Technische Universität Berlin

Integrating Embodied Music Cognition into Music Recommendation Services

Predicting Verbal Descriptions of Emotional Listening Experience with Gestural Motion Data

Master Thesis
am Fachgebiet Agententechnologien in betrieblichen Anwendungen und der Telekommunikation (AOT)
Prof. Dr. Dr. h.c. Sahin Albayrak
Fakultät IV Elektrotechnik und Informatik
Technische Universität Berlin

vorgelegt von
Melanie Irrgang

Betreuer:	Esra Acar, Dr. Andreas Lommatzsch, Dr. Hauke Egermann,
Gutachter:	Prof. Dr. Dr. h.c. Sahin Albayrak
	Prof. Dr. Stefan Weinzierl

Melanie Irrgang
[redacted]
Berlin, den 12.03.2015

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum

Unterschrift

Abstract

Music is often discussed to be perceived as emotional because it renders expressive movements into audible musical structures. Different theoretical accounts of embodied music cognition state that listeners internally mimic movements during music listening experiences. Thus, a valid approach to measure musical emotion could be to assess movement stimulated by music. That is why this Master Thesis tested if mobile-device generated acceleration data produced by free movement during music listening experience can be used to predict different degrees of the Geneva Emotion Music Scales (GEMS-9).

The study has been conducted in both lab (n = 22) and field (n=11). Participants were instructed to move a mobile device (smartphone or tablet) continuously while listening to music in order to describe their experience. After each embodied description they rated the perceived emotional qualities of the musical excerpts according to the GEMS on a 100-point, unipolar intensity scale initialized to '0'. For this study 10 musical excerpts of $\sim$ 40s duration were selected in advance by the 'field' participants covering various musical genres and GEMS states. During the experiment, participants were also asked how suitable they considered both embodied and GEMS descriptions for each excerpt.

In order to fit a model to predict ratings for each of the GEMS-9 states, spectral and temporal features were extracted from the motion data recorded by the device's acceleration sensors. These features have been related to the following categories: 'tempo', 'size', 'smoothness' and 'roughness' of the movement.

Best results have been achieved for *power*. The musical excerpts perceived as more *energetic* have also been those for which participants preferred the embodied description over the GEMS. And vice versa: The most *peaceful*, *nostalgic* and *tender* musical excerpts have been the ones for which participants preferred the GEMS over the embodied description. Another reason for the better performance of *power* over *transcendence* e.g., might be that *power* can be more easily captured by the rather rhythmic features of the method while *transcendence* would require additional directional features representing the contour of the movement. Last but not least, *transcendence* was most difficult to model because there had not been enough variance in its GEMS ratings between songs.

Zusammenfassung

Musik wird oft als emotional wahrgenommen, weil es expressive Bewegungen hörbar macht. Unterschiedliche theoretische Arbeiten zur Verkörperung von Musikwahrnehmung legen dar, dass Rezipierende während des Musikerlebens intern Bewegungen nachahmen. Folglich könnte ein valider Ansatz, um musikalische Emotionen zu messen, darin bestehen, Bewegung zu erfassen, die durch Musik stimuliert wurde. Daher wird im Rahmen dieser Masterarbeit getested, ob die Beschleunigungsdaten von mobilen Endgeräten, die durch freie Bewegung während des Musikerlebens produziert wurden, genutzt werden können, um unterschiedliche Ausprägunen der Geneva Emotion Music Scales (GEMS-9) hervorzusagen.
Die Studie wurde sowohl im Labor (N=22) als auch im Feld (N=11) durchgeführt. Teilnehmende wurden instruiert ein mobiles Endgerät (Smartphone oder Tablet) kontinuierlich zu Musik zu bewegen, um diese zu beschreiben. Nach jeder verkörperten Beschreibung bewerteten sie die wahrgenommenen emotionalen Qualitäten der musikalischen Ausschnitte entsprechend der GEMS auf einer unipolaren Skala von 0 bis 100, die mit '0' initialisiert wurde. Für diese Studie wurden im Vorfeld durch die 'Feld'-Teilnehmenden 10 musikalische Ausschnitte ausgesucht, die $\sim$ 40s dauerten und unterschiedliche musikalische Genres, sowie GEMS Zustände abdeckten. Während des Experiments wurden Teilnehmende auch für jeden Ausschnitt gefragt, wie geeignet sie die beiden Beschreibungsmöglichkeiten für diesen empfanden.
Um ein Modell für die Vorhersagen unterschiedlicher GEMS-Ausprägungen zu erstellen, wurden spektrale und temporale Features aus den Bewegungsdaten extrahiert, die von den Beschleunigungssensoren der Geräte aufgezeichnet wurden. Diese Features beziehen sich auf folgende Kategorien: 'Tempo', 'Größe', 'Glätte' und 'Regelmäßigkeit' der Bewegung.
Beste Ergebnisse sind für *Energie* erzielt worden. Die musikalischen Ausschnitte, die als am meisten *energetisch* wahrgenommen wurden, waren auch jene, für welche Teilnehmende die verkörperte Beschreibung den GEMS vorzogen. Und umgekehrt: Die am *friedlichsten*, *nostalgischsten* und *zärtlichsten* musikalischen Ausschnitte waren jene, für welche Teilnehmende die GEMS der verkörperten Beschreibung vorzogen. Ein weiterer Grund für das bessere Abschneiden von *Energie* im Vergleich zu *Transzendenz*, beispielsweise, kann auch sein, dass die eher rhythmischen Features der Methode *En-*

ergie besser abbilden können, während *Transzendenz* mehr Features erfordert, welche die Richtung der Bewegung und damit auch die musikalische Kontur beschreiben. Zu guter Letzt war *Transzendenz* am schwierigsten zu modellieren, da es nicht genug Varianz zwischen den GEMS-Bewertungen der Stücke gab.

Acknowledgements

This Master Thesis is the most interesting experience I ever got to share during my studies. I wanna thank everyone who made this experience that unique. In particular, I want to thank my advisors Andreas Lommatzsch, Esra Acar and Hauke Egermann who have been courageous and inquiring to work in this interdisciplinary setting with all its discussions about our methodological approach. Thank you for spending that much extra time to sort things out.

Special thanks also go to Petra Lucht and the participants from the GENDER PRO MINT program who encouraged me to choose a participatory approach and who also gave a lot of support to de-gender this project.

Most of all, I want to thank my amazing participants that kept me encouraged throughout this thesis and who made it relevant. They spend their time just for the sake of it. Here, I want to particularly thank L. who allowed me to hijack their rehearsals in order to work on the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

I am not so interested in how they move as in what moves them.[14]
Pina Bausch

Listening to music is highly linked to the perception and regulation of emotions. In particular, there is a close relationship between emotion and motion: Music is also perceived as emotional because it renders expressive movements into audible musical structures [30]. Pina Bausch, the probably most famous contemporary dancer and choreographer, once told about the dancers in her company that she was not so interested in *how they move as in what moves them.* This concept of being moved as a synonym for e(-)motion or feeling also suggests a more active approach towards the description of musical listening experience in the field of Music Information Retrieval (MIR) and Recommender Systems (MRS). Therefore, the goal of this Master Thesis is to investigate if there are regularities between corporeal articulations and verbal descriptions of music, but even more so if one can predict the emotional perception of music from its embodied descriptions. The following sections motivate an embodied approach towards MIR and MRS, and lay out this master's thesis approach to investigate the connection between motion and musical emotion.

1.1 Motivation

Music is an essential part of human culture and communication. According to the German 'Bundesverband der Musikindustrie' 83% of Germans (very much) like listening to music [1] for five hours per day on average [2]. This tendency is even higher for younger generations: Amongst the 14-19 year olds even 97% enjoy listening to music. Most of the time (80%) they listen to it on the way by using their smartphones for 192 minutes per day on average [2]. Lepa et al. [49] introduced the term 'digital mobilists' for this growing group who is characterized by: being born between 1979 and 1998; listening to music primarily on mobile devices like notebook (79%) or smartphone (70%); and for whom the music streaming service *YouTube* is the medium number one (88%) followed by radio (74%) and CD (74%). Even for the average population *YouTube* already comprises 41% of the usage next to the most important one, the radio (86%), and CD (66%). This trend clarifies that the mobile listening to music via smartphone becomes more and more important and that there is a growing market in this area. Besides, the spread of this new technology offers the opportunity to realize innovative forms of research about music and to integrate the outcomes into music recommender systems like *YouTube* or *Spotify*.

Music is often deployed to regulate emotions like getting rid of stress or to influence one's mood [41, 7]. This means that listening to music is highly linked to the experience of emotions as also studied in [40, 31, 23, 21]. These studies illustrate the particular meaning of emotion in music listening experience. Nonetheless, these subjective qualities of music do still play a minor role in the field of MIR and MRS, i.e. for the retrieval and recommendation of music offered by web-based services [69].

The approaches in MIR and MRS often used to split into those, describing content collaboratively [25], and those focusing on content analysis in order to semantically annotate it [73]. Collaborative approaches are built on *user generated tags*, i.e. they let users describe the content via key words; or they rely on statistical assumptions about similarity of music that co-occurs in a playlist for instance. Although collaborative approaches are in most cases superior to the content-based approaches in terms of recommendation quality, recent work [10] aims at combining both approaches to employ the best of both worlds. Claypool names the following phenomena motivating additional content-based filtering [13]: (1) *Early Rater Problem*, (2) *Sparsity Problem* and (3) *Grey Sheep*. (1) A 'good' recommendation requires a high quality of information from a diverse range of users. This cannot be guaranteed in particular when a new system is introduced or new music is added to the system. (2) Users annotate maximally 1-2 % of the content. Therefore it is difficult to find content for which information is sufficiently available. (3) Users are similarly diverse and complex in their preferences as contents are. That is why there will always be so called *grey sheeps* who will not profit from collaboratively generated recommendations. Hence, content-based filtering can particularly contribute to diversify recommendations and to explore content and preferences beyond the mainstream.

In the field of *Affective Computing* there is increasing effort [69]to connect the physical characteristics of music to emotional *valence* and *arousal* values of the circumplex model from Russell [64]. This is a new attempt to also consider content-based information when categorizing music. However, considering acoustical features of music and its implications on the perception of emotions does not yet consider the *motor origin hypothesis* of emotions [30]. Thus, music is also perceived as emotional because it renders expressive movements into audible musical structures [30, 50, 68, 75, 72] (cf.

Section 2.4). Rhythm is particularly relevant w.r.t. motion as all (musical) rhythms might be traced back to motor activities like the frequency of foot steps or heart beats [27, 60, 30]. Last but no least, the term *emotio* derives from the Latin *movere*, to move, and is being used as a synonym for *being moved.* That is also why Leman calls for new non-verbal, embodied possibilities to describe music and its experience [47]. Leman suggests corporeal articulations as a bridge between linguistic self-report measures and measurements of physical energy like *pitch*, *loudness* or *tempo* [47, p. 81] because "human action can realize the transformation from physical energy to cultural abstraction, and vice versa" [47, p. 77]. Referring to the paradigm shift from disembodiment to embodiment (cf. Section 2.1) Leman furthermore favors to engage "subjects in musical actions rather than preventing them from being active" and to see a subject as "active contributor rather than a passive receiver" [47, p. 236]. On the one hand the elaboration of such a bridge would enable to retrieve music by moving a mobile device like a smartphone, one of the most popular devices to listen to music as mentioned above. And on the other hand, missing emotional descriptions could be extracted from smartphone-assessed motion data. Hence, such a model would not only offer innovative, multimodal access to the retrieval of music, but also place additional semantic annotations about the emotional qualities of music at the disposal that also enrich the conventional verbal search.

The project is settled in three disciplines and accompanied by the *Distributed Artificial Intelligence Laboratory (DAI-Lab)*, the *Audio Communication Group* and the *Center for Interdisciplinary Research and Gender (ZIFG)* of the *TU Berlin.* This cooperation offers us to attach to concepts from the fields of *Semantic Search*, *Affective Computing*, *Machine Learning*, *Music Perception and Cognition*, as well as *Gender and Diversity Studies*, and opens a new interdisciplinary access to the research question.

1.2 Approach and Goals

A first approach from Amelynck et al. [5] that serves as a starting point for this thesis, already investigated in the lab how motion can be linked to emotion in the context of MIR. They let participants perform arm gestures while holding a wii remote control in order to describe the music. Afterwards the emotional qualities of the musical excerpts have been described using the cirumplex model (cf. Sections 2.3 and 2.4). Under the title "Integrating Embodied Music Cognition into Music Recommendation Services - Predicting Verbal Descriptions of Emotional Listening Experience with Gestural Motion Data" the thesis will be affiliated to the study from Amelynck [5] but will be using the *Geneva Emotion Music Scales (GEMS)* as emotional model. The goal is to explore which and how well each of the GEMS can be predicted by an embodied description of the music. Besides in contrast to the work of Amelynck et al., a part of this study will also be conducted in the field in order to capture diverse facets of experiencing music (cf. Section 2.2).

In order to avoid the pitfalls of *I-methodology*[1] the design of the instrument will be accomplished in a participatory design approach [8]. Participants in the development of an instrument will be, amongst others, participants of a *Mädchenrockband* class in Berlin-Neukölln. The participants of this girl rock band will not be representative of all users. However, as our volunteers in the department of computer science and audio communication are majoritarian 'white', 'academic' and 'male', I[2]considered them to introduce a certain kind of counterbalance.

In summary, I want to investigate how an embodied description might help to diversify the (technical) representation of musical content and how these corporeal articulations are correlated to 'traditional' verbal labels, first of all to the GEMS. The use of an active, embodied approach to describe music listening experiences, the deployment of the GEMS, as well as the conduction of the study in the field constitutes a knowledge gap that has not been investigated yet. Not only might the results be used to offer innovative ways of multimodal querying for music in the field of MIR and MRS, but it will also lead to new insights in the field of cognition and emotion.

1.3 Structure of the Thesis

This thesis is structured as follows: Chapter 2 covers essential background related to the concept of embodiment and emotion. It also presents related work in the field of "music and emotion" and "affective computing". Chapter 3 describes the method of this study w.r.t. the experimental design including the kind of corporeal and verbal descriptions, background questionnaires, music, participants and study procedure. It also explains the workflow to analyze the data including pre-processing, feature extraction and the fitting of regression models for each GEMS state. Chapter 4 presents the results for each model and discusses the method. In Chapter 5, I conclude how well the approach could predict emotional perception of music based on corporeal articulation and suggest adjustments that could be applied to optimize predictions in order to proceed towards the use case of multimodal querying in the future. Chapter 5.3 comprises additional plots of the results.

[1] Following the *I-methodology* is to consider oneself and one's preferences as representative and norm of all other subjects or users [6, 3].

[2] The use of the first instead of the third pronoun is quite spread in *Gender Studies*. It aims at disclosing the influence of the researcher on the project and to admit a certain inherent degree of subjectivity because there is no such thing as a neutral place in the world from which one could observe or conclude objectively. As Haraway concludes[33, p. 583]: "The moral is simple: only partial perspective promises objective vision. All Western cultural narratives about objectivity are allegories of the ideologies governing the relations of what we call mind and body, distance and responsibility. Feminist objectivity is about limited location and situated knowledge, not about transcendence and splitting of subject and object."

Chapter 2

Background

My body is a cage
that keeps me
from dancing with the one I love,
but my mind holds the key.

My body is a cage by Arcade Fire (from *Neon Bible* released 2007).

In the following chapters I want to summarize the fundamentals underlying this thesis starting with the concept of *embodiment* in Section 2.1. As it is quite contested and vague among musicologists what an emotion is, Section 2.2 will cover some of the complexity of the term *emotion* and it will clarify the kind of *emotion* this thesis is dedicated to. In Section 2.3 related work studying musical emotion will be presented and the two most widespread models (circumplex and GEMS) to verbalize musical emotion are compared. Section 2.4 presents approaches of related work to measure corporeal expression of emotion. Last but not least, Section 2.5 points out aspects concerning the compilation of a socio-biographical questionnaire with which the group of participants will be characterized.

2.1 The Burden of Descartes and the Concept of Embodiment

“Cogito ergo sum” - “I think, therefore I am” is one of Descartes most famous and recited sentences. This understanding of an immaterialized spirit lead to mind-body dualism and resides amongst a number of related dualisms like culture-nature, ratio-emotio, logos-intuitio, objective-subjective or male-female as discussed in Lenz [48]. These dualisms do not only span a domain but they even establish a dichotomic way of classifying subjects into two segregated spaces. One must either be male or female, a study is either objective or subjective, (real) nature cannot be in the (cultured space of a) city. Furthermore, for a subject being rational also implies and prerequisites being objective, being cultured and being male.

Considering music and the mind-body dichotomy Pelinski [61, p. 3] summarizes that "[...] listening to music were seen as disembodied activities obviously controlled by such superior instances as spirit, soul, or (if possible) pure reason". Peter Röbke [63] exemplifies how disembodiment also affected the pedagogical concepts of piano lessons in the 19th century: The body, especially the one of young females, had been drilled in order to move as less as possible. Not having a body, emotions or (*female*) gender seemed to be the only option for women when they wanted to also be musicians.

The paradigm of disembodiment also lead to disrespect of Other[1] music like rather rhythmic one. Susan McClary and Robert Walser [55] analyzed how researchers like Roman Ingarden shaped a definition of music that does not consider (African-based) music as music when it aims at animating the body in dance. They argue that "the mind and culture still remain the exclusive property of Eurocentric discourse, while the dancing body is romanticized as what is left over when the burdens of reason and civilization have been flung away."

Over the last few decades there has been a paradigm shift from the concept of disembodiment towards a model of embodiment assuming a unified body [15]. In "Descartes' Error: Emotion, Reason, and the Human Brain" Damásio [16, p. 252] argues that "the comprehensive understanding of the human mind requires an organismic perspective; that not only must the mind move from a nonphysical cogitum to the realm of biological tissue, but it must also be related to a whole organism possessed of integrated body proper and brain and fully interactive with a physical and social environment". Damásio is a neuroscientist who observed empirically that patients who suffered from brain damages in the subcortex, concerned with the processing of emotions, also had serious problems in rational decision making. Originally assuming emotions were rather obstacles towards rationality he discovered that the opposite holds: Emotions are indispensable in the process of rational decision making.

Among other related fields of research in medicine is psychosomatic medicine [4] that explores how psychological, social or behavioral factors influence health and the research of the stomach as "second brain" [29]. All these examples assume the body to be a network that is making a common effort towards thinking and that also the thinking constitutes to the experience or well-being of the body.

Whereas Descartes deeply impacted Western thinking in mind-body dichotomy, the East

[1] Othering means to distinguish between the self and the other and hence to identifying oneself in opposition to the other [38]. Here, Jensen also cites Simone de Beauvoir ("He is the Subject, he is the Absolute – she is the Other.") and explains how othering works to describe women as inferior deviation from the male norm. Furthermore, McClary concludes that "[...] the Other need not always be interpreted strictly as female – it can be anything that stands as an obstacle or threat to identity and that must, consequently, be purged or brought under submission for the sake of narrative closure." [54, p. 16]

seems to be ahead of our time already: “From a Vietnamese perspective, then, feeling involves thinking and thinking involves feeling, and the body is implicated in the expression of both thought and emotion” [59, p. 30]. Hence, there is no disembodied thinking and only by applying embodied approaches can we capture the whole picture.

2.2 What a feeling - Affect, Emotion and other Thoughts

If one wants to predict *emotion* from *motion*, it needs to be specified what an *emotion* is and which kind of *emotion* will be predicted by *motion*. First of all, there is no consensus about what an emotion is and how or even if emotions are elicited by music [42]. Confusion also emerges from using the term *emotion* as an umbrella term covering all kinds of differentiations like *affect*, *(musical) emotion*, *mood*, *feeling*, *preferences*, *cognitive appraisal* etc., see Table 2.1 [42, p. 561].

Juslin and Laukka [40] summarize the different competing approaches of gripping *emotions* into *categorical* [24], *dimensional*[64], *prototypical* [67] or *component-based* [65] and stress the necessity to distinguish between *push* and *pull* effects. Whereas *push* effects are spontaneous, non-reflected physiological reactions related to affect like respiration or muscle tension, “pull effects, on the other hand, involve external conditions, such as social norms, that may lead to strategic posing of emotional expression for manipulative purposes” [40, p. 774] as exemplified in [45].

