My music and I: Decoding Individual Differences via Musical Behaviour

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Computer Science and Engineering
by Research

by

Yudhik Agrawal
20161093
yudhik.agrawal@research.iiit.ac.in

International Institute of Information Technology
(Deemed to be University)
Hyderabad - 500 032, INDIA
June 2021

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "My music and I: Decoding Individual Differences via Musical Behaviour" by Yudhik Agrawal, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Adviser: Dr. Vinoo Alluri

To family and friends

Acknowledgments

Research is to see what everybody else has seen, and think what nobody has thought.

-Dr. Albert Szent-Gyorgyi (1893-1986)

This thesis has been an inconceivable journey from seemingly impossible challenges to occasional success. I feel privileged to have had the opportunity of meeting several people who've left an everlasting impact on me. Firstly, I would like to heartily thank my advisor/mentor/friend Dr. Vinoo Alluri for her infinite support, inspiration, and motivation throughout the journey. She provided me with complete freedom to express myself. I am thankful for her confidence in me and patience when laden with silly questions, for her being approachable at any time. Her constructive feedback, suggestions, and remarks made the learning experience thrilling and remarkable. It was indeed an honor for me to work under her. Next, I would also like to thank Dr. Petri Toiviainen and Dr. Emily Carlson for their assistance at every stage of the research project. It is genuinely the constant guidance and encouragement of all the aforementioned team that has made this thesis possible.

I would like to express my sincere gratitude to Dr. Avinash Sharma, who guided me in the initial days of my career in research. His guidance helped me build strong fundamentals from the start and laid a sound foundation for my growth.

I am deeply grateful to Abbhinav Venkat and Chaitanya Patel, who proved to be much more than co-authors for me. They are my go-to mentors who continue to help me in making insightful career decisions. Next, I would like to thank my colleagues Yash and Aayush, who played an essential role in my thesis by encouraging and showing unparalleled trust in me. Thanks to Guru R., Neeraj, Sai Soorya, and Sawar for being great lab-mates.

I would like to express my sincere gratitude to Samyak for believing in my ideas and providing technical support at various times, thanks to Mohit for 3 AM rides when I wanted to breathe fresh air the most, thanks to my first and last college roommate Ali for putting up with silly me for three whole years, and special thanks to all my dearest friends Aashish, Ali, Anvesh, Archit, Mohit, Mudit, Rohit, Sachin, and Samyak for their unwavering support and belief in me. I will cherish the moments when we used to play together, have group-study, and going out with them; all this made my college life more memorable than anyone can imagine.

This acknowledgment is incomplete without mentioning family. I am forever indebted for their continued moral support, backing me, cheering me when I was low, and providing me with the opportunities that has led me here.

Abstract

Music is omnipresent and has existed since time immemorial. In general, individuals' response to music, be it in terms of movements or listening strategies, tend to be dictated by features intrinsic to music such as rhythmic structure, genre, in addition to extrinsic factors such as mood states, socio-cultural norms, amongst others. However, at the same time, human differences in physical, psychological, and behavioural traits affect the decisions we make and the experiences we pursue in daily life. Individual differences can refer to gender, cognitive styles of thinking, personality, amongst others. In this thesis, we try to expand our understanding of how individual characteristics modulate our musical behaviour. The study and analyses presents various angles of approaching musical behaviour in an interdisciplinary manner, particularly in the areas of embodied music cognition and music consumption. We particularly investigate musical behaviour in two broader sub-domains: Active Music Engagement, which entails corporeal involvement and Passive Music Engagement, which involves listening to music.

In the first part, we look at decoding individual differences via active music engagement. The paradoxical balance between universality and individuality in human motoric responsiveness to music makes it interesting to note that all the individuals' traits are encoded in free dance movement. The study addresses this by identifying individual differences, specifically gender, cognitive styles, Big Five personality traits, and music preferences using free dance movements via our proposed Machine Learning model. We further demonstrate the robustness of the proposed model by testing its efficacy on another data set and further providing conclusive evidence about the learned models' generalizability. The results of the study support theories of embodied music cognition and the role of bodily movement in musical experiences by demonstrating the influence of gender, personality, and music preferences on embodied responses to heard music.

In the second study, which deals with passive engagement, we investigate out how individual differences, specifically personality, are associated with the kind of emotional experiences one seeks from lyrics. We chose to investigate lyrics specifically because they are under-explored in the field of music information retrieval despite them playing a crucial role in eliciting emotions. Firstly, we sought to identify emotional connotations of music based on lyrics following which we associate them with individual differences represented by personality traits. To this end, first we propose a novel deep learning architecture to identify emotional connotations of music based on lyrics, which we further compare with the existing deep learning and even traditional methods on relevant datasets and show state-of-the-art performance. Subsequently, we use this model to extract emotional preferences of individuals mined

via online music streaming platforms and associate them with inherent personality traits. Our findings validate our theory that various types of emotions conveyed by songs have a unique relationship with individual traits. This study contributes to a better understanding of the relationship between broad personality dimensions and the emotional experiences one naturally seeks on online music streaming platforms.

Both of our studies corroborate previous research in the field in addition to providing novel findings in both the aspect of musical behaviour being modulated by individual differences.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Human variations pertaining to physical, psychological, and behavioral characteristics manifest in the choices we make and experiences we seek in everyday life. No two persons are alike, not even twins. This differential psychology is linked with the study of individual differences[82, 41]. What are Individual Differences? Chen [37] writes, "Individual differences refer to enduring characteristics that distinguish one organism from another and that are stable over time and across situations." Individual differences influence not only our end goals but also the process of how one perceives it [86].

The use of individual differences as a guide to understanding both the uses and effects of the mass media has sparked a growing interest [205]. Investigations have been carried out to explore the effects of individual differences on online shopping and information systems to gain an understanding of how these differences impact online shopping attitudes [203]. Greenberg et al. [80], reported differences in cognitive style of thinking, revealing males to have more systemizing tendencies whereas females have more empathetic trait. Behavioral studies have reported gender differences with female participants scoring higher on measures linked to the affective component of social cognition than their male counterparts, such as emotion recognition [127], social sensitivity [13], and emotional intelligence [19]. Individual differences, like gender and personality, have been observed in the use of social media in general, from consuming information to socializing [70]. Various studies have found links between individual traits, specifically personality, towards choice in films [103]. For example, people open to experiences prefer imaginative rather than conventional forms of entertainment [146, 56]. Along with these general areas, music, which is an omnipresent medium consumed in everyday life (even via movies) would also be affected by individual differences.

1.1 Musical Behaviour

Musical behaviour, be it moving to music, choice of music, how we react to music, and how we consume music has been extensively researched in the fields of music psychology and music information retrieval and has been found to be modulated by individual differences. Musical behaviour results from the overall coalition of music, the individual, and the context as illustrated in Figure 1.1. Apart from the kind of music and individual differences, context additionally modulates our musical behaviour, albeit at various scales. Context can be at a global scale encapsulating cultural differences, or at a local scale encapsulating immediate surrounding in addition to transient mood states. For instance, moving to music, one might move a little or not at all when he or she is sad or tired when compared to a happy state. Hence, interconnection of these three elements gives rise to unique and potentially predictable Musical Behaviour.

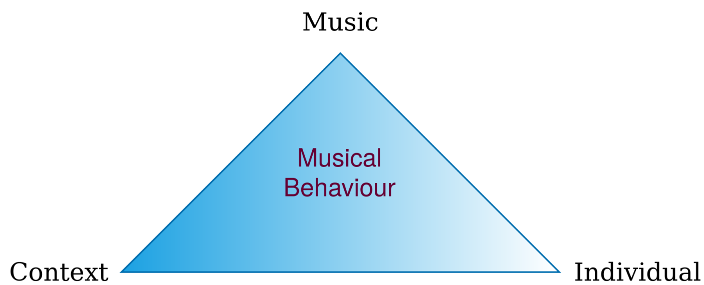


Figure 1.1: Musical Behaviour

Musical behaviour can be divided broadly into active and passive engagement. Active music engagement entails corporeal involvement, which can be either moving to music (i.e., music-induced movement), or producing music. Passive engagement on the other hand, involves listening to music which can range from attending to it , that is, 'listening carefully', or 'listening in the background' while performing other tasks.

Active engagement and Individual differences: Music-induced movements have been largely investigated when compared to music production, especially in light of individual differences. One potential reason is that music production related movements are less prone to individual differences owing to the physical constraints that the instrument poses during the production of sound and music [176]. For instance, an individual has greater degrees of freedom when moving to a musical beat; however, playing an instrument, leads to fewer degrees of freedom because the individual must engage their bodies in specific ways in order to produce music and perform. Several studies have found connection between music-induced movement and individual differences.

Individual differences have already been identified from gait studies. Troje et al. [193] showed that participants could easily learn to identify different individuals from point-light recordings of their gait. While gait is perhaps the most common way of analysing individual characteristics as they relate to bodily movement features, a model in which participants perform free, spontaneous dance movements has the capacity to provide greater individual variability of movement and also theoretical links to a variety of psychological and social functions. Given that gait studies with limited information from point-light displays have been effective in identifying gender, and that joint flexibility may have gender specificity [81], we should be able to identify gender well above chance level after having comparatively more information from all the joints.

According to the study by Luck et al. [121], there are several associations between dancers' personality traits and features of their free dance movements. Carlson et al. [35] found that individual movement signatures during free dance could be used to accurately classify individual dancers at a rate of 94 percent, suggesting that dance movements may be highly individualized. These studies suggest that it should be possible to predict participants' personalities from their dance movements, although this has not yet been done. Individuals who are more susceptible to stress and anxiety disorders are found to be associated with Neuroticism, hence, decoding music-induced movement is helpful in identifying underlying psychological states. Furthermore, dancers' self-reported cognitive styles of thinking and their reactions to various dance partners suggests individual differences to be also encoded in dance movements [32]. Hence, there is greater possibility to decode the information related to individual differences from freely moving to music which has not been done so far. The need to analyse EQ/SQ scores from music-induced movements is worth highlighting in light of recent work suggesting the existence of motor signatures unique to ASD, detectable from whole body movements as well as data drawn from participants' interaction with tablets [8, 212]. Therefore, the links between individual differences and music-induced movement have implications for music therapy as well as for music information retrieval.

Passive engagement and Individual differences: Individuals seek a varied range of experiences via music to satisfy their individual psychological needs and experience specific emotions. Listening to music has been repeatedly reported to be modulated by individual differences. Studies related to passive engagement have discovered that personality plays an important role in the perception of emotions represented by music. The impact of individual differences on music consumption has been widely investigated. Rentfrow and Gosling [160], discovered earliest links between music preferences and personality, implying that people who enjoy upbeat and conventional music are cheerful, socially outgoing, dependable, enjoy helping others, see themselves as physically attractive, and are relatively conventional. Personality traits were found to be strongly linked to preference ratings for music excerpts expressing different emotions, such as Agreeableness was strongly linked to a preference for happy and tender-sounding music, which is consistent with the definition of agreeableness as a pro-social trait [200]. Furthermore, a number of studies have looked into the connection between trait empathy and the enjoyment of sad music, finding that those who enjoyed sad music the most had high levels of empathic concern [201, 92]. The lack of external validity of previous research, which assessed individual differences in musical preferences using self-reported music-genre preferences, is among the most notable limitations. Music preferences and listening strategies have recently been found to be associated with ill-health, such as problem behavior [136] and internalizing symptomatology such as depression [109, 135].

To account for these limitations, [137, 7] have shown musical preferences to be associated with individual differences in personality using ecologically valid online music streaming data, such as user listening histories from music streaming platforms which are a better reflection of the users' true preferences and behaviours. Anderson et al. [7] in his experiments have shown people scoring high on Neuroticism are more likely to choose music which regulate their emotions, and users who score high on Conscientiousness may choose music based on goal-oriented behavior. Our innate tendencies that determine individual behaviour can thus manifest in musical behaviour. Recent studies using online streaming platform are examining the association between at-risk (risk for depression) individuals as it manifests in naturally occurring music listening behavior [185], which are further associated predominantly with the trait Neuroticism.

The overarching goal of the thesis is to examine the influence of individual differences in the context of Active and Passive music engagement.

1.2 Measuring Individual Differences

Music is an effective means of eliciting body movement; music-induced movement, in particular, might be an effective means of studying such relationships. We will discuss majorly about those Individual Differences that were selected to be investigated in this thesis research in greater detail. One could postulate that our bodily movements reflect, imitate, help to parse or support the understanding of the content of music.

Individual differences are measured using psychological constructs in addition to biological constructs. Towards psychological constructs are differences like personality, cognitive styles, etc, which are further associated with biological construct like gender.

1.2.1 Gender

A number of studies have explored perception of individual characteristics, like gender, from dance movement. Much of the research into perception of dance movement has focused on gender, both in terms of the perception of gender from dance movement and, drawing on hypotheses that human dance has played an evolutionary role in sexual selection, the perception of attractiveness and other qualities related to mate-selection from dance. Hufschmidt et al. [90], found that both children and adults could accurately identify dancer gender from avatar movements, while [206] found that female raters found the movements of male dancers with greater hand-grip strength to be more attractive. While it is not yet clear which movement features raters use to identify gender, differences in movement may arise from differences in average body structure and joint flexibility between genders [81].

1.2.2 Cognitive Styles of thinking

Humans have a remarkable soft skill that enables one to experience another person's perspective rather than just one's own. This pro-social ability to recognize and understand the feelings of another person comes from within. The German philosopher, Theodore Lipps conceptualized ”Einfühlung” (or ”feeling into”) as ”projecting oneself onto the object of perception.” The term ”Empathy” is an analogous term in English, introduced by the famous psychologist Edward Titchener (1867–1927) [188]. A great deal of disagreement for the term ”Empathy” arises because of its past usage; interchangeably to refer similar concepts like sympathy, a theory of mind, or emotional contagion. The Stanford Encyclopedia of Philosophy emphasizes to understand this phenomenon by referring to David Hume's dictum [91] "the minds of men are mirrors to one another," by stating "since in encountering other persons, humans can resonate with and recreate that person's thoughts and emotions on different dimensions of cognitive complexity" [181].

The majority of current research focuses on empathy, but it neglects to assess a related dimension that has been connected to empathy: systemizing. Empathy is the cognitive and affective capacity to perceive and respond adequately to the emotions and mental states of others, systemizing is the capacity to construct, analyze, and predict systems. These two dimensions of individual differences are the basis of the Empathizing–Systemizing (E-S) theory. The EQ measures participants' tendency to empathize with others [15], while the SQ measures the tendency to think in terms of systems [14]. These two measures were originally developed to increase understanding of people with ASD, as in this population trait systemizing tends to be very high while empathy tends to be low.

Listening to music engages several critical social functions, engaging core aspects of the human social cognitive system, that play a central role in social stimuli by coupling with the brain's system involved in the processing of emotions [107]. Music evokes emotions in humans, and contrariwise, emotions also influence people's choice of music. Individuals when feel low tend to listen to sad music or more upbeat music when they workout [140].

1.2.3 Personality

Earlier psychologists justified individuals' reaction to music based solely on the characteristics of the stimulus object [172, 204], which has been repealed by the modern theorists which accounts for individual reaction to music by crediting personality factors as the guiding factor for differential attention and perception of events. They tend to view human response to music from a phenomenological point of view. The current psychological theories in regards to human response have evolved from the perception of stimuli, and attribution of meaning of selected stimuli differs systematically for every individual.

The term Personality originates from the Latin word 'personalis,' meaning 'of a person.' In the Cambridge handbook of personality psychology, it is defined as "the characteristic sets of behaviors, cognitions, and emotional patterns that evolve from biological and environmental factors". Though there is no generally agreed-upon definition of personality, however, most theories view personality as relatively stable [42]. Personality pilots individuals to consistently think, feel, and behave in specific ways; essentially, this is what makes each individual unique. Though over time, these patterns strongly influence personal expectations, perceptions, values, and attitudes. Individuals with certain personality traits are expected to behave in a particular way depending on the situation. However, these stable qualities enable individuals to exhibit consistent behavior under different conditions [175, 69].

The characterization of personality as a set of facets or traits has a long history, but it was not well structured or organized. The trait theorists Cattell used the factor analysis techniques to create his own personality taxonomy to overcome the issue. Cattell identified 16 factors or dimensions of personality, which was the first factor-based system [36]. Over the next few years, many traits theorists independently found five factors to be sufficient and subsume most known personality traits which represent the basic structure behind all personality traits [144]. Many factor analyses found the five-factor model, also referred to as the Big Five [55], measuring the five general traits of (O)penness to Experience, (C)onscientiousness, (E)xtraversion, (A)greeableness, and (N)euroticism – is widely accepted [100] and considered to describe personality at the highest level of organization [75]. Furthermore, it has been validated across cultures [129, 130]. This model is commonly known as the OCEAN model and can be described as follows:

Openness: This is sometimes referred to as 'Openness to Experience,' which describes a person's tendency to think abstractly. This trait is associated with imagination, insight, curiosity, and enjoyment of arts and culture, and a tendency towards abstract thought [155, 101]. Those who score low on Openness are often much more traditional and may struggle with abstract thinking, whereas those who score high seem to have more connection between disparate brain regions [102], which explains their adaptability and their association with creativity and divergent thinking [113, 128], and are often looking for aesthetic experiences.

Conscientiousness: Conscientiousness describes person's tendency to be disciplined, responsible, and goal-oriented. Conscientiousness is characterized by high levels of thoughtfulness, good impulse control, efficiency, and goal-directed behaviors [155]. This trait is positively correlated with emotion regulation abilities and is related to academic success, job income [94, 53]. People high in Conscientiousness are found to live longer [125], and are less likely to commit crimes and have fewer problems with any drug abuse. Those who score low are impulsive, pay less attention to details, and easily sidetracked.

Extraversion: Extroverts have the habit of making bonds with other people; they can be friends, family members, co-workers, or even strangers. They are actively engaged in social activities, spending more time with other people. People scoring high in Extraversion are often happier and are less prone to certain psychological disorders [116]. People who are high in this trait are often assertive, talkative and are characterized by high amounts of emotional expressiveness [155]. In contrast, those who are low are fairly independent and are less likely to seek recognition from others in order to feel satisfied.

Agreeableness: Agreeableness, defined by cooperative and empathetic predisposition, feels empathy and concern for other people and tends to assists others who are in need of help. High Agreeableness is associated with increased activity in a part of the brain responsible for language processing and recognizing emotions in others [115]. People have attributes for trust, kindness, and other pro-social behaviors enabling better peer relationships in adolescence [155, 98]. People low in this trait are more competitive, manipulative, and are less empathic, putting their own concerns ahead of others.

Neuroticism: Many researchers use "emotional stability" to describe Neuroticism with a negative polarity. Neuroticism is a trait characterized by sadness, moodiness, anxiety, guilt, and emotional instability [155], and is associated with poor ability to manage psychological stress and a tendency to complain [143]. Persons with high Neuroticism tend to experience mood swings, anxiety, irritability and are more likely to react with fear, anger, and sadness to some situations. Low Neuroticism tends to be more stable by dealing with stress and emotionally resilient.

1.2.4 Musical Preference

Individuals exhibit specific tastes, which are unique for every individual. In order to study musical preferences, it is essential to categorize different music stimuli by labeling and providing descriptions onto which everyone agrees upon. In the past, researchers have introduced several models to categorize music stimuli into genres, but since they do not have clear boundaries, many limitations always prevailed. Limitations such as lack of consensus about which genres to study resulting in different researchers focusing on different music genres [52, 40]. Preferences are often discussed at the level of genre labels via Short Test Of Music Preferences (STOMP) [160], studies have shown both that these labels are inconsistently applied across large commercial databases [145], and that users' listening habits do not necessarily reflect their stated music preferences at the genre level [58, 64]. Further, assigning a single genre to every music is very ambiguous [194] hence relying on music genres as the unit for assessing preferences seems unfair.

Recent research has attempted to rectify these problems by developing a more refined assessment of music preferences. The solution offered was to investigate the pattern of reactions to audio excerpts by identifying a robust factor structure. The objective was to re-conceptualize music preference by including the preference for psychological as well as musical characteristics [18]. Rentfrow et al. [158, 159] generated a revised scale (STOMP-Revised) and developed preference factors based on self-reported ratings of heard stimuli followed by judges assigning attributes to the excerpts including musical and genre-based attributes.

1.3 Decoding Individual Differences via active engagement

Movement to music is a universal response to music. The urge to move in response to music, by means of head bobbing, foot tapping, or through elaborate dance moves, is seen to be an almost automatic response. According to Zentner and Eerola [217], infants displayed more rhythmic movement to music and metrical stimuli than to speech, implying a predisposition for rhythmic movement to music and other metrical regular sounds. Music-induced movement has predominantly been studied in the framework of embodied cognition which is described in the section below.

1.3.1 Embodied Cognition

The embodied model of cognition is built on the aspiration to provide a more holistic view on interaction of mind and matter. It aims to justify the phenomena of experience, that is, acknowledging the ways in which mind interacts with music is a simultaneous effect of movement, feeling and expression.

Recent view of Embodied Cognition suggests the importance of human body as mediator between subjective experience and the event [111, 198] is in contrast with the more traditional view of Embodied Cognition, which is less concerned with gestures and action, and more with mental processing. Traditional approach to cognition view mind as an abstract processing unit, which gives commands and receives input from a passive body. The philosopher and scientist René Descartes (1596-1650), in his Musicae compendium (1618), gave summary of the state of the art at that time which considers the impact of sound on listeners' emotion to be purely subjective, irrational element and therefore incapable of being measured scientifically. However, recent studies on embodied cognition views movement and sensation as integral to thought. [131] showed that people typically gesture when they speak to one another, and gesturing facilitates not just communication but language processing itself. In another study by [57] demonstrates, by using our bodies and surrounding environments to off-load storage, the simplifying nature of the cognitive processing. Cognition in the embodied view as corporeal, interactive, and real-time sets apart embodied cognition from more traditional approaches to cognition.

Embodied cognition refers to a framework for understanding human cognition and behavior that emphasizes the centrality of the form and functioning of the body, particularly sensorimotor processes, in shaping cognition. Rather than a linear process of perceive-compute-respond, action and perception are inherently and immediately linked and able to affect one another. Research has shown the characteristics of bodily movements to be influenced by the properties of the mind; An extrovert staying stuck to one spot on the floor while dancing, or carrying with slumped shoulders would he hard to imagine, or even a depressed individual skipping along briskly [171, 85, 134]. Experimental research has shown a relationship between motor feedback and perception; participants who were asked to hold a pen between their teeth, simulating a smile, found cartoons funnier than those who held a pen between their lips, simulating a frown [180]. In another study, men asked to rate their own assertiveness gave themselves higher ratings when asked to make a fist during the task rather than a neutral hand gesture [173]. This interwoven relation among the parts of environment, mind, corporeal, and sensorimotor capabilities forms the core of embodied cognition [154, 208, 209].

1.3.2 Embodied Music Cognition

Aristotle's mimesis theory states that music reflects "men in action" (Poetics, 1448a). Listener moving in a synchronous and ordered manner is a so natural tendency that it can actually be difficult to resist them [162, 114, 164]. Moving in response to heard music in an organized way, like rhythmically synchronizing with the pulse of the music by nodding their head, tapping their feet, moving their whole body, or mimicking instrumentalists' gestures, is a natural tendency [72, 112]. The irresistible quality of the engagement of people's motor systems while listening to music is the paradigm for Embodied Music Cognition, that cognition is deeply depends on aspects of the agent's body aside from the brain [174]. The studies have shown infants' ability to synchronize corporeal with musical stimuli showed that infants engage in significantly more rhythmic movement to music than to speech, suggesting a predisposition for rhythmic movement in response to music and other metrically regular sounds [191, 217, 60].

This innate ability of humans to respond to music not by mere minimal gestures but by showing affinity towards rhythm are elemental and forms the basis of Embodied Music Cognition.

Past research has shown support for the close relationship between sound and movement that exists in the brain [166, 110]. Bangert[11] in his experiment, where trained pianists listened to piano music, observed through fMRI studies that activations in areas of their cortexes were associated with motor control and not only with auditory processing. Recent research has shown direct links between musical features and human movement, including the reflection of hierarchical rhythmic structures in embodied eigen movements [189], the reflection of higher-level musical structures in group movement to Electronic Dance Movement [177], and reflection of spectral and timbral features of music in dance [26]. Bodily movement is one of the most commonly reported responses to music [114], and movement to music is one of the very few universal features of music across cultures [138].

