN: 1749-6632. DOI: 10.1111/j.1749-6632.2009.04592.x.

[56] B. M. Carlson, "The Nervous System", en-US, in *The Human Body*, Elsevier, 2018. DOI: 10.1016/B978-0-12-804254-0.00006-5.

[57] T. LeBouef, Z. Yaker, and L. Whited, "Physiology, Autonomic Nervous System", eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2026. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK538516/.

[58] S. L. Prescott and S. D. Liberles, "Internal senses of the vagus nerve", en, *Neuron*, vol. 110, no. 4, pp. 579–599, Feb. 2022, ISSN: 08966273. DOI: 10.1016/j.neuron.2021.12.020.

[59] S. A. Stevenson, "Effects of Music on Emotion, Heart Rate, Respiration, and Electrodermal Activity", en, *Proceedings of the National Conference on Undergraduate Research (NCUR)*, 2016. [Online]. Available: https://libjournals.unca.edu/ncur/wp-content/uploads/2021/06/2044-Stevenson-Sarah-FINAL.pdf.

[60] N. Pop-Jordanova and J. Pop-Jordanov, "Electrodermal Activity and Stress Assessment", eng, *Prilozi*, vol. 41, no. 2, pp. 5–15, Sep. 2020, ISSN: 1857-8985. DOI: 10.2478/prilozi-2020-0028.

[61] O. N. Rahma et al., "Electrodermal Activity for Measuring Cognitive and Emotional Stress Level", *Journal of Medical Signals and Sensors*, vol. 12, no. 2, pp. 155–162, May 2022, ISSN: 2228-7477. DOI: 10.4103/jmss.JMSS_78_20.

[62] F. Yasuma and J.-i. Hayano, "Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize With Respiratory Rhythm?", *Chest*, vol. 125, no. 2, pp. 683–690, Feb. 2004, ISSN: 0012-3692. DOI: 10.1378/chest.125.2.683.

[63] G. L. Ahern et al., "Heart Rate and Heart Rate Variability Changes in the Intracarotid Sodium Amobarbital Test", en, *Epilepsia*, vol. 42, no. 7, pp. 912–921, 2001, ISSN: 1528-1167. DOI: 10.1046/j.1528-1157.2001.042007912.x.

[64] S. C. Matthews, M. P. Paulus, A. N. Simmons, R. A. Nelesen, and J. E. Dimsdale, "Functional subdivisions within anterior cingulate cortex and their relationship to autonomic nervous system function", *NeuroImage*, vol. 22, no. 3, pp. 1151–1156, Jul. 2004, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2004.03.005.

[65] Y. Sattar and L. Chhabra, "Electrocardiogram", eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2026. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK549803/.

[66] E. V. E. Plunkett and M. E. Cross, "Einthoven's triangle and axis", in *Physics, Pharmacology and Physiology for Anaesthetists: Key Concepts for the FRCA*, 2nd ed., Cambridge: Cambridge University Press, 2014, pp. 241–243, ISBN: 978-1-107-61588-5. DOI: 10.1017/CBO9781107326200.102.

[67] D. E. Becker, "Fundamentals of Electrocardiography Interpretation", *Anesthesia Progress*, vol. 53, no. 2, pp. 53–64, 2006, ISSN: 0003-3006. DOI: 10.2344/0003-3006(2006)53[53:FOEI]2.0.CO;2.

[68] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms", *Frontiers in Public Health*, vol. 5, p. 258, Sep. 2017, ISSN: 2296-2565. DOI: 10.3389/fpubh.2017.00258.

[69] B. Brummett, S. Boyle, C. Kuhn, I. Siegler, and R. Williams, "Positive Affect is Associated with Cardiovascular Reactivity, Norepinephrine Level, and Morning Rise in Salivary Cortisol", *Psychophysiology*, vol. 46, no. 4, pp. 862–869, Jul. 2009, ISSN: 0048-5772. DOI: 10.1111/j.1469-8986.2009.00829.x.

[70] J.-E. Trihan et al., "Arterial Blood-Flow Acceleration Time on Doppler Ultrasound Waveforms: What Are We Talking About?", *Journal of Clinical Medicine*, vol. 12, no. 3, p. 1097, Jan. 2023, ISSN: 2077-0383. DOI: 10.3390/jcm12031097.

[71] G. Birdee et al., "Slow breathing for reducing stress: The effect of extending exhale", *Complementary Therapies in Medicine*, vol. 73, p. 102 937, May 2023, ISSN: 0965-2299. DOI: 10.1016/j.ctim.2023.102937.

[72] A. Tiwari, S. Narayanan, and T. H. Falk, "Breathing Rate Complexity Features for "In-the-Wild" Stress and Anxiety Measurement", in *2019 27th Euro-*

pean Signal Processing Conference (EUSIPCO), ISSN: 2076-1465, Sep. 2019, pp. 1–5. DOI: 10.23919/EUSIPCO.2019.8902700.