From a perspective of sociology Scherke [66, p. 31] differentiates between five dimensions of (musical) emotion that are closely linked to the terms from Table 2.1:

1. the physiological dimension comprising all physiological processes like heart rate or brain stem reflexes
2. the dimension of expression, i.e. the staging of emotions e.g. facial expressions
3. the dimension of experience comprises the experience of emotion and its verbalization as well as its regulation according to certain social constraints
4. the dimension of appraisal
5. the dimension of action tendency

The Master Thesis is settled in dimension (3), the dimension of experience. One could argue that dance also is an expression. However, expression in this context is rather something reflex and participants will be as active and conscious about their movements as about their verbalizations. As (3) is most relevant, using the term *emotion* in this

Affect	An umbrella term that covers all evaluative – or valenced (i.e., positive/negative) – states such as emotion, mood, and preference.
Emotions	Relatively intense affective responses that usually involve a number of sub-components – subjective feeling, physiological arousal, expression, action tendency, and regulation – which are more or less synchronized. Emotions focus on specific objects, and last minutes to a few hours.
Musical Emotions	A short term for "emotions that are induced by music."
Moods	Affective states that feature a lower felt intensity than emotions, that do not have a clear object, and that last much longer than emotions (several hours to days).
Feeling	The subjective experience of emotion (or mood). This component is commonly measured via self-report and reflects any or all of the other emotion components.
Arousal	Activation of the autonomic nervous system (ANS). Physiological arousal is one of the components of an emotional response but can also occur in the absence of emotions (e.g., during exercise).
Preferences	Long-term evaluations of objects or persons with a low intensity (e.g., liking of a specific music style).
Emotion Induction	All instances where music evokes an emotion in a listener, regardless of the nature of the process that evoked the emotion.
Emotion Perception	All instances where a listener perceives or recognizes expressed emotions in music (e.g., a sad expression), without necessarily feeling an emotion.
Cognitive Appraisal	An individual's subjective evaluation of an object or event on a number of dimensions in relation to the goals, motives, needs, and values of the individual.

Table 2.1: Working definitions of affective terms proposed by Juslin and Västfjäll [42, p. 561].

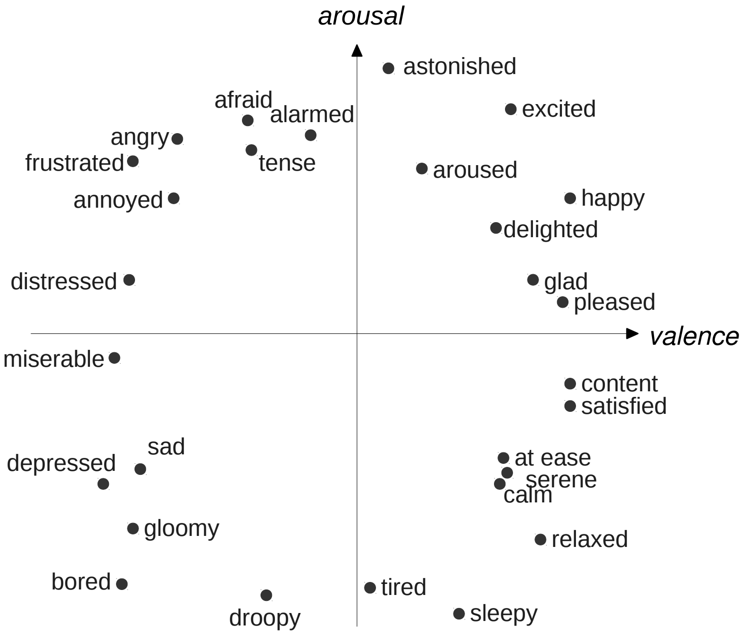


Figure 2.1: This figure shows the *dimensional* circumplex model from Russell [64]. Affective terms are mapped onto values of *valence* and *arousal*.

project's context refers to subjective *feeling*. These subjective feelings are influenced by other components like *mood*, *(musical) preferences* or *cognitive appraisal*. That is why these other components also need to be assessed by a background check in order to evaluate the interdependencies (cf. Section 3.3 and Section 3.4). Here, I also want to mention that although relevant, *push* effects will not be pursued because they are more related to *affect* and not *feeling*.
The dimensions of Scherke describe the potential levels of emotional perception and processing but they do not explain yet how emotions are induced by events like listening to music. Juslin et al. [42] suggest six mechanisms triggering emotions during musical listening experience stressing that "there is no single mechanism that can account for all instances of musically induced emotion" [42, p. 563]:

(1) Brain stem reflexes are activations of the brain corresponding to auditory stimuli. They reflect how the brain responds to sound on a low-level that is rather related to affects. "Brain stem reflexes can explain the stimulating and relaxing effects of music, and how mere sounds may induce pleasantness and unpleasantness. However, it is unclear how the mechanisms could explain the induction of specific emotions" [42, p. 564].

(2) Evaluative Conditioning in this context is the repeatedly pairing of a special kind or piece of music with other positive or negative stimuli. After the conditioning phase the effects of the positive or negative stimuli will also be triggered when only the neutral paired stimulus is presented, i.e. the piece of music. A famous movie showing an example for evaluative conditioning is "Clockwork Orange" from Stanley Kubrick. In this movie the protagonist was given drugs causing nausea without his knowledge. When the drugs started making him feel bad, music from Beethoven was being played amongst others. After some repetitions he also felt sick when only listening to Beethoven and without drugs being injected.

(4) Emotional Contagion is described as the "mimicking" of emotions expressed by music. This effect contributes to the ability to being able to perceive the emotional quality that is expressed by music. It does not necessarily mean that the person feels the expressed emotion. Emotional contagion is rather related to pushed affect and basic emotions and might hence be considered a bottom-up process. When more complex feelings are involved empathy that is considered a rather pulled top-down process, plays a very important role. Egermann and McAdams [22] investigated how identifying with a person and feeling empathy with the artist intensifies the degree to which the listening person also feels the expressed emotion.

(5) Visual imagery is the process of imagining pictures from places e.g., while listening to music. Sometimes visual imagery is intertwined with episodic memory but "[...] visual imagery is more strongly influenced or shaped by the unfolding structure of the music than is episodic memory [...]" [42, p. 567]. The importance of this mechanism is also pointed out when people talk about the "feeling of a place". As Magowan [53, p. 72] concludes about Australian Aboriginal rituals: "[...] the power and feeling of Country is elicited in references to this anatomy of the environment. Its sounds, movements, and colors evoke love, affection, desire, longing, fear, danger, or concern for those who know how to read its natural dimensions as emotionally and spiritually affecting."

(6) Episodic Memory is a phenomenon that links music to events or periods in the past during which we listened to it. Therefore music might also trigger a whole attitude to life from another phase in one's life. In addition, the triggered memory of a feeling might cause a whole set of further dissenting feelings like regretting that it was over.

(7) Musical Expectancy is closely linked to musical listening habits and preferences. Each musical genre usually obeys to certain syntactical rules about how the music needs to be structured. When listeners get used to these structures they might feel unpleasant about the structure being violated or be pleased when expectancies are fulfilled.

There are also distinctions being made between *perceived* and *felt* emotions that do not have to coincide [77]. As explicated below musical emotions can be very complex and multilayered. Music might be perceived as *happy* without having to induce this particular feeling on the listener. As this project is settled in the field of *Music Information Retrieval*, I will focus on a rather semantic and aesthetic description of the perceived (emotional) quality of music (cf. Section 3.7).

Due to the complexity of the mechanisms and their interdependencies Juslin et al. [42, p. 571] call for field studies: "[...] because if there are several mechanisms that can induce musical emotions, and their importance varies depending on the situation, only by sampling a wide variety of situations can we hope to capture all the mechanisms". In order to gain more insights about how music induces emotions Juslin et al. also point out the importance of carefully investigating which mechanisms have been involved in the particular situation.

In general, there are explicit and implicit rules about how to behave in order to conform to social norms in certain categories like *gender*, *milieu*, *culture* or *zeitgeist* as Butler [11, p. 521] puts it: "[...] the body is always an embodying of possibilities both conditioned and circumscribed by historical convention. In other words, the body is a historical situation, as Beauvoir has claimed, and is a manner of doing, dramatizing, and reproducing a historical situation." This is particularly important when it comes to *emotions* that are highly normalized as Degele et al. point out [18, p. 86]: "Das Normale und das Ideale stellen positives Erleben in Aussicht, und zwar dann, wenn die

Beteiligten eine Kongruenz zwischen sozialen Normalitätserwartungen und eigenem Empfinden herstellen können." Thus, according to Degele et al. *having emotions* is not a phenomenon independent of social processes, but it is highly influenced by social norms about how to feel when having a certain *age*, *gender*, *body* or *status*.
As mentioned above Juslin et al. [42] call for field studies in order to research the complex interplay of mechanisms that induce emotions during the listening of music. There is one more reason that commends an alternative approach to the collection of data: During controlled situations in a lab usually some artifacts occur because participants do not behave the way they usually do. Zentner et al. [77] address this phenomenon as *self-presentation bias*. As described above, the possibilities of how one can move or feel obey to various norms like *gender*, *status* or *zeitgeist*. When being in a supervised setting that is even being recorded, participants might behave *super-conform* to contemporary norms in order to appear as *normal* as possible.

In this section I gave an interdisciplinary overview of the term *emotion* and introduced the mechanisms involved in eliciting *musical emotion* according to Juslin et al. [42]. The perception of *emotions* is a complex process that also seems to always prerequisite a domain-specific access. And *emotions* are not a constant force of nature but also underly social norms and processes especially when it comes to feelings. That is why this chapter is introduced with lyrics from Arcade Fire: When they are singing "my body is a cage that keeps me from dancing with the one I love...", they are just referring to our emotions as embodiments of historical situations.

2.3 Verbal Articulation of Musical Emotion

This section presents related work in the field of "Music and Emotion" that is based on two of the most widespread models to describe the emotional qualities of music listening experience: The circumplex model [64] and the *Geneva Emotion Music Scales (GEMS)*. It also motivates the choice for the GEMS in the context of this project.

The circumplex model maps various emotions into a two-dimensional space (Figure 2.1) that is spanned by *valence* ("how pleasant or unpleasant is the experience?") and *arousal* ("how intense is the experience?").

Fernando et al. [21] compared psychophysiological responses to music of Canadians and Congolese Pygmies. They concluded that results "suggest that while the dimension of emotional *valence* might be mediated by cultural learning, changes in *arousal* might involve a more basic, universal response to low-level acoustical characteristics of music" [21, p. 1].

A similar study from Gomez and Danuser [31] investigated the "Relationships Between Musical Structure and Psychophysiological Measures of Emotion". They correlated musical features like *tempo*, *pitch* or *mode* to both physiological measures like *respiration* or *heart rate*, and self-reports of *valence* and *arousal*. Their findings have been concluded as follows: "Mode, harmonic complexity, and rhythmic articulation best differentiated between negative and positive valence, whereas tempo, accentuation, and rhythmic articulation best discriminated high arousal from low arousal" [31, p. 377].

Concerning *Musical Expectancy* (cf. Section 2.2) Egermann et al. [23] observed how *expectation violation* in a live concert setting causes psychophysiological emotional responses. They compared self-report measures of *valence* and *arousal*, as well as the level of *unexpectedness* to physiological responses like skin conductance or heart rate. Their results show that "musical structures leading to expectation reactions were manifested in emotional reactions at different emotion component levels (increases in subjective arousal and autonomic nervous system activations)" [23, p. 533].

In the field of *Affective Computing* participants of the MultimediaEval Benchmark's Task[2] "Emotion in Music" are working on finding appropriate physical features to describe music in order to distinguish between different degrees of *valence* and *arousal* [70] that have been assessed by self-reports of a *crowd*.

The most related work to this topic is probably the one from Amelynck et al. [5] who envision a use case in which music is queried by the movement of a mobile device. For "Toward E-Motion-Based Music Retrieval – A Study of Affective Gesture Recognition" they make use of acceleration sensor data generated by arm gestures while holding a wii remote control. The approach generates fairly good predictions for the dimension of *arousal* but performs poorer for the dimension of *valence*. They argue that this might be due to people also rating the perception of *sad* music as pleasant and conclude that the circumplex model might be unsuitable to be used with musical emotions. Indeed, *sad* music might also be perceived as pleasant as studied in [28]. That is why Amelynck et al. suggest the GEMS as alternative to the circumplex model. I also want to refer to the argumentation of Juslin et al. [42, p. 572]: "In any case the existence of mixed emotions speaks against using the *circumplex model* (Russell 1980, [64]) to study musical emotions, since it precludes feeling both sad and happy at the same time (Larsen et al. 2001, [46])".
The GEMS have been iteratively developed and evluated during four studies by Zentner, Grandjean and Scherer [78]. Studies were conducted in both lab and field during a live concert. The aim was to find a "more nuanced affect vocabulary and taxonomy than is provided by current scales and models of emotion" [78, p. 513], the GEMS. The original version of the GEMS comprises 45 terms amongst *feeling of transcendence*, *nostalgic*, *solemn* or *impatient* that are not part of other emotional models like *basic emotions*.

Torres-Eliard et al. [71] compared self-report measures of the GEMS from different

[2] www.multimediaeval.org

participants and suggest the GEMS as a more suitable model to assess musical emotion. They concluded that "the results indicate a high reliability between listeners for different musical excerpts and for different contexts of listening including concerts, i.e. a social context, and laboratory experiments." [71, p. 252]. A very similar study has been conducted by Lykartsis et al. [52] . In "The Emotionality of Sonic Events: Testing the Geneva Emotional Music Scales (GEMS) for Popular and Electroacoustic Music" they tested a German translation of the GEMS and compared the results of self-report measurements to those of the original version in English language. Their findings showed that also in German the "contextual meaning of the construct remains constant across different musical genres with a reasonable fit".

Even though GEMS highly diversify the term *emotion*, there is still one issue with this model: The first two studies that define the terms and thus the emotional vocabulary, are based on the participation of undergraduate psychology students from the University of Geneva. Thus, participants belong to the WEIRDest people in the world where WEIRD stands for Western Educated Industrialized Rich Democratic [36]. Henrich et al. question the representativity of findings based on university students and conclude that they "are among the least representative populations one could find for generalizing about humans".

Furthermore, participation has been obligatory for the students in order to finish their studies. Hence, participation has not been on a voluntary basis. These circumstances might have lead to some artifacts like compliant behavior and a bias towards psychological terms in the GEMS model. Findings might be different, even if other WEIRD people would have been participating. It is also not clear how the mechanism of visual imagery and the concept of Country in Aboriginal Culture (cf. Section 2.2) could be linked to the GEMS.

2.4 Corporeal Articulation of Musical Emotion

This section will introduce related work that aims at linking corporeal and verbal articulations of emotion. It will also give an overview over possibilities to measure embodiments of emotion. The importance of corporeal articulation to experiencing emotions is also highlighted by a study from Longhi [50] who investigated the role of multimodal interactions (auditory, visual, tactile, kinaesthetic) between mothers and infants for the child's emotional responses to music. 3-month-old infants showed much more emotional responses to music when mothers have been asked to also touch them when singing. This study among others highlights the importance of the body for the development of emotional responses to music. It also exemplifies that the body is describing (musical) experiences long before we are able to verbalize them.

In the field of *Affective Computing* Kapur et al. [43] aim at detecting emotions based on gestures generated by a motion capture system. They discriminate between *sadness*,

anger, *fear* and *joy* with a chance ranging from 84% - 92%. The most famous and contested application in affective computing is probably the one from *Google Glass* that uses facial recognition to detect the emotions *anger*, *joy*, *sadness* and *surprise* [3].
In musicology, there are a lot of possibilities to express musical listening experience in a corporeal way. Among them are tapping or moving parts of the body along with the beat, singing, imitating to play a musical instrument or dancing. Hedder also evaluated an approach based on facial expression as form of embodiment [34]. Drawing as described in De Bruyn [17] and Leman [47, p. 119-120] is another alternative as a means of graphical attuning to the experience.
Giordano et al. [30] studied the relationship between motion and emotion and its implications on the expression of musical performances. The emotions studied are *sadness*, *anger*, *fear* and *happiness*. Their findings "support the motor-origin hypothesis of musical emotion expression that states that musicians and listeners make use of general movement knowledge when expressing and recognizing emotions in music" [30, p. 29].
Another study investigating the connection between music, motion and emotion is from Sievers et al. who concluded that "music and movement share a dynamic structure that supports universal expressions of emotion" [68, p. 70]. Participants were asked to adjust the features rate, jitter, consonance/smoothness, step size and direction for each of the following five emotions: *angry*, *happy*, *peaceful*, *sad* and *scared*. For one group the adjustment of features lead to different movement and appearance of a bouncing ball. For the second group adjusting features changed the melody and expression of a piano piece. Experiments were both conducted in the U.S.A. and Cambodia.
As already mentioned in Section 2.3 the most related work to this topic is the one from Amelynck et al. [5] who make use of acceleration sensor data generated by arm gestures while holding a wii remote control in order to describe the emotional qualities of music.

2.5 Socio-biographical Questionnaires and the *Sex* Question

In order to characterize participants domain-specific socio-biographical questionnaires need to be designed. In the context of research on emotion, Degele et al. [18, p. 86] proposed the following factors: *milieu*, *age*, *gender*, *sexual orientation* and *biography of relationship/s*. In *Gender Studies*, it is also rather contested how to ask about *gender/ sex*. Döring points out that the binary distinction between *female* and *male* violates the statistical requirements for uniqueness, exclusivity and exhaustivity [19]. Unfortunately, there is no standard remedy yet to solve this problem. Another issue is the interdependence between *gender* and *sex* [12] that states that *sex* and *gender* are mutually influencing one another and that asking for only one of them will withhold

[3] http://www.iis.fraunhofer.de/en/ff/bsy/tech/bildanalyse/shore-gesichtsdetektion.html

important information. Additionally in Germany, there is only one word for *gender* and *sex*: *Geschlecht*. That is why, if no further specified, it is left to the interviewed person if ze[4] replies to parts of zer body or to zer *gender* identity.

[4] Ze is a gender neutral pronoun. It is introduced to account for trans* identities.

Chapter 3

Method

Having troubles telling how I feel
But I can dance, dance, dance
Couldn't possibly tell you how I mean
But I can dance, dance, dance

Oh, dance
I was a dancer all along
Dance, dance, dance
Words can never make up for what you do

Dance, Dance, Dance by Lykke Li (from *Youth Novels* released 2008)

The lyrics from Lykke Li exemplify the need for corporeal possibilities of describing an experience. Therefore, it motivates the aim of this study to search for embodied descriptions of musical listening experience that might be linked to verbal equivalents. In this chapter, I want to present the methodological approach of this study to achieve this goal. Section 3.1 motivates the choice for mobile devices as instruments on which the study is conducted. Section 3.2 explains the corporeal articulation and Section 3.3 explains the verbal model to describe the music with. There is a variety of background information that refers to sample, setting and the biography of the participant that also needs to be assessed (Section 3.4) in order to determine the scope of the results. One of the most important aspects, the music that is being used as mediator and stimulus, will be introduced in Section 3.5. The cast of participants is described in Section 3.6 and Section 3.7 will lay out the overall procedure and design of the study. It will furthermore illustrate how the APP-based instrument as main part of the experiment is going to look like and how data is processed and transferred from the local device on which the APP will be running to the central database server. Last but not least, Section

3.8 will describe how data is analyzed in terms of pre-processing, feature extraction and selection, as well as how the models are fit to predict GEMS states to different degrees.

3.1 Into the Fields – The Instrument

During the study, I want to minimize the effect of self-presentation and role playing, i.e. the performance of being 'normal' as discussed in Section 2.2, by an experimental setting in which the participant is not being watched and emotions are assessed anonymously. This is where the smartphones come in handy as they enable participants to conduct the experiment when and where they feel comfortable to do so. Smartphones are devices that are spread widely among the Western population: 50% of Germans are using smartphones; amongst the 14-29-year-olds even 78% are using smartphones[1]. Therefore it is a device that comes with some advantages compared to common instruments in the field [44]. First of all participants do not have to come to a lab in order to express their experience using non-off-shelve devices. It hence also enables us to collect data in the field, i.e. in the naturalistic and real environment of the participants. Secondly it is easier to get samples from a broader audience in the long term as devices do not have to be purchased for each participant; or it is less time-consuming when participants need to use the same instrument in the lab one after another. Last but not least it is a device that participants are already familiar with. This could lead to a more intuitive handling and enable the participants to focus on the experience.

As all 'field' participants except for one owned a device running on the Google's Android[2] operating system it was sensible to develop for Android. This OS is particularly popular in Germany where three fourths of mobile users opt for Google's operating system [3].

Alternative devices could be wristbands or anklets as described in Janssen[37] that are originally used in the field of the "Quantified Self". They would have the further advantage that participants did not have to hold a device. The drawbacks are that sensors of these devices are not as sophisticated as those of smartphones and that they are not as spread as smartphones.

This section motivated my choice for conducting the experiment on mobile devices running Android as they are widely spread among the Western population and therefore allow the realization of the study in the field. Besides, with regard to the use case of multimodal querying as discussed in Section 1.1, it accounts for an increasing tendency to listen to music using smartphones.