Embodied Music Cognition claim that bodily involvement is crucial in human interaction with music, and therefore, also in our understanding of that interaction. The embodied viewpoint holds that corporeal articulation shapes the way we perceive, empathize, experience, and grasp music. This is unique in relation to a disembodied way of dealing with music cognition, which considers it to be founded on the perception-based analysis of musical structure. In contrast with an embodied view, which understands it through bodily movements, which makes it corporeal rather than cerebral. During the musical activities, perception seems to induce a transition from musical stimuli to close interaction of auditory sensing in the brain with motor trajectories. Further, researchers have found several evidence that shows movement to be more strongly connected to the auditory system than the visual system. For instance, rhythmic movements are attracted more strongly to auditory than to visual rhythms [163]. Moreover, in the experiment conducted by (Patel et al.)[151], it was found that participants synchronized to auditory stimuli with ease, while they had difficulty in synchronizing to metrical non-isochronous visual stimuli.

In defining the concept of Embodied Music Cognition, Leman [111] argued that linguistic description of music is problematic, as their relationship to the physical, acoustic signal of music is often symbolic and ambiguous. He characterizes corporeal articulations as direct involvement with the music against indirect involvement via symbolic or linguistic descriptors. This view is further supported by a recent literature review [74], which draws a parallel between movement and musical sounds in a variety of domains. In on of the study involving the use of sound-tracing paradigms observed stark similarity between different listeners' embodiment of similar sounds, where participants were asked to trace unheard sounds as movements or as drawings. This theory further concretes the view of the movement as an integral part of the way we process and understand music on conscious and unconscious levels and not merely as a response to heard music. Janata [97] asked participants to tap to the music and found that participants not only moved the hand, but also other body parts, such as feet and head. A study by Burger et al. [23] focused on relationships between dance movements and timbral- and rhythmic-features, supports the theory as participants exhibited high values for all the movement for higher pulse clarity. Thompson [186] reported a study in which pianists were asked to perform with increased expressions, and the listeners found the performances with more movement were more expressive regardless of sound.

Spontaneous movements may be closely related to predictions of local bursts of energy in the musical audio stream, in particular to the beat and the rhythm patterns [111]. Perhaps, recruitment of the action system to move in synchrony with music is one of the most common manifestations of a listener's engagement with music. Furthermore, for achieving a pleasing psychological state for the individual, it is natural for our brain's action systems to engage while listening to music [97] and successfully parse the rhythmic structure [28]. This sensorimotor coupling is one of the most usual ways in which people engage with and enjoy music [67]. In fact, neuroimaging studies have also demonstrated a link between the sensorimotor and reward regions of the brain and modulated by individual differences, that is, level of musical expertise in this case [6]. We have seen that Embodied music cognition has been studied in regards with individual differences, and we know from cognition in general that individual differences drive our mental states and how we process the world. This means these individual differences should also manifest in movement.

Measuring Movement

Researchers have been trying to find ways to measure movement for a long time. Some of the earliest modern attempts came from Chronophotography towards direct motion capture. This involves capturing a number of phases of movements by taking many pictures very quickly using specially designed cameras. Th earliest attempt to study gait and biological motion influenced research on the perception of bodily movement in music performance. Amongst the first was Gunnar Johansson (1911-1998), and his method involved fixing a very bright light to key joints of the participants and record those in a dark room. In his study [99], he demonstrated that 5-10 elements in adequate combinations of proximal motion give the visual system highly efficient information about human motion. This enabled capturing the movement rather than the person's own characteristics. Moving forward, this work has been extended in modern optical marker capture setups, where multiple cameras are calibrated and simultaneous capturing of two-dimensional images enables the reconstruction of three-dimensional locations in the real space. Many studies have now enabled capturing the minute information up to skin/cloth level in recent times, but the process is often cumbersome and very expensive.

Various studies in the field of Embodied Music Cognition have successfully used the multi-camera setup to record humans' spontaneous or dance movements. Since the central part of the thesis focuses around embodied music cognition, the datasets from multi-camera setups have been used in the thesis' studies. Research has shown music-induced movements to influenced by various individual differences [122, 121, 24, 23, 21, 33]. The following topic introduces and discusses the influence of individual differences like personality, empathy, musical preferences, and their relationships with music and dance.

1.3.3 Individual Differences and movement

How we think, react, or even move can tell a lot about who we are. It seems intuitive that we'd be able to identify someone from their voice fairly easily, but what about identifying from just minimalistic movement information. Humans appear to have a remarkably fine-tuned ability to discern information about others based on bodily movement. Troje et al. [193] showed that participants could easily learn to identify different individuals from point-light recordings of their gait, even scoring three-times above chance when recordings were rotated and manipulated to remove information about size and speed. In a follow-up study, [207] also used a Fourier Transform to remove the most prominent frequencies from stimuli, found that identification was still above chance, and that participants were easily able to generalize information learned from different viewing angles. Being able to identify an individual from limited perceptual information has clear evolutionary advantages, particularly in the unique social context of early human cultures where identifying group members and non-members could be necessary for survival [84]. Along these same lines, it could also be considered adaptive to be able to assess other information from a person's bodily movement, such as their mood state or individual characteristics such as personality.

1.3.3.1 Gender and movement

Being able to identify an individual from limited perceptual information has clear evolutionary advantages. Along these lines, it could also be considered adaptive to be able to assess other information from a person's bodily movement, such as their gender, mood state, or individual characteristics such as personality. [12] demonstrated that, on average, observation of just two step cycles was sufficient for participants to correctly identify gender from point-light displays.

Computational analysis of gait has been used to identify individuals ([68]) to classify walkers according to gender ([139, 216]). Given that gait studies with limited information from point-light displays have been effective in identifying gender, and that joint flexibility may have gender specificity [81], we should be able to identify gender well above chance level after having comparatively more information from all the joints.

1.3.3.2 Empathy and movement

In the sections 1.3.3.3 and 1.3.3.4, we see how some of the individual differences can affect the characteristics of the movements which are elicited by music. First in the queue is empathy, Empathy has been studied in relation to musical engagement from the perspective of bodily movement. The emotional content of the music sometimes shapes Music-induced movements since emotions are an essential element of musical expression [22, 20]. Burger et al. [20], found significant correlations between the perceived emotional content of the music and the movement indicating musical emotions accounts for the effect of music on movements to some extent.

Leman [111] suggests that the body acts as a mediator of musical engagement and intention between the mind and physical environment. This form of engagement is influenced by components such as Synchronization, Embodied Attuning, and Empathy. Synchronization, in principle, is a natural inclination to move along with a pattern in the physical environment. He further suggests the term 'inductive resonance' for describing the first step in engaging with the music, referring to imitation and prediction of beat-related features in the music. Synchronization is involved with low-level sensorimotor mechanisms; on the contrary, embodied attuning refers to the corporeal imitation of more complex music features including melody, rhythm, and harmony. This suggests movements not only display beat-related features in the music but also reflect and imitates the musical structure in order to understand it. Lastly, the component 'Empathy' aids in linking music, or rather musical features, with expressivity and emotions. Previous research has found relationships between empathy and responsiveness to changes in heard music or in dance partner [10, 32], and between EQ/SQ scores and music preferences [34, 78], yet there has been no study that has tried to associate empathy with general movement patterns using dance movement, nor have patterns related to systemizing tendencies. Hence, this thesis aims to predict scores on the Empathy and Systemizing Quotients (EQ/SQ) from participants' free dance movements and further, investigate which bodily joints were most important in defining these traits.

1.3.3.3 Personality and movement

Identifying an individual from limited perceptual information has clear evolutionary advantages. It could also be considered adaptive to be able to assess other information from a person's bodily movement, such as their mood state or individual characteristics such as personality. Satchell et al., [169] used motion capture of participants walking on treadmills to show that Big Five personality traits including Agreeableness, Conscientiousness, and Extraversion, as well as aggression, can be predicted from patterns of gait related to speed and range of movement of the torso. Koppensteiner [108] has shown that observers can use even limited movement information from the head and hands to judge Extraversion and Neuroticism. Thoresen et al. [187], found that observers made reliable judgments about personality from gait cues but that these judgments generally did not align with the self-reported personalities of the walkers, suggesting that the relationship between how personality is encoded and decoded in bodily movement is not necessarily straightforward.

1.3.3.4 Musical Preference and movement

From the perspective of most computational analysis, music can be defined as sound, its essential features yielding to the decomposition of waveforms. However, for the vast majority of history, musical sound could not be separated from its source; to whatever degree it may have evolved biologically to serve various human functions, music must be regarded as an embodied and socially embedded phenomenon [17, 45, 165]. It has been shown that musical compositions falling under the same genre share several musical characteristics, like rhythmic-structure, instrumentalness, and harmonics [196].

In music information retrieval, the detection of these audio features for automatically classifying music by their genre is an indispensable research application [9, 196]. It is postulated that such detectable acoustic cues correlate to very specific movements [73]. Research has shown intimate links between musical features and human movement, including the reflection of hierarchical rhythmic structures in embodied eigen movements [189], the reflection of higher-level musical structures in group movement to Electronic Dance Movement [177], and reflection of spectral and timbral features of music in dance [26]. Different genres show positive predilection towards specific movements, like listening to jazz music tends to make listeners nod their head, tap their foot, or head-banging associated with rock or stay still during classical music. Despite these associations and relationships between genres and bodily movement evident in empirical work, there appears to be a lack of research in predicting musical preferences from music-induced movements, a topic being addressed in this thesis.

Accordingly the objectives to investigate the influence of individual differences in the context of Active engagement are:

- How well individual differences can be decoded from music-induced movements?
- Which body joints contribute to this decoding?
- If the transitivity property holds, then how well individual music preferences can be decoded from music-induced movement?

1.4 Decoding Individual Differences via passive engagement

Cognitive psychology combines acquiring knowledge, storing the knowledge in memory, and retrieving it in a highly dynamic manner. The eminent neuroscientist, Joaquin M. Fuster, describes the perception-action cycle as the circular flow of information that takes place between the organism and its environment in the course of a sensory-guided sequence of behaviour towards a goal [66]. The cycle of perception & action acts as a feedback loop, which helps us build an understanding of how the world works. While the brain perceives the outcome of our actions, it either rewires the brain's existing neural networks or begins to create a new understanding of how the world works. The importance of perception extends beyond identifying objects or helping us take action within our environment. Various studies have shown that there are substantial individual differences in the way perceivers respond to feedback [95, 210, 211].

1.4.1 Individual Differences and Musical Preferences

Individual Differences play a crucial factor in modulating the way we perceive the world, and that in turn determines the way we act. From the scope of this thesis, action can manifest as 'movement' to heard music, and "choice of music". Choice of music has been assessed traditionally in lab settings wherein people use questionnaires like STOMP [160] to report their genre preferences. Some studies have shown that not only are these labels inconsistently applied across large commercial databases [145], but also the users' listening habits do not necessarily reflect their stated music preferences at the genre level. Other than the available self-reported means of presenting the choice of music, it can also be mined using Big-Data. However, with the advent of Big-Data, the studies have become more result-oriented as opposed to finding the noble cause for the choice of music. Research suggests that individual differences influence the (musical) experiences we seek [126]. Hence it becomes essential to look into the emotional aspect of music, as genres may indeed be limited in measuring musical preferences.

Individuals seek a varied range of emotional experiences via music. Identifying emotions from music helps in organization, retrieval, and music recommendation to satisfy an individual's personal needs. Hence, the task of identifying emotions from a given music track has been an active pursuit in the Music Information Retrieval community for years. However, music emotion recognition has been limited to the usage of acoustic content, social tags, and metadata [59, 30]. Lyrics have been largely neglected despite them being a vital factor contributing to musical reward, and the crucial role they play in eliciting emotions [218, 77]. Few studies have even reported the superior performance of music emotion classifiers based on features extracted from lyrics than audio [87, 213], but still the role of lyrics in music emotion recognition remains under-appreciated. Hence, we choose to identify emotion from music via lyrics and evaluate the association with individual differences thereof.

Several findings suggest that individual differences influence the selection of events and environments to maximize compatibility; thus, the experiences or circumstances they are in often relate to their personality traits [5, 27, 93]. Individuals spend around one-seventh of their waking time listening to music [132]. This significant amount of time to listen to music indicates that individuals may use music, particularly linguistic cues in lyrics, to create listening experiences that are compatible with their personality [156]. Qiu et al. [156] have shown associations between favorite song lyrics and personality traits, suggesting that listeners use them to satisfy their individual psychological needs and experience specific emotions. This however has never been investigated as it occurs naturally on online music streaming platforms. Hence, we further aim to understand the relationship between personality traits and emotions, as represented in the Valence-Arousal (VA) space, conveyed through lyrics of their preferred music.

Accordingly the objectives to investigate the influence of individual differences in the context of Passive engagement are:

- How individual differences are associated with the kind of emotional experiences one seeks from lyrics?

1.5 The Scope and Contributions

The central problem of this thesis is "Decoding Individual Differences via Musical Behaviour". Based on the concepts discussed above, we will talk about the influence of individual differences will be studied as to how they affect Embodied Cognition. In the scope of the thesis, we will look at how individual differences affect Embodied Cognition, and that can be observed in the way people move, the music they like via movement, and the emotional experiences we seek from lyrics. To that end, the following are the major contributions of this thesis -

- We propose a Machine-Learning model that predicts individual traits, specifically Big Five personality traits, and scores on the Empathy and Systemizing Quotients (EQ/SQ) from participants' free dance movements.
- We demonstrate the robustness of the proposed model by using another dataset and further providing conclusive evidence about the learned models' generalizability.
- We explored the relationship between musical induced movement and musical experience by successfully exploring the unseen space for musical preferences from participants' free music-induced movements.
- We propose a deep learning architecture to identify emotional connotations of music based on lyrics. We further compare our method with the existing deep learning and even traditional methods on relevant datasets and show state-of-the-art performance.
- We perform a study to understand the relationship between personality traits and the emotions, as represented in the Valence-Arousal (VA) space, conveyed through lyrics of their preferred music.

1.6 Thesis Roadmap

The remainder of this thesis is organized as follows -

- Chapter 2 summarises the work done on decoding individual differences via active engagement.
- Chapter 3 sums up the work done on decoding individual differences via passive engagement.
- Finally, Chapter 4 presents the conclusions and future directions to look out for.

Chapter 2

Decoding Individual Differences via active engagement

As the field of Music Information Retrieval continues to expands, it is vital to consider the multimodality of music and how aspects of musical engagement such as movement and gesture can be regarded. Bodily movement is universally associated with music and reflects important individual traits associated with music preference such as personality, mood, and empathy. Taking these aspects into consideration can benefit future multimodal MIR systems. Although the concept of an interactive music system that allows music playback to be regulated and altered through human movements has long been proposed [182], human-movement based interaction techniques and devices are rapidly gaining prominence in the field of HCI [71]. In this context, it makes decoding aspects of a individual via human movements a key and useful endeavor, which would then aid in the design of more personalized experiences.

2.1 Background and Motivation

Humans tend to have an astonishingly fine-tuned and robust ability to discern information about others based on bodily movement. For example, Cutting et al. [47] demonstrated that friends can recognise each other from their walk with only point-light (stick figure) displays of movement, without the need for other distinguishing features. Troje et al. [193] showed that participants were able to quickly learn to distinguish various individuals from point-light recordings of their gait, even scoring three times above chance accuracy when recordings were rotated and manipulated to suppress knowledge about size and speed. In a follow up study, Westhoff et al. [207] additionally used a Fourier Transform to remove the most prominent frequencies from point-light stimuli, and found that identification was still above chance, and that participants were easily able to generalize information learned from different viewing angles. Being able to recognise an individual based on limited perceptual information has significant evolutionary benefits, particularly in the uniquely social context of early human societies, where distinguishing between group members and non-members may be crucial for survival [84]. In the same lines, being able to assess other information from a person's bodily movement, such as their gender, mood state, or individual characteristics such as personality, may be considered adaptive. Barlcay et al.

[12] demonstrated that, on average, observation of just two step cycles was sufficient for participants to correctly identify gender from point-light displays. Observers using even limited movement details from the head and hands were able to judge extraversion and neuroticism, according to Koppensteiner [108]. Thoresen et al. [187] found that observers made reliable judgements about personality from gait cues, although these judgements did not always align with the self-reported personalities of the walkers.

Although it is not yet clear which features of human movement allow such information to be decoded, computational analysis of complex movement provides a way to investigate how knowledge about individual differences is encoded in subtle ways that are difficult to detect with the naked eye. Computational analysis of gait has been used to distinguish individuals [68] to classify walkers according to gender [139, 216], and to identify individual differences of personality [169, 184] and emotion [167].

While gait is perhaps the most common method for examining individual characteristics as they relate to features of bodily movement, a paradigm in which participants undergo free, spontaneous dance movements does have the potential benefit of greater individual variability of movement as well as theoretical connections to a range of psychological and social functions. Music and dance are found in every known human culture and play important roles in social contexts, and in most cultures represent largely inseparable phenomena [138]. Music and dance, according to Cross et al. [43, 44, 46], may have played a role for early humans in negotiating times of intra- and inter-group uncertainty, such as the changes marked by wedding or coming-of-age ceremonies, because of their capacity to convey non-specific but individually interpretable meaning, which he refers to as floating intentionality. Christensen et al. [38] have proposed six distinct neural and bio-behavioral functions of human dance, such as communication, self-intimation, and social cohesion.

A variety of experiments have been conducted to investigate the perception of individual characteristics from dance movement. For example, Van Dyck et al. [197] found that, when presented with side-by-side avatars of dancers expressing happiness or sadness, participants could correctly recognise which emotion was being expressed. Most of the research into how people perceive dance movement has focused on gender, both in terms of the perception of gender from dance movement and, in light of theories that human dance has played an evolutionary role in sexual selection, the perception of attractiveness and other attributes relevant to mate-selection from dance. Hufschmidt et al., [90] found that both children and adults could accurately identify dancer gender from avatar movements, while Weege et al. [206] found that female raters found the movements of male dancers with greater hand-grip strength to be more attractive. Although it's unclear which movement features raters use to determine gender, variations in movement between genders may bearise due to differences in average body structure and joint flexibility. Fink et al. [65] have hypothesized that, on an individual level, dance movements serve to signal information related to mate selection, while on a group level, dance movement serves such functions as coalition signaling and supporting group coordinated action.

Individual differences in dance movement have become the subject of computational research, which has started to reveal how such information is encoded in kinematic features. Carlson et al. [35] found that individual movement signatures during free dance, as captured by the three-dimensional co-variance of movement between joints, could be used to accurately classify individual dancers at a rate of 94 percent, suggesting that dance movements may be highly individualized. Individual differences in movement patterns have also been investigated in the context of disorders that have altered or impaired movement [50, 8, 190]. The EQ measures participants' tendency to empathize with others [15], while the SQ measures the tendency to think in terms of systems [14]. These two measures were originally developed to increase understanding of people with ASD, as in this population trait systemizing tends to be very high while empathy tends to be low. However, the EQ/SQ has also been used in the past to assess how these traits are distributed across the population. Although previous work has found relationships between empathy and responsiveness to changes in heard music or in dance partner [10, 32], and between EQ/SQ scores and music preferences [34, 78], general movement patterns associated with empathy have not been explored using dance movement, nor have patterns related to systemizing tendencies.

However, studies have found links between dancers' personality traits and aspects of their free dance movements, implying that group-level individual differences are often encoded in dance movements. Luck et al. [121], used principal component analysis to identify five components of free dance movement, showing that neuroticism, for example, was negatively correlated with global movement (movement of the whole body across the dance floor) but positively correlated with local movement (movement within the body), while Openness was positively correlated with local movement, and extraversion was positively correlated with all five components, suggesting extraverted participants tend to move more in general. Carlson et al. [32] found that dancers' self-reported trait empathy related to their responses to different dance partners. These studies suggest that it should be possible to predict participants' personalities from their dance movements, although this has not yet been done.

Although Carslon et al. [35] were able to identify individuals via dance movement with high accuracy, in the same study their attempt to classify which of the eight genres the participants were dancing to only had a 23 percent accuracy. One possible explanation for this can be that individual factors, such as personality and dancers' preferences for some genres over others, had a greater influence on their movements than did the specific characteristics of the musical stimuli. However, research into how music preferences are embodied in dance is limited. Luck et al. [120] discovered that participants' preferences for different musical excerpts resulted in a U-shaped curve on a variety of kinematic features, suggesting that both low and high preferences for specific musical excerpts resulted in more movement. However, more research is required to investigate the interaction of preference and movement at the level of musical genres, which, despite being a fuzzy concept [145], are nevertheless a common way music preferences are discussed and related to individuals' self-concepts and judgements of others [160, 161]. Previous work additionally provides broad support for the existence of relationships between music preferences and individual differences, including personality trait empathy, and trait systemizing [79, 160]. Greenberg et al. [78] suggested that even music may play a role in increasing empathy in people with empathy-related disorders by exploring the relationship between music preference and empathizing-systematizing theory. In light of the transitivity property, if movement patterns are associated with personality and personality with music preferences, then movement patterns should be associated with music preferences. Indeed, it may be logical to hypothesize that listening to music of a preferred genre leads to the recruitment of cognitive faculties related to greater attention and involvement that in turn results in potentially identifiable implicit movement patterns.

2.2 Objectives and Hypothesis

The current study attempts to address these many potential influences on an individual's music-induced movements, by using free dance movement to predict dancers' gender, cognitive styles of thinking, personality traits, and music preference. To this end, we use data sets from [33, 121] and employ machine learning techniques to predict Gender, Personality, and musical preferences from music-induced movement. We follow Carlson et al. [35], in making use of co-variance between joints as a kinematic feature, as [193, 207] has suggested that phase relationships between joints may have perceptual validity as a feature used in perceiving human movement. We aim to assess the robustness of our proposed machine learning architecture by comparing its accuracy on the two different data sets, allowing us to explore the degree to which findings generalize. Additionally, we attempt at proposing a *Joint Importance profiles* which is novel measure to evaluate the importance of specific joints and their relative movement in characterising personality and music preferences.

We expect, based on previous research showing that gender can accurately be classified from gait, that it will be possible to accurately classify gender from free dance movement as well. The current study also focuses on identifying scores on the Empathy Quotient (EQ) and Systemizing Quotient (SQ), from participants' free dance movements to various genres of music. We also expected, based on previous research, that personality will be possible to predict from joint co-variance. Given that joint co-variance relates to local movement (as opposed to global movement across a dance floor), in light of Luck et al.'s [121] findings, our predictions may be more accurate for neuroticism and less accurate for openness and extraversion. Due to evidence of relationships between personality and musical preferences, we additionally expect that music preferences may also be predicted using participants' free dance movements.

2.3 Methods

2.3.1 Datasets

We use two datasets in our study. They comprised spontaneous movement data of participants moving to musical stimuli in response to musical stimuli of various genres. Music induced movements of an individual is influenced by the genre of the music they are listening to. Various movement patterns are more representative of particular genres [73], hence it is pertinent to include music from different genres in any investigation of music-induced movement. The data were obtained via Motion Capture systems.

	Dataset-1	Datatset-2
Participants	58	60
Gender	41 Females, 17 Males	43 Females, 17 Males
Age	Mean: 26.8 years, Std: 4.7 years	Mean: 24 years, Std: 3.3 years
Personality (BFI)	✓	✓
Music Preferences (STOMP-R)	✓	✗

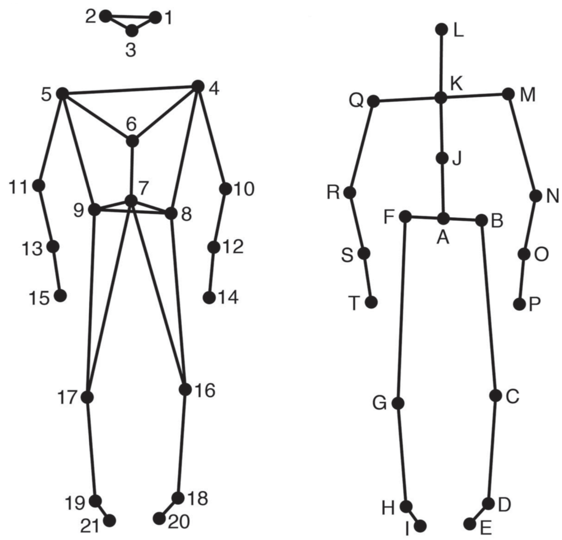
Table 2.1: Dataset Comparison

2.3.1.1 Dataset - 1

The first dataset is from the study by [33] comprising data from 58 university students (41 females; mean age = 26.8 years, std = 4.7 years). Thirty-six reported having received formal musical training, while twenty participants reported having received formal dance training. The stimuli comprised sixteen 35-second excerpts from eight genres, in randomized order: Blues, Country, Dance, Jazz, Metal, Pop, Rap, and Reggae. Participants were instructed to listen to the music and move as freely as they desired that felt natural with regards to the stimuli presented. Additionally, participants were encouraged to dance if they so desired, but only within the designated capture area. The aim of these instructions was to create a naturalistic setting, such that participants would feel free to behave as they might in a real-world situation.