[73] K. A. Kim, I. K. Lee, S. S. Choi, S. S. Kim, T. S. Lee, and E. J. Cha, "Wearable transducer to monitor respiration in a wireless way", in *2007 6th International Special Topic Conference on Information Technology Applications in Biomedicine*, Nov. 2007, pp. 174–176. DOI: 10.1109/ITAB.2007.4407372.

[74] M. Chu et al., "Respiration rate and volume measurements using wearable strain sensors", en, *npj Digital Medicine*, vol. 2, no. 1, p. 8, Feb. 2019, ISSN: 2398-6352. DOI: 10.1038/s41746-019-0083-3.

[75] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on Psychological Stress Detection Using Biosignals", *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, Jan. 2022, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2019.2927337.

[76] D. Ayata, Y. Yaslan, and M. Kamasak, "Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems", *Journal of Medical and Biological Engineering*, vol. 40, Jan. 2020. DOI: 10.1007/s40846-019-00505-7.

[77] B. M. Appelhans and L. J. Luecken, "Heart Rate Variability as an Index of Regulated Emotional Responding", EN, *Review of General Psychology*, vol. 10, no. 3, pp. 229–240, Sep. 2006, ISSN: 1089-2680. DOI: 10.1037/1089-2680.10.3.229.

[78] W. Kim, "Personalization: Definition, Status, and Challenges Ahead", en, *Jounal of Object Technology*, vol. 1, no. 1, pp. 29–40, 2002. DOI: doi.org/10.5381/jot.2002.1.1.c3.

[79] Y.-K. Lin, H. Chen, R. A. Brown, S.-H. Li, and H.-J. Yang, "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multi-

task Learning Approach", *MIS Quarterly*, vol. 41, no. 2, pp. 473–496, 2017, Publisher: Management Information Systems Research Center, University of Minnesota, ISSN: 0276-7783. Accessed: Jan. 7, 2026. [Online]. Available: https://www.jstor.org/stable/26629723.

[80] D. B. Olawade, O. Z. Wada, A. Odetayo, A. C. David-Olawade, F. Asaolu, and J. Eberhardt, "Enhancing mental health with Artificial Intelligence: Current trends and future prospects", *Journal of Medicine, Surgery, and Public Health*, vol. 3, Aug. 2024, ISSN: 2949-916X. DOI: 10.1016/j.glmedi.2024.100099.

[81] A. P. Kefauver and D. Patschke, *Fundamentals of Digital Audio*. Middleton (WI), UNITED STATES: A-R Editions, Inc., 2007, ISBN: 978-0-89579-611-0. Accessed: Jan. 8, 2026. [Online]. Available: http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=3115106.

[82] M. R. Steenbergen and B. S. Jones, "Modeling Multilevel Data Structures", *American Journal of Political Science*, vol. 46, no. 1, pp. 218–237, 2002, Publisher: [Midwest Political Science Association, Wiley], ISSN: 0092-5853. DOI: 10.2307/3088424.

[83] T. Sandhan, S. Sonowal, and J. Y. Choi, *Audio Bank: A High-Level Acoustic Signal Representation for Audio Event Recognition*, arXiv:2304.05067 [eess], Apr. 2023. DOI: 10.48550/arXiv.2304.05067.

[84] E. Bagiella, R. P. Sloan, and D. F. Heitjan, "Mixed-effects models in psychophysiology", eng, *Psychophysiology*, vol. 37, no. 1, pp. 13–20, Jan. 2000, ISSN: 0048-5772.

[85] P. Kilian, S. Ye, and A. Kelava, "Mixed effects in machine learning – A flexible mixedML framework to add random effects to supervised machine learning regression", en, *Transactions on Machine Learning Research*, Jan. 2023,

ISSN: 2835-8856. [Online]. Available: https://openreview.net/forum?id=MKZyHtmfwH.

[86] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, en, ISBN: 9780511790942, Dec. 2006. DOI: 10.1017/CBO9780511790942.

[87] M. R. Berthold, C. Borgelt, F. Höppner, and F. Klawonn, *Guide to Intelligent Data Analysis* (Texts in Computer Science). London: Springer, 2010, ISBN: 978-1-84882-259-7 978-1-84882-260-3. DOI: 10.1007/978-1-84882-260-3.

[88] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995, ISSN: 0035-9246.

[89] F. Sohil, M. U. Sohali, and J. Shabbir, "An introduction to statistical learning with applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, $41.98, eISBN: 978-1-4614-7137-7", en, *Statistical Theory and Related Fields*, vol. 6, no. 1, pp. 87–87, Jan. 2022, ISSN: 2475-4269, 2475-4277. DOI: 10.1080/24754269.2021.1980261.

[90] J. I. Agbinya, *Applied Data Analytics - Principles and Applications*. Milton, UNITED KINGDOM: River Publishers, 2019, ISBN: 978-1-000-79553-0.

[91] G. C. McDonald, "Ridge regression", en, *WIREs Computational Statistics*, vol. 1, no. 1, pp. 93–100, 2009, ISSN: 1939-0068. DOI: 10.1002/wics.14.