[1] http://www.bitkom.org/de/presse/81149_79598.aspx

[2] https://www.android.com/

[3] http://de.statista.com/infografik/1097/marktanteile-der-smartphone-betriebssysteme/

3.2 “...in any case without words!” - The Corporeal Articulation

During pre-tests of two methods, i.e. drawing lines according to DeBruyn (cf. Section 2.4) or moving devices like in Amelynck et al. [5], it turned out that participants felt more comfortable with moving a smartphone than with drawing lines. They argued it offered them more differentiated means of expression and it was more intuitive to move along with the music. Furthermore, the screen on a smartphone is very limited to draw lines on and hence, participants had to make rather small movements in order to being able to differentiate between different levels of intensity. They also reported it felt like having to artificially contain themselves when listening to more energetic music like Metal*. Besides, in settings similar to drawing lines, it is usually difficult to express the experience adequately if you do not know previously what kind of music needs to be differentiated. This also holds for methods like EMujoy [57] when you move a joystick in a rather restricted space. Participants do not know if there will be an even more arousing musical excerpt and rather tend to stay “in the middle” or at “the extremes”, i.e. it is difficult to use the available space appropriately.

When moving a device, movements do not have to be as fine and controlled as with drawing lines. Participants can usually use the whole space around them and even jump if they feel like it. That is why making use of the available motion sensors in smartphones seemed to be more suitable than drawing lines.

There are three kinds of motion sensors that are built in smartphone devices: Accelerometer, Gyroscope and Magnetometer [4]. Accelerometer measures the device’s acceleration in all three dimensions. Gyroscopic sensors measure changes in orientation. They are not built in older devices. Last but not least, the magnetometer measures gravity and might hence be used to determine the absolute position in world space of the device. In order to track motions one would ideally have to fuse all three sensors. As this is far beyond the scope of this thesis, I went with acceleration that is most strongly linked to motion (see Figure 3.1). Besides, acceleration sensors are available in all devices and are well supported by the Android Software Development Kit (SDK). Figure 3.2 illustrates the meaning of acceleration data in 3D space.

In this section, I explained that participants chose free movement of a smartphone over drawing lines as corporeal articulation as it was a more intuitive description. Furthermore, I also gave reason for why to employ the acceleration sensor data of the devices: As of all sensors, they are most related to movement.

[4] http://www.uni-weimar.de/medien/wiki/images/Zeitmaschinen-smartphonesensors.pdf

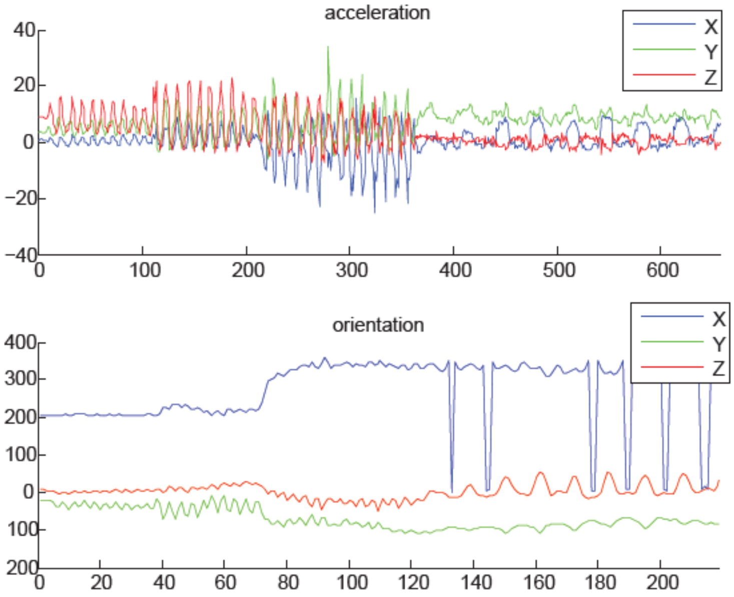


Figure 3.1: These graphs show the acceleration and orientation data from basic hand shake movements with different sizes of movement (0 - ∼360) and speed (∼ 360 – end shows a slow movement of same size as in between (∼ 210 - ∼360)). Acceleration contains more information about the overall movement.

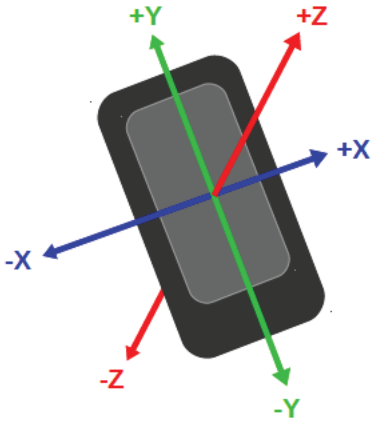


Figure 3.2: A sketch illustrating the meaning of acceleration values in 3D space.

3.3 “...and if we had to find words?” - The Verbalization

Even though there might not always be a verbal equivalent for an embodied description, it is difficult to study the embodiment of musical emotion without reference to existing verbal models. Therefore, I chose the model from Zentner, Granjean and Scherer [78], the *Geneva Emotion Music Scales* in the compressed version of the GEMS-9 as described and for the reasons discussed in Section 2.3. The GEMS-9 are a lower-dimensional version of the Geneva Emotion Music Scales (GEMS). Although Zentner et al. recommend to ideally use the full range of 45 scales, they offer this version for which scales are grouped into 9 categories in order to provide a model that is more suitable for experiments in which participants do not have a lot of time to make decisions. Besides the issue to reduce the workload for participants in this project, the GEMS-9 are also more suitable because GEMS-25 or GEMS-45 would not have fit on a smartphone’s or tablet’s display. Participants would have had to scroll and would never have gotten the full picture. The GEMS-9 and their German translations in this experimental setting are as shown in Table 3.1.

In preparation of this project, I asked participants to talk about how they describe music. Although terms from the GEMS like *melancholic* or *happy* have also been part of the vocabulary, all interviewees also mentioned places and linked the feeling of places to music. To account for these deviating verbal descriptions and to also test the suitability of the GEMS for the purpose, I wanted to add two other options to express the experience on a verbal basis. The following verbal descriptions will be available to the participants:

> (1) a completely open text field in which participants can write creatively about the music, how the music is linked to events of their past or what they would like to do when listening to this music.
>
> (2) a semi-standardized option that offers participants to choose tags from a folksonomy[5] to describe the music. They are also allowed to add their own tags to the folksonomy in order to make it grow.
>
> (3) standardized scales based on the GEMS-9.

As this is a rather qualitative approach to experience, the most open option will be first, i.e. (1), because closed questions in the run-up might limit the openness of the open questions [35]. Although, participants will know about the other options after one sample and thus, for following songs (2) and (3) will be influencing the open answer, I

[5] Folksonomies are user-created vocabularies, sometimes also referred to as ‘tag clouds’. They help users to develop a common understanding of a subject by creating a vocabulary that assemble generality without loosing the specificity of a subject. Generality is usually implemented by assigning bigger font size to tags that are used more often and hence serve as a kind of super-category.

wonder	**Verwunderung**
filled with wonder, dazzled, allured, moved	voller Verwunderung, geblendet, verlockend, bewegt
transcendence	**Transzendenz**
fascinated, overwhelmed, feelings of transcendence and spirituality	fasziniert, überwältigt, Gefühl von Transzendenz und Spiritualität
power	**Energie**
strong, triumphant, energetic, fiery	stark, triumphierend, energetisch, glühend
tenderness	**Zärtlichkeit**
tender, affectionate, in love, mellowed	liebevoll, zärtlich, verliebt, weich
nostalgia	**Nostalgie**
nostalgic, dreamy, sentimental, melancholic	nostalgisch, verträumt, sentimental, melancholisch
peacefulness	**Friedlichkeit**
serene, calm, soothed, relaxed	still, ruhig, sanft, entspannt
joyful activation	**Freude**
joyful, amused, animated, bouncy	freudig, vergnügt, lebhaft, hüpfend
sadness	**Traurigkeit**
sad, sorrowful	traurig, sorgenvoll
tension	**Anspannung**
tense, agitated, nervous, irritated	angespannt, aufgeregt, nervös, irritiert

Table 3.1: This table shows the GEMS-9 and its German translation.

still wanted to keep the order this way because it gives more weight to the openness and makes it less likely to just skip it at the end. The open format also gives us the chance to learn a little bit more about the mechanisms triggering a particular feeling in the field as participants are also asked to write a little bit about their associations with the music.

In this section, I introduced the three verbal options to describe the music: An open text field, a folksonomy and GEMS-scales.

3.4 "Are you WEIRD, too?" - Background Questionnaires

There will be three questionnaires in this study related to the musical excerpt (Sample Questionnaire), the experimental Setting (Setting Questionnaire) and participant (Sociobiographical Questionnaire). While the first two questionnaires will be integrated into the instrument, the last one will be handed out in paper subsequent to the experiment.

3.4.1 Sample Questionnaire

During the experiment after each sample the participant is asked how suitable it was to describe the experience by motion and by words. Besides important factors for the evaluation are also how much the person liked the sample of music and how well the music has already been known before[22].

3.4.2 Setting Questionnaire

As the study will be conducted in the field as well, there are additional effects that need to be assessed in order to determine their influence on the results. The following six questions will be posed at the end of the experiment:

1. Where did you conduct the experiment? (open)
2. Did you have company, e.g. friends, family? (open)
3. How comfortable did you feel in your surroundings during the experiment? (open)
4. How did you listen to the music, e.g. speakers of your device, headphone with(out) cable? (open)
5. Which device did you use, manufacturer and model? (open)

6. In which mood have you been during the experiment? (calm/ruhig, confident/zuversichtlich, bugged/genervt, bored/gelangweilt, happy/glücklich, sad/traurig, annoyed/verärgert, in love/verliebt, hungry/hungrig, frustrated/frustriert, lonely/einsam, depressed/deprimiert, impatient/ungeduldig, confused/durcheinander, insecure/unsicher, tired/müde, shy/schüchtern, irritated/irritiert)

3.4.3 Socio-biographical Questionnaire

After the experiment participants will be asked to fill out a one page paper questionnaire about their biographical background. The first part of the questionnaire will comprise questions that are particularly related to the experiment whereas the second part will consist of general questions about the person. All questions are in open format. Both since I did not want to assume anything to be 'normal' and because the amount of participants still allows to evaluate it openly.
The questionnaire first asks which kind of experiences the participant acquired in playing music (e.g. classes in school, music schools, choir, band). Second, the person is asked which genres had an impact on their listening preferences in the past. Third: "Music that you like sounds like...(name of artist or band)." The fourth and last question of the first part asks about their experience in movement, e.g. dancing class, club, eurythmy. Eurythmy is an art that aims at expressing language in an embodied way. It can also be considered a sign language that is performed with the whole body. For the second part it turned out to be rather difficult to find a compromise. On the one hand we had been worried about the workload and anonymity of the participants, on the other hand we wanted to assess the diversity of the participants in order to qualify the results. The designed questions refer to Degele et al. [18, p. 86] as discussed in Section 2.5. For 'milieu' we would have had to ask several questions like educational background of parents, income etc. That is why we did not assess it regarding the timely resources of the participants. The question for 'sexual orientation' seemed too personal, especially when dealing with 14-year old participants in a sensitive environment. Concerning 'gender' we aimed for a question that is as less suggestive as possible and ended with the following formulation: "Do you feel you belong to a 'Geschlecht'? If yes, to which one?" The other questions ask for their professional and non-professional activities, whether they parent children or care for relatives, if they are in partnership/s, their year of birth and if they regularly use a smartphone or tablet. All participants have been told that they don't have to answer questions they do not feel comfortable with or if they fear the answer could risk their anonymity. The German questions of the questionnaire are:

1. Wo haben Sie bisher Erfahrungen im Musizieren gesammelt, z.B. Schule, Musikschule, Band, Chor?

2. Welche Genres haben Ihre Hörgewohnheiten geprägt?

3. Musik, die Sie mögen, klingt wie...(Name der Künstler_innen oder Band)?

4. Welche Erfahrungen haben Sie bisher in Bezug auf Bewegung gemacht, z.B. Tanzkurs, Club, Eurythmie?

   ————————————————————————————-

5. Welcher/en Tätigkeit/en gehen Sie nach, z.B. Beruf, Ehrenamt, Schüler_in, Student_in (bitte Studiengang angeben)?

6. Übernehmen Sie derzeit Verantwortung für die Betreuung von Kindern oder pflegen Sie Angehörige?

7. Befinden Sie sich in Partnerschaft/en?

8. In welchem Jahr sind Sie geboren?

9. Nutzen Sie regelmäßig ein Smartphone oder Tablet?

10. Fühlen Sie sich einem Geschlecht zugehörig? Wenn ja, welchem?

In this section, I presented the three relevant questionnaires during this study that concern music, setting and participant. Although, most of the factors cannot be empirically evaluated in the scope of this thesis, they allow for further evaluations in the future and the formulation of new hypotheses.

3.5 “It is just a list of samples, right?” - The Music

The music has been selected in a participatory approach as well in order to make participants feel more comfortable in the setting and to reduce the gap between *perceived* and *felt* emotions as mentioned in Section 2.2. Therefore, I first collected song proposals from the group of ‘field’ participants and compiled a bigger playlist. From this list I selected songs according to the following criteria:

1. account for a variety of participants’ preferences

2. cover the range of the GEMS-9

3. keep the balance between ‘female’ and ‘male’ composers

4. do not let emotions be covered in a stereotypical way like ‘tenderness’ by ‘female’ artists or tension by ‘male’ artists

5. artists of different color

6. cover a variety of genres

Artist	**Title**	**Genre**
L7	Wargasm	Punk
Souad Massi	Raoui	Folk Rock
Adele	Rolling in the Deep	Pop
Bruno Mars	Count on Me	Pop
Air	La Femme d'Argent	Electronica
Two Fingers	Sweden	Grime
Bob Marley	Corner Stone	Reggae
Mel Bonis	Berceuse Op.23 No.1	Classic
Andy Allo	People Pleaser	Funk
David Bowie	Rebel Rebel	Rock 'n Roll

Table 3.2: This table shows the final list of samples.

Finally, 'field' participants had a vote over the songs that should be chosen from this smaller list. Participants were told to vote for five songs having high inter-heterogeneity but low intra-heterogeneity w.r.t. the GEMS-9.
Subsequently we agreed on having another five songs to have more data. Hence, I added two more songs from the participants' list: "Corner Stone" from Bob Marley (to have another Reggae song) and "Count on me" from Bruno Mars (to have one more song from the girl rock band). Additionally, I added "Berceuse" from Mélanie Hélène Bonis to have another classic and peaceful song, "People Pleaser" from Andy Allo to have a funky and happy song and "Rebel, Rebel" from David Bowie to have a Rock 'n Roll song that is both 'tender' and 'wild'. Table 3.2 shows the final list of musical excerpts that were of ~ 40s duration and chosen such that the excerpt was as homogeneous as possible w.r.t GEMS during this time. The list of samples was initialized in random order for each participant. They were then free to choose the preferred order in which to assess the samples.

This section explained the criteria and procedure for compiling the list of samples. Letting participants compile the list of samples is a first step towards having subjects assess their favorite music only.

3.6 Participants

Besides the girl rock band (from whom only one person participated in the final phase), there have been another 10 participants from my circle of acquaintances who participated in the field. As the collection of data in the field turned out to be very time-consuming, another part of the study was conducted in the lab to guarantee that the amount of data is enough. For this lab study, 22 participants have been recruited from the Master's program of Audio Communication and Technology to conduct the study in the MediaLab. They got credits for their studies for being part of the experiment. For a more detailed characterization of participants see the evaluation of the socio-biographical questionnaires in Section 4.1.5.

3.7 Procedure and Study Design - The Orange Tour

Participants in the field either had the APP installed on their own devices and could conduct the experiment anytime they wanted to or we scheduled an appointment during which they used the device I provided (Nexus 2013). After a test round the participating person would conduct the experiment in another room (door closed) all alone except for one participant from the girl rock band who conducted the experiment in the rehearsal room dancing and singing along with the other band mates.
81.8% used the provided tablet (Nexus 2013), another 9.1% each used their personal smartphones (Samsung Galaxy S3 Mini and Sony Xperia Z2). 91% of participants wore headphones, the other 9% listened to the music via the device's speakers. Some weeks later they were asked to fill out the socio-biographical questionnaire (cf. Chapter 3.4.3).

For the collection of data in the lab the APP has been modified in order to keep the duration of the experiment in the scope of an hour. Therefore, the open verbal description and the folksonomy have been removed. After a test round participants were asked to conduct the experiment in the MediaLab. The MediaLab had been illuminated only slightly and offered them enough space to move freely. They have been all alone in the lab during the experiment (door closed) but could come out to ask questions. This only occurred three times and only for the last activity when they had to enter additional information about the experimental setting. In the lab the experiment had been conducted on a Motorola Moto G with Android Version 4.4.4. Participants wore headphones with long cable from AKG. Subsequent to the experiment they were asked to fill out the one page paper socio-biographical questionnaire, see Section 3.4.3.

Figure 3.3 visualizes the general procedure of the APP (orange) and its communication with the local and central database (green and pink). After both 'field' and 'lab' participants were logged in, see Figure 3.4a, they were instructed to select a musical excerpts from a randomly initialized list in the order they preferred as you can see in Figure 3.4b. The order in which participants selected a song is remembered in Step 21. After one sample has been selected it will be disabled and highlighted in terms of color for the next round. Steps 4 to 21 are repeated until all samples have been annotated. Participants are prohibited to go back to a previous step in order to ensure the comparability of results among participants.

After a song has been selected participants are redirected to Step 7, the activity in which they are to express the listening experience in a corporeal way as shown in Figure 3.6. Here the first step is to listen to the song on trial in order to get into it. However, it is possible to abbreviate this listening phase for listeners who already know the song by pressing 'pause' or 'stop'. In the second phase the movements are actually recorded in synchrony to the music as visualized by Figure 3.5b. For this part of the study participants were instructed as follows: "Please move now with the device according to the music. It is important that you stand and don't sit during motion capturing. You can move freely, i.e. all parts of the body but keep in mind that only movement of the device can be captured".

When the sample finishes playing participants are automatically directed to Step 10, see Figure 3.6a. Here the sample might be described very openly. What kind of associations are elicited by the song? What did the person want to do while listening to the sample?

By pressing 'weiter' the person is directed to Step 14, the folksonomy, see Figure 3.6b. In this step, participants can 'tag' the sample. They might either choose from a list that former participants already created or add their own tags. The original tags are from a list of tags participants announced when asked for words to describe a piece of music. New tags are added to the database on the central server and will be visible for all users consequently.

After tagging the sample the person is asked to describe the song by using the GEMS-9 in Step 18, see Figure 3.8. For this step participants were instructed to "Please rate the perceived emotional quality of the music according to the GEMS-9. Do not rate how you *felt* during listening". For each of the GEMS-9 a unipolar slider is provided ranging form '0' to '100'. Its default position is set to '0'. Each dimension is also described by adjectives belonging to this category. In Step 21 participants are asked to provide some more information about how suitable they consider the corporeal and the verbal options to express the experience for this particular sample, see Figure 3.7. Besides they are asked if they knew the song already and how much they liked it.
When the last sample from the list has been assessed they continue with Step 24, else they are directed to Step 4 where they can pick another sample. Step 24 is the final activity in which participants should provide information about the setting in the field, cf. Section 3.4.2. Finally, in Step 26 participants are asked to hold on via a load screen while the data is being transferred to the central server. When transmission is complete, they can close the application.