Participants' movements were recorded using a twelve-camera optical motion-capture system (Qualisys Oqus 5+), tracking at a frame rate of 120 Hz, the three-dimensional position of 21 reflective markers attached to each participant (see Figure 1a). The MATLAB Motion Capture (MoCap) Toolbox [25] was used to analyze this data. Data were first trimmed to the duration of each stimulus and, Following this, to simplify analysis and reduce redundancy, the original markers were transformed to yield a secondary set of markers subsequently referred to as joints, the locations of which are depicted in Figure A.5b.

The locations of joints B, C, D, E, F, G, H, I, M, N, O, P, Q, R, S, and T are identical to the locations of one of the original markers, while the locations of the remaining joints were obtained by averaging the locations of two or more markers; Joint A: midpoint of the two back hip markers; J: midpoint the shoulder and hip markers; K: midpoint of shoulder markers; and L: midpoint of the three head markers. The data were then transformed to a local coordinate system, in which the location of each joint was expressed in relation to the root joint (Figure A.5b, marker A), which is defined as the origin, and the line connecting the hip markers as the mediolateral axis, allowing for comparison between dancers regardless of their orientation within the original mocap space. Further, using the Mocap Toolbox, the instantaneous velocity of each marker in each direction was calculated by time differentiation followed by the application of a 2nd-order Butterworth filter with a cutoff frequency of 24Hz.



(a) (b)

Figure 2.1: Marker and joint locations (a) Anterior view of the marker locations a stick figure illustration; (b) Anterior view of the locations of the secondary markers/joints used in animation and analysis of the data

2.3.1.2 Dataset - 2

For the second dataset, sixty-four participants took part in the motion capture data collection in a study by [121]. Four participants were excluded from further analysis due to incomplete data. Thus, 60 participants were retained (43 females; mean age = 24 years, std = 3.3 years). They were recruited based on a database of 952 individuals that contained their scores of the Big Five Inventory [100]. The aim of the original study involved recruiting high- and low-scoring individuals for each of the five dimensions. Six participants had received formal music education, while four participants had a formal background in dance. Participants were presented with 30 randomly ordered musical stimuli. Among the 30 musical stimuli, five stimuli belonged to each of the following popular music genres: Jazz, Latin, Techno, Funk, Pop, and Rock. All stimuli were 30 seconds long, non-vocal, and in 4/4 time, but differed in their rhythmic complexity, pulse clarity, and tempo. As described for Dataset-1, the participants were directed to move freely to the music.

A similar process was followed as described in section 2.3.1.1 to transform the the 28-marker data into a set of 20 secondary markers, referred to hereafter as joints, displayed in Figure 2.2. The locations of these 20 joints are depicted in Figure 2.2. The locations of joints C, D, E, G, H, I, M, N, P, Q, R, and T are identical to the locations of one of the original markers, and the locations of the remaining joints were obtained by averaging the locations of two or more markers. Data were trimmed and transformed to a local coordinate system.

The kinematic variables, position and velocity were estimated using the Savitzky–Golay smoothing FIR filter [170] with a window length of seven samples and a polynomial order of two. In the time

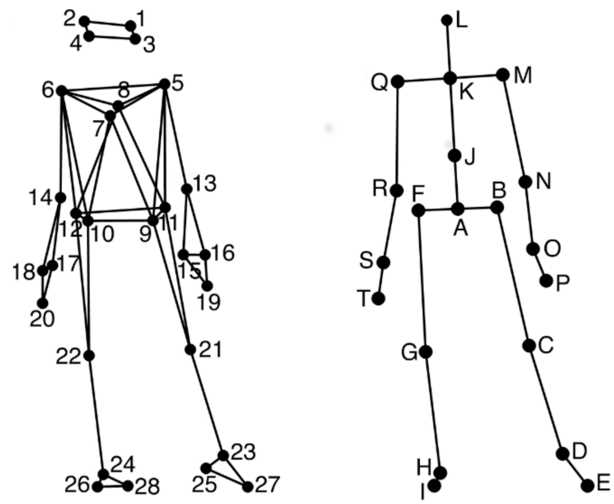


Figure 2.2: Anterior view of the location of the markers attached to the participants' bodies; Anterior view of the locations of the secondary markers/joints used in the analysis.

derivatives, these values were discovered to provide the optimal combination of precision and smoothness.

2.3.1.3 Personality and Trait Empathy Measures

The Big Five Inventory (BFI) was used to capture the five personality dimensions, namely, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN) [101]. This data is available for both the datasets. Trait Empathy was measured using the EQ- and SQ-short form version, developed and validated by Wakabayashi et al. [202], as a result giving an empathizing-quotient and systemizing-quotient score per participant. Data for Trait Empathy is available for Dataset-1.

2.3.1.4 Music Preferences Measure

Self-report measures may be too presumptive of shared genre definitions between listeners, while listener ratings of expert-selected music may fail to reflect typical listeners' genre-boundaries. A possible solution to this problem has been the creation of genre-free models of music preference. Rentfrow et al. [158] sought to re-conceptualize music preference by focusing on underlying musical features. The authors first developed preference factors based on participants' ratings of heard stimuli and then had judges assign attributes to the excerpts including musical and genre-based attributes.

For Dataset-1, the participants' music preferences were assessed. A revised and updated version of the "Short Test Of Music Preferences" (STOMP)[160], was used as a starting point for genre selection, also know as the STOMP-Revised (STOMP-R) [158]. This version includes genres not found in the original STOMP, thus allowing a larger pool of genres to choose from. Genres that were not suitable for dancing (e.g., Classical, Opera) were eliminated. After careful considerations regarding the ”Danceability” of stimuli belonging to genres and several other factors [34] the total of 48 tracks from 12 genres were considered in the online listening experiment. Participants rated their liking for the heard stimuli on a seven-point Likert scale. On a seven-point Likert scale, participants rated their liking for the heard stimuli. An excerpt could be listened to more than once by participants. Participants then completed a version of the STOMP-R that only included the 12 genres used in the experiment after rating all 48 excerpts.

2.3.2 Movement Features

The pipeline for movement feature extraction is illustrated in Figure 2.4. We calculated the position and instantaneous velocity for the set of 20 markers as specified in Sections 2.3.1.1 and 2.3.1.2, using the MoCap Toolbox ([25]). We calculated the marker by marker covariance matrix for each participant, for each of the stimuli for position and velocity data separately. The covariance between all the marker time series data (position or instantaneous velocity) in each direction (X, Y, and Z) for each stimuli measured the degree to which the movement of any two markers in any direction covaried with each other across the entire stimulus. We used correntropy [118], a non-linear measure to calculate covariance between the marker time series x_i and x_j , given by:

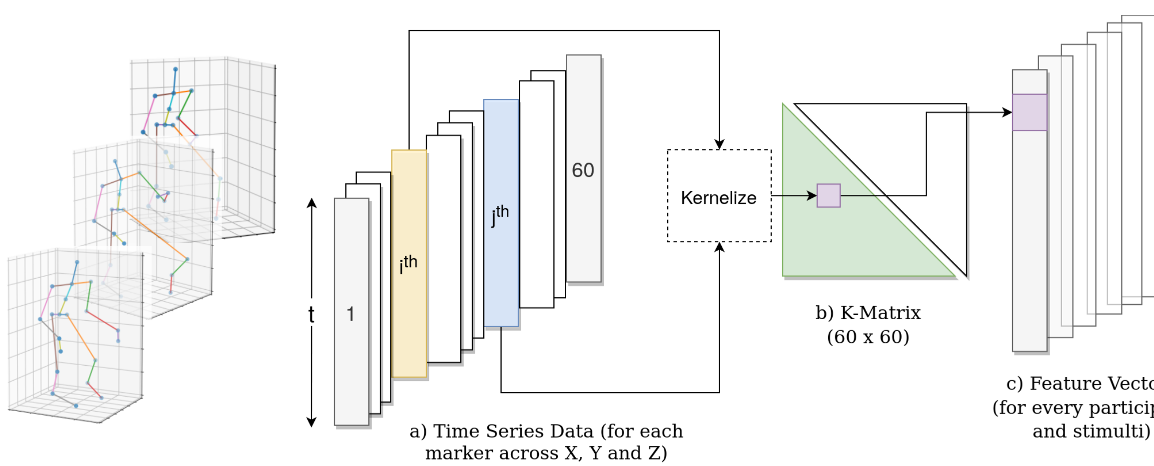


Figure 2.3: Overview of our Feature Extraction pipeline. Given the position of joints across time frames in 3D Euclidean space(a), we apply pairwise correntropy between time series x_i and x_j and calculate the K-matrix (b). Then, taking the lower triangular part of the symmetric covariance matrix, we get the feature vectors (c).

$$K(x_i, x_j) = e^{\frac{-||x_i - x_j||_2^2}{2\sigma^2 T^2}} \tag{2.1}$$

where $||x_i - x_j||_2$ is the L2-norm between x_i and x_j, σ is a constant (12.0 in our case), and T is the length of the time-series. The L2-norm is normalized according to the length of the time-series to take account of different lengths of stimuli. Since the number of joints is 20 and each joint has three coordinates, the dimension of the covariance matrix K turns out to be 60×60. Owing to the symmetric property of K, the lower triangular part excluding the diagonal elements was vectorised to produce a feature vector of length 1770 for each participant and for each stimulus. The features extracted for Position and Velocity data were used to train the regression models.

We also run our experiments using the Normalized feature vectors calculated by using Position and Velocity, we employed standard Gaussian Normalization technique:

$$\hat{X} = \frac{X - \mu(X)}{\sigma(X)} \tag{2.2}$$

where $\hat{X}$ is the feature vector, $\mu(X)$ is the mean and $\sigma(X)$ is standard deviation.

2.3.3 Classifying Gender from Movement

In order to identify gender from movement, we used Linear Support Vector Machines (SVM), a classification technique that identifies the optimal solution for separating the classes in some hyperspace. SVM creates the largest possible buffer space between the two classes which is defined as the Optimal Separating Hyperplane, or OSH. The OSH minimises the risk of incorrectly classifying any new data ([179]). The Euclidean Distance or L2-norm was used as a penalty measure to identify the optimal class boundary. The tolerance of $1e^{-5}$ was used as the stopping criteria; this tells model to stop looking for a more optimal solution once some tolerance is achieved.

2.3.4 Predicting Individual Differences from Movement

We aimed to predict Empathizing Quotient, Systemizing Quotient, Personality and STOMP-related Musical preferences from movement patterns. To this end, we used Bayesian Regression and Principal Component Regression(PCR) for training our computational models. The most common regression model for value prediction tasks used is Linear Regression. The goal here is to find an optimal line that minimizes the total prediction error. But this model treats its parameters as unknown constants whose values must be derived. Moreover, the weights become sensitive when the dataset size is large. So to prevent the model from overfitting, we took principal components of the features to train the model. We also approached this problem by using Bayesian Regression other than Principal Component Regression (PCR)[1].

In Bayesian Regression the parameters are treated as random variables belonging to an underlying distribution. Depending on the dataset and its size, we can be more or less certain about the model weights. The parameters of the Bayesian model belong to a distribution, unlike Linear Regression

[1]Detailed analysis of Principal Component Regression (PCR) and Bayesian Regression are discussed in Appendix A

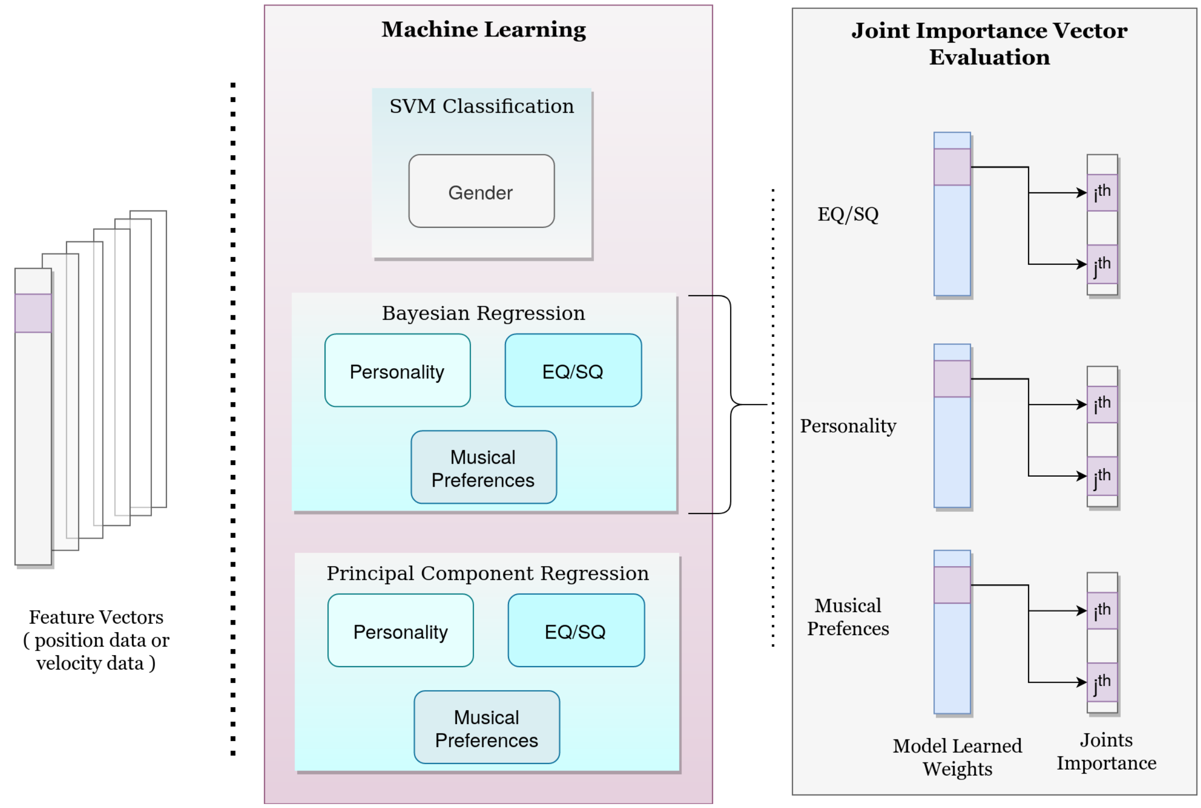


Figure 2.4: Overview of our ML tasks and *Joint Importance vector* evaluation pipeline. Given the extracted feature vectors, we apply SVM for Gender Classification, and Bayesian Regression & PCR for predicting EQ/SQ, personality, and musical preferences. Then, using the trained regression model, we evaluate trait-wise *Joint Importance vectors*.

model where we get the single estimate of the model parameters. Thus, the predictions of the model also belong to a distribution. Moreover, Bayesian Regression provides confidence bounds for our prediction which enable us to evaluate the uncertainty of the predictions. We used the Python based scipy toolkit ([199]) to perform our analyses.

2.3.4.1 Evaluation Metrics

To evaluate the accuracy of the models, we use variance explained (i.e., R^2) and Root-mean-square-error (RMSE) as performance measures.

(a) Root-Mean Square Error (RMSE): If $\hat{y}_i$ is the predicted value of the i^{th} sample and y_i is the corresponding true value for total n samples, then the RMSE estimated is defined as:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.3}$$

We can interpret the model's performance using RMSE as it estimate the deviation of an observed value from our model's prediction

(b) R^2 Score: If $\hat{y}_i$ is the predicted value of the i^{th} sample, y_i is the corresponding true value for total n samples, and $\bar{y}$ is the mean of the ground truth data, the estimated R^2 is defined as:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.4}$$

RMSE estimates the deviation of an observed value from our model's prediction, and has the useful property of being in the same units as the variable being predicted. The R^2 is a measure of variance explained in the data and represents goodness-of-fit.

RMSE and R^2 are estimated as an average of 5-fold cross-validation. To this end, we split the data randomly in 5 equal chunks, where each chunk is used as a testing set at some point. In the first iteration of 5-fold Cross Validation, the first chunk (that is, 20 % of total data) was used to evaluate the model while the remaining 80 % of the data was used to train the model. This process was repeated five times such that all data was used at least once as the testing set. This process helps in obtaining a more generalized estimate of classification accuracy.

2.3.5 Key movement patterns associated with predicting individual differences

In order to identify the key movement patterns associated with predicting individual differences, we propose a novel *Joint Importance* metric. *Joint Importance* is a measure of the contribution of each of the joints in predicting each class (e.g., how important the shoulder or elbow joint is in predicting a given personality trait). For each iteration, we obtain a trained model with weights associated with each of the 1770 input feature vector as mentioned above. Each of these 1770 elements is associated with pairwise correlation of joints in each of the three dimensions in Euclidean space. We first map the 1770 weights to the respective 20 joints. Then, for each joint, we find the *Joint Importance* by summing the absolute weights of those associated with the respective joint across all the 5 iterations. We sum the absolute weights because the magnitude preserves the importance of that joint-pair correlation. The pseudo-code to get the *Joint Importance* is as follows:

Algorithm 1 Joint Importance

Result: Calculate a vector J of 20 dimensions representing importance of each joint.

W is the weight vector; J is the importance vector initialised with 0; S contains lower triangular indices excluding diagonal indices; 0-indexing is followed;

```
S ← LowerTriangularIndices(60 × 60)
N ← S.length()
for k = 0 : N − 1 do
    (i, j) := S(k)
    (î, ĵ) ← IndexToJoint(i, j)
    J(î) := J(î) + |W(k)|
    J(ĵ) := J(ĵ) + |W(k)|
end for
return J
```

This process results in the *Joint Importance vector* of length 20. In order to visualize and compare *Joint Importance* across classes, we perform Min-Max Normalization. Min-Max Normalization is a standard procedure where the values the minimum is transformed to 0, while the maximum gets set to 1.

$$\overline{JI}[i] = \left(\frac{JI[i] - min(JI)}{max(JI) - min(JI)}\right) \forall JI[i] \tag{2.5}$$

where $\overline{JI}$ represent the normalized *Joint Importance vector*. Algorithm 1 describes the pseudo-code to get the importance of joints from the weights of the trained regression model.

Min-Max Normalization is a standard procedure where the values the minimum is transformed to 0, while the maximum gets set to 1.

2.4 Analysis and Results

2.4.1 Gender Classification

The results for Gender classification on Dataset-1 and Dataset-2 can be found in Table 2.2.

As can been seen from Table 2.2, clearly position data gives a higher accuracy than velocity data for both the datasets. We achieved slightly higher accuracy for the second dataset, which can be attributed to the fact that Dataset-2 is almost twice as big compared to Dataset-1.

Classification Accuracy (in %)	Dataset-1	Dataset-2
Position	96.53	98.76
Velocity	84.59	86.33

Table 2.2: Gender classification Results using Position data and Velocity data for Five Personality Traits using Bayesian Regression on both the datasets.

2.4.2 EQ and SQ Prediction

The results for EQ prediction are in Table 2.3 and SQ prediction are in Table 2.4. The results are calculated using 5-fold cross validation. The range of EQ and SQ is 0-80. The boldface values represent the best score.

Input	PCR		Bayesian Ridge	
	RMSE	R^2	RMSE	R^2
Position	3.071	0.708	**2.722**	**0.771**
Position(N)	3.201	0.684	2.733	0.765
Velocity	4.938	0.249	4.343	0.423
Velocity(N)	4.583	0.353	4.015	0.503

Table 2.3: Prediction Results for Empathizing Quotient

Input	PCR		Bayesian Ridge	
	RMSE	R^2	RMSE	R^2
Position	2.398	0.781	**2.161**	**0.867**
Position(N)	2.363	0.786	2.502	0.838
Velocity	4.448	0.252	3.832	0.469
Velocity(N)	4.211	0.329	3.714	0.552

Table 2.4: Prediction Results for Systemizing Quotient

The 'N' in the tables denote that Gaussian Normalization was applied on the features. We trained and evaluated two different models for each of the aforementioned tasks. We can see that using position data, instead of velocity data, to generate the feature vectors, gave us the best results. Also, we can see that the Bayesian Regression gave better results than Principal Component Regression on both metrics. So from here on, we will be using Bayesian Regression for other prediction and analysis tasks.

2.4.2.1 Joints' Importance

For evaluating *Joint Importance* we used learned weights of the model using position data across the different prediction tasks. In order to get an overview of the relative importance of joints, we averaged those joints which occur in pairs (e.g., L and R shoulder). This was also done for hips, knee, ankle, toe, elbow, wrist, and finger. Thus reducing the total number of joints to 12, this is referred to as *Joint*

Importance profile. Altogether the results in characterizing an individual trait is dominated by the limbs than the core of the body.

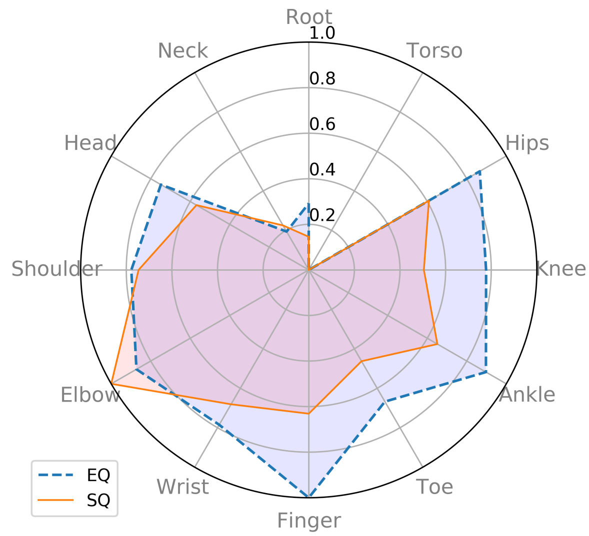


Figure 2.5: Relative importance of Joints in EQ and SQ Tasks using the Position Data.

From the relative *Joint Importance* depicted in Figure 2.5, we observe that 'Ankle', Elbow' and 'Shoulder' play an important role in determining EQ and SQ of an individual, whereas 'Neck' and 'Torso' have a negligible contribution. We also infer that 'Finger', 'Hip', and 'Knee' are more crucial joints for predicting EQ than for SQ whereas 'Elbow' holds significantly higher importance for predicting SQ than for EQ.

2.4.3 Personality Prediction

Overall, Bayesian Regression demonstrated superior performance over Principal Component Regression, hence we only report those results. Moreover, Bayesian Regression provides confidence bounds for our prediction which enable us to evaluate the uncertainty of the predictions. The results for personality prediction on Dataset-1 and Dataset-2 can be found in Tables 2.5 and 2.6 respectively.[2] Moreover, using position data as input features provided superior prediction when compares to velocity data. Hence we report here detailed results of Bayesian Regression based on position data.

Bayesian Regression performed considerably well for both the data sets as evidenced by the high proportion of variance explained, that is, average R^2 score of 76.3% and 89.0% across all traits for Dataset-1 and Dataset-2 respectively using the position data. On the other hand, using velocity data resulted in a lower average R^2 score of 44.4% and 39.0% for Dataset-1 and Dataset-2 respectively. Furthermore, the considerably low RMSE values when compared to the range of personality values (i.e.,

[2]The results using Principal Component Regression can be found in supplementary material.

Input	Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Position	**0.197**	**0.776**	**0.317**	**0.760**	**0.384**	0.743	**0.252**	**0.776**	**0.384**	**0.758**
Position(N)	0.227	0.740	0.332	0.690	0.414	**0.756**	0.273	0.716	0.390	0.739
Velocity	0.332	0.464	0.487	0.415	0.556	0.523	0.440	0.335	0.557	0.483
Velocity(N)	0.304	0.527	0.426	0.543	0.501	0.623	0.408	0.442	0.461	0.654

Table 2.5: Dataset-1: Prediction Results for Five Personality Traits using Bayesian Regression

Input	Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Position	**0.192**	**0.897**	**0.218**	**0.898**	**0.251**	**0.883**	**0.173**	**0.892**	**0.239**	**0.887**
Position(N)	0.213	0.886	0.237	0.875	0.280	0.861	0.204	0.848	0.271	0.859
Velocity	0.488	0.424	0.556	0.360	0.571	0.449	0.413	0.384	0.612	0.343
Velocity(N)	0.342	0.717	0.369	0.722	0.417	0.706	0.294	0.692	0.418	0.681

Table 2.6: Dataset-2: Prediction Results for Five Personality Traits using Bayesian Regression

1.0 - 5.0) reflects high model accuracy. The average RMSE scores for Dataset-1 and Dataset-2 using the position data was found out to be 0.31 and 0.21 respectively, which was considerably smaller than the avg. RMSE scores of 0.47 and 0.53 using the velocity data for Dataset-1 and Dataset-2 respectively. We can see that using position data to extract features gave the best results on predicting all five personality traits on the dataset. We can concur that using position data instead of velocity data in the kernelized space is better for these regression tasks.