[92] A. Rabinowicz and S. Rosset, "Cross-Validation for Correlated Data", *Journal of the American Statistical Association*, vol. 117, no. 538, pp. 718–731, Apr. 2022, ISSN: 0162-1459. DOI: 10.1080/01621459.2020.1801451.

[93] C. A. Holt and S. P. Sullivan, "Permutation tests for experimental data", en, *Experimental Economics*, vol. 26, no. 4, pp. 775–812, Sep. 2023, ISSN: 1386-4157, 1573-6938. DOI: 10.1007/s10683-023-09799-6.

[94] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, *FMA: A Dataset For Music Analysis*, Sep. 2017. DOI: 10.48550/arXiv.1612.01840.

[95] World Medical Association, *WMA declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants*, en-US. Accessed: Apr. 24, 2026. [Online]. Available: https://www.wma.net/policies-post/wma-declaration-of-helsinki/.

[96] P. M. Scheufele, "Effects of Progressive Relaxation and Classical Music on Measurements of Attention, Relaxation, and Stress Responses", en, *Journal of Behavioral Medicine*, vol. 23, no. 2, pp. 207–228, Apr. 2000, ISSN: 1573-3521. DOI: 10.1023/A:1005542121935.

[97] K. Pope, "The Effects of Jazz and Classical Music on Recall", en, *Journal of Health Education Research & Development*, vol. 05, no. 02, 2017, ISSN: 23805439. DOI: 10.4172/2380-5439.1000215.

[98] G. Gerra et al., "Neuroendocrine responses of healthy volunteers to 'techno-music': Relationships with personality traits and emotional state", *International Journal of Psychophysiology*, vol. 28, no. 1, pp. 99–111, Jan. 1998, ISSN: 0167-8760. DOI: 10.1016/S0167-8760(97)00071-8.

[99] J. Gordon and M. C. Gridley, "Musical Preferences as a Function of Stimulus Complexity of Piano Jazz", *Creativity Research Journal*, vol. 25, no. 1, pp. 143–146, Jan. 2013, ISSN: 1040-0419. DOI: 10.1080/10400419.2013.752303.

[100] P. Gomez and B. Danuser, "Relationships between musical structure and psychophysiological measures of emotion", eng, *Emotion*, vol. 7, no. 2, pp. 377–387, May 2007, ISSN: 1528-3542. DOI: 10.1037/1528-3542.7.2.377.

[101] K. Motoki, N. Takahashi, C. Velasco, and C. Spence, "Is classical music sweeter than jazz? Crossmodal influences of background music and taste/flavour on healthy and indulgent food preferences", *Food Quality and Preference*, vol. 96, Mar. 2022, ISSN: 0950-3293. DOI: 10.1016/j.foodqual.2021.104380.

[102] C. G. Pusey, T. Haugen, R. Høigaard, A. Ivarsson, A. W. Røshol, and A. Laxdal, "Put Some Music on: The Effects of pre-Task Music Tempo on Arousal, Affective State, Perceived Exertion, and Anaerobic Performance", EN, *Music & Science*, vol. 6, Jan. 2023, ISSN: 2059-2043. DOI: 10.1177/20592043231174388.

[103] D. Dellacherie, M. Roy, L. Hugueville, I. Peretz, and S. Samson, "The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance", eng, *Psychophysiology*, vol. 48, no. 3, pp. 337–349, Mar. 2011, ISSN: 1469-8986. DOI: 10.1111/j.1469-8986.2010.01075.x.

[104] D. Makowski et al., "NeuroKit2: A Python toolbox for neurophysiological signal processing", en, *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, ISSN: 1554-3528. DOI: 10.3758/s13428-020-01516-y.

[105] F. Gan, G. Ruan, and J. Mo, "Baseline correction by improved iterative polynomial fitting with automatic threshold", *Chemometrics and Intelligent Laboratory Systems*, Selected Papers from the International Conference on Chemometrics and Bioinformatics in Asia, vol. 82, no. 1, pp. 59–65, May 2006, ISSN: 0169-7439. DOI: 10.1016/j.chemolab.2005.08.009.

[106] A. F. Ruckstuhl, M. P. Jacobson, R. W. Field, and J. A. Dodd, "Baseline subtraction using robust local regression estimation", *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 68, no. 2, pp. 179–193, Jan. 2001, ISSN: 0022-4073. DOI: 10.1016/S0022-4073(00)00021-2.

[107] J. Braithwaite, D. Watson, R. Jones, and M. A. Rowe, "Guide for Analysing Electrodermal Activity & Skin Conductance Responses for Psychological Experiments", *CTIT technical reports series*, 2013. [Online]. Available: https://www.semanticscholar.org/paper/Guide-for-Analysing-Electrodermal-Activity-%26-Skin-Braithwaite-Watson/b99d1f004e4194ac6ef86a86bb0918a1115

[108] S. G. K. Patro and K. K. Sahu, *Normalization: A Preprocessing Stage*, Mar. 2015. DOI: 10.48550/arXiv.1503.06462.