3.7.1 Data Flow – The Green and Pink Tour

All data will finally be stored at a central MySQL-Server (pink). In order to prevent a loss of data in case of interrupted internet connections, the data is first cached on a local

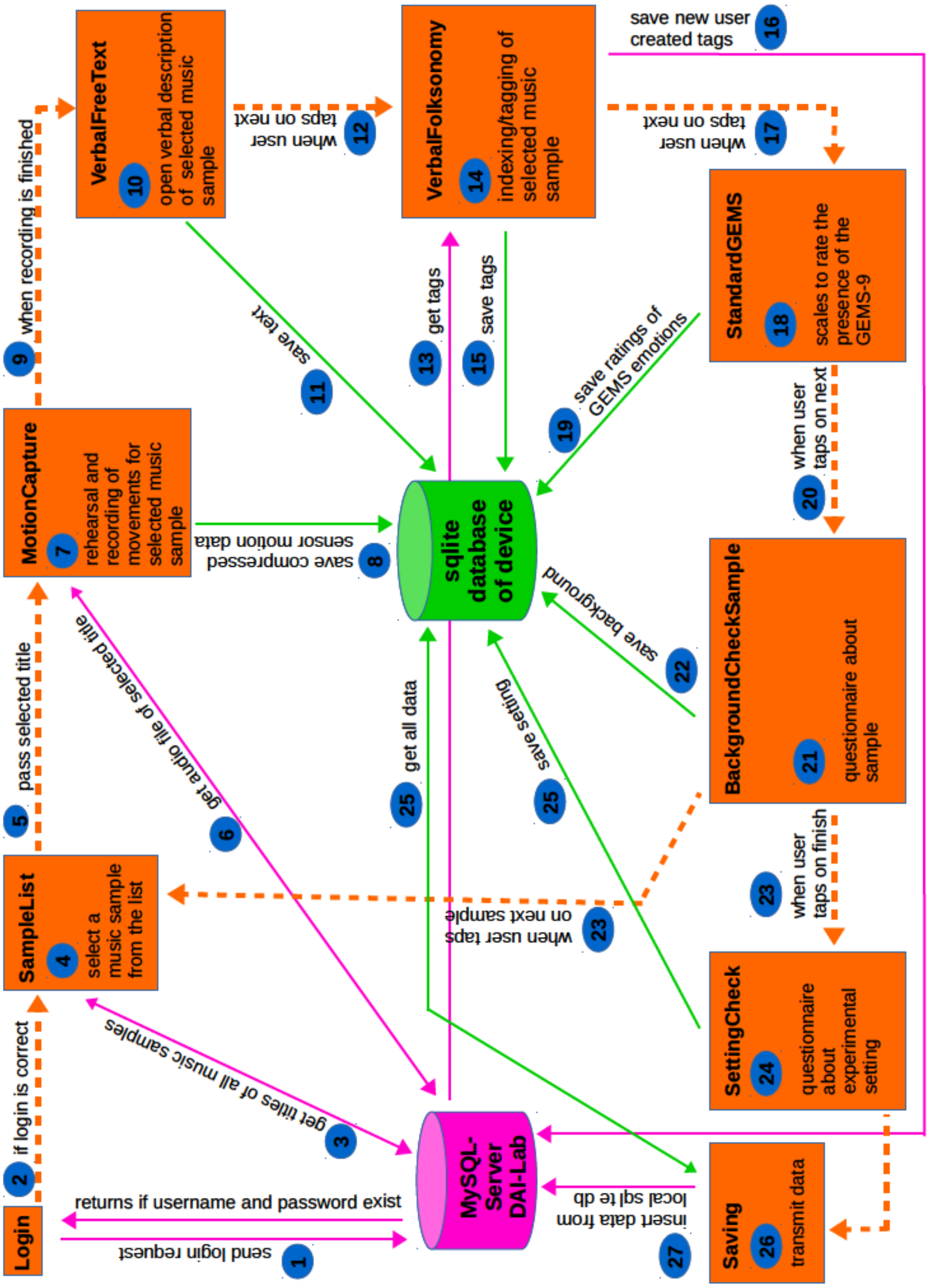
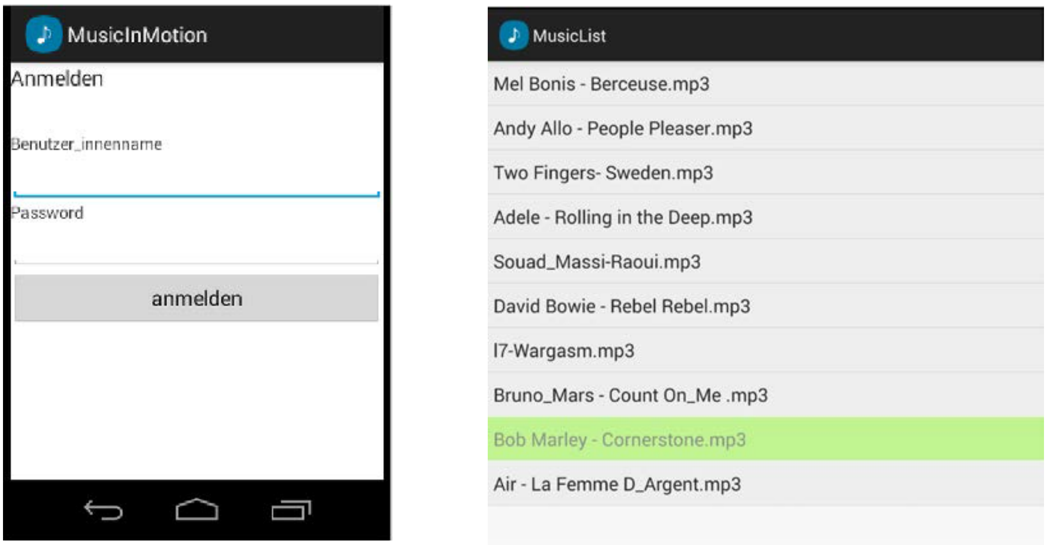

Login
1 send login request
returns if username and password exist
2 if login is correct
SampleList
4 select a music sample from the list
3 get titles of all music samples
5 pass selected title
MotionCapture
7 rehearsal and recording of movements for selected music sample
6 get audio file of selected title
8 save compressed sensor motion data
9 when recording is finished
VerbalFreeText
10 open verbal description of selected music sample
11 save text
12 when user taps on next
VerbalFolksonomy
14 indexing/tagging of selected music sample
13 get tags
15 save tags
16 save new user created tags
17 when user taps on next
StandardGEMS
18 scales to rate the presence of the GEMS-9
19 save ratings of GEMS emotions
20 when user taps on next
BackgroundCheckSample
21 questionnaire about sample
22 save background
23 when user taps on finish
23 when user taps on next sample
SettingCheck
24 questionnaire about experimental setting
25 save setting
25 get all data
Saving
26 transmit data
27 insert data from local sql te db
sqlite database of device
MySQL-Server DAI-Lab

Figure 3.3: This figure shows the APP flow and its communication with the databases.



Mel Bonis - Berceuse.mp3

Andy Allo - People Pleaser.mp3

Two Fingers- Sweden.mp3

Adele - Rolling in the Deep.mp3

Souad_Massi-Raoui.mp3

David Bowie - Rebel Rebel.mp3

I7-Wargasm.mp3

Bruno_Mars - Count On_Me .mp3

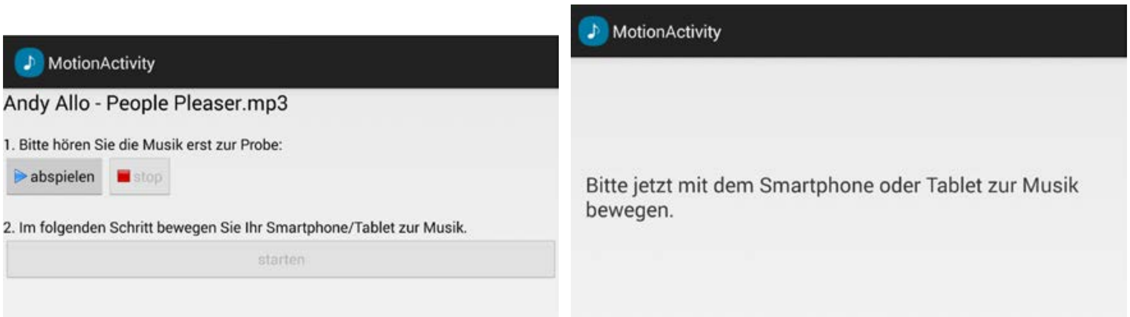
Bob Marley - Cornerstone.mp3

Air - La Femme D_Argent.mp3

(a) Login

(b) List of Samples

Figure 3.4: Figure 3.4a shows the Login-Activity. In Figure 3.4b the activity is shown that is active when the participating person successfully logged in: The list of samples. Samples that have already been annotated are highlighted in green.



Andy Allo - People Pleaser.mp3

1. Bitte hören Sie die Musik erst zur Probe:

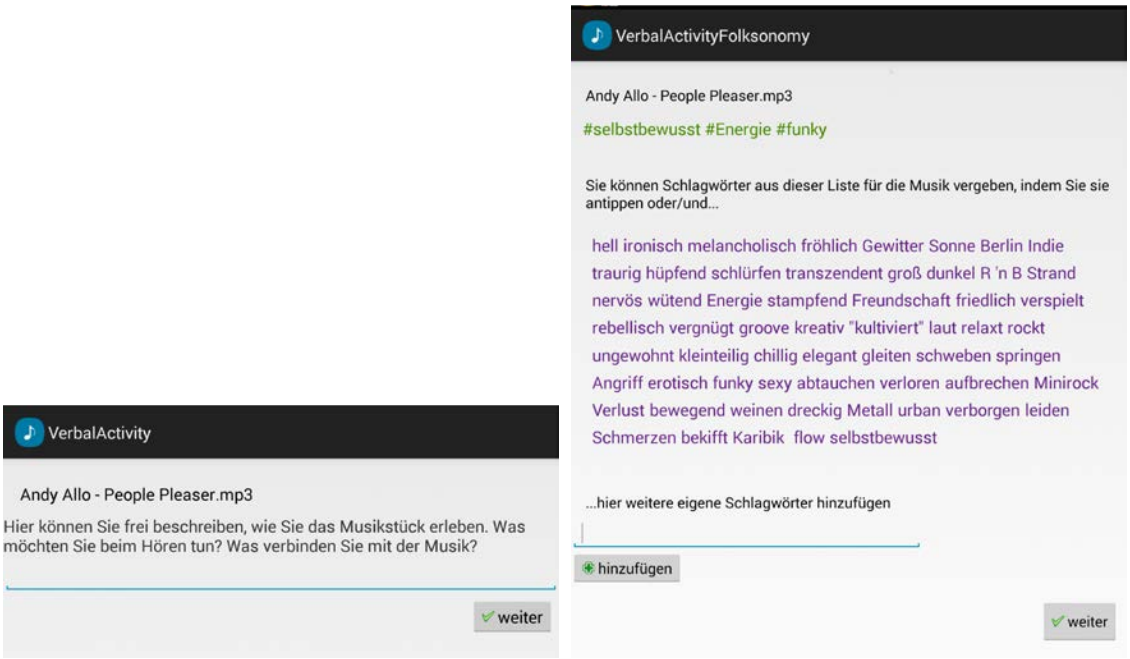
2. Im folgenden Schritt bewegen Sie Ihr Smartphone/Tablet zur Musik.

Bitte jetzt mit dem Smartphone oder Tablet zur Musik bewegen.

(a) Test Listening Phase

(b) Motion Capture View

Figure 3.5: Corporeal Expressions: Figure 3.5a shows the view for the test listening phase while Figure 3.5b displays the view when the capturing of corporeal articulations has been started.



Andy Allo - People Pleaser.mp3

Hier können Sie frei beschreiben, wie Sie das Musikstück erleben. Was möchten Sie beim Hören tun? Was verbinden Sie mit der Musik?

Andy Allo - People Pleaser.mp3

#selbstbewusst #Energie #funky

Sie können Schlagwörter aus dieser Liste für die Musik vergeben, indem Sie sie antippen oder/und...

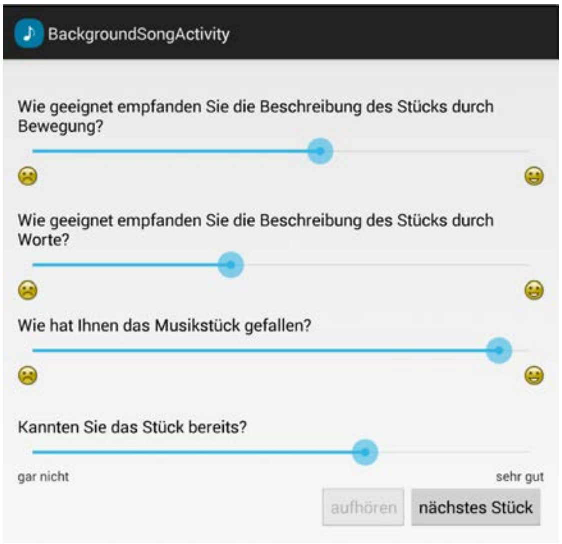
hell ironisch melancholisch fröhlich Gewitter Sonne Berlin Indie traurig hüpfend schlürfen transzendent groß dunkel R 'n B Strand nervös wütend Energie stampfend Freundschaft friedlich verspielt rebellisch vergnügt groove kreativ "kultiviert" laut relaxt rockt ungewohnt kleinteilig chillig elegant gleiten schweben springen Angriff erotisch funky sexy abtauchen verloren aufbrechen Minirock Verlust bewegend weinen dreckig Metall urban verborgen leiden Schmerzen bekifft Karibik flow selbstbewusst

...hier weitere eigene Schlagwörter hinzufügen

(a) Verbal Open Description

(b) Verbal Folksonomy Description

Figure 3.6: Figure 3.6a shows the view in which participants can describe the experience completely open. Figure 3.6b shows the view in which participants can describe the experience by tags.



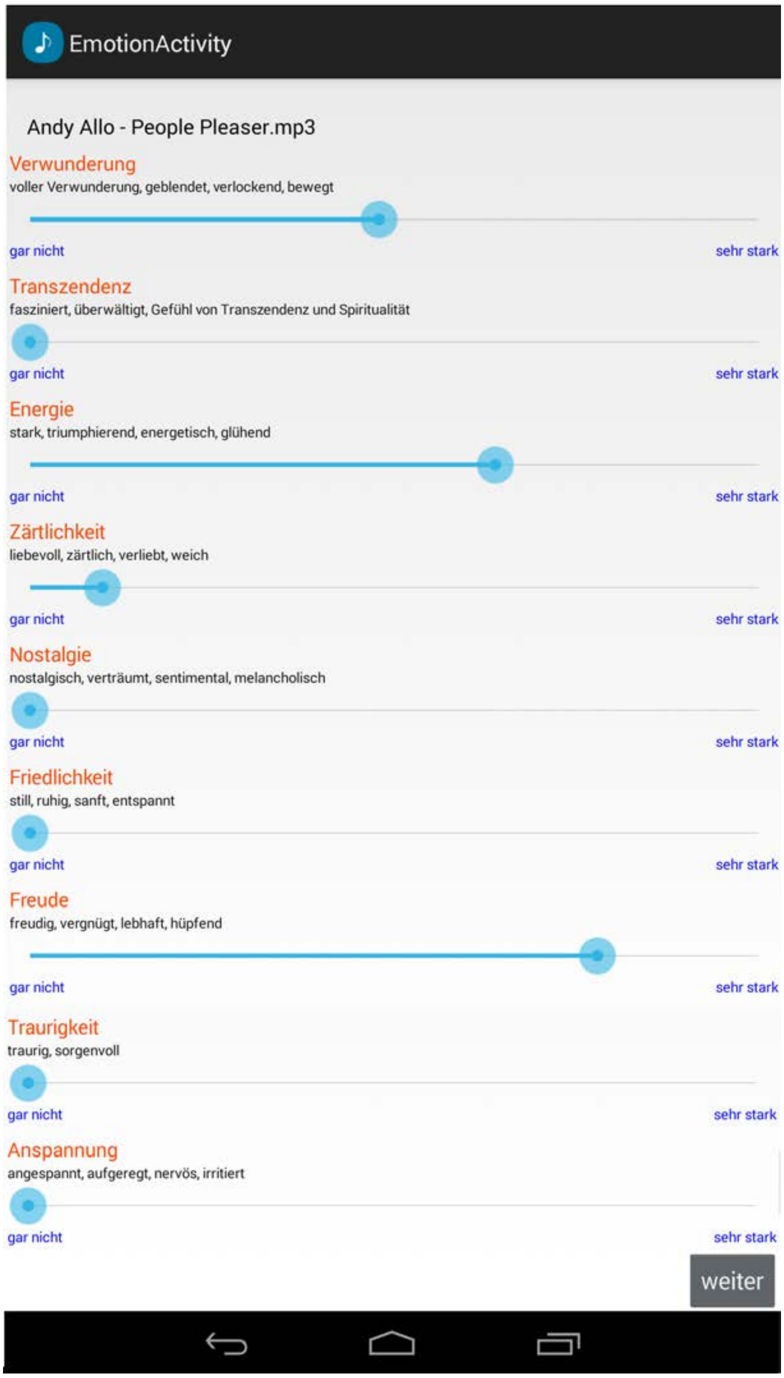
Wie geeignet empfanden Sie die Beschreibung des Stücks durch Bewegung?

Wie geeignet empfanden Sie die Beschreibung des Stücks durch Worte?

Wie hat Ihnen das Musikstück gefallen?

Kannten Sie das Stück bereits?

Figure 3.7: This figure shows the view in which participants provide background information about the sample.



Andy Allo - People Pleaser.mp3

Verwunderung
voller Verwunderung, geblendet, verlockend, bewegt
gar nicht sehr stark

Transzendenz
fasziniert, überwältigt, Gefühl von Transzendenz und Spiritualität
gar nicht sehr stark

Energie
stark, triumphierend, energetisch, glühend
gar nicht sehr stark

Zärtlichkeit
liebevoll, zärtlich, verliebt, weich
gar nicht sehr stark

Nostalgie
nostalgisch, verträumt, sentimental, melancholisch
gar nicht sehr stark

Friedlichkeit
still, ruhig, sanft, entspannt
gar nicht sehr stark

Freude
freudig, vergnügt, lebhaft, hüpfend
gar nicht sehr stark

Traurigkeit
traurig, sorgenvoll
gar nicht sehr stark

Anspannung
angespannt, aufgeregt, nervös, irritiert
gar nicht sehr stark

Figure 3.8: This figure shows the view in which participants describe the experience in terms of the GEMS-9.

SQLite database of the device (green). The following steps describe the communication between the APP and the two databases in more detail.

In Step 1 it is checked whether the given *user name* and *password* are valid and exist in the central database. To accomplish Step 2 a list of all samples is requested from the MySQL-Server in order to display it for the user. In the following step, the file belonging to the selected song is requested from the central server and hence locally saved on the device. As mentioned above all subsequently assessed data is first saved to the device's local sqlite database (green arrows). Acceleration data is compressed and zipped before saving it as blob data into the database.

The folksonomy's tags are loaded from the central server and hence saved on the local device. User created tags are saved directly to the central server after the user left the Folksonomy-Activity in order to make them directly available for the next round and for other users. After having finished Step 24 all data is transferred from the SQLite database to the central MySQL database.

3.8 Data Analysis - From Motion to Emotion

In order to predict the emotional qualities of music from the movement it evoked, features need to be extracted that characterize the motion as good as necessary. In the second step these features can be used to fit a linear regression model and to evaluate the quality of its predictions with new data.
The workflow from pre-processing to fitting and evaluating a model is as follows:

1. get raw acceleration data from motion sensor for x, y and z in 3D space
2. cut beginning and end to standardize duration of signals
3. resample with sample rate $\sim$ 5.7 Hz
4. apply PCA to x, y and z
5. extract features
6. split data set into training (85%) and test (15%) set
7. stepwise select features on training set and fit linear regression model
8. evaluate model quality on test set

Sections 3.8.1 to 3.8.3 will describe this workflow in more detail.

3.8.1 Pre-Processing

The musical excerpts had different lengths ranging from 40 seconds to about 50 seconds. That is why in the first step the motion sensor data is cut to fit 35 seconds. Five seconds are cut-off at the beginning because the raw signals showed that participants needed some seconds to get ready. Thus the end is cut in the second step to fit the standardized length. As sample rates varied highly depending on the speed or the jerkyness at the beginning of the recording, they have all been resampled in the third step to the lowest one, i.e. to 5.7 Hz.
In the fourth step a PCA is applied to the pre-processed signals in order to enhance the comparability of movements between participants. The Principle Component Analysis (PCA) detects the directions of maximum covariance in X by computing the eigenvectors, the Principle Components (PCs), and eigenvalues of the covariance matrix of X [39]. Hence by mapping X onto its PCs and into its eigenspace, X is decorrelated. Since no directional features are extracted, for which it would be difficult to interpret PCA processed data, but rather rhythmic ones, it turned out to be beneficial for the prediction results during pre-tests. Amelynck et al.[5] also apply a PCA to account for different ways to hold a device among participants. Figure 3.9 shows sample motion data with and without PCA. The different subplots also show that participants feature a large variety of describing the experience corporeally.

3.8.2 Feature Extraction

Popular features when it comes to capturing emotions in either music or motions are usually related to *tempo*, *size*, *regularity*, *smoothness* and *direction* [40, 68, 30, 5]. Table 3.3 shows an overview over the features extracted to characterize the movement for the scope of this thesis w.r.t. these categories. Besides the features can be classified into *spectral* and *temporal* ones. Spectral features are extracted by applying a Fast Fourier Transform (FFT). The FFT is a faster version of the Discrete Fourier Transform (DFT). It transforms a temporal signal into frequency domain and is applied to detect the present frequencies of a signal and their corresponding weights. FFT is commonly used in signal processing in order to eliminate noise or to compress data[32].