2.4.3.1 Joints' Importance

Figures 2.6 and 2.7 display relative personality-wise *Joint Importance* for Dataset-1 and Dataset-2 respectively. The black line plotted in each sub-figure indicates the mean of *Joint Importance* across personality traits for the respective dataset. The farther away from the mean the *Joint Importance* value is, the more important that joint is in characterizing that trait.

Figure 2.6 displays the relative *Joint Importance* of personality along with the mean plotted in each sub-figure. The farther away from the mean the *Joint Importance* value for an individual joint is, the more important it is in characterizing that trait. Some similarities in the *Joint Importance profiles* across the personality traits can be attributed to the inherent correlation that exists among them [3]. We observe that it is the 'Finger', 'Elbow', and 'Knees' that contribute to Feature Importance whereas 'Root', 'Neck' and 'Torso' have negligible contribution. For characterizing Conscientiousness, 'Shoulders', 'Knees' and 'Neck' play a crucial role while 'Head' and 'Toe' plays an important role for Extraversion. For Agreeableness, 'Neck' and 'Wrists' have relatively less importance as compared to other joints whereas, 'Wrists' play an important role in Openness. Finally, there are no significant defining features

[3]The table for Spearman Correlation between the personality traits is provided in the supplementary material.

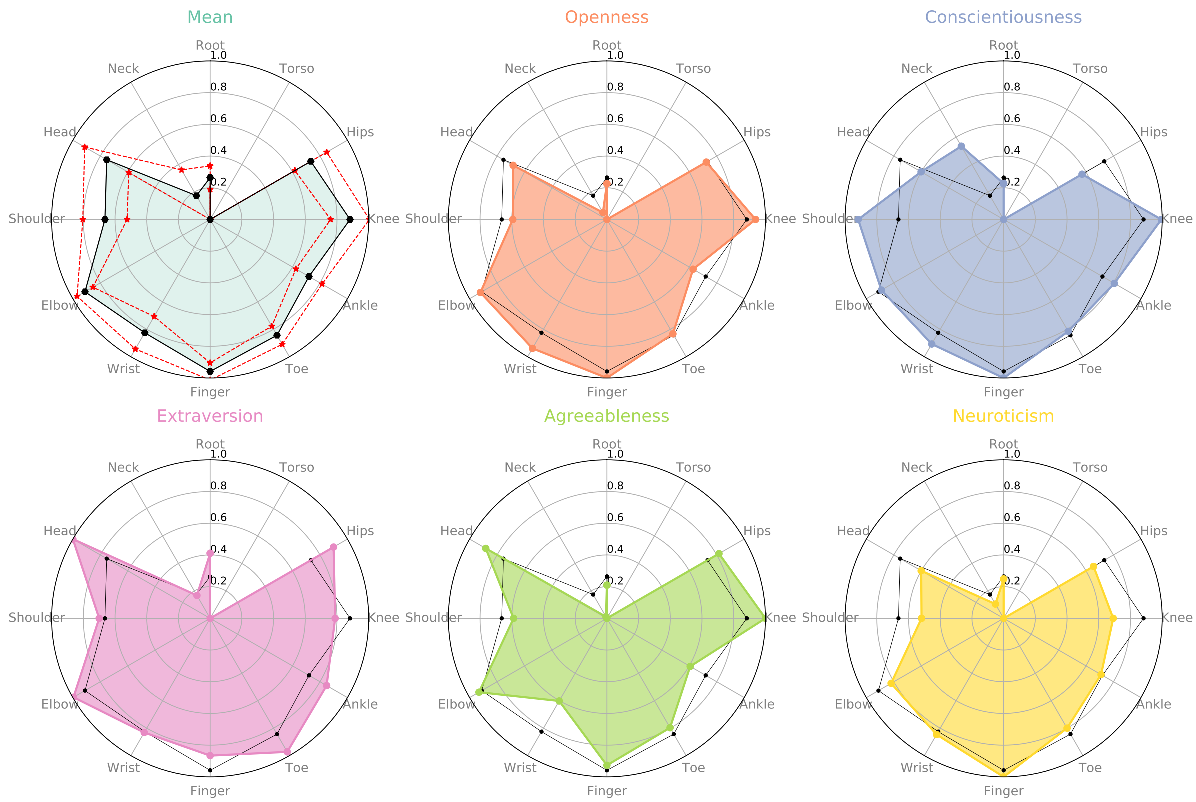


Figure 2.6: Dataset-1: Relative importance of Joints of the five personality traits(Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) using the Position Data. The black line indicates the mean importance of the corresponding joint marker. The red dotted line in the top left sub-figure indicates the standard deviation about the mean.

for Neuroticism, which indicates that their expression in Dance Movements through Music-Induced Movements are very limited.

As seen from the Figures 2.6 and 2.7, Extraversion and Conscientiousness shows some similarity across the datasets, reflecting the importance of specific joints. For Extraversion, the 'Head', 'Shoulder', 'Elbow', 'Knee', and 'Hips' are consistently more important across datasets. Similarly for Conscientiousness, the 'Head', 'Shoulder', and 'Elbow' are consistently more important. This is reflected in the Spearman correlation between the *Joint Importance profiles* across datasets; Conscientiousness exhibited significant correlation (*r = .73, p < .01*) while Extraversion demonstrated borderline significance (*r = .52, p = .08*).

2.4.4 Generalization of trait-wise movement patterns

To investigate consistency of movement patterns for individual traits across datasets, we perform Agglomerative Hierarchical clustering on the *Joint Importance profiles*. Agglomerative Hierarchical

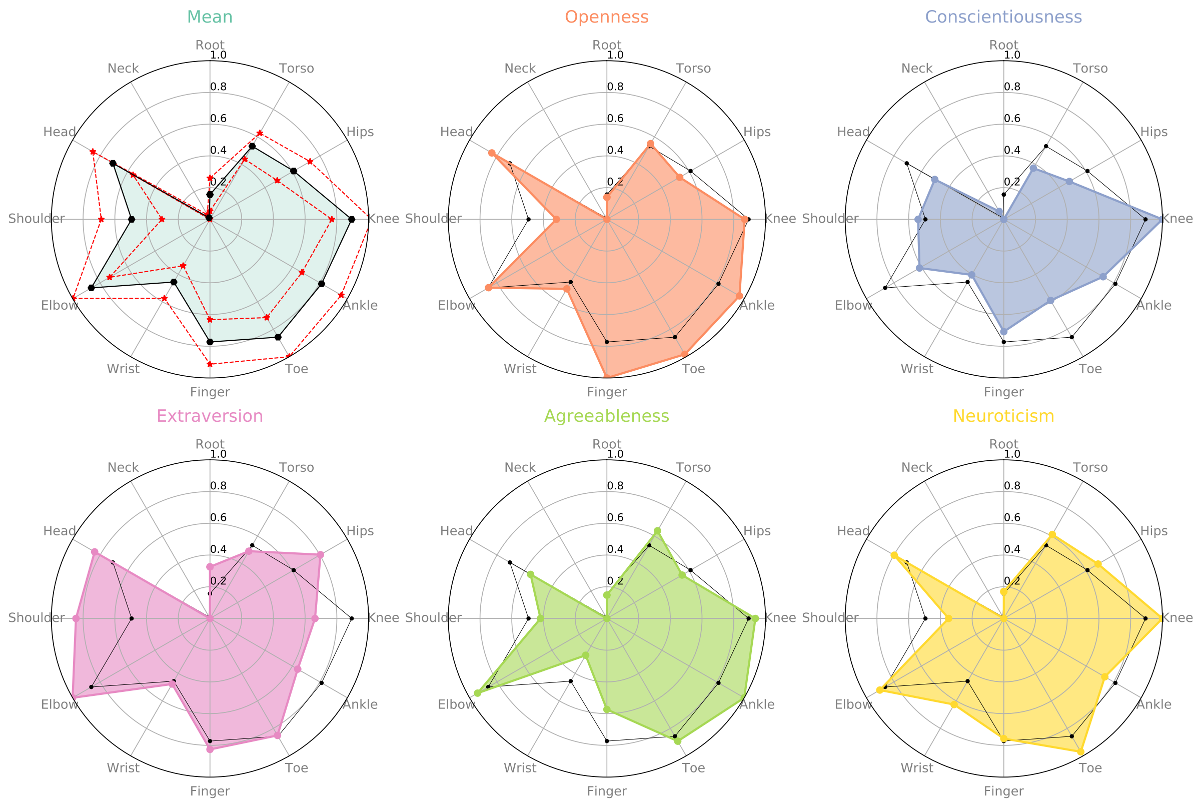


Figure 2.7: Dataset-2: Relative importance of Joints of the five personality traits using the Position Data. The black line indicates the mean importance of the corresponding joint marker. The red dotted line in the top left sub-figure indicates the standard deviation about the mean.

clustering allows to cluster the profiles in a hierarchical manner by capturing inherent similarities. We apply the commonly used ward's linkage method to compute the distance between the clusters, as it tends to produce homogeneous cluster and yields compact spherical clusters when compared to other approaches, [96]. Figure 2.8 shows the dendrogram of the learned *Joint Importance profiles* of different personality traits across data sets.

As seen in Figure 2.8, traits Extroversion, Conscientiousness, and Agreeableness, demonstrate high inherent similarity in *Joint Importance profiles* suggesting that individual prediction models are similar. In other words, this implies that there exist trait-specific movement patterns. On the other hand, Openness and Neuroticism cluster together suggesting similarities in their overall movement patterns. However, agglomerative clustering only reveals similarities at a cluster level but not how each of the elements within a cluster relate to each other. Hence, we perform Multidimensional Scaling (MDS), a common technique that is used to visualize interrelationships within high-dimensional data in a lower dimensional space. We perform MDS by creating the dissimilarity matrix based on the Euclidean distance between the *Joint Importance profiles* and projecting them onto a 3-dimensional space. Figure 2.9

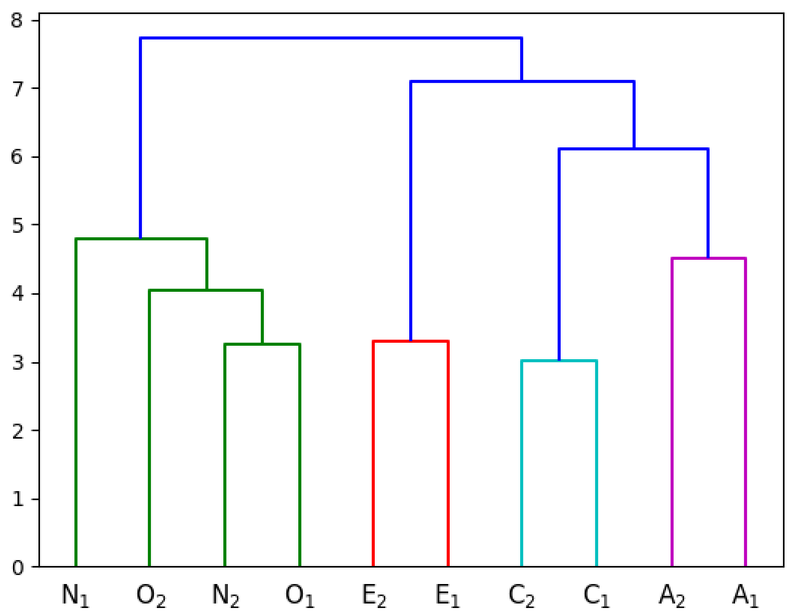


Figure 2.8: Dendrogram of the learned *Joint Importance profiles* for Personality traits in different datasets. Number $\{1, 2\}$ in the subscript denotes the dataset number.

provides the 3D representation of *Joint Importance profiles* similarity as a result of MDS revealing high trait-wise similarity across data sets.

2.4.5 Prediction of Music Preference

As was the case in personality prediction models, Bayesian Regression outperformed Principal Component Regression; thus, we report the results using Bayesian Regression. In addition, using position data as input features provided superior prediction when compares to velocity data. The results of music preference prediction using Bayesian regression for Dataset-1, with position data as input features, can be found in Table 2.7.

Metric	Input	Blues	Country	Dance	Funk	Jazz	Metal	Oldies	Pop	Rap	Reggae	Rock	Soul
R^2	P	**0.806**	**0.727**	**0.840**	**0.719**	**0.772**	**0.838**	**0.751**	**0.853**	**0.721**	**0.733**	**0.794**	**0.773**
	V	0.445	0.490	0.485	0.462	0.464	0.524	0.438	0.495	0.420	0.405	0.566	0.420
RMSE	P	**0.583**	**0.962**	**0.634**	**0.649**	**0.720**	**0.850**	**0.635**	**0.534**	**0.943**	**0.807**	**0.492**	**0.717**
	V	1.000	1.324	1.139	0.897	1.103	1.460	0.951	0.989	1.373	1.240	0.717	1.049

Table 2.7: The Results for 12 Stomp Traits using Bayesian Regression
Default Scaling at 1 - 7

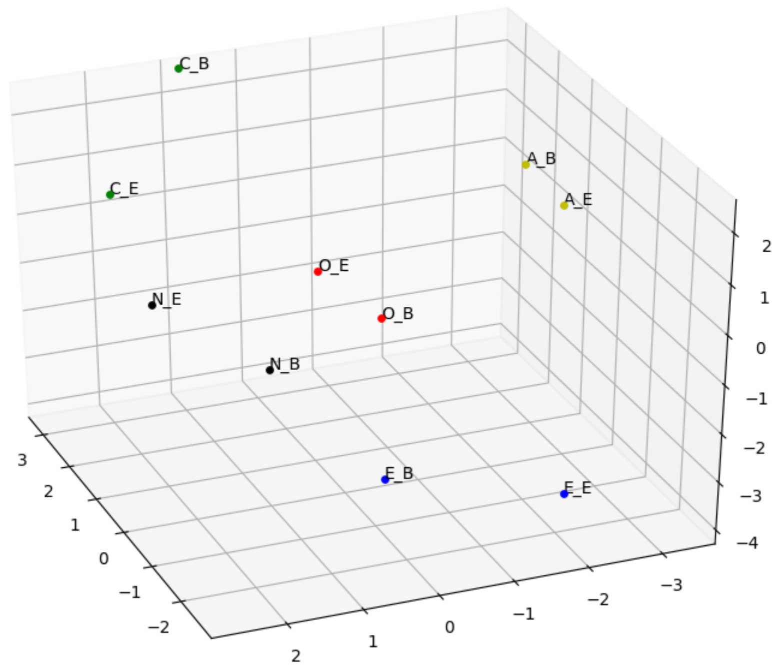


Figure 2.9: Multi-Dimension Scaling (MDS) results for the learned *Joint Importance profiles* for personality traits on both the datasets. Number $\{1, 2\}$ in the subscript denotes the dataset number.

From Table 2.7, we can see that the Bayesian regression model was able to predict musical preferences with high accuracy. Utilizing (P)osition data to extract features gave the best results on predicting musical preferences compared to (V)elocity data for the dataset. The range for R^2 score varies between 72% and 85% for different genres with an avg. R^2 score of 77.5% using the position data. Additionally, the low RMSE values ranging from 0.49 to 0.96 when compared to the range of self-reported music preference values (i.e., 1.0 - 7.0) demonstrates high model precision.

2.4.5.1 Joints' Importance

Figure 2.10 displays Spearman Correlations between *Joint Importance profiles* that contribute to the prediction of Musical Preferences. Based on our hypothesis, we expected to see similar *Joint Importance profiles* for genres that are jointly preferred per trait. As can be seen in Figure 2.10, significant positive Spearman correlation was observed between the *Joint Importance profiles* of Jazz and Blues (r = 0.60, $p < .05$), Soul and Funk (r = .7, $p < .05$), Pop and Rock (r = 0.62, $p < .05$), and Rock and Oldie (r = 0.66, $p < .05$), supporting our hypothesis. Additionally, we also observe significant negative correlations between the *Joint Importance profiles* of Rock and Jazz (r = -0.75, $p < .01$), and Metal and Rap (r =

-0.75, p < .01), Funk and Pop (r = -0.68, p < .05), Blues and Country (r = -.6, p < 0.05), and Pop and Soul (r = -0.58, p < .05).

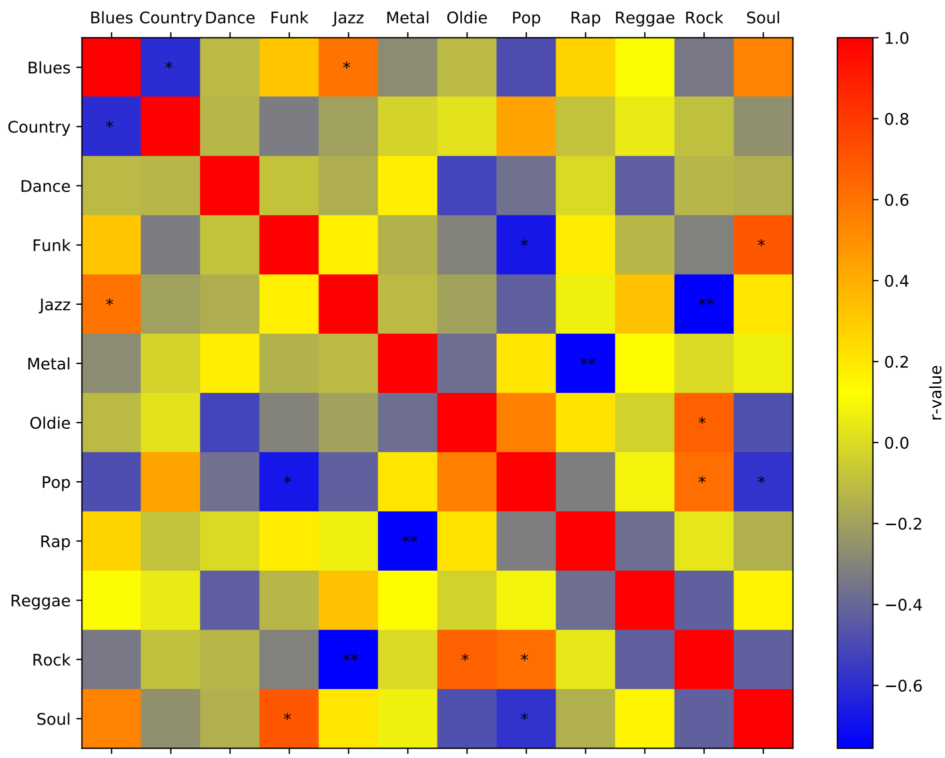


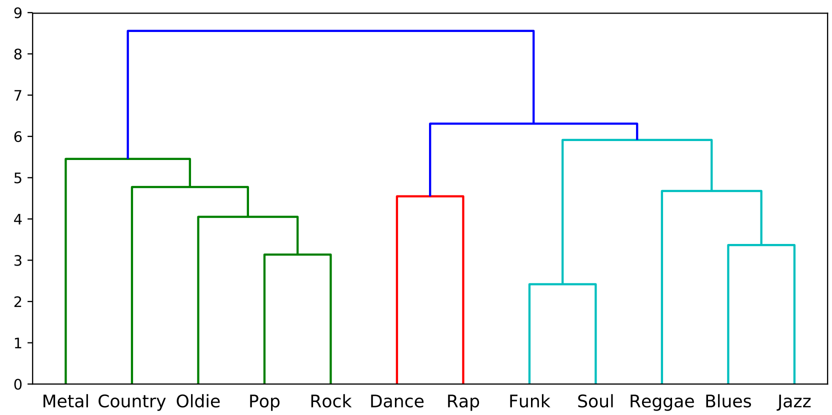
Figure 2.10: Spearman Correlations between *Joint Importance profiles* for different Musical Preferences. (*p < 0.05) & (**p < 0.01)

In order to better capture the intrinsic similarities in *Joint Importance profiles*, we performed agglomerative hierarchical clustering on the learned *Joint Importance profiles* for different genres. The clustering was performed using the same approach as explained in the section above. Figure 2.11a represents the dendrogram of hierarchical clustering on the *Joint Importance profiles*.

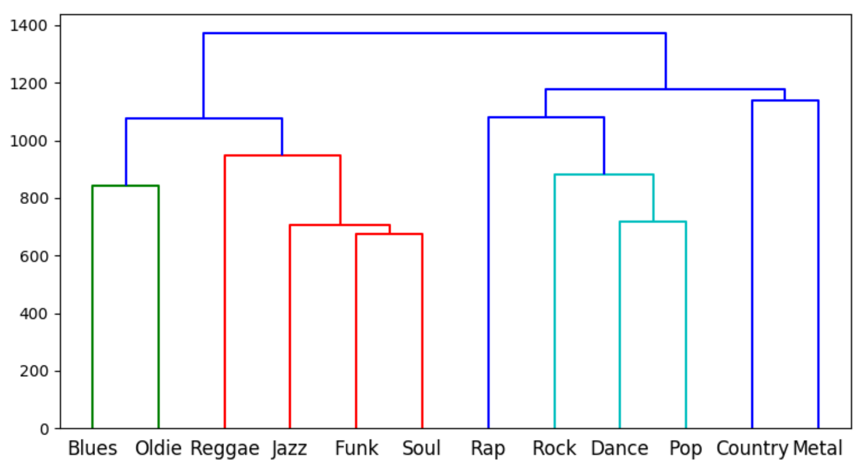
We observe from Figure 2.11a, the similar genres such as Funk & Soul, and Blues & Jazz, cluster together early in the hierarchy tree. This clustering pattern reveals similar genres being merged in the beginning suggesting similar overall movement patterns for those genres.

2.4.5.2 Similarity between Self-reported music preferences and Joint Importance profiles of predicted preferences

We further compare the similarities between the self-reported participants' Music preferences with learned *Joint Importance profiles*. Agglomerative clustering of self-reported ratings revealed an intrinsic structure fairly similar to those of movement patterns. Figure 2.11b represents the dendrogram of hierarchical clustering on the self-reported ratings.



(a)



(b)

Figure 2.11: Dendrogram of hierarchical clustering performed on the learned *Joint Importance profiles* (a) and self-reported ratings for STOMP music preferences (b)

Comparing Figure 2.11a and Figure 2.11b, we find two broad clusters. Specifically, we observe that similar clustering for the genres Funk and Soul, and Blues, Jazz and Reggae in one of the clusters. In the other cluster, the genres Rock, Pop, Metal and Country exhibited high similarity. The one exception was that of the genres Dance and Rap showing high similarity in *Joint Importance profiles* but father away in terms of user preference.

2.5 Discussion

Music experiences are highly embodied, making it necessary to consider individual embodied responses to music in developing more advanced personalized user experiences. The current study is among the first to the authors' knowledge to use participants' free dance movements to predict personality traits, music preferences, and both the Empathizing and Systemizing Quotients (EQ/SQ). The results demonstrate that the classification model was able to classify gender with near perfect accuracy and the regression model was able to predict personality, empathy, systemizing, and music preferences with high accuracy. The results of the current study provide broad support for the influence of individual differences over embodied responses to heard music, corroborating ideas of embodied music cognition and the role of bodily movement in musical experiences. Furthermore, these findings support the hypothesis that dance movement expresses socially relevant information and may therefore represents an evolved adaptation. In addition, we devise a novel measure to evaluate the importance of specific joints and their relative movement in characterising the individual traits. Furthermore, these results are the first to demonstrate the generalization of an analytical architecture across two data sets collected using two differing sets of musical stimuli. Although neither data set used in our analysis is large enough to make a commercial or industrial grade model, we were able to utilize two data sets of modest size to expand theoretical understanding and provide support for the robustness of our computational model.

As hypothesized, we were able to predict gender with near perfect accuracy of 96.53% and 98.76% for Dataset-1 and Dataset-2 respectively. As Troje et al. [192] was able to achieve high accuracy of gender minimal information, such as a single eigen-postures or as few as four principle components, it is not surprising that we were able to achieve such high accuracy. Identifying gender with such high accuracy from music-induced movement suggests that there are gender specific movements to music. These gender differences may be the result of socio-cultural influences as well as differences in natural movement patterns influenced by gender-specific joint flexibility.

Besides the higher accuracy achieved by the classification model for gender classification, the bayesian regression model for prediction of Empathizing and Systemizing Quotient demonstrated the R^2 score of 77.1%, and 86.7% respectively using the position data. Whereas, results with velocity data were far inferior with R^2 score of 50.3%, and 55.2% for EQ and SQ respectively. The proposed novel measure to evaluate the importance of specific joints and their relative movement in characterizing individual traits demonstrated that the limbs of the body to have more significance in predicting individual traits than the core body. The 'Finger,' 'Elbow,' and 'Knees' were found to have the greater impact on *Joint Importance*, while the 'Root,' 'Neck,' and 'Torso' had an insignificant impact. This is consistent with the fact that gestures are necessary for communication ([76]). We also deduce that the joints 'Finger, Hip, and Knee' are more crucial in determining EQ than SQ, while the 'Elbow' is significantly more important for predicting SQ than EQ.