The statistical features *skewness*, *median* and *standard deviation* are computed to get a time compressed representation of the features extracted from the time series of motion data: *Skewness* expresses the asymmetry of a distribution. For positive values of skewness the distribution has a ‘long tail’ on the right and for negative values its tail stretches towards the left side of its center of mass. Skewness captures the shape of distributions that are not perfectly normally distributed and can hence not be described sufficiently by mean and standard deviation. For skew distributions the *median* is a better representation of the average value than *mean* since outliers might influence the

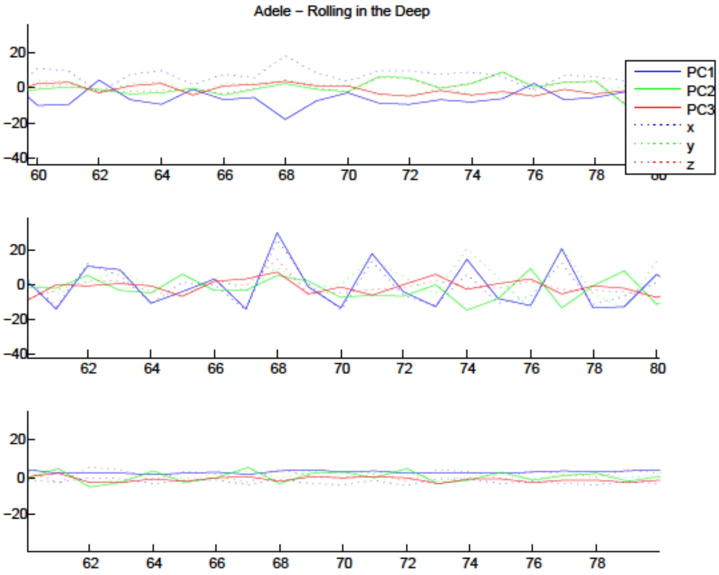


Figure 3.9: The figure shows an excerpt of the signals from three different participants after PCA has been applied. From the plots one can observe that there are large inter-individual differences in the corporeal description.

mean, such that it does not represent the center of the distribution. Having all values of a distribution sorted in ascending or descending order, the *median* is the value in the middle of the list. *Standard Deviation (std)* describes the variance of a Gaussian distribution, i.e. the standard deviation from the distribution's *mean*.

Spectral Features As mentioned above spectral features are computed by applying a FFT to each pre-processed signal (x). Afterwards the absolute value ($|X|$)is computed from the results of the FFT to get their magnitude, i.e. the amplitude of each frequency from the signal's spectrum:
$|X| = abs(fft(x))$.

The first spectral features are simply the most dominant frequency (max_freq_hz) and its corresponding magnitude (max_freq_mag). The frequency is related to the tempo or rhythm of the movement. Sometimes the highest frequency magnitude does not occur in the first PC. That is why only the maximum magnitude of all three dimensions is taken. The spectrum of all dimensions are correlated though and a higher magnitude for frequency X in PC1 usually also results in higher frequency magnitudes in PC2 and PC3. That is also why only one feature is extracted to avoid a violation of the additivity prerequisite for linear regression as described in Menard [56].

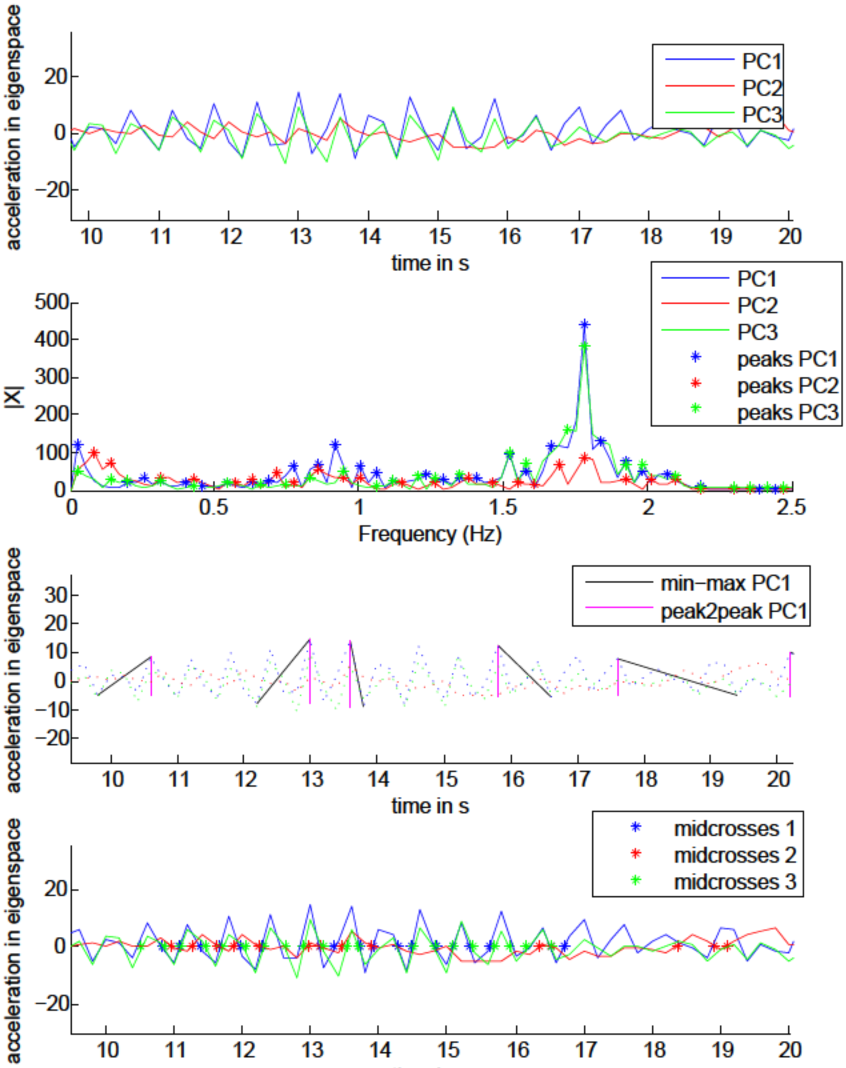


Figure 3.10: The figure shows the features for *Rolling in the Deep* from one participant.

For the third spectral feature (rel_max_freq_mag) the peaks of each spectrum are computed (see 2nd subplot of Figures 3.10 - 5.15). Rel_max_freq_mag is then the maximum magnitude relative to the median of the spectrum's peaks. This feature also is selected such that it is maximum. Thus it is chosen from the same PC as the maximum frequency magnitude. Rel_max_freq_mag is an indicator for the dominance of the highest frequency in contrast to other present frequencies. The bigger it is the, the more regular the movement.

Temporal Features Most features have not been normally distributed. That is why median and skewness have been computed instead of mean. The standard deviation still served as a robust feature to represent the degree of variance in the distribution.
As in Amelynck et al. [5] the volume of the acceleration data point cloud is extracted as an indicator for the size of the movement. Note that acceleration does not cover absolute positions in space and that the size of slow but large gestures might not be captured appropriately by this feature. It is computed from all three PCs by applying point triangulation (*delaunayTriangulation*) and computing the size of the body enclosed by the triangles (*convexHull*).
Another temporal feature related to size is peak2peak that reflects the amplitude of the signal (see 3rd subplot of Figures 3.10 - 5.15). The amplitude of acceleration data refers to the size of the movement w.r.t. time. From peak2peak the median, standard deviation and skewness is computed for each PC (= 9 Features). A temporal feature related to tempo is the distance of midcrosses. It is the distance between one crossing of the mid-reference level and another. The mid-reference level is usually the x-axis when data is centered (see Figures 3.10 - 5.15). For the distance of midcrosses median, standard deviation and skewness are computed for all three PCs (= 9 Features).
Smoothness of movement is described by the edginess of a signal. Signals with longer rise (attack) and fall (release) times are less edgy and hence smoother. From rise and fall times median, standard deviation and skewness are computed for all three dimensions.

3.8.3 Model and Feature Selection

As mentioned in Chapter 3.3 tags and open verbalizations are only used to evaluate the scope of the GEMS as emotional model. Hence, the statistical model simply relies on the GEMS variables.
As participants worked on ten samples in a row during one experiment one would assume an effect from the ordering of samples like an 'emotional afterglow' and recommend a repeated measures model. However, no such effect could be observed for test plots on autocorrelation of residuals and test plots that map 'order' against residuals. That is why the simpler linear regression model is chosen to be fitted to the data for which the features extracted from motion are the independent or predictor variables x

category	**name**	**dimensions**	**description**	**Matlab function**
tempo	max_freq_hz	1	max. frequency in Hz	fft
	median_dist_midcrosses	3	median of distances between midcrosses	midcross
	skewness_dist_midcrosses_PC1/2/3	3	skewness of midcrossings' distribution	midcross
size	median_peak_PC1/2/3	3	median amplitude	peak2peak
	skewness_peak_PC1/2/3	3	skewness of amplitude distribution	peak2peak
	volume	1	volume of acceleration point cloud	delaunay-Triangulation, convexHull
regularity	max_freq_mag	1	max. frequency magnitude	fft
	std_peak_PC1/2/3	3	std of amplitude distribution	peak2peak
	std_dist_PC1/2/3	3	std of midcrossing's distribution	midcross
	std_fall_PC1/2/3	3	std of releases' distribution	falltime
	std_rise_PC1/2/3	3	std of attacks' distribution	risetime
	rel_max_freq_mag	1	max. frequency magnitude relative to median of spectrum's peak magnitudes	fft
smoothness	median_fall_PC1/2/3	3	median of releases	falltime
	skewness_fall_PC1/2/3	3	skewness of releases' distribution	falltime
	median_rise_PC1/2/3	3	median of attacks	risetime
	skewness_rise_PC1/2/3	3	skewness of attacks' distribution	risetime

Table 3.3: This table shows the list of extracted features.

and the GEMS are the dependent or response variables y. Furthermore, no strong multicollinearity between motion features could be observed indicated by the fact, that for any predictor, the *Variance Inflation Factor* was smaller than 10 as recommended in Brosius [9]. Before selecting the features and fitting the model, the data was partitioned into training and test set ~15% in order to evaluate each model's ability to generalize with unseen data. The test set was compiled by randomly selecting five observations from each music excerpt.

The model for linear regression describes fixed effects, the coefficient estimates β, that map x, independent or predictor variables, onto y, the dependent or response variables, where ε describes the error of the model:

$$y = X\beta + \varepsilon \tag{3.1}$$

$$\begin{aligned} y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \\ \varepsilon_i \sim NID(0, \sigma^2) \end{aligned} \tag{3.2}$$

For the linear regression model to be suitable, the error term ε must be Gaussian noise with zero mean. The fit of the model is evaluated using various measures such as the *Root Mean Squared Error* (RMSE) and the *Coefficient of Determination* R^2 that measures the quality of fit between data and model. The R^2 indicates the degree to which the variance of the dependent response variable can be described by the model where $\hat{Y}_i$ is the predicted response variable, Y_i the actual GEMS rating and $\bar{Y}$ the ratings' mean. Values of R^2 range from 0 to 1 whereas the closer to 1, the better the data fits the model [58].

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{3.3}$$

The RMSE measures the distance between observations and the ideal regression line where $\hat{y}_i$ refers to the predicted GEMS value and y_i to the actual GEMS rating.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.4}$$

Besides t-statistic and p-value are both computed to test the null-hypothesis that assumes the variable or model to have no predictive influence on the response observation. If $p - value < 0.05$, i.e. it is below the significance level of 5%, the null hypothesis can be rejected. The t-statistic determines the coefficient estimate of the predictive variable relative to its standard error. The higher the t-statistic, the lower the p-value and the more significant the variable or model.

For each GEMS state, features are selected using the forward stepwise algorithm for linear regression as described in Draper [20]. In an interactive manner features with

highest significance (lowest p-value) are added to the model or removed when their significance decreases again until no further improvement can be achieved. Before selecting the features and fitting the model, the data is partitioned into train and test set in order to evaluate the predictive quality of each model with unseen data as already mentioned above.
In order to determine the relatedness of GEMS-states in this experimental setting *Spearman Correlation* is computed as described in Yule et al. [76]. Correlation measures the effect the change in one variable has onto the change in another variable. In contrast to the common *Pearson Correlation Coefficient*, Spearman applies a rank normalization and follows a non-parametric approach to determine correlation. Therefore it is more robust to outliers and suitable for distributions that are less Gaussian. As for Pearson the closer the correlation coefficient to 1 or -1, the stronger variables are correlated. A correlation coefficient of 0 indicates the absence of correlatedness.

In this section I explained the workflow that is applied to predict GEMS to different degrees based on smartphone-assessed motion data. The next chapter will present the results of the method and discuss them.

Chapter 4

Evaluation

Kick me
Kike me
Don't you
Black or white me
They don't care about us by Michael Jackson (from *HIStory* released 1995)

The following chapters are summarizing the results and discuss the approach on various levels. These levels comprise technical aspects of the approach w.r.t. the suitability and scope of the extracted features or the selected regression model, as well as conceptual ones, e.g. how suitable are GEMS as emotional base model? Besides, I will also discuss the particular benefits of the interdisciplinary and participatory approach in Sections 4.2.4 and 4.2.5. Finally, there will also be a reflection on my perspective's influence on the project in Chapter 4.2.6.

4.1 Results

In this section, I want to present this study's results. First, verbal descriptions and GEMS ratings, as well as questionnaire outcomes are summarized. Secondly, I will present the results for the fitted GEMS models and interpret the selection of features and how they relate to the performed movement of participants.

4.1.1 Verbal Open Descriptions

As described in Chapter 3.3 participants did have the opportunity to describe the listening experience completely open right after the embodied description. The open text field did not only serve as a means of evaluating the GEMS-9-model but it also offered to get to understand a little bit more about the mechanisms that have been triggered by the stimulus and that will be influencing the emotional perception. Most participants in the field just wrote some keywords, for some samples a few skipped the open description and some participants even wrote short stories about the situation they associate with the music.

Descriptions (translated from German) might be sorted into the categories of Table 4.1. The categories illustrate how multi-layered participants experience and describe the music. It seems to trigger a wide variety of actions and moods in particular. Especially the dimension of ‘action’, however, is not at all covered by the GEMS and demands for new embodied approaches.

4.1.2 Folksonomy: Verbal Tags

The tags generated by the participants are summarized in Table 4.2. They are distinguished between tags that are also part of the GEMS-9 and tags that are not covered by the GEMS-9. Note that only tags are listed that appeared more than once and that only participants from the field annotated the content using tags (11 participants).
For most of the samples the given GEMS-related tags and their frequency also corresponds to the distribution of GEMS (cf. Chapter 4.1.3). For sample song #39, #43 and #47 GEMS-related tags have been more popular to describe the sample. All other samples, especially #42, #44, #45 and #48, have been annotated more frequently with non-GEMS related tags i.a. from the categories of ‘aesthetic descriptions’, ‘body’, ‘embodied character’ and ‘place’.

4.1.3 GEMS

The following sections summarize the results for the GEMS-9: The histograms of emotional qualities over all samples, the histograms of emotions for each sample and the correlation between GEMS. As the data is not Gaussian distributed, but with a strong bias towards the minimum, the values are visualized as histograms rather than plotting mean and standard deviation.

Histograms of Emotional Qualities over all Samples The histograms in Figure 4.1 show the distribution of intensity values for each emotion. There are emotions like *energy*, *joy* and *tenderness* for which the values are closer to a uniform or Gaussian distribution. Especially *tension*, *sadness* and *transcendence* have not been perceived as often. In general all distributions are very skew towards the default, the absence of emotion, i.e. 0.

category	**descriptions**
appraisal	exhausting, flavorless, pleasant, boring, wow, earworm, easy, beautiful, superficial, bad text, wannabe emo-guy, too common chords, captivating, monotonous, arty, unfamiliar
action	car ride, bicycling, going out, clubbing, having a drink, dancing, Pogo, chatting, jumping, imitating instruments, smoking, concert, destroy things, diving, singing, cleaning up, village fête, walking in rhythm, clapping hands, being in trance, rebel, empathize emotionally, go and protest, crying, go to the movies, education, observing the forest, dreaming romantically
memory	music my parents used to listen to, youth, 80s, 90s, separation, acrobatics, after having an all-night party, 1969, music class in elementary school
genre	rock, soul, jazz, live, blues, punk, reggae, rock ’n roll, oriental, Indian
instrument	sax, synthesizer
mood	goosing, cultivated boredom, desire to consume mind-expanding drugs, “makes me nervous”, longing, post-love, being angry, good mood, atmosphere of departure, love sick, cozy, harmony, relaxed, melancholy, “feel sheltered by the mood of the woman”, sadness
physical reaction	butterflies in my stomach, heartache, disgust, animates to move, pain
embodied character of music	flowing, exhilarating, seesawing, touching, intellectual, feather, tender and wide
aesthetic descriptions of musical features	groovy, heavy, aggressive, chilled, calm, powerful, French, hot, strong, agile, pensive, airy, energetic, dynamic, sweeping, fascinating, sensual, minimalistic elegance
body (parts)/ person	long hair, erotic voice, Adele
attitude	insubordination, without obligation, self-confidence, homophobic, revolutionary, freedom
clothes	hat, skirt
place	bar, beach, Caribbean, park, London, Berlin, Club, “Warschauer Brücke”, charts, Festival

Table 4.1: This table shows the open verbal descriptions and their suggested categories.

sample	**GEMS-related**	**non-GEMS-related**
#39 Rolling in the Deep	Energie (5), fröhlich (3), vergnügt (3), melancholisch (2), hüpfend (2), wütend (2) (not in GEMS but might be grouped to tension), Stärke (2)	laut (3), groß (2), Angriff (2), leiden (2)
#40 Sweden	Energie (4), nervös (3), transzendent (2)	Berlin (4), dunkel (3), groß (2), groove (2), abtauchen (2), urban (2)
#41 Count on Me	fröhlich (4), friedlich (3), vergnügt (3), melancholisch (2)	Freundschaft (4), hell (2), Strand (2), verspielt (2), elegant (2)
#42 People Pleaser	friedlich (2), hüpfend (2)	sexy (5), funky (4), groove (4), relaxt (3), chillig (3), erotisch (3), hell (2), Strand (2), verspielt (2), kreativ (2), elegant (2), selbstbewusst (2)
#43 La Femme d'Argent	transzendent (4), friedlich (4), gleiten (3), fröhlich (2), traurig (2), nervös (2)	chillig (2), schweben (2), erotisch (2), sexy (2), abtauchen (2)
#44 Wargasm	Energie (6), wütend (3)	rebellisch (6), rockt (5), stampfend (3), laut (3), Verlust (2), dreckig (2), Metall (2), aggressiv (2)
#45 Cornerstone	friedlich (4), vergnügt (3), melancholisch (2), fröhlich (2)	Sonne (5), chillig (5), relaxt (4), schweben (3), Strand (2), Freundschaft (2), gleiten (2), abtauchen (2), bekifft (2)
#46 Rebel, Rebel	Energie (3), vergnügt (3), hüpfend (2)	rebellisch (6), rockt (3), sexy (2), Freundschaft (2), dreckig (2)
#47 Raoui	traurig (6), melancholisch (3), bewegend (3), transzendent (2), friedlich (2), nachdenklich (2)	leiden (4), weinen (4), Verlust (3), elegant (2)
#48 Berceuse	friedlich (5), transzendent (2), bewegend (2)	verspielt (5), hell (4), schweben (4), abtauchen (3), Sonne (2), elegant (2), Klassik (2)

Table 4.2: This table shows the verbal tag descriptions and their frequencies.

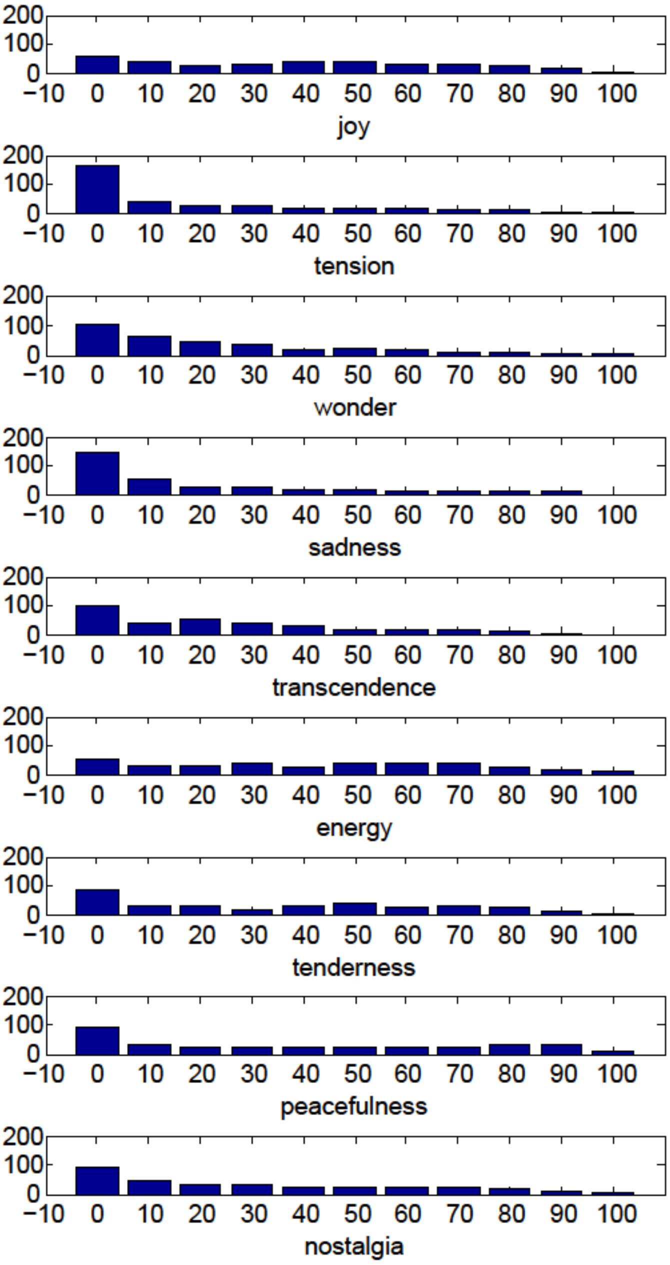


Figure 4.1: This figure shows the distribution of GEMS over all samples. There are emotions like *energy*, *joy* and *tenderness* for which the values are closer to a uniform or Gaussian distribution. Especially *tension*, *sadness* and *transcendence* have not been perceived as often. In general all distributions are very skew towards the default, the absence of emotion, i.e. 0.

Distribution of Emotional Qualities per Sample Figures 5.1- 5.10 show the emotional distribution for each sample over all participants, i.e. field and lab. *Raoui* is the "saddest" and most "transcendent" song. However, for "transcendence" the variance between songs has not been very high. *People Pleaser* is the most "joyful" and *Wargasm* the most 'tense' song. *Sweden* has been most "filled with wonder". *Rolling in the Deep* has been perceived as most "energetic". *Berceuse* is the most "tender" and "peaceful" one. The most "nostalgic" excerpt was *Count on Me*. Except for "transcendence" for which I would have expected "la femme d'argent" to be most "transcendent", the perception of participants matched our expectations and the list of samples covered the range of GEMS to a reasonable degree.

Correlation between Emotions Figure 4.2 visualizes the correlation between the GEMS-9. Since the values for the GEMS emotions are not Gaussian distributed, *Spearman* rank correlation is computed that does not presume any distribution (cf. Section 3.8.3). There is high positive correlation between *peacefulness* and *tenderness*, as well as between *tenderness* and *nostalgia*. Besides *peacefulness* is correlated to *nostalgia* and slightly to *transcendence*. *Transcendence* is noticeably correlated to wonder and partly also to *tenderness*, *peacefulness* and *nostalgia*. Stronger correlation also exists between *power* and *joy*, *sadness* and *nostalgia* and *tension* and *power*. The strongest negative correlation is present for *peacefulness* and *tension*. Although these results do not come as a surprise, they also reflect the choice of samples and should not be considered universally valid.

4.1.4 Background Evaluation on Sample

Only for few songs the distribution for the variables below is Gaussian. Often it is also very skew. That is why the median has been extracted instead of the mean. Especially for *liking* there are songs that polarize strongly like #39, #40 and #43 for which participants either like or dislike the score strongly. From the results (cf. Table 4.3) one can observe that participants preferred the verbal description for samples annotated as *peaceful* like #47, #41, #43 and vice versa, i.e. for those described as very *energetic* like #44, #42, #40 the embodied description has been considered more suitable. Another interesting observation is that for the funky and classic score the suitability for the embodied description is noticeably higher in the lab than for the field, although the difference in liking is not very high. One might hypothesize that the musical background has a big impact on the movement or whether someone perceives the inherent motion in music respectively. Five of the songs have been rather unknown. Still this factor did not have an obviously negative impact on the liking as the majority of unknown songs is liked above-average. This is probably also due to a familiarity with the genres of the unknown songs, see Participant's listening preferences of the subsequent section. In

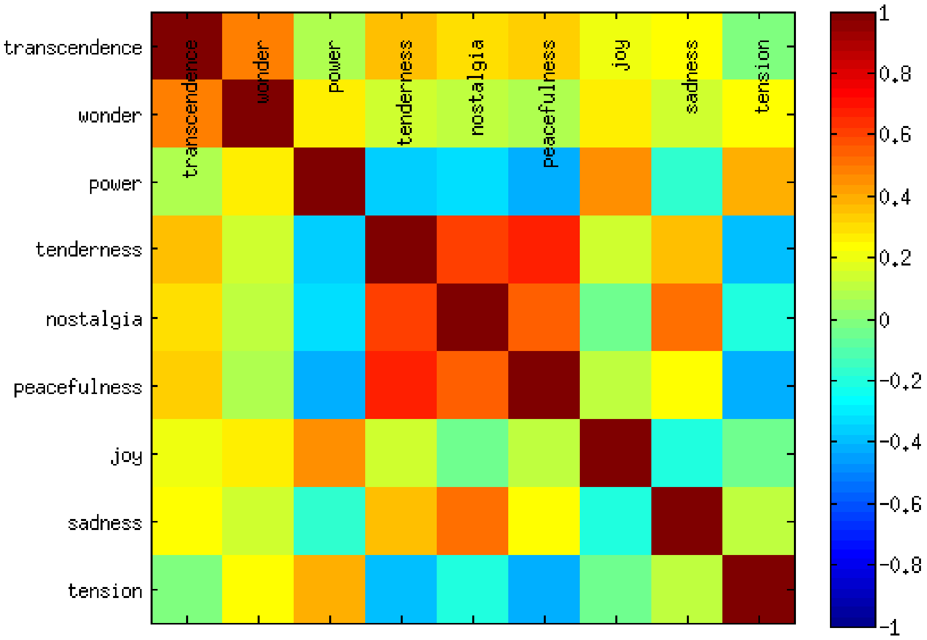


Figure 4.2: This figure shows the Spearman correlation between GEMS states.

sample	familiarity	liking	suitability	suitability	order
			embodiment	verbalization	
	(median)	(median)	(median)	(median)	(median)
	all;field;lab	all;field;lab	all;field;lab	all;field;lab	all;field;lab
#39	68;78;58	71;80;53	62;76;55	59;68;48	5;4;6
#40	0;0;0	68;68;64	75;80;74	42;51;39	7;7;8
#41	37;78;35	37;47;35	50;54;45	49;61;44	5;3;6
#42	0;0;0	68;62;68	76;49;81	49;61;44	5;3;6
#43	0;12;0	57;66;56	48;29;53	54;54;55	5;6;4
#44	0;0;0	33;34;28	57;64;53	38;44;36	7;6;7
#45	45;45;38	60;60;63	57;60;54	57;58;56	4;3;4
#46	41;29;49	64;48;70	59;56;64	43;59;31	5;6;5
#47	0;0;0	51;54;40	37;38;36	65;64;67	5;8;5
#48	0;0;0	58;66;57	50;22;52	56;56;53	5;8;4
mean	19;24;18	57;59;53	57;53;57	51;58;47	

Table 4.3: This table shows how suitable participants considered embodied and verbal (GEMS) descriptions for each sample where rating scales ranged from 0 (= unsuitable) to 100 (= very suitable).

general participants from the field tended to differentiate more and had a bigger range for values of 'liking' or 'suitability'.

Figure 4.3 shows that there is a lot of variance in the corporeal and verbal articulation between participants. The figure also suggests an even closer connection between the indicated *embodied suitability* and the size of the movements than with the perceived *power*. It furthermore does not seem to be necessary to like the music a lot in order to perform very pronounced movement (cf. *liking* = 18 from Participant B). For more figures visualizing movement of different participants and its corresponding GEMS see Figures 5.11- 5.13.

4.1.5 Participants and Experimental Setting

This section evaluates the experimental settings in both field and lab and provides a characterization of participants assessed as described in Section 3.4.3. In the *field* 73% of the participants described their surroundings as very comfortable; the other 27% noted it was whether particularly comfortable nor uncomfortable. 36.4% had been in a "calm", 18.2% in "confident" and another 9.1% each have been in a "sad", "distracted", "bored" or "happy" mood during the study. In the *lab* 59% did feel very or quite

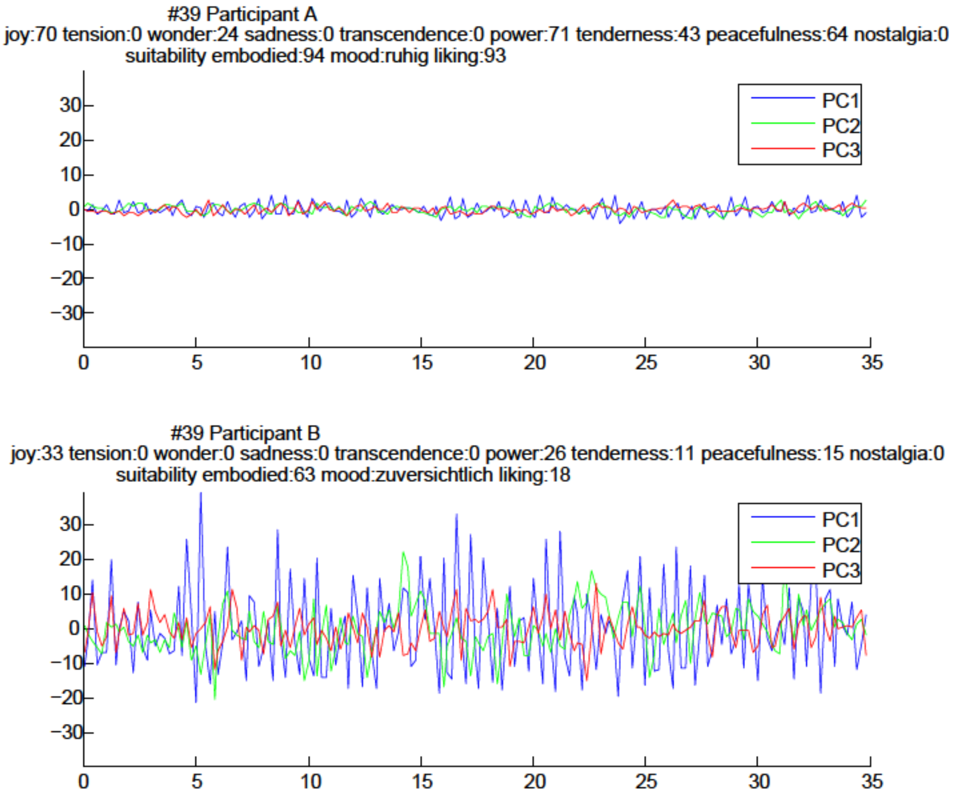


Figure 4.3: The figure shows the corporeal and verbal articulations of two participants for *Rolling in the Deep*.

comfortable; 27.3% concluded that the surroundings were "okay" or fitted the purpose and 13.7% did feel slightly uncomfortable because the lab was too neutral. 59.1% of participants were in a "calm" mood during the study. "Confident" and "insecure" have been indicated by 9.1% of the persons for each. Another 4.55% were chosen per "hungry", "happy", "impatient" and "annoyed".

Field Participants had an average age of 34 years(standard deviation = 11.5). 27.3% identified as "male", 45.4% as "female" and 27.3% did not identify with any gender. 45.5% of the participants did not learn any instrument, another 45.5% did play an instrument in the past or are still actively playing an instrument or singing in a choir and 9% are professional musicians. 64% of participants are experienced in dance classes or similar activities, another 27% is occasionally dancing in clubs or on concerts and another 9% are professional dancers.
36% do not parent children or care for relatives while 64% do. 55% of participants are in partnership/s while 27% are not and 18% are "a little". 45% are regularly using a smartphone while 55% do not. All participants follow different activities from the fields of physics, biology, music, gender studies, computer science, linguistics, science of art, gardening, physiotherapy, audio communication, sociology and booktrade.
Among the most influencing genres have been Rock* (55%), Pop* (55%), Punk (27%) and Hip Hop* (18%).

Lab Participants had an average age of 27 years(standard deviation = 2.36). 72.7% identified as "male", 18.2% as "female", 4.55% as "rather male" and another 4.55% did not identify with any gender. 90.91% are experienced in playing an instrument, the production of music or singing in a choir. 4.55% only have short term experience in making music beyond classes in school and 4.55% indicated to have none experience at all. 36.4% already participated in dancing classes or similar activities for which movement is related to music. 50% are only dancing occasionally in clubs or on concerts. 13.6% wrote they had no experience at all in movement w.r.t. music.
90.9% do not parent children or care for relatives while 9.1% do. 45.5% of participants are in partnership/s while 40.9% are not and 13.6% are "a little". 81.8% are regularly using a smartphone, 9.1% are experienced in using a smartphone but do not use one now and 9.1% are not using one at all. All participants from the lab are studying Audio Communication and Technology or/and are working in this area. 9.1% are also experienced in the field of computer science and 4.55% originally come from the area of audiology.
Among the most influencing genres have been Rock* (68.2%), Electro* (54.6%), Hip Hop* (31.8%), Classic (27.3%), Metal (22.7%), Jazz (22.7%), Punk (22.7%), Pop* (18.2%), Dub(step) (18.2%) and Reggae (13.6%).

GEMS Feeling	rmse (training)	R^2 (training)	p-value (training)	rmse (test)	R^2 (test)
transcendence	24.0	0.0211	0.0142	24.4	0.0092
wonder	23.6	0.0660	0.0003	26.1	0.0628
power	23.9	0.3570	0.0000	26.6	0.3334
tenderness	27.5	0.1440	0.0000	28.6	0.1085
nostalgia	28.1	0.0728	0.0000	27.7	0.0715
peacefulness	32.5	0.0808	0.0000	34.5	0.0581
joyful activation	27.3	0.1500	0.0000	27.5	0.1942
sadness	23.1	0.1650	0.0000	28.9	0.2963
tension	24.6	0.112	0.0000	23.7	0.1391

Table 4.4: Prediction results for each emotion where $R^2(training)$ was computed for the training set only and $R^2(test)$ for the test set, respectively, dito for rmse. As R^2 and rmse do not vary noticeably between training and test set, models were not overfitted. Best results were achieved for *power*. For *transcendence*, *wonder*, *nostalgia* and *peacefulness* only little variance could be explained by the rather rhythmic approach. In general, models for which more significant features could be found, also achieved better prediction results.

Summarizing participants felt more comfortable in the field on average. The field group is more heterogeneous concerning all characteristics except for "genre preferences". Participants from the lab had been influenced by a broader range of musical genres.

4.1.6 Regression Models

As described above, the data set was split into training and test set before selecting features and fitting a linear regression model for each GEMS state. Features had been selected for each of the models individually and on the training set only. Tables 4.5-4.13 summarize the estimates for the model's features. The models have been evaluated on the unseen test set that constitutes $\sim 15\%$ of the whole dataset. Table 4.4 shows the prediction results for each emotion.

The results on the test set do not diverge noticeably from the results on the training set. Hence, the model has not been overfitted to the training data. The 'emotion' for which variables could explain most of the variance is *power*. The most difficult

feature	estimate	SE	t-statistic	p-value
intercept	20.8470	2.0353	10.2430	0.0000
std_dist_midcrosses_PC2	2.26718	0.9191	2.4668	0.0142

Table 4.5: This table shows the results for *transcendence*. It is only related to regularity.

feature	estimate	SE	t-statistic	p-value
intercept	13.3200	3.5638	3.7375	0.0002
std_rise_PC1	-13.9977	6.3915	-2.1901	0.0293
std_peak_PC1	2.4446	0.7621	3.2077	0.0015
skewness_fall_PC3	2.4418	1.1347	2.1520	0.0322

Table 4.6: This table shows the results for *wonder*. It is related to regularity and smoothness.

'emotion' to capture by features extracted from motion is *transcendence* followed by *wonder*, *nostalgia* and *peacefulness*. The following sections interpret the selection of features and their weights for each GEMS emotion and how they relate to the performed movement.

Transcendence TRANSCENDENCE is related to regularity in this experimental setting (cf. Table 4.5). The less regular the movement, the more *transcendent* it is estimated.

Wonder WONDER is related to regularity and smoothness (cf. Table 4.6) The higher the variance in the size of the movement, the more surprising the music (std_peak_PC1). For the std in smoothness (std_rise_PC1) there is high deviation according to its estimate and t-statistic. The relative suggestion is that the more irregularity in smoothness, the less *wonder*. Last but not least, *wonder* is also explained by a tendency towards shorter fall times (skewness_fall_PC3). Note that *wonder* also is positively correlated to *power*.

Power POWER is related to smoothness, size, tempo and regularity (cf. Table 4.7). The less irregular the movement (the smaller the variance in rise time), the more *energetic* it is predicted. Also, the bigger the movement (median_peak_PC1) is, the higher *power*. Skewness_rise_PC1, skewness_dist_midcrosses and median_fall_PC2 suggest a music to be more *powerful* when it its corresponding movements have tendency towards shorter rise and fall times, as well as a higher frequential movement. Hence, movements that are faster and less smooth, are stimulated by music that had been perceived as more *energetic*.

feature	**estimate**	**SE**	**t-statistic**	**p-value**
intercept	13.3420	4.1746	3.3399	0.0010
std_rise_PC1	-13.3940	6.4214	-2.0858	0.0379
skewness_rise_PC1	4.0498	1.4189	2.8543	0.0046
median_peak_PC1	1.7712	0.1950	9.0816	0.0000
median_fall_PC2	-13.0480	4.9662	-2.6274	0.0091
skewness_dist_midcrosses_PC3	4.5266	1.2491	3.6239	0.0003

Table 4.7: This table shows the results for *power*. It is related to smoothness, size, tempo and regularity.

feature	**estimate**	**SE**	**t-statistic**	**p-value**
intercept	63.1820	4.5383	13.9220	0.0000
max_freq_mag	-0.0511	0.0098	-5.2129	0.0000
skewness_dist_midcrosses_PC1	-3.9272	1.4566	-2.6962	0.0074
skewness_fall_PC2	-4.4398	1.5270	-2.9076	0.0039

Table 4.8: This table shows the results for *tenderness*. It is related to tempo, smoothness and regularity.

Tenderness TENDERNESS is related to tempo, smoothness and regularity (cf. Table 4.8). The higher the tendency of the distributions towards slow tempo (skewness_dist_midcrosses_PC1) and smoothness (skewness_fall_PC2), the more *tender* the motion is predicted. There also is a small tendency for very regular movement (max_freq_mag) not to be *tender*. Note that *tenderness* also is correlated to *peacefulness* and *nostalgia*.

Nostalgia NOSTALGIA is related to size and smoothness (cf. Table 4.9). The smaller the movement (median_peak_PC1) and the more it tends to fall smoothly (skewness_fall_PC3), the more *nostalgic* it is predicted.

Peacefulness PEACEFULNESS is related to size (cf. Table 4.10). The smaller the movement, the more *peaceful* it is predicted (median_peak_PC2).

feature	**estimate**	**SE**	**t-statistic**	**p-value**
intercept	44.6110	3.4860	12.7970	0.0000
median_peak_PC1	-0.7762	0.2225	-3.4880	0.0006
skewness_fall_PC3	-3.6484	1.3495	-2.7035	0.0073

Table 4.9: This table shows the results for *nostalgia*. It is related to size and smoothness.

feature	estimate	SE	t-statistic	p-value
intercept	53.2610	3.5690	14.9230	0.0000
median_peak_PC2	-5.0709	1.0184	-4.9791	0.0000

Table 4.10: This table shows the results for *peacefulness*. It is only related to size.

feature	estimate	SE	t-statistic	p-value
intercept	22.6710	5.0241	4.5124	0.0000
std_dist_midcrosses_PC1	-2.6099	1.1782	-2.2151	0.0276
skewness_fall_PC1	3.3866	1.6085	2.1054	0.0361
skewness_dist_midcrosses_PC2	3.3359	1.4929	2.2346	0.0262
median_peak_PC3	0.9187	0.2485	3.6966	0.0002

Table 4.11: This table shows the results for *joy*. It is related to regularity, tempo and size.

Joy JOY is related to the categories of regularity, tempo, smoothness and size (cf. Table 4.11). The less variance in tempo (std_dist_midcrosses) and the bigger the movement (median_peak_PC3), the happier the estimate. Also, the less smooth (skewness_fall_PC1) and the faster (skewness_dist_midcrosses_PC2) the movements, the happier the predictions.

Sadness SADNESS is related to tempo, regularity and smoothness (cf. Table 4.12). The higher the frequency (max_freq_hz), the faster the movement, the less *sad* it has been perceived. The more irregularity in the movement (std_dist_midcrosses), the bigger the tendency towards slow movement (skewness_rise_PC1) and the longer the fall time (median_fall_PC3), the *sadder* it is predicted.

Tension TENSION is primarily distinguished by the size of the movement (cf. Table 4.13). The bigger the movement, the more *tension*.

feature	estimate	SE	t-statistic	p-value
intercept	23.2370	3.3207	6.9976	0.0000
max_freq_hz	-4.3490	1.7447	-2.4927	0.0133
std_dist_midcrosses_PC1	2.1565	0.9939	2.1698	0.0309
skewness_rise_PC1	-4.5439	1.4198	-3.2005	0.0015
median_fall_PC3	7.2240	2.4100	2.9975	0.0030

Table 4.12: This table shows the results for *sadness*. It is related to tempo, regularity and smoothness.

feature	**estimate**	**SE**	**t-statistic**	**p-value**
intercept	6.0348	2.6980	2.2368	0.0261
median_peak_PC3	1.2561	0.2107	5.9626	0.0000

Table 4.13: This table shows the results for *tension*. It is related to the size of the movement.