Our Bayesian regression model for prediction of personality demonstrated an average R^2 score of 76.3% and 89.0% across all traits for Dataset-1 and Dataset-2 respectively using the position data. Previous studies ([121]) have shown Extraversion to be positively related to global movement as opposed to Neuroticism which relates negatively to it. As hypothesized, given that joint co-variance is related to local movement, Neuroticism was found to have slightly better classification accuracy than Extraversion. The current findings are especially interesting in light of the findings of Carlson et al. [35], who used covariance matrices derived from motion capture data in an analysis in which individual dancers were correctly classified using SVM classification with an accuracy of over 80 percent, suggesting that covariance between markers encodes individual differences both on the level of individual identity and on group-level features such as personality.

The *Joint Importance profile* learned for personality traits using the position data illustrated several similarities. Extraversion and Conscientiousness have some closeness in terms of *Joint Importance profiles* across data sets, indicating the importance of specific joints. 'Head, Shoulder, and Knee' play a crucial role in characterizing Conscientiousness, and these joints were consistently more important across data sets. For Extraversion, the 'Head,' 'Hips,' 'Shoulder,' 'Elbow,' and 'Knee' are consistently more important across the data sets. Luck et al. [121] discovered a link between Extraversion and head movement speed, confirming the current theory that the head is particularly important in determining Extraversion. Carlson et al. [31] found that in relation to Extraversion, the core body was more significant in responsiveness to musical tempo than Conscientiousness, which is partly confirmed by the marginally higher importance of the finger and wrist markers in our analysis but partly refuted by the importance of the shoulder marker in Extraversion. Overall, the disparity in results may be due to the current study's use of position data rather than velocity or acceleration data. As a result, the specific markers relevant to predicting individual traits sometimes contradict and largely corroborate previous research in some cases.

Agglomerative clustering and dendrogram plots of the *Joint Importance profiles* for personality provided evidence for the ability of our regression model to generalize across data sets by demonstrating that the weights learned by different models are similar. We have further shown the generalization by visualizing the learned profiles using Multi-Dimensional Scaling (MDS) based on Euclidean distance. The five mini clusters formed for each personality trait from the different data sets provides further corroboration that the weights learned by different model for two data sets are similar. As hypothesized, we were able to successfully regress the personality values for the second data set without tweaking any hyper parameters, which further demonstrates the robustness of our model. We were able to achieve higher accuracy for the second data set, which can be attributed to that set's larger sample size. While we did not use a single model to train and evaluate the different data sets, due to the dissimilarities of genres and participant demography, our model nevertheless showed similarities between the two data sets.

Related to prediction of STOMP-based Music Preferences, our regression model performed considerably well, demonstrating an average R^2 score of 77.5% with the score varying from 71.9% for Funk to 85.3% for Pop at the higher side for Music Preferences. To a great extent, the *Joint Importance profile* learned from the model using the position data and participants' self-reported ratings of music preference show similar clustering pattern. From Figs. 2.11a and 2.11b it is clearly visible that music preferences like Funk and Soul, and Blues, Jazz and Reggae are clustered similarly in both dendrograms in one of the higher-level clusters. In the another higher-level cluster, there were also similar clustering between Rock, Pop, Metal and Country, as well as between Dance and Rap. This is a notable finding, as self-reported music preferences are made consciously using verbal representation, which Leman [111] has argued are problematically distant from direct (embodied) engagement with music. That we could find similarities in clustering between reported preferences and movement patterns suggests that, despite being at a far remove from conscious, verbal descriptions, embodied responses to specific, heard music can indeed be seen to encode music preference at the genre level.

Another observation related to this point from Figure 2.11a is that Oldie and Country both belong to the "Upbeat and Conventional" factor as suggested by Rentfrow et al. [160], and clustering via *Joint Importance* represents a closer relationship for these compared to self-reported ratings; that is, in this case, movement patterns were more closely associated with Rentfrow and Gosling's model than were verbal self-reports. This may be due to similarity of movement in response to acoustic similarities between these genres. This result may also reflect differences in participants' perceptions of genre compared to industry-standard labeling, which have been shown to be ambiguous and inconsistent, as well as subject to cultural differences in the distinction between more granular categories [145]. Overall, these results indicate a relationship between embodied responses to music to specific stimuli and more general music preferences that is detectable despite the many other factors that may have influenced dancers' movements, including specific musical features. Taken together with our results regarding personality, this analysis demonstrates that abstract psychological and psychosocial concepts such as personality and preference are indeed evidenced by concrete features of complex dance movement, corroborating the idea of dance as a useful paradigm in exploring how socially relevant information is encoded in human movement.

The achieved prediction accuracy using the position data was about twice as good compared to the velocity data. The results of the current analysis show covariance to be a useful feature extraction mechanism in predicting individual differences. The current findings are especially interesting in light of the findings of Carlson et al. [35], who used covariance matrices derived from motion capture data in an analysis in which individual dancers were correctly classified using SVM classification with an accuracy of over 80 percent, suggesting that covariance between markers encodes individual differences both on the level of individual identity and on group-level features such as personality and music preference. This study represents an early step to make this approach applicable to personalized gesture-based retrieval systems, it can be extended to monocular video captured by accessible devices such as a mobile phone camera. This would then allow future recommendation systems to take embodied processes into account, resulting in better and more responsive personalized experiences.

Several limitations of the current study should be noted. First, the majority of participants were from European or North American countries, and all eight music stimuli were of Western origin, limiting the degree to which results can be generalized cross-culturally. Secondly, There may exist potential bias due to gender imbalance. Future work could include separate analysis performed within gender categories.

And lastly, participants' preferences for heard stimuli were not included in our model. This would be an important feature to focus on in future work, as preference and enjoyment are highly relevant to personalized MIR.

Future research could help clarify the relationship between the current results and those movement features which are perceptually relevant to judgements of gender, empathy, systemizing, personality and music preference. While analysis showed interpretable relationships within and between sets of *Joint Importance profiles*, further research is needed to determine whether, how, and in what contexts this encoded information may be decoded by human observers. Such research may have particular relevance to our understanding of social perception of human movement, and therefore to the understanding and treatment of disorders involving social deficits, such as autism.

Further extension of this work could help to make music recommendation systems more robust. Previous work has considered the relationship between personality and music preference [34, 160], while Greenberg et al. [78] explored the relationship between music preference and empathizing-systematizing theory, suggesting even that music may play a role in increasing empathy in people with empathy-related disorders, such as ASD. This study represents an early step towards multimodal MIR and to make this approach applicable to personalized gesture-based retrieval systems, it can be extended to monocular video captured by accessible devices such as a mobile phone camera. This would then allow future recommendation systems to take embodied processes into account, resulting in better and more responsive personalized experiences.

2.6 Key Findings and Contribution

- The conducted study is among the first to the authors' knowledge to use participants' free dance movements to predict personality traits, music preferences, and both the Empathizing and Systemizing Quotients (EQ/SQ).

- We proposed a Machine-Learning model that predicts individual traits from participants' free dance movements. We further demonstrate the robustness of the proposed model by using another dataset and further providing conclusive evidence about the learned models' generalizability.

- We introduced a novel method to carry out investigations to explore the relative importance of joints in defining human traits. We explored the relationship between music induced movement and musical experience by successfully exploring the unseen space for EQ/SQ, personality and musical preferences from participants' free music-induced movements.

- Co-variance between joint velocities has previously been used to identify an individual from their free dance movements. However, we identify the use of position data instead of velocity data in the kernelized space to perform better for the regression tasks.

Chapter 3

Decoding Individual Differences via passive engagement

In this chapter we sought to investigate the influence of individual differences in the context of passive engagement. We explore and find out how individual differences, specifically personality, are associated with the kind of emotional experiences one seeks from lyrics.

3.1 Background and Motivation

Information retrieval and recommendation, whether in the form of news, music, goods, or images, is crucial in e-commerce and on-demand content streaming applications. Indeed, due to the Covid pandemic, the music streaming platform Spotify has witnessed a staggering rise of six million subscribers in the first quarter of the year, and has now crossed a total of 144 million paid subscribers and counting [1]. This has led to increased need and relevancy for MIR systems. The content-based aspect of MIR systems, on the other hand, has focused primarily on acoustic content [59, 147], social tags [30], and emotions and moods derived from them. Lyrics have been largely neglected despite the crucial role they play in especially eliciting emotions [77], a key factor in musical reward [126], as well as reflecting user traits and tendencies [156], which are linked to musical preferences [133]. Despite a handful of studies showing that music emotion classifiers based on features extracted from lyrics perform better than audio [87, 213], the significance of lyrics in music emotion recognition remains underappreciated.

Analyzing lyrics and its emotional connotations using advanced Natural Language Processing (NLP) techniques would make for a natural choice. However, NLP has been used in MIR for topic modelling [105], song structure identification via lyrics [62], and mood classification [87]. In the context of Music emotion recognition [123, 213], typically traditional NLP approaches have been used, which are restricted to word-level representations and embeddings, as opposed to more modern NLP techniques that are based on context and long-term dependencies such as transformers [54, 215]. Lyrics can be treated as narratives rather than isolated words or sentences, using transformers to mine affective connotations is indeed a natural option.

Lyrics are crucial in attracting listeners to songs and induce various emotions such as joy, sorrow, nostalgia, amongst others [218] which may lead to shifts in mood states. Musicology has remained a medium of expressing and stimulating emotions, but inferring these emotions from a musical fragment is complex. Favorite song lyrics have been liked to personality traits as listeners use them to satisfy their own psychological needs and experience specific emotions [156]. This however has never been investigated as it occurs naturally on online music streaming platforms. Emotions are also a core ingredient to both music and personality individually and in combined circumstances. The individuals tend to satisfy their reflections of personality by choosing specific emotions [200].

3.2 Objectives and Hypothesis

The further Chapter has been divided into two sections. In Section 3.3, we propose a deep neural network based method for identifying emotional connotations of music based on lyrics. Accordingly the objective is to investigate the influence of individual differences in the context of Passive engagement.

We expect, based on previous research showing that individual differences are associated with lyrics from their favourite music as listeners use them to satisfy their individual psychological needs, to find relation between personality traits and the emotions conveyed through lyrics of their preferred music. The study-2 in Section 3.4 aims to examine individual tendencies for listening to specific emotion songs. Thus we hypothesize that, individual differences should be associated with the emotional experiences one seeks as it occurs naturally on online music streaming platforms. The study will help in understanding the relationship between personality traits and their choice of emotions conveyed through lyrics.

3.3 Study-1

Sentiment analysis, or analysing affective connotations from text, has been actively attempted in brief contexts such as reviews [16, 148], tweets [3, 39], news articles [157] amongst others with limited application to lyrics. In recent times, sentiment analysis has also been extended to several emotion classes in the field of longer narratives [117] but the dataset is limited to passages with an average length of 100 words. Sentiment analysis has come a long way from its inception based on surveys and public opinions [106] to use of linguistic features like character n-grams [83], bag-of-words [16] and lexicons like SentiWordNet [141] to state-of-the-art that employ context-based approaches [54, 153] for capturing the polarity of a text. The task of sentiment analysis has been approached using traditional Machine Learning techniques like Naive Bayes [149], Maximum Entropy Classification [149], SVM [149, 83] and unsupervised learning methods [63, 195]. Further, deep learning techniques like RNN [39, 150], CNN [39], and transformers [54, 89] have also been applied in doing sentiment analysis tasks. These recent techniques have shown to outperform conventional machine-learning approaches by a significant margin [104].

Music psychology follows two widely used taxonomies for music emotion classification: categorical and dimensional. The categorical approach has different categories as representations like joy, bitterness, peace, and tension. Since it has limited categories and language usage is subject to interpretation, this approach does not capture the richness of human emotions. Russell proposed a two-dimensional Valence-Arousal circumplex model [168] of affect where an emotion is a point in a two-dimensional continuous space. Valence describes pleasantness and Arousal represents energy level. The precise location in the continuous valence-arousal space can be used to determine discrete emotional categories such as happy and sad, among others. For instance, the quadrant associated with Happiness is represented by positive arousal and positive valence, whereas the quadrant associated with Sadness is represented by negative arousal and negative valence. Some of the existing studies and datasets based upon lyrics are [49], which classifies songs into 5 predefined mood clusters. We identified the use of MoodyLyrics [29] and MER Dataset [124] which uses Russell classification for classification of songs in the V-A plane.

Music emotion classification using lyrics has been performed at a word-level based using traditional NLP methods. The word level embeddings were obtained via emotions-based lexicons [87, 88]. The lexicons not only have very limited vocabulary but also the word-level emotion values are aggregated without using any contextual information. In recent years the use of pre-trained models like GloVe word embeddings [152], ELMO contextualized word embeddings [153], transformers [54, 183] are fast gaining importance for large text corpus has shown impressive results in downstream several NLP tasks such as text classification, semantic textual similarity, and many others. Authors in [51, 2] perform emotion classification using lyrics by applying RNN model on top of word-level embedding. The MoodyLyrics dataset [29] was used by [2] who report an impressive $\mathcal{F}_1$-score of 91.00%. Recurrent models, such as LSTMs, operate on the Markov principle, in which data from previous steps is processed via a sequence of computations to predict a future state. These sequential models which fail to take overall information while evaluating the context. Meanwhile, the transformer architecture eschews recurrence nature and introduces self-attention, which establishes longer dependency between each step with all other steps. Since we have direct access to all the other steps (self-attention) ensures negligible information loss. In this study, we employ Multi-task setup, using XLNet as the base architecture for classification of emotions and evaluate the performance of our model on several datasets that have been organized by emotional connotations solely based on lyrics. We demonstrate superior performance of our transformer-based approach compared to RNN-based approach [51, 2]. In addition, we propose a robust methodology for extracting lyrics for a song.

3.3.1 Methods

3.3.1.1 Datasets

MoodyLyrics [29]**:** This dataset comprises 2595 songs that are evenly distributed across the four quadrants of the Russell's Valence-Arousal (V-A) circumplex model [168] of affect where emotion is a point in a two-dimensional continuous space which has been reported to sufficiently capture musical emotions [61]. The authors assigned V-A values at the word level using a combination of existing lexicons such as ANEW, WordNet, and WordNet-Affect, accompanied by song-level averaging of these values. These were further validated by using subjective human judgment of the mood tags from AllMusic

Dataset [124]. Finally, the authors only held songs in each quadrant if their Valence and Arousal values were above specific thresholds, thereby rendering them to be highly representative of those categories.

MER Dataset [124]**:** This dataset contains 180 songs distributed uniformly among the four emotion quadrants of the 2-D Russell's circumplex model. Several annotators assigned the V-A values for each song solely based on the lyrics displayed without the audio. The Valence and Arousal for each song were computed as the average of their subjective ratings. Furthermore, this dataset has been stated to have high internal consistency, making it highly perceptually important.

AllMusic Dataset [124]**:** This dataset comprises 771 songs with close to uniform distribution of songs in each of the four quadrants. Expert-annotated mood tags from the AllMusic website were projected onto the V-A space using the ANEW lexicon for each song. Following that, three annotators validated this quadrant classification based on their initial agreement on the mood tags in the first place.

3.3.1.2 Lyrics Extraction

The first step for the task at hand, is to extract lyrics as the datasets do not include lyrics due to copyright issues. One approach for mining the lyrics is to build a crawler for each of the websites included in the datasets. However, some of those URLs were broken. As a result, the role of extracting lyrics becomes extremely important. The URLs from different lyric websites are provided in each of the datasets. All the existing APIs, including Genius require the exact Artist and Track name for extracting the lyrics. However, if the artist or track names are misspelled in the dataset, the API fails to extract the lyrics. Hence, in order to address this concern, we provide a robust approach for extracting lyrics using the Genius website. We handled this issue by introducing a web crawler to obtain the Genius website URL for the lyrics of the song instead of hard-coding the artist and track name in Genius API. Our approach was able to handle the issues like spelling mistakes and missing terms for the lyrics of the song. Using the web crawler, we were able to considerably improve the number of songs extracted from 60% - 80% for the different datasets to $\sim$99% for each dataset.

3.3.1.3 Proposed Architecture

We propose a deep neural network architecture that, given the lyrics, outputs the classification of Emotion Quadrants, in addition to Valence and Arousal Hemispheres. The entire network is trained jointly on all these tasks using weight-sharing, an instance of multi-task learning. Multi-task learning acts as a regularizer by introducing inductive bias which favours hypotheses that explain all of the tasks. It boosts the model's ability to accommodate random noise during training and eliminates the risk of overfitting while achieving faster convergence [219]. Figure 3.1 displays the architecture of our proposed method

We use XLNet [215] as the base network, which is a large bidirectional transformer that uses improved training methodology, larger data and more computational power. XLNet improves upon

BERT [54] by using the Transformer XL [48] as its base architecture. The added recurrence to the transformer enables the network to have a deeper understanding of contextual information.

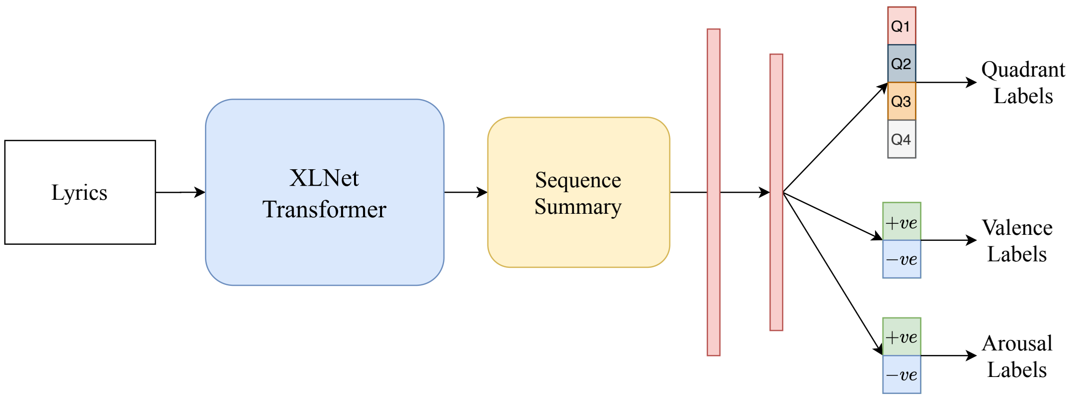


Figure 3.1: Overview of our method

The XLNet transformer Model generates raw hidden states, which are then passed on to $Sequence$-$Summary$ block, which computes a single vector summary of a sequence of hidden states, followed by one more hidden Fully-Connected (FC) layer which encodes the information into an 8-dimensional vector. Finally, this layer branches into three complementary tasks via a single FC layer on top for classification of Quadrant, Valence, and Arousal separately. As we feed input data, the entire pre-trained XLNet model and the additional untrained classification layers are trained for all three tasks. We use the following loss function to train our network.

$$L = (\lambda_1 * L_Q) + (\lambda_2 * L_V) + (\lambda_3 * L_A) \tag{3.1}$$

where L_Q,L_V, and L_A represents the classification loss on Quadrants, Valence, and Arousal, respectively.

3.3.1.4 Implementation Details

We use the AdamW optimizer [119] with an initial learning rate of $2e^{-5}$ and a dropout regularization with a 0.1 discard probability for the layers. We use Cross-Entropy Loss for calculating loss.The batch size was set at eight. We also limit the length of the lyrics to 1024 words. More than 99 percent of the songs' lyrics had less than 1024 words. We leverage the rich information of pre-trained (XLNet-base-cased) model as they have been trained on large corpora. As the pre-trained model layers already encode a rich amount of information about language, training the classifier is relatively inexpensive [183]. We also train our network on single-task classification and compare the results in a later section as part of our ablation study.

3.3.2 Experiments and Results

3.3.2.1 Evaluation Measures

For evaluating the effectiveness of our proposed model, we use the standard recall, precision, and F_1 measures. Recall is defined as the ratio of correct assignments by the system divided by the total number of correct assignments. Precision is the ratio of correct assignments by the system divided by the total number of the system's assignments. The $\mathcal{F}_1$-score can be interpreted as a weighted average of the precision and recall. We provide results for both macro-averaged F_1 and micro-averaged F_1. The micro-average F_1 is also the classifier's overall accuracy. We use Macro-averaged F_1($\mathcal{F}_1$-score) [214] as given in Equation 3.2. The scores are first computed for the binary decisions for each individual category and then are averaged over categories.

$$F1_x = 2\frac{P_x R_x}{P_x + R_x}; \qquad \mathcal{F}_1 = \frac{1}{n}\sum_x F1_x \tag{3.2}$$

where $F1_x$, P_x, R_x denote F1-score, precision and recall with respect to class x. This metric is significantly more robust towards the error type distribution as compared to the other variants of the Macro-averaged F_1 [142].

3.3.2.2 Results

To compare our performance on different datasets, we use the multi-task setup. We perform the data splits for respective datasets, as mentioned in respective studies, to conduct a fair evaluation of our method. All the results reported hereon are the average of multiple data splits. Tables 3.1 and 3.2 compares the results of our approach on MoodyLyrics and MER dataset respectively. These findings show that our method outperforms previous studies that attempted the music emotion recognition task on the same datasets. For the MoodyLyrics dataset, we see a significant improvement in accuracy and F1-score of about 4%. The quadrants are more representative in MoodyLyrics dataset as the authors used certain thresholds for Valence and Arousal to have high confidence for the categorization process which justifies the better performance.

Approach	**Accuracy**	**Precision**	**Recall**	**$\mathcal{F}_1$-score**
Naive Bayes [2]	83.00%	87.00%	81.00%	82.00%
BiLSTM + Glove [2]	91.00%	92.00%	90.00%	91.00%
Our Method	94.78%	94.77%	94.75%	94.77%

Table 3.1: Results of classification by Quadrants on MoodyLyrics dataset.

We also compare the performance of our approach by validating on the AllMusic dataset [124]. We follow the same procedure of training on the MER dataset and evaluating on the AllMusic dataset as mentioned by the authors. Table 3.3 shows comparable results with other methods. We get an improved $\mathcal{F}_1$-score of 75.40% compared to their reported 73.60% on single-task Quadrant classification

Classification	Approach	Accuracy	Precision	Recall	$\mathcal{F}_1$-score
Quadrant	CBF + POS tags, Structural and Semantic features [124]	-	-	-	80.10%
Quadrant	**Our Method**	88.89%	90.83%	88.75%	88.60%
Valence	CBF + POS tags, Structural and Semantic features [124]	-	-	-	90.00%
Valence	**Our Method**	94.44%	92.86%	95.83%	93.98%
Arousal	CBF + POS tags, Structural and Semantic features [124]	-	-	-	88.30%
Arousal	**Our Method**	88.89%	90.00%	90.00%	88.89%

Table 3.2: Results of classification on MER dataset.

in addition to improved Accuracy of 76.31% when compared to the reported Accuracy of 74.25%, albeit on a subset of the AllMusic dataset, in[29]. Our Multi-task method demonstrated comparable $\mathcal{F}_1$-score and accuracy of 72.70% and 73.95% when compared to our single-task Quadrant classification.

Approach	Accuracy	Precision	Recall	$\mathcal{F}_1$-score
CBF + POS tags, Structural and Semantic features [124]	-	-	-	73.60%
V-A from ANEW, WordNet and WordNet-Affect [29]	74.25%	-	-	-
Our Baseline	76.31%	75.86%	75.21%	75.40%

Table 3.3: Results of classification by Quadrants on AllMusic dataset. Trained on MER and validated on AllMusic dataset using our baseline network.

3.3.2.3 Ablation Study

We conduct extensive analysis with different architecture types and sequence lengths on the Moody-Lyrics dataset owing to its large size and quadrant representativeness. In the initial set of experiments, we aimed to find the best model where we compared our baseline model with BERT transformer with same sequence length of 512, and Table 3.4 shows that XLNet transformer outperforms BERT to show a superior performance of $\mathcal{F}_1$-score by around 1.3%. Lyrics of around 91% of the songs were less than 512. Truncating inputs by the shorter sequence length decreased the performance since the model cannot capture the global information ranging the complete lyrics. We also compare the performance of our baseline model with our multi-task setup. Table 3.4 shows that we perform similar to our baseline method, but we saw a huge improvement in training speed as the latter converge faster. This necessitates training new tasks from scratch each time, which is highly inefficient.