4.2 Discussions

The guiding question of this Master Thesis was whether corporeal articulations could predict the perception of musical emotions. Furthermore, if the GEMS-9 are suitable to be predicted from embodied descriptions. The results suggest that embodied and verbal articulations are linked to each other. However, the degree of variance that can be explained by the model varies strongly between emotional GEMS-states. The following sections will be discussing how the single factors of the method contributed to the results and why some states like *power* can be predicted better than others like *nostalgia*. It starts with the participants' feedback on the instrument and experimental setting in Section 4.2.1. The next part (Section 4.2.2) will be discussing how feature extraction influenced the prediction quality of the method. Finally, Section 4.2.3 is going to name the inter-individual differences which have been assessed during the experiment but not been considered by the predicting model, yet.

4.2.1 Feedback on Instrument and Experimental Setting

There are some general remarks about the experiment that occurred for a majority of participants:
First of all the music is a very big issue. They reported that they enjoyed the experiment because they liked the music or they felt rather uncomfortable with the experiment because they usually listen to 'other' music. One fourth to one third of the participants also felt that the samples did not cover the whole space. That is why some also reported they did not use the whole range of the GEMS. Indeed the set of samples has been quite small w.r.t. genre preferences of the participants and the high dimensionality of the emotional model. In a bigger scope like an online-survey it would be possible to let participants only work on 3-5 of their most favorite songs.
Secondly, many reported hardware issues like the cable of the headphones being annoying, the device being too big (some experiments in the field have been conducted on a tablet) or the headphones not fitting well enough (the headphones in the MediaLab did not work well for people with smaller heads). Consequently, it seems to be important to use a handy smartphone with suitable headphone in order to avoid distracting participants from the experience as much as possible.
Especially in the field when participants accomplished more verbalizations most felt

exhausted after the experiment and recommended to only work on five songs in order to spend the same amount of energy on all songs. Some felt insecure or nervous about using a smartphone beforehand but did not have any difficulties during the experiment. However, there have been some uncertainties about: how much or how they should move; which scales to move for the verbalization of the emotional qualities when none mirrored their way to express the experience; or also generally whether they contributed "enough". For example, there were comments like "I sometimes only moved some sliders because I felt like I had to indicate at least something although it did not fit to how I perceived it". Another remark concerning the GEMS has been that sometimes it felt like categories had been redundant while at the same time some others were missing. This might also coincide with the selection of samples and it is not completely clear whether these categories would still be redundant when different songs are selected.

Irritation has also been caused by the initialization of sliders. In general, there are many possibilities to set the default position like at the min, max, in the middle or randomly. The choice was to set it at the min position and to make the absence of all emotions the default status. Some participants would have assumed the default position to be in the middle, so that they can move the slider in both directions. One person also remarked that it would invite persons to leave the slider at the min and not move it at all. Indeed it increases the effort to move it to the maximum and might hence have led to a bias towards the minimum.

Participants also wished for more variations of 'mood', e.g. "sick" was missing. They also wanted to being able to select multiple moods and indicate the change in mood over the experiment. Asking participants for their mood before or during the experiment, however, could influence the results but the possibility to select more than one mood is certainly desirable.

Another important aspect has been that participants reported a learning process for the description of music. Once they learned about the GEMS and tags from other users, they indicated different emotional qualities than in the first runs. This observation might also explain why the GEMS have also been very present in the folksonomy. Still, there was no systematic error related to the order in which samples had been described that would have called for modeling 'order' as random effect (cf. Section 3.8.3).

In the field participants who spend a lot of time on the open and tag-based verbalizations felt like they wanted to listen to the sample again when assessing the GEMS. Some also needed to getting used to holding a device while dancing. They suggested to use wristbands or bands for the feet as well. This option has already been discussed in Section 3.1.

In general some people are more rhythmic oriented which is stronger associated with the movement of feet and some rather stick to the contour of the melody by moving their hands. Movement from the feet might only partly be captured by the smartphone sensors in the way that other parts of the body move too when feet are moving.

4.2.2 GEMS and Embodiment

As already mentioned above, the results showed that some emotional states of the GEMS-9 like *power* are captured more appropriately than others like *nostalgia* by the approach. There are several aspects or possible explanations for this observation:

> (1) Participants preferred the embodied description for the more *energetic* and *tense* scores, those that could be predicted more easily. For the most *nostalgic* and *peaceful* scores like *La Femme d'Argent*, *Raoui* and *Berceuse* participants preferred the GEMS-9 over the embodied description.
>
> (2) The GEMS-9 afford a longer learning phase from participants: The poor performance of *transcendence* comes from a lack of variance in the perception of *transcendence* between songs. During pre-tests *transcendence* has been the emotional state that caused most of the questions. Participants did not really know what *transcendent* music sounded like or the kind of 'feeling' it described.
>
> (3) There also have not been enough samples to sufficiently cover all states of the GEMS-9. Especially *sadness* and *tension* have not been perceived by the participants very often. It also needs to be evaluated if this fact had been due to other reasons as well: Some participants reported that they missed *aggressive* or *angry* when it came to *tension*. Hence, it could also have been due to a missing understanding of this emotional state when it had been perceived that rarely. *Sadness* is often more intensively perceived by extrovert people[74]. Therefore, it had to be evaluated if this factor influenced the GEMS-ratings as well.
>
> (4) Participants did not enjoy moving to some of the scores because they did not match their personal preferences (cf. 4.1.5).
>
> (5) Participants are not trained to express emotions like *nostalgia* in a corporeal way.
>
> (6) Emotions like *nostalgia* and *transcendence* call for features that are less bound to rhythm but to the musical contour like directional and time series features.
>
> (7) Hence, participants also had to be explicitly instructed to describe the musical contour of the score in particular and not to mimic its rhythm when aiming

at capturing these emotional GEMS-states. Though, this might not meet their personal preferences or individual capabilities to describe the experience.

(8) Acceleration data is not sufficiently covering *nostalgic* or *transcendent* gestures. As acceleration does not determine the absolute position in space, it is particularly difficult for slow but big gestures to be captured. This calls for applying different and more sophisticated motion sensor fusion techniques.

(9) Pre-processing of sensor motion signals had to be improved in terms of normalization and adaptation to drifts to further improve the extraction of rhythmic features.

(10) Inter-individual differences had to be adapted, i.e. motion signals from persons who accomplish more expressive movements should not be compared to those of less expressive movements without adaptation.

(11) The correlation between GEMS states could suggest a lower dimensional model to be more suitable. It needs to be further examined whether the correlation arises from the particular musical dataset and will be weaker with different scores.

In general, models performed better for which more features could be selected because they explained parts of the variance in emotional perception. For models like *nostalgia* and *transcendence* it had to be evaluated if directional features could explain more variance.

4.2.3 Performativity and Randomness

The last section already discussed some of the issues that occur when comparing data from different participants. Although, there are some general tendencies, both the perception of musical emotion as well as the corporeal articulation of the listening experience are assumedly highly dependent on individual factors like 'personality', 'status' or 'musical background' (see Figures 4.3 and 5.11- 5.13). This issue has been discussed already in Sections 2.2 and 3.1.
As the regression model does not yet evaluate the influence of random effects such as 'musical background' or 'personality', it would be desirable to apply a *Mixed Effects Model* [62] in a future evaluation and to determine the degree to which the error might be predicted systematically. Nonetheless, the GEMS-9 and the corporeal articulation offer a lot of variance for the participants to describe the experience that will always create some surprises.

4.2.4 Why Care About Participation? – We Are the Experts, Aren't We?

Besides this project's research question, one major question was whether a participatory approach would change the course of this thesis to one that is less likely to walk into the trap of I-methodology. So, what's the participants' part in it?

During the first pre-test that had been completely open, participants from the girl rock band did not come up with a huge variety of possibilities but rather agreed on the aspect that they did not want to describe their experience using words. Therefore, I let them pre-test two more specified prototypical ways that had also been discussed in a common meeting with my advisors, namely drawing lines and moving their smartphones according to the music. During this pre-test they opted for moving the device because it offered them more space and ways to differentiate between different songs.

It was only when they finally tested the APP that a whole new bunch of ideas came up like making soap bubbles burst according to the music, interact with a frog, make the frog look like the music, assign color to music etc. Therefore I conclude that the APP was necessary as a basis in order to get alternative clues. By the time, participants also had been more into the topic and it seemed like the former meetings had established a certain kind of trust that enabled participants to be more confident to answer in a less compliant way and to address their own opinions and ideas.

Another aspect was that they had been irritated because the APP only served as a tool to collect data and did not give them any feedback about their inputs. They expected the multimodal querying to be implemented already and to get song recommendations based on their inputs. The most interesting part was that they did not know anything about Leman's vision of multimodal querying[47] but suggested to get music according to how they move the device or according to the expression on their face when they look into the camera.

There have been another two major conflicts that arose during this participatory approach. First, although the majority of participants have been rather young, i.e. between 13 and about 35, they had different priorities. While the younger ones wanted to discover the APP on their own and did not want to be taught how to use it, the elder ones wanted to avoid as much confusion as possible. This need for more clarity might also be due to being less native to these kinds of devices as the younger ones. That is why I ended up with an APP that is less exploratory in order to make sure everyone understands what needs to be done. Secondly, the comparability of results and the timely scope of this project demanded for less freedom. Therefore, I finally had to standardize the times participants are allowed to listen to the sample and to make them assess all samples and not only a few of their favorite ones.

Now, how would the APP have looked like without participatory feedback? First of all, there would not have been an APP but a web application. It was only for the smartphone-philia of the participants that I desperately wanted it to work on a mobile device. The participatory approach also motivated an approach that worked in the field, e.g. participants started dancing together while moving their smartphones. This suggests that people behave and express their experience in different ways according to the situation. By offering ways of expression that will only work in a lab you will never be able to capture the whole spectrum. And smartphones can go anywhere.

One more detail is that there would not have been a way to stop or pause the sample during test listening. Participants got very annoyed when having to listen to a score they already knew very well. It is a little thing that still assigns more competence to people.

During the compilation of samples it turned out that participants who are socialized as 'male' are more likely to suggest 'male' artists and 'femally' socialized personalities rather mentioned 'female' artists. That is why it has been very beneficial for the sample list to have a diverse range of participants, especially since I tend to be one of those people who rather stick to 'male' artists.

Besides, the appreciation for the GEMS-9 that all participants shared let me overcome my reluctance to use such a standardized approach. They still also appreciated the open format and the folksonomy claiming it was offering them to give more individual feedback and express their own ideas about the music. Hence, participation had also been very crucial to me to evaluate the different means of verbalization and their importance.

Last but not least, without all these pre-tests the APP would probably have been rather crappy and incomprehensible.

Since the first meeting with the girl rock band, I have become a big sympathizer of "Citizen Science"[26]. I have really been stunned by how much they wanted to be asked and by how much they wanted to contribute. Science is never neutral, it is a product of our everyday politics [33, 51]. That is why I believe people should be allowed to participate in all processes that are going to influence their lives. The experience with the girl rock band also taught me that it is very important to let them participate early in order to keep them engaged in the democratic idea. By the way, they also started being interested in fields like computer science they might never have gotten insights into otherwise. Therefore, letting young adults participate might also be a promotion for new talents and keep them motivated throughout school.

Summarizing, it has not been easy to mediate between conflicting interests in this participatory and interdisciplinary approach, above all because I never had all parties negotiate on the same table. Sometimes, I did not feel comfortable of being the person who still decides in the end.

Participation takes time, it slows down your progress. Participants need time to understand the topic. The researching persons need time to understand the participants' perspectives. If you want to quickly produce some beautiful numbers, it is probably not the way to go. Thus, I really appreciated to discuss the idea with the participants and to get a lot of detailed feedback on the instrument and the setting. I also highly appreciated the participants to know that much about the topic, so they could give me more integral

feedback. It tremendously helped me to finish the puzzle to which each and everyone contributed a piece. I could get rid of irritations and artifacts I would never have known about if I would have had to rely on myself only. The pre-tests also taught me that this kind of integral feedback is not possible in a single session and prerequisites regular meetings.
The experience to get feedback from the "real world" that highly motivated the research question has been the most amazing thing to me. It kept me motivated throughout the whole time to have so many people to talk to about the idea and it helped me to reflect about it myself. And I guess it pushed things much father than without the dynamics that arose by participation.

4.2.5 "The Whole is Greater than the Sum of its Parts." - What Are the Interdisciplinary Parts in It?

As already mentioned in the introductory part this project has been accompanied by three disciplines, namely in the field of semantic content analysis and information retrieval, in the field of music perception and cognition and in the field of gender studies. What would this thesis have looked like without the perspective from musicology and gender and diversity studies? I would probably have done some kind of machine learning on a given dataset, e.g. the one provided by MediaEval[1]. The data is annotated based on Russell's Circumplex Model [64]. It is annotated based on a crowd of only ten people whose backgrounds are completely intransparent [70]. What are these ten people like? Are they computer scientists? Are they musicians? Are they representative? We would not have known.
I would have missed most of the fundamentals from musicology that offered a deeper glimpse into how emotions are elicited. I would not have used a participatory approach and developed a new instrument. I would not have investigated the correlation between corporeal and verbal expressions.
But now that we did this interdisciplinary project, how did the perspective from computer science contribute to the whole? The knowledge and competence from computer science enabled us to actually create a different instrument, an instrument that is able to collect data in the field, the APP. It enabled us to collect the data via the Internet and to make use of the most common acceleration sensors, the sensors in smartphones. Another contribution is the creation of a folksonomy in order to annotate the content. Folksonomies are very common in the field of Semantic Search and have also been highly appreciated by the participants. Although they only serve as a means of evaluating the GEMS-9 model in this project, the provided tags could be used to enrich the GEMS in future work.
What would this thesis have looked like without the perspective from audio communi-

[1] http://multimediaeval.org/mediaeval2014/emotion2014/

cation? Audio communication proposed an embodied approach to the topic in the first place. It was only by this proposition that I dived deeper into the concept of embodiment and the parallels to gender studies in overcoming mind-body dualism. This comprises most of the theoretical fundamentals this project is built on as well. The standardized view of the GEMS-9 also is a contribution from musicology that allows for a more differentiated expression of the emotional quality of music than a tag being present or not, i.e. it allows to express participants to determine the degree to which a score is *sad* e.g. At the same time it provides a vocabulary that makes the verbalizations more comparable to related work. There are also some adjustments concerning the standardization of the experiment w.r.t. instruction and the rehearsal of songs proposed by musicology that this project benefited from, see Section 3.7. It brought in a lot more comparability and hence facilitated the evaluation.

It has also been audio communication who proposed a second round of data collection after the first one, so we could collect more data. The collection in the lab has been a compromise between the need for more data and leaving the scope of a master thesis by collecting more data in the field.

What would this thesis have looked like without the perspective from Gender and Diversity Studies? Most of all the project would not have applied a participatory and interdisciplinary approach. As already discussed in the previous section participation pushed this project much further and improved its quality tremendously. From Gender Studies also comes a lot of care for the tiny little details starting with the list of samples over the selection of participants to the biographical questionnaires but there has also been a lot of motivation from Gender Studies to the overall design of the instrument and the collection of data in the field.

One aspect concerning the verbalizations that also comes from Gender Studies is to start with the most open question and to reduce the complexity in the following steps via the semi-standardized folksonomy to the completely standardized GEMS. This sequence from open to closed formats is a major change towards common designs that only feature a small open question at the end after everything has been said and done. Gender Studies also motivated the use of unipolar scales instead of bipolar ones. This has been another major motivation for using the GEMS-Model over any other as Zentner et al. also argue against the use of bipolar scales[77] because it might unwillingly mutually exclude the presence of two emotional states on opposite sides of the scale and hence lead to interpretation errors. One last remark is about the question for ‘gender’. We have been discussing how to best ask for it for weeks as it is a violent thing to force people into a binary system. This might have been the most difficult example but it also exemplifies that we did not take anything for granted during this project.

In summary, I want to conclude that this cooperation has really been extraordinary. And though, it has been difficult sometimes to find a way to go, this struggle also contributed a lot to getting the “big picture”.

4.2.6 So, this Thesis is Absolutely Free of Any I-Methodology?

Now is the time for confessions. One might think, this thesis is absolutely free of any I-methodology since it is participatory. I must admit: To me it looks like the I-methodology per se.

It is always difficult to tell whether participants would have replied in the same way to a different person. I doubt this and I guess I did implement some reflections that are highly influenced by my own perspective. Besides playing the piano and bass guitar, I have also been a dancer for as long as I can remember. Hence, this project offered me the chance to combine all of these aspects. It also explains my obsession about embodiment.

Additionally, there is one song in the list of samples that sneaked its way into it after knowing I could select another five songs to enlarge the list and after visiting the David Bowie Exhibition in Berlin: Rebel Rebel. I could make up a lot of other arguments to justify why this song had to be in the list but the truth is: I simply love it and I wanted Bowie to be in the list to show my appreciation for this amazing artist. Finally, in order to qualify the degree of I-methodology in this project a little bit: I am absolutely innocent of the decision for the APP. I am not a smartphone user and might never be.

During this project. there has been one quote from Damasio [16, p. 252] that accompanied my struggle between creating meaning, i.e. creating difference, and embracing complexity at the same time that I want to close my personal reflections with:

Perhaps the most indispensable thing we can do as human beings, every day of our lives, is remind ourselves and others of our complexity, fragility, finiteness, and uniqueness. And this is of course the difficult job, is it not: to move the spirit from its nowhere pedestal to a somewhere place, while preserving its dignity and importance; to recognize its humble origin and vulnerability, yet still call upon its guidance.

Chapter 5

Conclusion and Future Work

Meine Stücke wachsen nicht von vorne nach hinten, sondern von innen nach außen.
My pieces grow from the inside out.[14]

Pina Bausch

When I read this quote from Pina Bausch, it reminded me very much of this project and our approach. We moved from the inside to the outside like emotions do when they are transformed into movement and then also tried to find our way back to the inside in order to detect the motions' emotional origins. The following chapters will be summarizing this process, drawing conclusions and thereupon proposing next steps for future work.

5.1 Summary

Starting off with the question if there were similarities between how people move and the emotions that moved them, the goal of this study has been to investigate if one could predict the perceived emotional qualities of music from the movements it stimulates.
In Chapter 2, I presented the fundamental theory this thesis is build on including the concept of embodiment, the term 'emotion' and the motivation for choosing the Geneva Emotion Music Scales over other emotion models. Chapter 3 explained the method of this study, i.e. how to get from smartphone-assessed motion data to predictions of GEMS states to different degrees ranging from the musical stimuli, participants, experimental setting and procedure as well as the data analysis. After data had been collected in both 'field' and 'lab', spectral and temporal features were extracted and selected, and a linear regression model was fitted for each GEMS state according to the workflow described in Section 3.8. Extracted features were linked to the following categories: 'size', 'frequency', 'smoothness' and 'regularity' of the movement.

In Chapter 4, I summarized the final results and discussed them. The results shown in Section 4.1 suggest a connection between motion and emotion whereas the quality of prediction varies between different states of GEMS as follows: power ($r^2 = .36$), sadness ($r^2 = .17$), tenderness ($r^2 = .14$), joy ($r^2 = .15$), tension ($r^2 = .11$), peacefulness ($r^2 = .08$), nostalgia ($r^2 = .07$), wonder ($r^2 = .07$) and transcendence ($r^2 = .02$). Furthermore, the open verbal descriptions comprised a whole range of different categories like ‘action’, ‘emotion’, ‘place’ or ‘aesthetic character’. For many songs the GEMS were also very dominant in the annotated tags but there were also songs like “People Pleaser” for which the GEMS did not seem to cover the experience appropriately.

Section 4.2 discussed the participants’ feedback on the instrument, the prediction results for the different GEMS states and the suitability of the linear regression model w.r.t. striking inter-personal differences in the perception of emotion and the corporeal articulation. It also praised the participatory and interdisciplinary approach to the project and reflects on how they contributed to the overall design of the project. Last but not least, Section 4.2.6 contained some personal reflections about how much my own perspective might have influenced the general course of this project.

5.2 Conclusion

The results showed that there are similarities between how people move and how they emotionally perceive the music. The evaluation also confirmed that the GEMS offered participants an appropriate means of describing their experience verbally. Although, the variety of verbalizations (open and tag-based) in the field has been much higher and could not be covered by the GEMS. At the same time, it needs further research if all GEMS states might be connected to distinctive movement. How huge is the resolution and scope of motion that is possible to be performed by participants w.r.t. both technical limitations of the device’s sensors as well as individual capabilities and preferences to describe the experience in a corporeal way? So, some emotional states like *nostalgia* might be more difficult or less suitable to be expressed in an embodied way according to participants’ ratings.

The extracted features could explain more variance when emotions had been primarily linked to *power* or rhythm and turned out to be less suitable when GEMS states were probably stronger linked to musical contour and the direction of the movement and gestures as for *nostalgia*. This observation suggests that there might still be a lot of hidden potential in additional features that could capture direction, as well as position, and also describe the course of the movement (time series features). Additionally, the perception of emotions also underwent a learning process and some emotions like *transcendence* still had to be learned in order to be perceived or felt.