Approach	Classification	Model	Max Length	Accuracy	$\mathcal{F}_1$-score
Our Method	Quadrant	XLNet	1024	94.78%	94.77%
	Valence			95.73%	95.67%
	Arousal			94.38%	94.23%
Our Baseline	Quadrant	XLNet	1024	95.68%	95.60%
	Valence	XLNet	1024	96.51%	96.46%
	Arousal	XLNet	1024	94.38%	94.35%
Our Baseline	Quadrant	XLNet	512	94.96%	94.90%
Our Baseline	Quadrant	BERT	512	93.80%	93.62%

Table 3.4: Ablation Study on MoodyLyrics

3.3.3 Discussion & Future Works

We have demonstrated the robustness of our novel transformer-based approach for music emotion recognition using lyrics on multiple datasets when compared to hitherto used approaches. Our multi-task setup helps in faster convergence and reduces model overfitting, however, the single-task setup performs marginally better albeit at the expense of computational resources. We used a robust methodology to enhance web-crawlers' accuracy for extracting lyrics. This study has important implications in improving applications involved in playlist generation of music based on emotions in addition to improving music recommendation systems. This study can help revolutionize the user recommendation systems as it can change the way an individual selects and listens to their favored musical tracks. Also, hybrid music recommendation systems, which utilize predominantly acoustic content-based and collaborative filtering approaches can further benefit from incorporating emotional connotations of lyrics for retrieval. This approach can be extended in future to multilingual lyrics.

3.4 Study-2

This study aims to understand the relationship between personality traits and the emotions, as represented in the Valence-Arousal (VA) space, conveyed through lyrics of their preferred music.

3.4.1 Methods

We use the dataset by Melchiorre et al.[133] which is a subset of MyPersonality dataset[178]. This dataset contains music listening histories of 1470 Last.fm users and their respective personality scores. Users took psychometric questionnaires to assess their personality according to the five-factor model. A user's listening history comprises of listening events, where each event is defined by a username, artist name, track name, amongst others. The listening history was further complemented by content-based information of the music tracks acquired from Spotify.

Majority of the listening history had songs with lyrics. Further, we only considered those users (N=1120) for whom lyrics for more than 70% of the songs could be extracted. To identify emotion from lyrics, we used our proposed approach as described in 3.3.1.3 to calculate the four quadrant scores in the continuous Valence-Arousal space associated with joy, anger, sadness, and tenderness. We aggregate the *emotion quadrant scores* for every user by averaging the quadrant score for their complete listening history.

Subsequently, Spearman correlations were performed between *emotion quadrant scores* and personality traits. There is an increase in the likelihood of obtaining a significant result by pure chance while conducting multiple analyses on the same dependent variable. This is also referred to as Type I error. To account for Type I error, we performed Bonferroni correction ensuring that the observed differences are not due to chance.

3.4.2 Results

Results for Spearman Correlation between *emotion quadrant scores* and personality traits can be found in Table 3.5.

	Joy	**Angry**	**Sad**	**Relax**
Openness	0.090**	-0.173***	0.085**	0.185***
Conscientiousness	0.054	-0.035	-0.056	0.039
Extraversion	0.074*	-0.084**	0.037	0.045
Agreeableness	0.083**	-0.120***	0.076*	0.010***
Neuroticism	-0.043	0.068*	0.009	-0.092**

Table 3.5: Spearman Correlation between *emotion quadrant scores* and personality traits. (*$p < 0.05$, **$p < 0.01$, & ***$p < 0.001$)

Tenderness correlated positively with traits Openness (r=.18), Agreeableness(r=0.1) and negatively with Neuroticism (r=-0.1) (all $p < .001$). Anger showed significant correlation with Openness(r=-0.17) and Agreeableness(r=-0.12) (all $p < .001$). Joy correlated with Openness (r=.089, $p<.002$). Without Bonferroni correction, sadness correlated with Openness (r=0.09, p=0.005) and Agreeableness (r=0.08, p=0.011), anger with Extraversion(r=-0.09) and Neuroticism(r=0.07), and joy with Extraversion(r=0.074) and Agreeableness(r=0.08) (all $.0025<p<0.05$). No significant correlations were found between Conscientiousness and EQS.

3.4.3 Discussion & Future Works

Openness and Agreeableness demonstrate similar patterns with a predilection for positively valenced lyrics in addition to those representing sadness. This is in line with such individuals open to seeking various emotional experiences and enjoyment of sad music [200]. Agreeableness, defined by cooperative and empathetic predisposition appears to be reflected in their liking for joyful, tender, and sadness associated with lyrics and dislike for negative words represented by anger. Neuroticism is associated with preference for lyrics depicting positive valence and low arousal thereby acting as a potential balancing mechanism for their already aroused negative states.

For the field of psychology of music, this study help us gain insight into the relationship between individual differences and preferences for certain kinds of emotionally-laden lyrics. We can say that people with certain dominant personality trait prefer particular emotions songs over others. So, emotions from lyrics can be used in music recommendation by recommending songs similar in emotions to songs in a session.

3.5 General Discussion

Music Emotion Recognition has gained prominence over the recent years in the field of MIR, albeit relying on acoustic features, social tags, and other metadata to identify and classify music emotions. The role of lyrics in music emotion recognition remains under-appreciated in spite of several studies reporting superior performance of music emotion classifiers based on features extracted from lyrics. The significant technological developments in the field of NLP have led to novel improved ways of analyzing text and extracting relevant features such as topics, sentiments, and emotions. With modern NLP techniques, modeling of affective connotations has moved form a word-level to context-level using transformers. This is the first study to use a transformer based approach in identifying emotional connotations of lyrics. Our model beats all the existing state-of-the-art methods. The model is robust owing to its superior performance on multiple datasets. We used improved methodology to extract lyrics using a crawler.

Also, the novel results highlighting the associations between individual differences like personality traits and preferences for certain kinds of emotionally-laden lyrics, suggests such individuals may rely on linguistic cues in regulating their states via music. Hybrid music recommendation systems, which utilize predominantly acoustic content-based and collaborative filtering approaches can further benefit from incorporating emotional connotations of lyrics for retrieval.

3.6 Key Findings and Contribution

- Through this study, we sought to identify emotional connotations of music based on lyrics. Lyrics play a crucial role in eliciting emotions, despite that Music Emotion Recognition has been limited to the usage of acoustic content, social tags, and metadata.
- We provide an improved approach for extracting lyrics using the Genius website and an added web-crawler. Our approach is able to handle the issues like spelling mistakes and missing terms for the lyrics of the song. This helps in improving the performance for extraction of from mere 60%-80% to 99%.
- We propose a deep learning architecture to identify emotional connotations of music based on lyrics. We further compare our method with the existing deep learning and even traditional methods on relevant datasets and show state-of-the-art performance.
- For the first time we demonstrate the associations between the emotions conveyed through lyrics and individual differences as they occur ”in the wild” (i.e, online music streaming platforms).
- The results were mostly in line with previous research, in addition to revealing novel findings.
- Our findings validate our theory that various types of emotions conveyed by songs have a unique relationship with individual characteristics. This study contributes to a better understanding of the relationship between personality traits (OCEAN) and the emotional experiences one naturally seeks on online music streaming platforms.

Chapter 4

Conclusions

The work presented in this thesis sought to investigate the influence of individual differences in active and passive musical engagement. In the first part of the thesis we explore how well individual differences can be decoded in the context of active engagement. We proposed a Machine-Learning model to achieve notably accurate results on classifying gender and predicting individual traits, specifically the Big Five personality traits. In addition we also achieve high prediction accuracy for the Empathy and Systemizing Quotients (EQ/SQ), representative of Cognitive Styles of thinking from participants' free dance movements. The results of this part of the thesis provide support for theories of embodied music cognition and the role of bodily movement in musical experiences by demonstrating the influence of gender, cognitive styles, personality, and music preferences on embodied responses to heard music.

The results of the first study provide broad support for the influence of individual differences over embodied responses to heard music, corroborating ideas of embodied music cognition and the role of bodily movement in musical experiences. In addition, these results provide support for the hypothesis that dance movement provides socially relevant information, and may therefore represent an evolved adaptation. We also demonstrate the robustness of the proposed model by using another dataset and further providing conclusive evidence about the learned models' generalizability. Additionally, we attempt at proposing a Joint Importance profiles, which is novel measure to evaluate the importance of specific joints and their relative movement in characterising human traits. The joints that were significant in predicting individual traits in most cases were in agreement with previous findings but in some cases contradicted them. Differences between our results and previous results could be related to a number of factors, such as cultural differences between dancers, but are also likely to relate to the use of different musical stimuli.

The implications of this study are socially an psychologically relevant. For instance, individuals who are more susceptible to stress and anxiety disorders are found to be associated with Neuroticism. Hence, predicting such a trait from music-induced movement is helpful in identifying underlying potential psychological distress and may aid in early identification and intervention. The need to analyse EQ/SQ scores from music-induced movements is worth highlighting in light of recent work suggesting the existence of motor signatures unique to ASD, detectable from whole body movements as well as data drawn from participants' interaction with tablets. In the field of MIR, this study represents an early step toward applying this approach to personalized gesture-based retrieval systems. With recent advancements in the field of 3D human pose prediction, which can predict the human body joint coordinates from a monocular video, it can be now be extended to monocular videos captured by hand-held devices like mobile camera. This would then encourage future recommendation systems to account for embodied processes, resulting in better and more personalized experiences.

In the later half, we investigate the influence of individual differences in the context of passive engagement. In order to effectively study emotions from the lyrics of songs our first task was to devise a method to identify emotional connotations of music based on lyrics. We proposed a deep learning architecture to identify emotional connotations of music based on lyrics. To further comment on the models' robustness we compared our method with the existing deep learning and even traditional methods on relevant datasets and show state-of-the-art performance. Further, in order to find out how individual differences are associated with the kind of emotional experiences one seeks from lyrics, we perform a study to understand the relationship between personality traits and the emotions conveyed through lyrics of their preferred music. This however has never been investigated as it occurs naturally on online music streaming platforms. The results revealed Openness and Agreeableness to demonstrate predilection for sad music, which corroborates previous literature. Additionally, the study revealed several novel findings, Such as Neuroticism being linked to a preference for positive valence and low arousal lyrics, potentially acting as a balancing mechanism for their already aroused negative states.

These novel results highlighting the associations between individual differences like personality traits and preferences for certain kinds of emotionally-laden lyrics, suggests such individuals may rely on linguistic cues in regulating their states via music. This study has important implications in improving applications involved in playlist generation of music based on emotions in addition to improving music recommendation systems. This study can help revolutionize the user recommendation systems as it can change the way an individual selects and listens to their favored musical tracks. Also, hybrid music recommendation systems, which utilize predominantly acoustic content-based and collaborative filtering approaches can further benefit from incorporating emotional connotations of lyrics for retrieval. This approach can be extended in future to multilingual lyrics.

To conclude, this thesis set out to investigate individual differences associated with musical behaviour. Overall, the results of this thesis corroborate ideas of embodied music cognition and the role of bodily movement in musical experience and preferences. Furthermore, the thesis highlights the role of individual differences in musical behaviour. Taken together, these differences and related musical preferences thereof manifest as both active music-induced movement and passive engagement on online music streaming platforms.

Appendix A

Supplementary Material

Chapter 2

Principal Component Regression

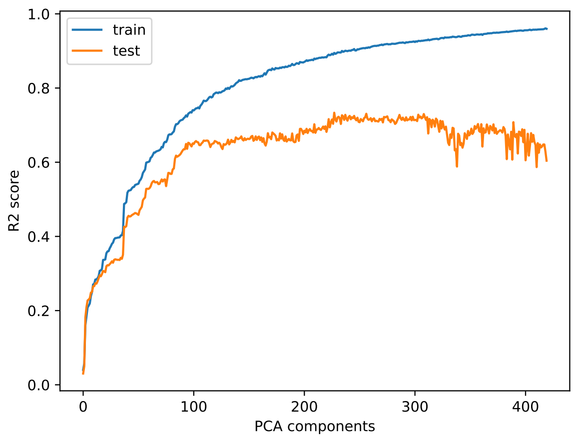
The Linear Regression model was used to predict the Empathizing Quotient (EQ) and Systemizing Quotient (SQ) values but we found that the model was highly overfitting. So, we took Principal Components of the features for this model. We calculated the RMSE and R^2 scores for both of the aforementioned tasks after taking the Principal Components. We repeated this experiment by varying the number of principal components.

Empathizing Quotient

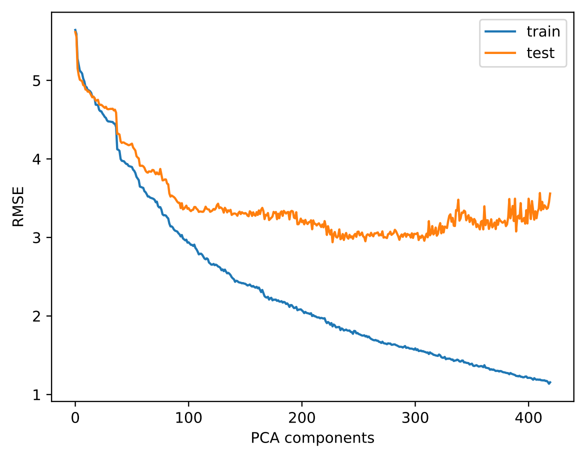
From Figs. A.1a and A.1b we can say that using position data for feature extraction, the model started to overfit after having more than around 240 principal components since the RMSE started increasing and R^2 score started decreasing for testing data. Similarly, from Figure A.1c and A.1d, we can say that using velocity data for feature extraction, the model started to overfit after having more than around 140 principal components.

Systemizing Quotient

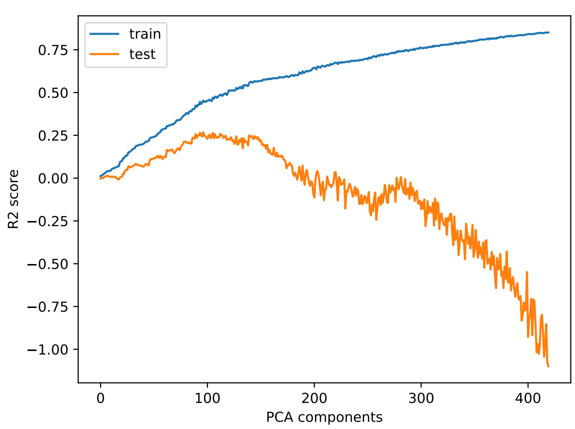
From Figs. A.2a and A.2b, we can say that using position data for feature extraction, the model started to overfit after having more than 260 principal components. Similarly from Figure A.2c & A.2d when we use velocity data for feature extraction, the model started to overfit after taking more than 170 principal components.



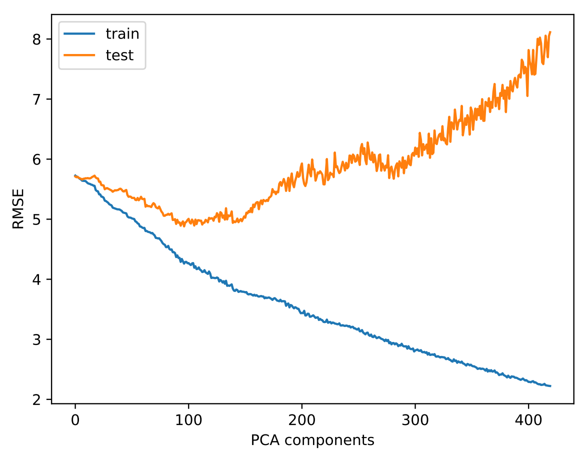
(a) R^2 score on Position Data



(b) RMSE on Position Data

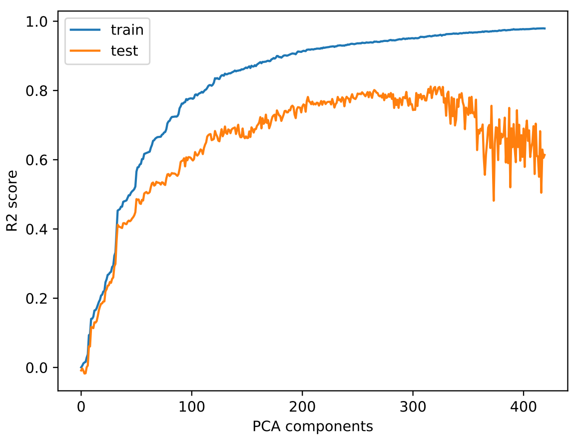


(c) R^2 score on Velocity Data

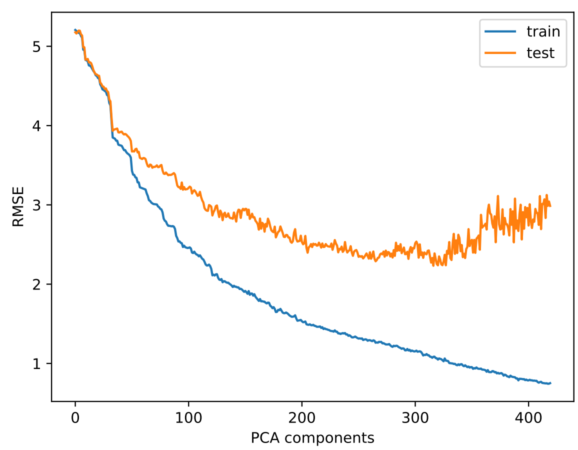


(d) RMSE on Velocity Data

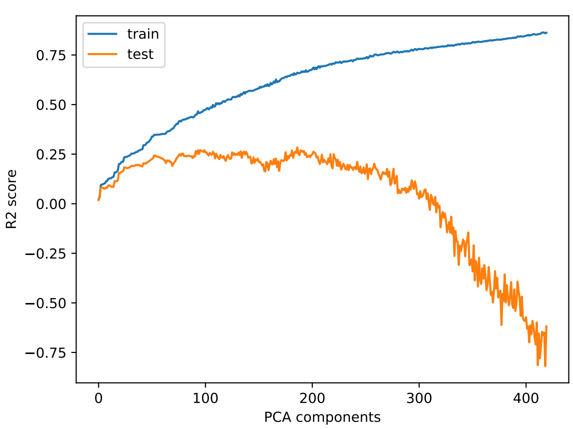
Figure A.1: Principal Component Regression results on EQ



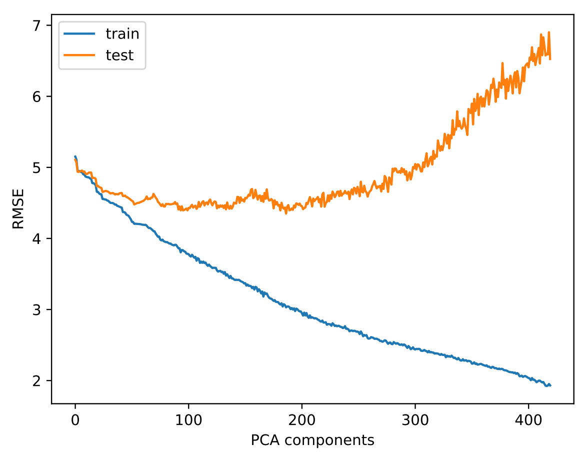
(a) R^2 score on Position Data



(b) RMSE on Position Data



(c) R^2 score on Velocity Data



(d) RMSE on Velocity Data

Figure A.2: Principal Component Regression results on SQ

Bayesian Regression

The model for the Bayesian Regression with the response sampled from a normal distribution is

$$y \sim N(\beta^T x, \sigma^2 I) \tag{A.1}$$

where β is the weight vector, x is the feature vector and σ is the standard deviation. Here β and σ are the model parameters. The goal is not to find the single best value of model parameters but to determine the posterior distribution for the model parameters. The formulation of model parameters as distributions encapsulates the Bayesian worldview, i.e, we use non-informative priors for the parameters such as a normal distribution to start with, followed by some prior, lastly, the likelihood washes out the prior as the amount of data points increases making our model becomes less wrong.

The posterior probability of the model parameters can be defined as

$$p(\beta|D) \propto p(D|\beta)p(\beta) \tag{A.2}$$

$$\beta \sim N(0, \sigma_\beta^2 I_d) \tag{A.3}$$

where $p(\beta)$ is the initial probability distribution, also known as prior distribution and $p(D|\beta)$ is known as the likelihood function. Using these approaches, we attempted two tasks 1. EQ and SQ Prediction 2. Personality Prediction.

Due to its robustness, it is evident that the model did not overfit on the dataset as the R^2 score kept on increasing with the number of principal components. The R^2 scores for training and testing set are 0.92 and 0.86 when all the features are considered. We also repeated this experiment by varying the number of principal components incrementally to test its robustness.

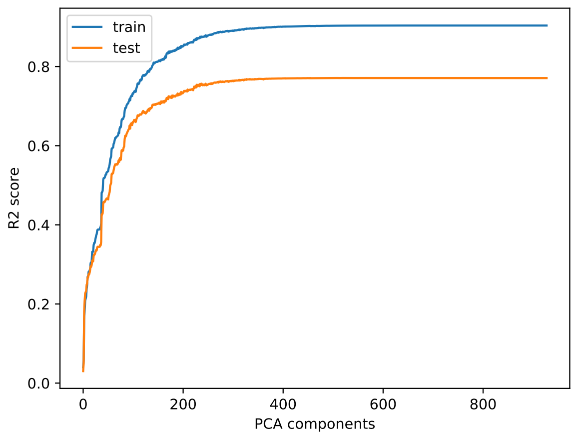
Empathizing Quotient

From Figs. A.3a & A.3b, we can say that using position data, the maximum R^2 score achieved is 0.76 and minimum RMSE is 2.73. From Figs. A.3c & A.3d, we can say that using velocity data, the maximum R^2 score is 0.42 and minimum RMSE is 4.35. The RMSE and R^2 scores become somewhat saturated at some point but still gets better marginally with the increase in principal components.

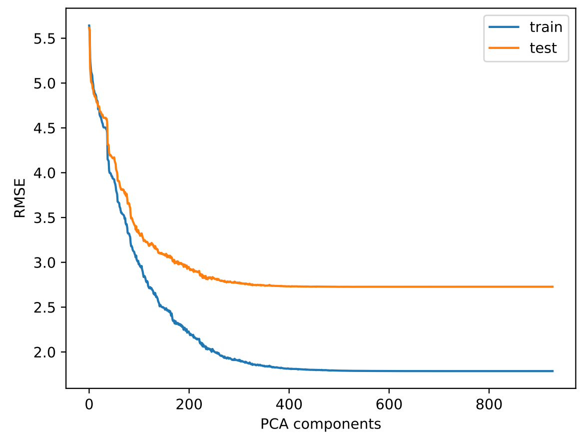
Systemizing Quotient

From Figs. A.4a & A.4b, we can say that using position data, the maximum R^2 score is 0.82 and minimum RMSE is 2.18. From Figs. A.4c & A.4d, we can say that using velocity data gained maximum R^2 score of 0.44 and minimum RMSE of 3.82. In the paper, we reported the RMSE and R^2 scores of Bayesian Regression without taking any principal components.

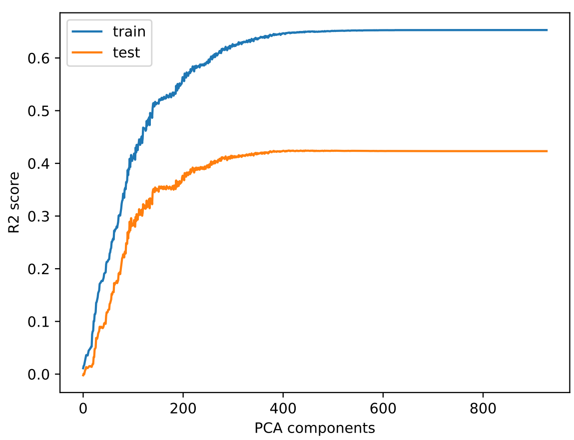
In EQ, using position data, the RMSE and R^2 is 2.72 and 0.77. Similarly, using velocity data, the RMSE and R^2 is 4.43 and 0.42. In SQ, using position data, the RMSE and R^2 is 2.16 and 0.86. Similarly, using velocity data, the RMSE and R^2 is 3.83 and 0.46. Since, the results were close but better than that of after performing the PCA, we present all the results in our paper for Bayesian Regression without performing PCA.



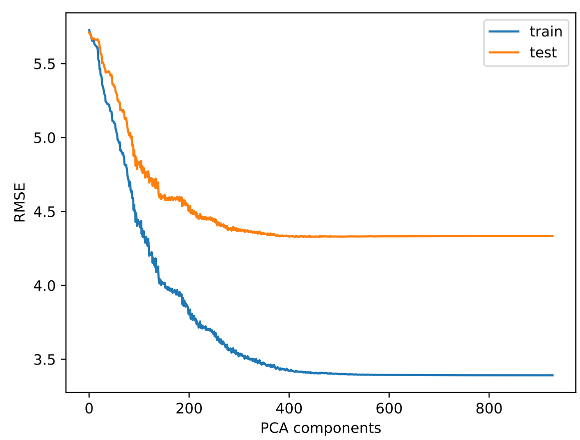
(a) R^2 score on Position Data



(b) RMSE on Position Data

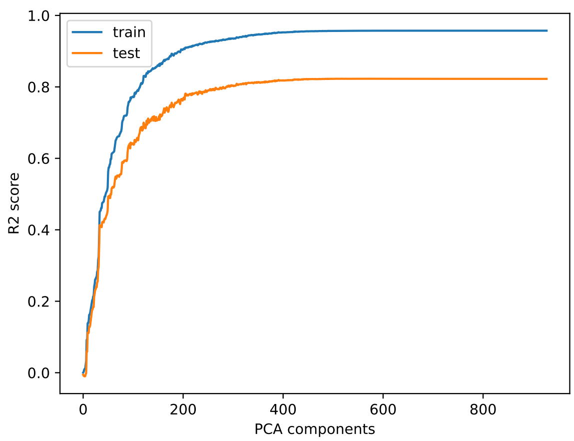


(c) R^2 score on Velocity Data

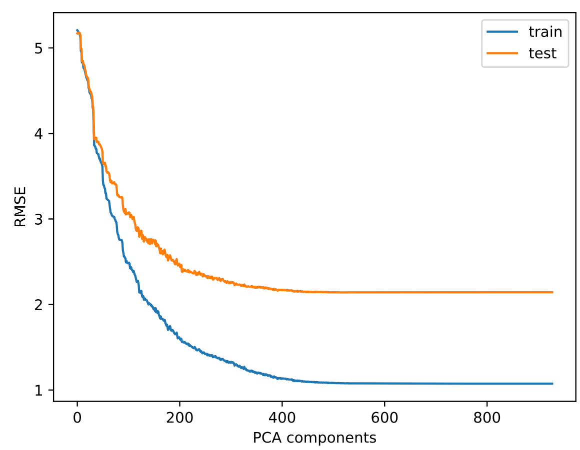


(d) RMSE on Velocity Data

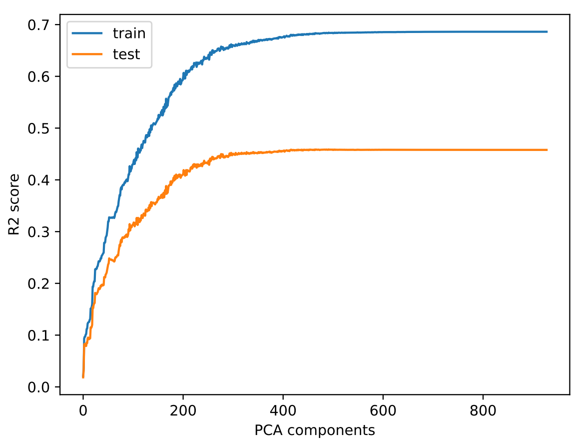
Figure A.3: Bayesian Regression Results on EQ



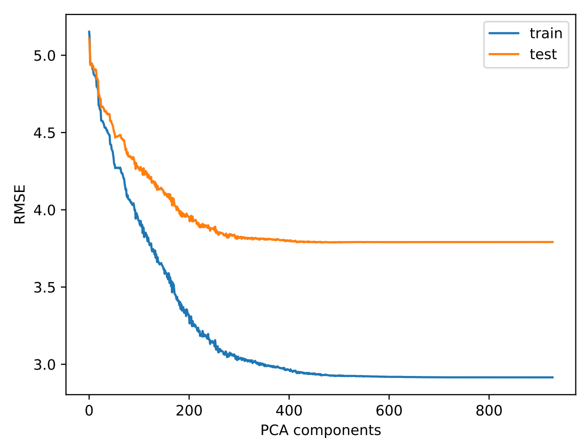
(a) R^2 score on Position Data



(b) RMSE on Position Data



(c) R^2 score on Velocity Data



(d) RMSE on Velocity Data

Figure A.4: Bayesian Regression Results on SQ

Principal Component Regression results for Personality Prediction

Principal Component Regression		Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
		RMSE	R2	RMSE	R2	RMSE	R2	RMSE	R2	RMSE	R2
Dataset-1	Position	0.21	0.76	0.36	0.67	0.39	0.77	0.29	0.72	0.40	0.72
	Velocity	0.38	0.31	0.53	0.29	0.62	0.41	0.48	0.21	0.59	0.40
Dataset-2	Position	0.26	0.84	0.25	0.88	0.32	0.83	0.21	0.84	0.27	0.85
	Velocity	0.52	0.36	0.60	0.25	0.63	0.33	0.44	0.31	0.53	0.35

Table A.1: Prediction Results for Five Personality Traits using Principal Component Regression on both the datasets.

We took Principal Components of the features for the model against Linear Regression model to avoid overfitting. Table A.1 contains the calculated RMSE and R^2 scores for individual personality traits from both the datasets after taking the Principal Components. We repeated this experiment by varying the number of principal components. The results show comparatively inferior performance as compared to the Bayesian Regression.

In order to demonstrate the effect of varying components, Figure A.5 contains the graph of R^2 and RMSE for the performed PCR on Openness from both Datasets.

Spearman Correlations between trait-wise personality across datasets.

In order to unearth underlying associations in *Joint Importance profiles* across datasets, we evaluated Spearman correlations between *Joint Importance profiles* for individual personality traits across datasets which is provided in Table A.2.

We calculated the correlation coefficient between trait-wise *Joint Importance profiles* within a dataset and pool them across datasets. We performed pooling of r-values by averaging them, in addition to averaging the Fisher-Z transformed ([4]) r-values. The resulting correlations were found to be near identical, hence we provide the results of the corresponding r-values of the averaged Z-values in Table A.3.

Joint Importance Profile for Genre Prediction

Figure A.6 displays the relative joint importance of Musical Preferences along with the mean plotted in each sub-figure. The farther away from the mean the joint importance value for an individual joint is, the more important it is in characterizing that trait. Some similarities in the joint importance profiles across different genres can be attributed to the inherent correlation that exists among them.

	O_1	C_1	E_1	A_1	N_1	O_2	C_2	E_2	A_2	N_2
O_1	1.0					0.11				
C_1	-0.08	1.0				-0.17	0.73*			
E_1	-0.50	-0.61*	1.0			0.01	-0.63*	0.52		
A_1	0.29	-0.51	0.1	1.0		-0.08	-0.1	-0.06	0.32	
N_1	0.17	-0.14	-0.35	-0.29	1.0	0.35	0.01	-0.03	-0.12	-0.2
O_2	0.11					1.0				
C_2	-0.08	0.73*				-0.49	1.0			
E_2	-0.48	-0.43	0.52			-0.27	-0.24	1.0		
A_2	0.17	0.04	-0.10	0.32		-0.34	0.18	-0.56	1.0	
N_2	0.31	-0.24	0.23	0.03	-0.20	0.28	-0.57	-0.15	-0.23	1.0

$*p < 0.05$

Table A.2: Results of the Spearman Correlation between the Mean-centered *Joint Importance* Vectors learned for personality traits.

	O	C	E	A	N
C	-0.24				
E	-0.41	-0.53			
A	-0.08	-0.42	-0.16		
N	0.15	-0.14	-0.24	-0.31	

Table A.3: Average of Spearman Correlations via Fisher-Z transformation between *Joint Importance profiles* for personalities across both datasets.

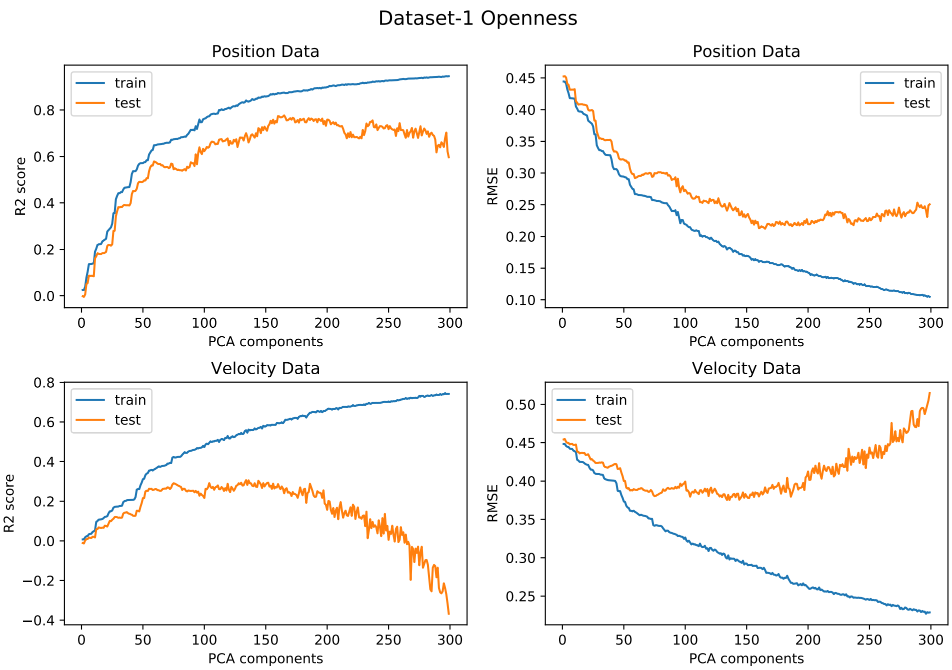
Ward's Method

The distance between two clusters, A and B, using the ward's method is given by:

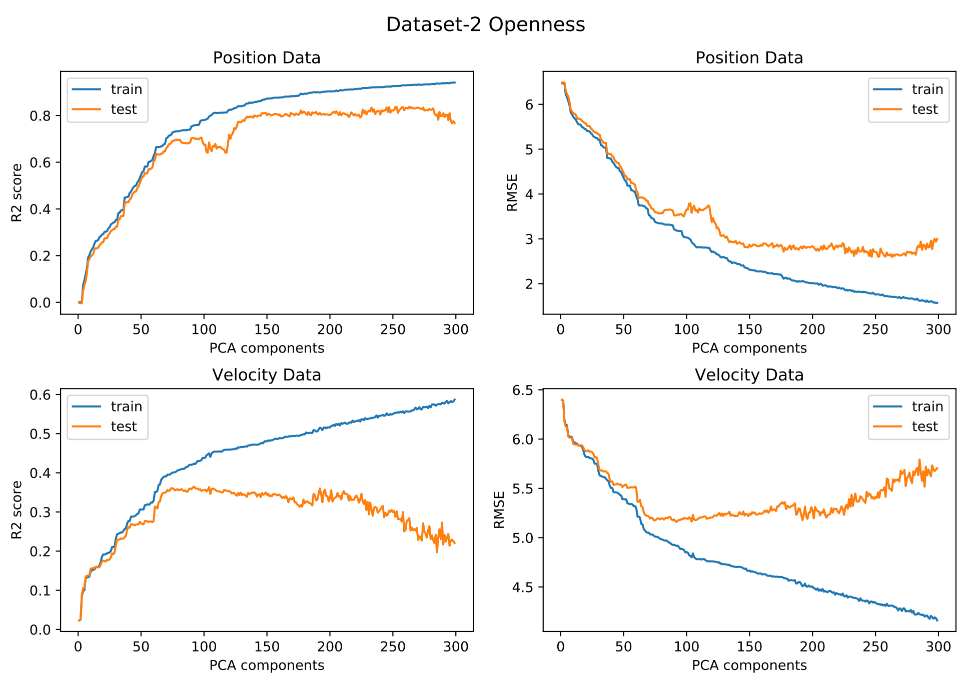
$$\Delta(A,B) = \sum_{i \in A \cup B} ||\vec{x}_i - \vec{m}_{A \cup B}||^2 - \sum_{i \in A} ||\vec{x}_i - \vec{m}_A||^2 - \sum_{i \in B} ||\vec{x}_i - \vec{m}_B||^2 \qquad \text{(A.4)}$$

$$\Delta(A,B) = \frac{n_A n_B}{n_A + n_B} ||\vec{m}_A - \vec{m}_B||^2 \qquad \text{(A.5)}$$

where $\vec{m}_j$, n_j represents the center of cluster, the number of points in cluster j. $\Delta(A, B)$ is called the merging cost of combining the clusters A and B. Ward's method keeps the merging cost as small as possible.



(a)



(b)

Figure A.5: Principal Component Regression results on Openness personality trait from Dataset-1 and Dataset-2.

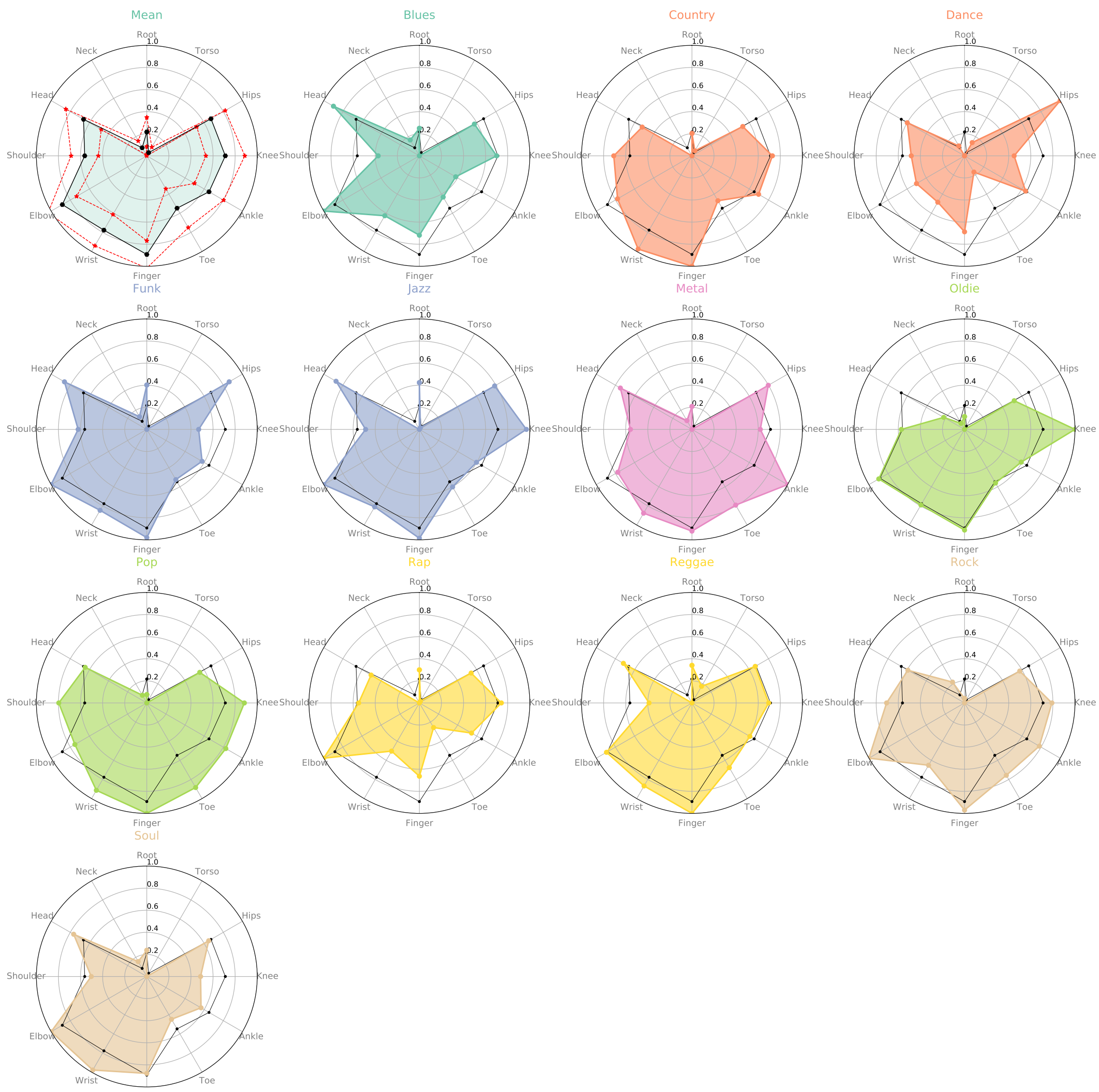


Figure A.6: Relative importance of Joints of the 12 music preferences (Blues, Country, Dance, Funk, Jazz, Metal, Oldies, Pop, Rap, Reggae, Rock, and Soul.) using the Position Data. The black line indicates the mean importance of the corresponding joint marker. The red dotted line in the top left sub-figure indicates the standard deviation about the mean.

Related Publications

- **Yudhik Agrawal**, Samyak Jain, Emily Carlson, Petri Toiviainen, Vinoo Alluri.“Towards Multimodal MIR: Predicting individual differences from music-induced movement.” In 21st International Society for Music Information Retrieval Conference (ISMIR). 2020.

- **Yudhik Agrawal**, Ramaguru Guru Ravi Shanker, Vinoo Alluri. “Transformer-based approach towards music emotion recognition from lyrics.” In 43rd European Conference On Information Retrieval (ECIR). 2021.

- **Yudhik Agrawal**, Vinoo Alluri. “Personality correlates of Preferred Emotions through Lyrics.” Accepted as Oral Presentation in 16th International Conference on Music Perception and Cognition (ICMPC-ESCOM). 2021.

- **Yudhik Agrawal**, Emily Carlson, Petri Toiviainen, Vinoo Alluri. “Dance who you are: Decoding Individual differences and musical preference via music-induced movement.” In submission with Scientific Reports. 2021

Other Publications

- Abbhinav Venkat, Chaitanya Patel, **Yudhik Agrawal**, Avinash Sharma. “HumanMeshNet: Polygonal Mesh Recovery of Humans.” In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV-W). 2019.

- Neeraj Battan*, **Yudhik Agrawal***, Veeravalli Saisooryarao, Aman Goel, Avinash Sharma. “GlocalNet: Class-aware Long-term Human Motion Synthesis.” In IEEE Winter Conference on Applications of Computer Vision (WACV). 2021.

- Ramaguru Guru Ravi Shanker, **Yudhik Agrawal**, Vinoo Alluri. “Preference for instrumental music on online music streaming platforms associated with individual differences.” Accepted as Poster Presentation in 16th International Conference on Music Perception and Cognition (ICMPC-ESCOM). 2021.

- Kunal Vaswani, **Yudhik Agrawal**, Vinoo Alluri. “Multimodal Fusion based Attentive Networks for Sequential Music Recommendation.” Submitted in 15th ACM Conference on Recommender Systems (RecSys). 2021.

Fellowships

- Huawei Scholarship of Excellence, 2020-21

Bibliography

[1] *Spotify reaches 144 million paid subscribers amid COVID-19 pandemic*. https://in.finance.yahoo.com/news/spotify-music-144-million-paid-subscribers-amid-covid-19-pandemic-120816933.html.

[2] J. Abdillah, I. Asror, Y. F. A. Wibowo, et al. Emotion classification of song lyrics using bidirectional lstm method with glove word representation weighting. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4):723–729, 2020.

[3] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau. Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 30–38, 2011.

[4] R. A. Alexander. A note on averaging correlations. *Bulletin of the Psychonomic Society*, 28(4):335–336, 1990.

[5] G. W. Allport. Pattern and growth in personality. 1961.

[6] V. Alluri, E. Brattico, P. Toiviainen, I. Burunat, B. Bogert, J. Numminen, and M. Kliuchko. Musical expertise modulates functional connectivity of limbic regions during continuous music listening. *Psychomusicology: Music, Mind, and Brain*, 25(4):443, 2015.

[7] I. Anderson, S. Gil, C. Gibson, S. Wolf, W. Shapiro, O. Semerci, and D. M. Greenberg. "just the way you are": Linking music listening on spotify and personality. *Social Psychological and Personality Science*, page 1948550620923228, 2020.

[8] A. Anzulewicz, K. Sobota, and J. T. Delafield-Butt. Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific reports*, 6(1):1–13, 2016.

[9] J.-J. Aucouturier and F. Pachet. Representing musical genre: A state of the art. *Journal of new music research*, 32(1):83–93, 2003.

[10] J. M. S. Bamford and J. W. Davidson. Trait empathy associated with agreeableness and rhythmic entrainment in a spontaneous movement to music task: Preliminary exploratory investigations. *Musicae Scientiae*, 23(1):5–24, 2019.

[11] M. Bangert, T. Peschel, G. Schlaug, M. Rotte, D. Drescher, H. Hinrichs, H.-J. Heinze, and E. Altenmüller. Shared networks for auditory and motor processing in professional pianists: evidence from fmri conjunction. *Neuroimage*, 30(3):917–926, 2006.

[12] C. D. Barclay, J. E. Cutting, and L. T. Kozlowski. Temporal and spatial factors in gait perception that influence gender recognition. *Perception & psychophysics*, 23(2):145–152, 1978.

[13] S. Baron-Cohen, M. O'riordan, V. Stone, R. Jones, and K. Plaisted. Recognition of faux pas by normally developing children and children with asperger syndrome or high-functioning autism. *Journal of autism and developmental disorders*, 29(5):407–418, 1999.

[14] S. Baron-Cohen, J. Richler, D. Bisarya, N. Gurunathan, and S. Wheelwright. The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1430):361–374, 2003.

[15] S. Baron-Cohen and S. Wheelwright. The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, 34(2):163–175, 2004.

[16] J. Barry. Sentiment analysis of online reviews using bag-of-words and lstm approaches. In *AICS*, pages 272–274, 2017.

[17] J. C. Bispham. *The Human Faculty for Music: What's special about it?* PhD thesis, University of Cambridge, 2018.

[18] A. Bonneville-Roussy, P. J. Rentfrow, M. K. Xu, and J. Potter. Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of personality and social psychology*, 105(4):703, 2013.

[19] M. A. Brackett and P. Salovey. Measuring emotional intelligence with the mayer-salovery-caruso emotional intelligence test (msceit). *Psicothema*, 18:34–41, 2006.

[20] B. Burger, J. Polet, G. Luck, M. R. Thompson, S. Saarikallio, and P. Toiviainen. Investigating relationships between music, emotions, personality, and music-induced movement. In *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013*. University of Jyväskylä, Department of Music, 2013.

[21] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen. Emotions move us: Basic emotions in music influence people's movement to music. In *Proceedings of the 12th international conference on music perception and cognition and 8th triennial conference of the european society for the cognitive sciences of music*, pages 177–182. Aristotle University of Thessaloniki Thessaloniki, Greece, 2012.

[22] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen. Relationships between perceived emotions in music and music-induced movement. *Music Perception: An Interdisciplinary Journal*, 30(5):517–533, 2012.

[23] B. Burger, M. R. Thompson, G. Luck, S. Saarikallio, and P. Toiviainen. Influences of rhythm-and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology*, 4:183, 2013.

[24] B. Burger, M. R. Thompson, S. Saarikallio, G. Luck, and P. Toiviainen. Influence of musical features on characteristics of music-induced movements. In *Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC)*, pages 425–428, 2010.

[25] B. Burger and P. Toiviainen. Mocap toolbox-a matlab toolbox for computational analysis of movement data. In *10th Sound and Music Computing Conference, SMC 2013, Stockholm, Sweden*. Logos Verlag Berlin, 2013.

[26] B. Burger and P. Toiviainen. Embodiment in electronic dance music: Effects of musical content and structure on body movement. *Musicae Scientiae*, page 1029864918792594, 2018.

[27] D. M. Buss. Selection, evocation, and manipulation. *Journal of personality and social psychology*, 53(6):1214, 1987.

[28] D. J. Cameron and J. A. Grahn. The neuroscience of rhythm. *The Oxford handbook of music psychology*, pages 357–368, 2016.

[29] E. Çano and M. Morisio. Moodylyrics: A sentiment annotated lyrics dataset. In *Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*, pages 118–124, 2017.

[30] E. Çano, M. Morisio, et al. Music mood dataset creation based on last. fm tags. In *2017 International Conference on Artificial Intelligence and Applications, Vienna, Austria*, 2017.

[31] E. Carlson, B. Burger, J. London, M. R. Thompson, and P. Toiviainen. Conscientiousness and extraversion relate to responsiveness to tempo in dance. *Human movement science*, 49:315–325, 2016.

[32] E. Carlson, B. Burger, and P. Toiviainen. Dance like someone is watching: A social relations model study of music-induced movement. *Music & Science*, 1:2059204318807846, 2018.

[33] E. Carlson, B. Burger, and P. Toiviainen. Empathy, entrainment, and perceived interaction in complex dyadic dance movement. *Music Perception: An Interdisciplinary Journal*, 36(4):390–405, 2019.

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

[35] E. Carlson, P. Saari, B. Burger, and P. Toiviainen. Dance to your own drum: Identification of musical genre and individual dancer from motion capture using machine learning. *Journal of New Music Research*, pages 1–16, 2020.

[36] R. B. Cattell, H. W. Eber, and M. M. Tatsuoka. *Handbook for the sixteen personality factor questionnaire (16 PF): In clinical, educational, industrial, and research psychology, for use with all forms of the test*. Institute for personality and ability testing, 1970.

[37] E. Chen and G. E. Miller. Social context as an individual difference in psychoneuroimmunology. *Psychoneuroimmunology*, 4:497–508, 2007.

[38] J. F. Christensen, C. J. Cela-Conde, and A. Gomila. Not all about sex: neural and biobehavioral functions of human dance. *Annals of the New York Academy of Sciences*, 1400(1):8–32, 2017.

[39] M. Cliche. Bb_twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms. *arXiv preprint arXiv:1704.06125*, 2017.

[40] A. Colley. Young people's musical taste: Relationship with gender and gender-related traits 1. *Journal of applied social psychology*, 38(8):2039–2055, 2008.

[41] C. Cooper. *Individual differences*, volume 2. Arnold London, 2002.

[42] P. J. Corr and G. Matthews. *The Cambridge handbook of personality psychology*. Cambridge University Press, 2020.

[43] I. Cross. Music, cognition, culture, and evolution. *Annals of the New York Academy of sciences*, 930(1):28–42, 2001.

[44] I. Cross. Musicality and the human capacity for culture. *Musicae Scientiae*, 12(1_suppl):147–167, 2008.

[45] I. Cross. The evolutionary nature of musical meaning. *Musicae scientiae*, 13(2_suppl):179–200, 2009.

[46] I. Cross. Music and communication in music psychology. *Psychology of music*, 42(6):809–819, 2014.

[47] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society*, 9(5):353–356, 1977.

[48] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019.

[49] T.-T. Dang and K. Shirai. Machine learning approaches for mood classification of songs toward music search engine. In *2009 International Conference on Knowledge and Systems Engineering*, pages 144–149. IEEE, 2009.

[50] M. J. de Dreu, A. Van Der Wilk, E. Poppe, G. Kwakkel, and E. E. van Wegen. Rehabilitation, exercise therapy and music in patients with parkinson's disease: a meta-analysis of the effects of music-based movement therapy on walking ability, balance and quality of life. *Parkinsonism & related disorders*, 18:S114–S119, 2012.

[51] R. Delbouys, R. Hennequin, F. Piccoli, J. Royo-Letelier, and M. Moussallam. Music mood detection based on audio and lyrics with deep neural net. *arXiv preprint arXiv:1809.07276*, 2018.

[52] M. J. Delsing, T. F. Ter Bogt, R. C. Engels, and W. H. Meeus. Adolescents' music preferences and personality characteristics. *European Journal of Personality: Published for the European Association of Personality Psychology*, 22(2):109–130, 2008.

[53] J. J. Denissen, W. Bleidorn, M. Hennecke, M. Luhmann, U. Orth, J. Specht, and J. Zimmermann. Uncovering the power of personality to shape income. *Psychological science*, 29(1):3–13, 2018.

[54] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[55] J. M. Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.

[56] S. J. Dollinger, L. A. Orf, and A. E. Robinson. Personality and campus controversies: Preferred boundaries as a function of openness to experience. *The Journal of psychology*, 125(4):399–406, 1991.

[57] M. Donald. *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press, 1993.

[58] P. G. Dunn, B. de Ruyter, and D. G. Bouwhuis. Toward a better understanding of the relation between music preference, listening behavior, and personality. *Psychology of Music*, 40(4):411–428, 2012.

[59] T. Eerola, O. Lartillot, and P. Toiviainen. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In *Ismir*, pages 621–626, 2009.

[60] T. Eerola, G. Luck, P. Toiviainen, et al. An investigation of pre-schoolers' corporeal synchronization with music. In *Proceedings of the 9th international conference on music perception and cognition*, pages 472–476. Citeseer, 2006.

[61] T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, 2011.

[62] M. Fell, Y. Nechaev, E. Cabrio, and F. Gandon. Lyrics segmentation: Textual macrostructure detection using convolutions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2044–2054, 2018.

[63] M. Fernández-Gavilanes, T. Alvarez-López, J. Juncal-Martinez, E. Costa-Montenegro, and F. J. González-Castano. Gti: An unsupervised approach for sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 533–538, 2015.

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

[65] B. Fink, B. Bläsing, A. Ravignani, and T. K. Shackelford. Evolution and functions of human dance. *Evolution and Human Behavior*, 2021.

[66] J. M. Fuster. Upper processing stages of the perception–action cycle. *Trends in cognitive sciences*, 8(4):143–145, 2004.

[67] A. Gabrielsson and S. L. Wik. Strong experiences related to music: adescriptive system. *Musicae scientiae*, 7(2):157–217, 2003.

[68] A. Galajdová, D. Šimšík, and R. Rákay. An automated procedure for identification of a person using gait analysis. *International Journal of Advanced Robotic Systems*, 13(5):1729881416667627, 2016.

[69] R. J. Gerrig and P. G. Zimbardo. Psikoloji ve yaşam. *Çev. Ed. Sart, G.). Nobel Yayıncılık. Ankara*, 2012.

[70] H. Gil de Zúñiga, T. Diehl, B. Huber, and J. Liu. Personality traits and social media use in 20 countries: How personality relates to frequency of social media use, social media news use, and social media use for social interaction. *Cyberpsychology, Behavior, and Social Networking*, 20(9):540–552, 2017.

[71] M. Gillies. Understanding the role of interactive machine learning in movement interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(1):1–34, 2019.

[72] R. I. Godøy, E. Haga, and A. R. Jensenius. Playing "air instruments": mimicry of sound-producing gestures by novices and experts. In *International Gesture Workshop*, pages 256–267. Springer, 2005.

[73] R. I. Godøy and A. R. Jensenius. Body movement in music information retrieval. 2009.

[74] R. I. Godøy, M. Song, K. Nymoen, M. R. Haugen, and A. R. Jensenius. Exploring sound-motion similarity in musical experience. *Journal of New Music Research*, 45(3):210–222, 2016.

[75] L. R. Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.

[76] S. Goldin-Meadow. Talking and thinking with our hands. *Current directions in psychological science*, 15(1):34–39, 2006.

[77] A. Greasley and A. Lamont. Musical preferences. *Oxford handbook of music psychology*, pages 263–281, 2016.

[78] D. M. Greenberg, S. Baron-Cohen, D. J. Stillwell, M. Kosinski, and P. J. Rentfrow. Musical preferences are linked to cognitive styles. *PloS one*, 10(7):e0131151, 2015.

[79] D. M. Greenberg, M. Kosinski, D. J. Stillwell, B. L. Monteiro, D. J. Levitin, and P. J. Rentfrow. The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science*, 7(6):597–605, 2016.

[80] D. M. Greenberg, V. Warrier, C. Allison, and S. Baron-Cohen. Testing the empathizing–systemizing theory of sex differences and the extreme male brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences*, 115(48):12152–12157, 2018.

[81] T. Gummerson. *Mobility training for the Martial Arts*. A & C Black, 1990.

[82] J.-E. Gustafsson and J. O. Undheim. Individual differences in cognitive functions. 1996.

[83] Q. Han, J. Guo, and H. Schuetze. Codex: Combining an svm classifier and character n-gram language models for sentiment analysis on twitter text. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 520–524, 2013.

[84] E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello. Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, 317(5843):1360–1366, 2007.

[85] H. Hicheur, H. Kadone, J. Grèzes, and A. Berthoz. The combined role of motion-related cues and upper body posture for the expression of emotions during human walking. In *Modeling, simulation and optimization of bipedal walking*, pages 71–85. Springer, 2013.

[86] J. R. Hollenbeck and A. P. Brief. The effects of individual differences and goal origin on goal setting and performance. *Organizational Behavior and Human Decision Processes*, 40(3):392–414, 1987.

[87] X. Hu and J. S. Downie. When lyrics outperform audio for music mood classification: A feature analysis. In *ISMIR*, pages 619–624, 2010.

[88] Y. Hu, X. Chen, and D. Yang. Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In *ISMIR*, 2009.

[89] Y.-H. Huang, S.-R. Lee, M.-Y. Ma, Y.-H. Chen, Y.-W. Yu, and Y.-S. Chen. Emotionx-idea: Emotion bert–an affectional model for conversation. *arXiv preprint arXiv:1908.06264*, 2019.

[90] C. Hufschmidt, B. Weege, S. Röder, K. Pisanski, N. Neave, and B. Fink. Physical strength and gender identification from dance movements. *Personality and Individual Differences*, 76:13–17, 2015.

[91] D. Hume. *A treatise of human nature*. Courier Corporation, 2003.

[92] D. Huron and J. K. Vuoskoski. On the enjoyment of sad music: Pleasurable compassion theory and the role of trait empathy. *Frontiers in psychology*, 11:1060, 2020.

[93] W. Ickes, M. Snyder, and S. Garcia. Personality influences on the choice of situations. In *Handbook of personality psychology*, pages 165–195. Elsevier, 1997.

[94] Z. Ivcevic and M. Brackett. Predicting school success: Comparing conscientiousness, grit, and emotion regulation ability. *Journal of research in personality*, 52:29–36, 2014.

[95] D. M. Jacobs, S. Runeson, and C. F. Michaels. Learning to visually perceive the relative mass of colliding balls in globally and locally constrained task ecologies. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5):1019, 2001.

[96] A. K. Jain and R. C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[97] P. Janata, S. T. Tomic, and J. M. Haberman. Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, 141(1):54, 2012.

[98] L. A. Jensen-Campbell, R. Adams, D. G. Perry, K. A. Workman, J. Q. Furdella, and S. K. Egan. Agreeableness, extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression. *Journal of Research in Personality*, 36(3):224–251, 2002.

[99] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, 14(2):201–211, 1973.

[100] O. P. John, L. P. Naumann, and C. J. Soto. Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. 2008.

[101] O. P. John, S. Srivastava, et al. *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives*, volume 2. University of California Berkeley, 1999.

[102] R. E. Jung, J. M. Segall, H. Jeremy Bockholt, R. A. Flores, S. M. Smith, R. S. Chavez, and R. J. Haier. Neuroanatomy of creativity. *Human brain mapping*, 31(3):398–409, 2010.

[103] A. Kallias. *Individual differences and the psychology of film preferences*. PhD thesis, Goldsmiths, University of London, 2012.

[104] D. Kansara and V. Sawant. Comparison of traditional machine learning and deep learning approaches for sentiment analysis. In *Advanced Computing Technologies and Applications*, pages 365–377. Springer, 2020.

[105] F. Kleedorfer, P. Knees, and T. Pohle. Oh oh oh whoah! towards automatic topic detection in song lyrics. In *Ismir*, pages 287–292, 2008.

[106] A. L. Knutson. Japanese opinion surveys: the special need and the special difficulties. *Public Opinion Quarterly*, 9(3):313–319, 1945.

[107] S. Koelsch. Towards a neural basis of music-evoked emotions. *Trends in cognitive sciences*, 14(3):131–137, 2010.

[108] M. Koppensteiner. Motion cues that make an impression: Predicting perceived personality by minimal motion information. *Journal of experimental social psychology*, 49(6):1137–1143, 2013.

[109] E. Lacourse, M. Claes, and M. Villeneuve. Heavy metal music and adolescent suicidal risk. *Journal of youth and adolescence*, 30(3):321–332, 2001.

[110] A. Lahav, E. Saltzman, and G. Schlaug. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27(2):308–314, 2007.

[111] M. Leman et al. *Embodied music cognition and mediation technology*. MIT press, 2008.

[112] M. Leman and R. I. Godøy. Why study musical gestures. *Musical gestures: Sound, movement, and meaning*, pages 3–11, 2010.

[113] J. A. LePine, J. A. Colquitt, and A. Erez. Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel psychology*, 53(3):563–593, 2000.

[114] M. Lesaffre, L. D. Voogdt, M. Leman, B. D. Baets, H. D. Meyer, and J.-P. Martens. How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology*, 59(5):695–707, 2008.

[115] T. Li, X. Yan, Y. Li, J. Wang, Q. Li, H. Li, and J. Li. Neuronal correlates of individual differences in the big five personality traits: evidences from cortical morphology and functional homogeneity. *Frontiers in neuroscience*, 11:414, 2017.

[116] T. Lischetzke and M. Eid. Why extraverts are happier than introverts: The role of mood regulation. *Journal of personality*, 74(4):1127–1162, 2006.

[117] C. Liu, M. Osama, and A. De Andrade. Dens: a dataset for multi-class emotion analysis. *arXiv preprint arXiv:1910.11769*, 2019.

[118] W. Liu, P. P. Pokharel, and J. C. Príncipe. Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Transactions on Signal Processing*, 55(11):5286–5298, 2007.

[119] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[120] G. Luck, S. Saarikallio, B. Burger, M. Thompson, and P. Toiviainen. Emotion-driven encoding of music preference and personality in dance. *Musicae Scientiae*, 18(3):307–323, 2014.

[121] G. Luck, S. Saarikallio, B. Burger, M. R. Thompson, and P. Toiviainen. Effects of the big five and musical genre on music-induced movement. *Journal of Research in Personality*, 44(6):714–720, 2010.

[122] G. Luck, S. Saarikallio, and P. Toiviainen. Personality traits correlate with characteristics of music-induced movement. In *ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, 2009.

[123] R. Malheiro, R. Panda, P. Gomes, and R. Paiva. Music emotion recognition from lyrics: A comparative study. 6th International Workshop on Machine Learning and Music (MML13). Held in ..., 2013.

[124] R. Malheiro, R. Panda, P. Gomes, and R. P. Paiva. Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, 9(2):240–254, 2016.

[125] L. R. Martin, H. S. Friedman, and J. E. Schwartz. Personality and mortality risk across the life span: the importance of conscientiousness as a biopsychosocial attribute. *Health Psychology*, 26(4):428, 2007.

[126] E. Mas-Herrero, J. Marco-Pallares, U. Lorenzo-Seva, R. J. Zatorre, and A. Rodriguez-Fornells. Individual differences in music reward experiences. *Music Perception: An Interdisciplinary Journal*, 31(2):118–138, 2012.

[127] E. B. McClure. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological bulletin*, 126(3):424, 2000.

[128] R. R. McCrae. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology*, 52(6):1258, 1987.

[129] R. R. McCrae. Neo-pi-r data from 36 cultures. In *The five-factor model of personality across cultures*, pages 105–125. Springer, 2002.

[130] R. R. McCrae and A. Terracciano. Personality profiles of cultures: aggregate personality traits. *Journal of personality and social psychology*, 89(3):407, 2005.

[131] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.

[132] M. R. Mehl and J. W. Pennebaker. The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, 84(4):857, 2003.

[133] A. B. Melchiorre and M. Schedl. Personality correlates of music audio preferences for modelling music listeners. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 313–317, 2020.

[134] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte. Embodiment of sadness and depression—gait patterns associated with dysphoric mood. *Psychosomatic medicine*, 71(5):580–587, 2009.

[135] D. Miranda and M. Claes. Music listening, coping, peer affiliation and depression in adolescence. *Psychology of music*, 37(2):215–233, 2009.

[136] J. Mulder, T. Ter Bogt, Q. Raaijmakers, and W. Vollebergh. Music taste groups and problem behavior. *Journal of youth and adolescence*, 36(3):313–324, 2007.

[137] G. Nave, J. Minxha, D. M. Greenberg, M. Kosinski, D. Stillwell, and J. Rentfrow. Musical preferences predict personality: evidence from active listening and facebook likes. *Psychological Science*, 29(7):1145–1158, 2018.

[138] B. Nettl. An ethnomusicologist contemplates universals in musical sound and musical culture. *The origins of music*, 3(2):463–472, 2000.

[139] C. B. Ng, Y. H. Tay, and B. M. Goi. Vision-based human gender recognition: A survey. *arXiv preprint arXiv:1204.1611*, 2012.

[140] A. C. North, D. J. Hargreaves, and J. J. Hargreaves. Uses of music in everyday life. *Music perception*, 22(1):41–77, 2004.

[141] B. Ohana and B. Tierney. Sentiment classification of reviews using sentiwordnet. In *9th. it & t conference*, volume 13, pages 18–30, 2009.

[142] J. Opitz and S. Burst. Macro f1 and macro f1. *arXiv preprint arXiv:1911.03347*, 2019.

[143] J. Ormel, H. Riese, and J. G. Rosmalen. Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect? *Clinical Psychology Review*, 32(1):71–79, 2012.

[144] B. P. O'Connor. A quantitative review of the comprehensiveness of the five-factor model in relation to popular personality inventories. *Assessment*, 9(2):188–203, 2002.

[145] F. Pachet, D. Cazaly, et al. A taxonomy of musical genres. In *RIAO*, pages 1238–1245. Citeseer, 2000.

[146] P. Palmgreen, L. A. Wenner, K. E. Rosengren, et al. Uses and gratifications research: The past ten years. *Media gratifications research: Current perspectives*, 1:l–37, 1985.

[147] R. Panda, R. Malheiro, B. Rocha, A. Oliveira, and R. P. Paiva. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis. In *International Symposium on Computer Music Multidisciplinary Research*, 2013.

[148] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*, 2004.

[149] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*, 2002.

[150] A. Patel and A. K. Tiwari. Sentiment analysis by using recurrent neural network. In *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019.

[151] A. D. Patel, J. R. Iversen, Y. Chen, and B. H. Repp. The influence of metricality and modality on synchronization with a beat. *Experimental brain research*, 163(2):226–238, 2005.

[152] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[153] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

[154] R. Pfeifer and C. Scheier. *Understanding intelligence*. MIT press, 2001.

[155] R. A. Power and M. Pluess. Heritability estimates of the big five personality traits based on common genetic variants. *Translational psychiatry*, 5(7):e604–e604, 2015.

[156] L. Qiu, J. Chen, J. Ramsay, and J. Lu. Personality predicts words in favorite songs. *Journal of Research in Personality*, 78:25–35, 2019.

[157] P. Raina. Sentiment analysis in news articles using sentic computing. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 959–962. IEEE, 2013.

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

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

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

[161] P. J. Rentfrow and S. D. Gosling. The content and validity of music-genre stereotypes among college students. *Psychology of music*, 35(2):306–326, 2007.

[162] B. H. Repp. Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, 12(6):969–992, 2005.

[163] B. H. Repp and A. Penel. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological research*, 68(4):252–270, 2004.

[164] B. H. Repp and Y.-H. Su. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review*, 20(3):403–452, 2013.

[165] J. Richter and R. Ostovar. "it don't mean a thing if it ain't got that swing"–an alternative concept for understanding the evolution of dance and music in human beings. *Frontiers in human neuroscience*, 10:485, 2016.

[166] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996.

[167] C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese. Critical features for the perception of emotion from gait. *Journal of vision*, 9(6):15–15, 2009.

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

[169] L. Satchell, P. Morris, C. Mills, L. O'Reilly, P. Marshman, and L. Akehurst. Evidence of big five and aggressive personalities in gait biomechanics. *Journal of nonverbal behavior*, 41(1):35–44, 2017.

[170] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[171] M. Sawada, K. Suda, and M. Ishii. Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and motor skills*, 97(3):697–708, 2003.

[172] M. Schoen. The psychology of music: a survey for teacher and musician. 1940.

[173] T. W. Schubert and S. L. Koole. The embodied self: Making a fist enhances men's power-related self-conceptions. *Journal of Experimental Social Psychology*, 45(4):828–834, 2009.

[174] L. Shapiro. *Embodied cognition*. Routledge, 2019.

[175] M. E. Shaw. *Group dynamics: The psychology of small group behavior*. McGraw-Hill College, 1981.

[176] J. A. Sloboda. Individual differences in music performance. *Trends in cognitive sciences*, 4(10):397–403, 2000.

[177] R. T. Solberg and A. R. Jensenius. Pleasurable and intersubjectively embodied experiences of electronic dance music. *Empirical Musicology Review*, 11(3-4):301–318, 2017.

[178] D. J. Stillwell and M. Kosinski. mypersonality project: Example of successful utilization of online social networks for large-scale social research. *American Psychologist*, 59(2):93–104, 2004.

[179] M. Stitson, J. Weston, A. Gammerman, V. Vovk, and V. Vapnik. Theory of support vector machines. *University of London*, 117(827):188–191, 1996.

[180] F. Strack, L. L. Martin, and S. Stepper. Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology*, 54(5):768, 1988.

[181] K. Stueber. Empathy. *International Encyclopedia of Ethics*, 2013.

[182] M. Subotnick. Interactive music playback system utilizing gestures, Nov. 1 2001. US Patent App. 09/835,840.

[183] C. Sun, X. Qiu, Y. Xu, and X. Huang. How to fine-tune bert for text classification? In *China National Conference on Chinese Computational Linguistics*, pages 194–206. Springer, 2019.

[184] J. Sun, P. Wu, Y. Shen, Z. Yang, H. Li, Y. Liu, T. Zhu, L. Li, K. Zhang, and M. Chen. Relationship between personality and gait: predicting personality with gait features. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1227–1231. IEEE, 2018.

[185] A. Surana, Y. Goyal, M. Shrivastava, S. Saarikallio, and V. Alluri. Tag2risk: Harnessing social music tags for characterizing depression risk. *arXiv preprint arXiv:2007.13159*, 2020.

[186] M. R. Thompson and G. Luck. Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, 16(1):19–40, 2012.

[187] J. C. Thoresen, Q. C. Vuong, and A. P. Atkinson. First impressions: Gait cues drive reliable trait judgements. *Cognition*, 124(3):261–271, 2012.

[188] E. B. Titchener. *Lectures on the experimental psychology of the thought-processes*. Macmillan, 1909.

[189] P. Toiviainen, G. Luck, and M. R. Thompson. Embodied meter: hierarchical eigenmodes in music-induced movement. *Music Perception*, 28(1):59–70, 2010.

[190] E. B. Torres, M. Brincker, R. W. Isenhower III, P. Yanovich, K. A. Stigler, J. I. Nurnberger Jr, D. N. Metaxas, and J. V. José. Autism: the micro-movement perspective. *Frontiers in integrative neuroscience*, 7:32, 2013.

[191] C. Trevarthen, T. Kokkinaki, and G. A. Fiamenghi Jr. What infants' imitations communicate: With mothers, with fathers and with peers. 1999.

[192] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision*, 2(5):2–2, 2002.

[193] N. F. Troje, C. Westhoff, and M. Lavrov. Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, 67(4):667–675, 2005.

[194] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.

[195] P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*, 2002.

[196] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[197] E. Van Dyck, P. Vansteenkiste, M. Lenoir, M. Lesaffre, and M. Leman. Recognizing induced emotions of happiness and sadness from dance movement. *PloS one*, 9(2):e89773, 2014.

[198] F. J. Varela, E. Thompson, and E. Rosch. *The Embodied Mind, revised edition: Cognitive Science and Human Experience*. MIT press, 2017.

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

[200] J. K. Vuoskoski and T. Eerola. The role of mood and personality in the perception of emotions represented by music. *Cortex*, 47(9):1099–1106, 2011.

[201] J. K. Vuoskoski, W. F. Thompson, D. McIlwain, and T. Eerola. Who enjoys listening to sad music and why? *Music Perception*, 29(3):311–317, 2011.

[202] A. Wakabayashi, S. Baron-Cohen, S. Wheelwright, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, and L. Weil. Development of short forms of the empathy quotient (eq-short) and the systemizing quotient (sq-short). *Personality and individual differences*, 41(5):929–940, 2006.

[203] J. Wang, L. Gu, and M. Aiken. A study of the impact of individual differences on online shopping. *International Journal of E-Business Research (IJEBR)*, 6(1):52–67, 2010.

[204] K. B. Watson. The nature and measurement of musical meanings. *Psychological Monographs*, 54(2):i, 1942.

[205] J. B. Weaver III. Exploring the links between personality and media preferences. *Personality and individual differences*, 12(12):1293–1299, 1991.

[206] B. Weege, M. N. Pham, T. K. Shackelford, and B. Fink. Physical strength and dance attractiveness: further evidence for an association in men, but not in women. *American Journal of Human Biology*, 27(5):728–730, 2015.

[207] C. Westhoff and N. F. Troje. Kinematic cues for person identification from biological motion. *Perception & psychophysics*, 69(2):241–253, 2007.

[208] M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.

[209] R. A. Wilson and L. Foglia. Embodied cognition. 2011.

[210] R. Withagen and C. F. Michaels. The role of feedback information for calibration and attunement in perceiving length by dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6):1379, 2005.

[211] R. Withagen and M. Van Wermeskerken. Individual differences in learning to perceive length by dynamic touch: Evidence for variation in perceptual learning capacities. *Perception & Psychophysics*, 71(1):64–75, 2009.

[212] D. Wu, J. V. José, J. I. Nurnberger, and E. B. Torres. A biomarker characterizing neurodevelopment with applications in autism. *Scientific reports*, 8(1):1–14, 2018.

[213] Y. Xia, L. Wang, and K.-F. Wong. Sentiment vector space model for lyric-based song sentiment classification. *International Journal of Computer Processing Of Languages*, 21(04):309–330, 2008.

[214] Y. Yang and X. Liu. A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 42–49, 1999.

[215] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5753–5763, 2019.

[216] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Transactions on image processing*, 18(8):1905–1910, 2009.

[217] M. Zentner and T. Eerola. Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107(13):5768–5773, 2010.

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

[219] Y. Zhang and Q. Yang. An overview of multi-task learning. *National Science Review*, 5(1):30–43, 2018.