GEMS ratings have also been correlated to each other. This might be due to the selected musical excerpts but could also mean that GEMS ratings had to be decorrelated, e.g. by

applying a PCA in order to determine the principal (uncorrelated) components.
The approach compared participants' movements without appropriate adjustments to their inter-individual differences. Some participants accomplished bigger movements than others, some had been in a different mood than others, some might have preferred particular states of the GEMS to describe the experience, some might have employed the whole range of intensity while others stayed in the middle or at the minimum of the scales for the GEMS. Hence, it would probably be necessary to check the influence of these inter-individual differences and to apply appropriate pre- and post-processing algorithms in order to achieve better or more personalized predictions. People playing an instrument and being familiar with a various range of genres, in particular preferred the embodied way of describing the experience. The bigger difference among field participants also recommends a broader range of participants in order to being able to make general assumptions between movement and emotional perception.
Generally, for the complexity of the GEMS and the variance in perception and movement between participants, the data has not been enough in order to generalize about the observations and motivates further research with a tremendously bigger set of samples and participants.
The participatory approach and the conduction in the field also strongly motivated the use case of retrieving recommendations for music by dancing while holding a mobile device. The majority of participants reported to having had a lot of fun and would pretty much like to have such a feature. Thus, the participatory approach could also already confirm the acceptance for such a technology and be legitimated by potential 'real' users.

5.3 Future Work

There are several propositions for how to proceed in future work concerning the emotional model, the experimental setting, the feature extraction and regression model. User-generated tags from the folksonomy could be integrated into the GEMS-model. As the default position of the sliders might have led to a bias towards the absence of any emotion, it should be evaluated how a different default position impacts the results.
Concerning the experimental setting, first of all, it must be mentioned that the experiment should be conducted in the field again but at large scale. There are several smaller adjustments that had to be made for the selection of samples in order to make it work for the participant's collection of music and to avoid a pre-compiled list of samples. Furthermore, participants had to be asked what *transcendent* means to them and what a *transcendent*, *nostalgic* or *tender* gesture would look like. If there was some consensus about these issues, one could think of suitable features and more sophisticated sensor fusion algorithms in order to capture it. In case participants report a lot of difficulties in expressing such emotions by motion it might be worthwhile reconsidering a lower

dimensional model than the GEMS-9.
As it turned out to be insufficient to detect *nostalgia* with a rhythmic approach, another study had to evaluate how instructing people to describe the experience according to its musical contour could enhance the predictions for *nostalgia* or *tenderness*. Furthermore, it needed to be determined if people preferred to corporeally describe the music in terms of rhythm or musical contour. If musical contour turned out be unpopular, it would not make sense to further follow an approach that is based on this individual capability or preference. The motion signals, however, suggest that participants apply both ways and that there are tendencies for the more *energetic* and *tense* excerpts to be described more rhythmically while the more *tender* and *nostalgic* ones are more likely to be assessed by their musical contour.

Hence, the most important improvement could be achieved by complementing the set of features for more directional and time series ones.

The inclusion of inter-individual factors like 'musical background' or 'mood' into a mixed random effects model also seems to be quite important to determine the influence of these factors as next step.

Finally, participants want to receive recommendations. So, a proof of concept would consist of a prototype that generates recommendations based on a motion query from the user. The participant could then rate the result in order to evaluate the model's quality. Here, criteria could be the quality of fit, i.e. how well did it fit the intended query; the innovativeness of the recommendation, i.e. Did you know this song before; and also the liking of the match? For this purpose, one also must think about ways to integrate the query feature into existing algorithms that consider the individual preferences of users. The recommendations could first of all be based on a similarity match of motion data, i.e. the motion query is compared to existing motion sensor data from other songs and one song corresponding to the most similar motion signals is recommended. This is one example for how acceleration sensor data can be employed to integrate embodied music cognition into Music Recommender Systems. This study contributed to this goal.

Bibliography

[1] Jahreswirtschaftsbericht 2010 des Bundesverbandes Musikindustrie. `http://www.musikindustrie.de/jwb-musiknutzung-10/` (8. Jan. 2015).

[2] Musikindustrie in Zahlen 2013. `http://www.musikindustrie.de/fileadmin/piclib/statistik/branchendaten/jahreswirtschaftsbericht-2013/download/140325_BVMI_2013_Jahrbuch_ePaper.pdf` (8. Jan. 2015).

[3] AKRICH, M. User representations: Practices, methods and sociology. In *Managing Technology in Society: The Approach of Constructive Technology Assessment*, A. Rip, T. Misa, and J. Schot, Eds. Pinter, 1996, pp. 167–184.

[4] ALEXANDER, F. *Psychosomatic Medicine*. W W Norton & Company Incorporated, 1965.

[5] AMELYNCK, D., GRACHTEN, M., VAN NOORDEN, L., AND LEMAN, M. Toward E-Motion-Based Music Retrieval a Study of Affective Gesture Recognition. *IEEE Transactions on Affective Computing 3*, 2 (Apr. 2012), 250–259.

[6] BATH, C. De-Gendering von Gegenständen der Informatik: Ein Ansatz zur Verankerung von Geschlechterforschung in der Disziplin. In *Gender und Diversity in den Ingenieurwissenschaften und der Informatik*, B. Schwarze, M. David, and C. B. Belker, Eds. Bielefeld: Webler, 2008, pp. 166–182.

[7] BEHNE, K. E. The development of “Musikerleben” in adolescence: How and why young people listen to music. In *Perception and cognition of music*, S. A. J. Deliége, I., Ed. Psychology Press, 1997, pp. 143–159.

[8] BERGOLD, J., AND THOMAS, S. Participatory research methods: A methodological approach in motion. *Forum: Qualitative Social Research 13*, 1 (2012), 191–222.

[9] BROSIUS, F. *SPSS 8 Professionelle Statistik unter Windows*. mitp, 1998.

[10] BURKE, R. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction 12*, 4 (Nov. 2002), 331–370.

[11] BUTLER, J. Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal 40*, 4 (1988), 519–531.

[12] BUTLER, J. *Bodies that matter, on the discursive limits of 'sex'*. Routledge New York & London, 1993.

[13] CLAYPOOL, M., GOKHALE, A., MIRANDA, T., MURNIKOV, P., NETES, D., AND SARTIN, M. Combining conten-based and collaborative filters in an online newspaper. `http://web.cs.wpi.edu/~claypool/papers/content-collab/content-collab.pdf`. (8. Jan. 2015), 1999.

[14] CLIMENHAGA, R. *Pina Bausch (Routledge Performance Practitioners)*. Routledge Chapman & Hall, 2008.

[15] CORNESS, G. The Musical Experience through the Lens of Embodiment. *Leonardo Music Journal 18* (2008), 21–24.

[16] DAMASIO, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain*, 1 ed. Harper Perennial, 1995.

[17] DE BRUYN, L., MOELANTS, D., AND LEMAN, M. An embodied approach to testing musical empathy in subjects with an autism spectrum disorder. *Music and Medicine* (2011).

[18] DEGELE, N., AND BETHMANN, S. Gewusst wie: Richtig lieben und leiden. In *Emotionen in Geschlechterverhältnissen: Affektregulierung und Gefühlsinszenierung im historischen Wandel*, S. Flick and A. Hornung, Eds. transcript Verlag, 2009, pp. 83–103.

[19] DÖRING, N. Zur Operationalisierung von Geschlecht im Fragebogen: Probleme und Lösungsansätze aus Sicht von Mess-, Umfrage-, Gender- und Queer-Theorie. *GENDER*, 2 (2013), 94–113.

[20] DRAPER, N., AND SMITH, H. *Applied Regression Analysis*, 2 ed. John Wiley and Sons, Inc., 1981.

[21] EGERMANN, H., FERNANDO, N., CHUEN, L., AND MCADAMS, S. Music induces universal emotion-related psychophysiological responses: Comparing canadian listeners to congolese pygmies. *Frontiers in Psychology: Emotion Science* (2015).

[22] EGERMANN, H., AND MCADAMS, S. Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception 31*, 2 (2013), 139–156.

[23] EGERMANN, H., PEARCE, T. M., WIGGINS, A. G., AND MCADAMS, S. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience 13*, 3 (2013), 533–553.

[24] EKMAN, P. An argument for basic emotions. *Cognition & Emotion 6* (1992), 169–200.

[25] EKSTRAND, D. M., RIEDL, T. J., AND KOSTAN, A. J. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction 4*, 2 (2010), 81–173.

[26] FINKE, P. *Citizen Science - Das unterschätzte Wissen der Laien.* oekom Verlag, 2014.

[27] FRAISSE, P. *The Psychology of Music.* Academic Press, Orlando, FL., 1982, ch. Rhythm and tempo., pp. 149–180.

[28] GARRIDO, S., AND SCHUBERT, E. Negative emotion in music: What is the attraction? a qualitative study. *Empirical Musicology Review 6*, 4 (2011), 214–230.

[29] GERSHON, M. *The Second Brain: A Groundbreaking New Understanding of Nervous Disorders of the Stomach and Intestine.* HarperCollins, 1999.

[30] GIORDANO, L. B., EGERMANN, H., AND BRESIN, R. The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity. *PLoS ONE 9* (2014).

[31] GOMEZ, P., AND DANUSER, B. Relationships between musical structure and psychophysiological measures of emotion. *Emotion 7*, 2 (2007), 377–387.

[32] GONZALES, C. R., AND WOODS, E. R. *Digital Image Processing*. Prentice Hall, 2007.

[33] HARAWAY, D. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies 14*, 3 (1988), 575–599.

[34] HEDDER, K. M. A new non-verbal measurement tool towards the emotional experience of music. `http://essay.utwente.nl/60499/1/MSc_Hedder%2C_M.K..pdf` (8. Jan. 2015), 2010.

[35] HELFFERICH, C. *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews.*, 4 ed. VS Verlag für Sozialwissenschaften — Springer Fachmedien Wiesbaden, 2011.

[36] HENRICH, J., HEINE, J. S., AND NORENZAYAN, A. The weirdest people in the world? *The Behavioral and Brain Sciences 33*, 2-3 (2010), 61–135.

[37] JANSSEN, J.-K. Quantified self: Körpermessgeräte. *C't Heft 18, heise* (2012).

[38] JENSEN, Q. S. A. Preliminary notes on othering and agency - marginalized young ethnic minority men negotiating identity in the terrain of otherness. In *CASTOR Seminar Logstor* (2009), forfatterne og forskningsgruppen CASTOR.

[39] JOLLIFFE, T. I. *Principal Component Analysis*, 2 ed. Springer Series in Staistics. Springer, 2002.

[40] JUSLIN, N. P., AND LAUKKA, P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin 129*, 5 (2003), 770–814.

[41] JUSLIN, N. P., AND LAUKKA, P. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research 33* (2004), 217–238.

[42] JUSLIN, N. P., AND VÄSTFJÄLL, D. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences* (2008), 559–621.

[43] KAPUR, A., VIRJI-BABUL, N., TZANETAKIS, G., AND DRIESSEN, P. F. Gesture-based affective computing on motion capture data. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction* (Berlin, Heidelberg, 2005), ACII'05, Springer-Verlag, pp. 1–7.

[44] KOPIEZ, R., DRESSEL, J., LEHMANN, M., AND PLATZ, F. *Vom Sentographen zur Gänsehautkamera - Entwicklungsgeschichte und Systematik elektronischer Interfaces in der Musikpsychologie*. Tectum Verlag, 2011.

[45] KREBS, J. R., AND DAWKINS, R. *Animal signals: Mind-reading and manipulation*, 2 ed. Oxford, England: Blackwell, 1984, pp. 380–402.

[46] LARSEN, J. T., MCGRAW, A. P., AND CACIOPPO, J. T. Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology 81* (2001), 684–96.

[47] LEMAN, M. *Embodied Music Cognition and Mediation Technology*. MIT Press, London, 2008.

[48] LENZ, I. Geschlecht, herrschaft und internationale ungleichheit. In *Das Geschlechterverhältnis als Gegenstand der Sozialwissenschaften*, R. Becker-Schmidt and A.-G. Knapp, Eds. Campus Verlag, 1995.

[49] LEPA, S., HOKLAS, A.-K., GULJAMOW, M., AND WEINZIERL, S. Wie hören die Deutschen heute Musik? Trends und Basisdaten zur musikbezogenen Audiomediennutzung 2012 in Deutschland. *Media Perspektiven*, 11 (2013), 545–553.

[50] LONGHI, E. Emotional responses in mother-infant musical interactions: A developmental perspective. In *Emotional Responses to Music: The Need to Consider Underlying Mechanisms. Behaviroal and Brain Sciences*, V. D. Juslin, N. P., Ed. 2008, pp. 586–587.

[51] LUCHT, P. Usability und Intersektionalitätsforschung - Produktive Dialoge. In *Fachtagung Gender-UseIT 2014 (#GUI2014): HCI, Web-Usability und UX unter Gendergesichtspunkten*, N. Marsden and U. Kempf, Eds. DE GRUYTER Oldenbourg, 2014.

[52] LYKARTSIS, A., PYSIEWICZ, A., VON COLER, H., AND LEPA, S. The emotionality of sonic events: Testing the Geneva Emotional Music Scale (GEMS) for popular and electroacoustic music. *Music & Emotion (ICME3)* (2013), n.n.

[53] MAGOWAN, F. Performing emotion, embodying country. In *Performing Gender, Place, and Emotion in Music - Global Perspectives*, F. Magowan and L. Wrazen, Eds. University of Rochester Press, 2013.

[54] MCCLARY, S. *Feminine Endings - Music, Gender, and Sexuality*. University of Minnesota Press, 1991.

[55] MCCLARY, S., AND WALSER, R. Theorizing the Body in African-American Music. *Black Music Research Journal 14*, 1 (1994), 75–84.

[56] MENARD, W. S. *Logistic regression: From introductory to advanced concepts and applications*. SAGE Publications Inc., 2010.

[57] NAGEL, F., KOPIEZ, R., GREWE, O., AND ALTENMÜLLER, E. Emujoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods 39*, 2 (2007), 283–290.

[58] NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., AND WASSERMAN, W. *Applied Linear Statistical Models*, 4 ed. McGraw-Hill, 1996.

[59] NORTON, B. Engendering emotion and the environment. In *Performing Gender, Place, and Emotion in Music - Global Perspectives*, F. Magowan and L. Wrazen, Eds. University of Rochester Press, 2013.

[60] PARNCUTT, R. *The perception of pulse in musical rhythm.* Royal Swedish Academy of Music, 1987, pp. 127–138.

[61] PELINSKI, R. Embodiment and Musical Experience. *Trans. Revista Transcultural de Música* (2005).

[62] PINHERIO, J. C., AND BATES, D. M. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing 6* (1996), 289–296.

[63] RÖBKE, P. Musizieren als Affektgestaltung. In *Kultur der Gefühle - Wissen und Geschlecht in Musik, Theater, Film*, I. D. Ellmaier, A., Ed. Böhlau Verlag, 2012.

[64] RUSSELL, A. J. A Circumplex Model of Affect. *Journal of Personality and Social Psychology 39*, 6 (1980), 1161–1178.

[65] SCHERER, K. R. Vocal correlates of emotional arousal and affective disturbance. In *Handbook of Social Psychophysiology: Emotion and Social Behavior*, H. Wagner and A. Manstead, Eds. Wiley New York, 1989, pp. 165–197.

[66] SCHERKE, K. Auflösung der Dichotomie von Rationalität und Emotionalität? Wissenschaftssoziologische Anmerkungen. In *Emotionen in Geschlechterverhältnissen: Affektregulierung und Gefühlsinszenierung im historischen Wandel*, S. Flick and A. Hornung, Eds. transcript Verlag, 2009, pp. 23–42.

[67] SHAVER, P., SCHWARTZ, J., KIRSON, D., AND O'CONNOR, C. Emotion knowledge: Further explorations of a prototype approach. *Journal of Personality and Social Psychology 52* (1987), 1061–1086.

[68] SIEVERS, B., POLANSKY, L., CASEY, M., AND WHEATLEY, T. Music and movement share a dynamic structure that supports universal expression of emotion. *PNAS 110*, 1 (2013), 70–75.

[69] SOLEYMANI, M., CARO, N. M., AND SCHMIDT, M. E. The mediaeval 2013 brave new task: Emotion in music. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop Barcelona*. 2013.

[70] SOLEYMANI, M., CARO, N. M., SCHMIDT, M. E., SHA, C.-Y., AND YANG, Y.-H. 1000 songs for emotional analysis of music. *Proceedings of CrowdMM' 2013, ACM* (2013).

[71] TORRES-ELIARD, K., LABBÉ, C., AND GRANDJEAN, D. Towards a Dynamic Approach to the Study of Emotions Expressed by Music. In *Intelligent Technologies for Interactive Entertainment - 4th International ICST Conference*, A. Camurri and C. Costa, Eds. Springer Berlin Heidelberg, 2012, pp. 252–259.

[72] TROST, W., SCHÖN, D., LABBÉ, C., PICHON, S., GRANDJEAN, D., AND VUILLEUMIER, P. Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage 103* (2014), 55–64.

[73] TZANETAKIS, G. Audio feature extraction. In *Music Data Mining*, T. Li, M. Ogihara, and G. Tzanetakis, Eds. Boca Raton : CRC Press, 2012, pp. 43–74.

[74] VUOSKOSKI, J. K., AND EEROLA, T. The role of mood and personality in the perception of emotions represented by music. *Cortex 47*, 9 (2011), 1099–1106.

[75] WALLBOTT, G. H. Bodily expression of emotion. *European Journal of Social Psychology 28* (1998), 879–896.

[76] YULE, G. U., AND KENDALL, M. G. *An Introduction to the Theory of Statistics*, 14 ed. Charles Griffin & Co, 1968.

[77] ZENTNER, M., AND EEROLA, T. Self-report measures and models. In *Handbook of music and emotion: Theory, Research, Applications*, V. D. Juslin, N. P., Ed. Oxford University Press, 2010, pp. 187–221.

[78] ZENTNER, M., GRANDJEAN, D., AND SCHERER, K. R. Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion (Washington, D.C.) 8*, 4 (2008), 494–521.

Appendices

Appendix A: Abbreviations

APP	Application running on a Mobile Device
FFT	Fast Fourier Transform
GEMS	Geneva Emotion Music Scales
MIR	Music Information Retrieval
MP3	MPEG-1 Audio Layer 3
MRS	Music Recommender Systems
PCA	Principle Component Analysis

GEMS Results per Sample

Visualization of Features

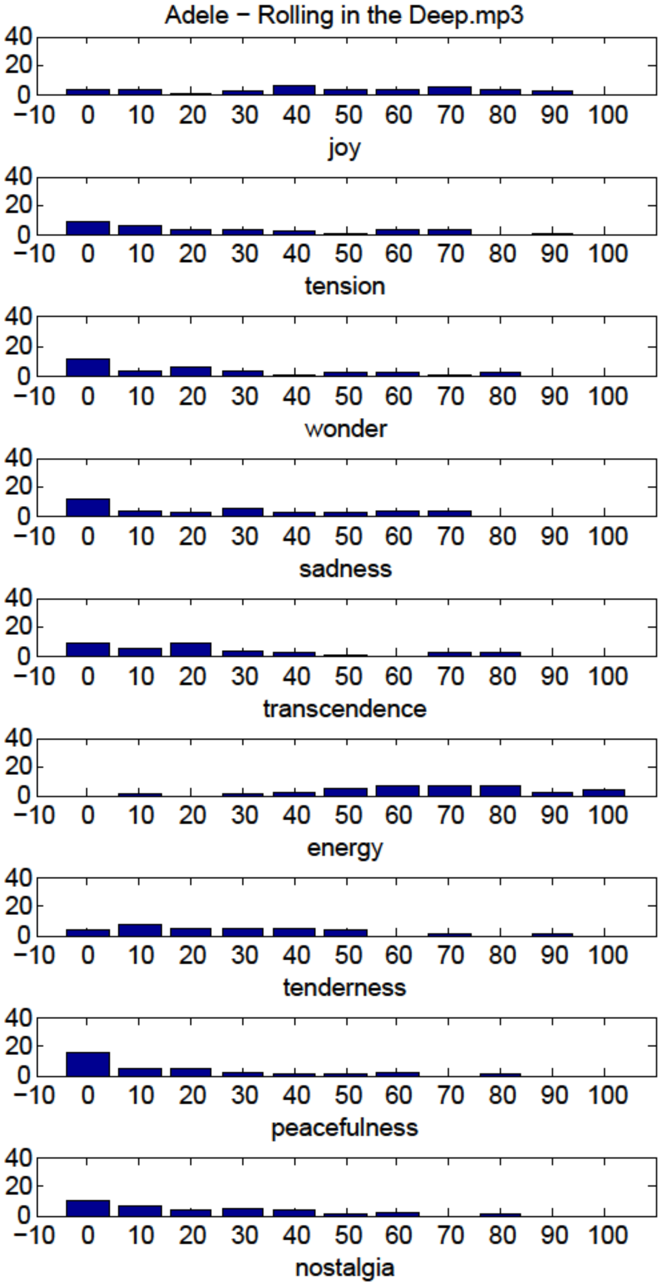


Figure 5.1: This figure shows the distribution of GEMS for *Rolling in the Deep*.

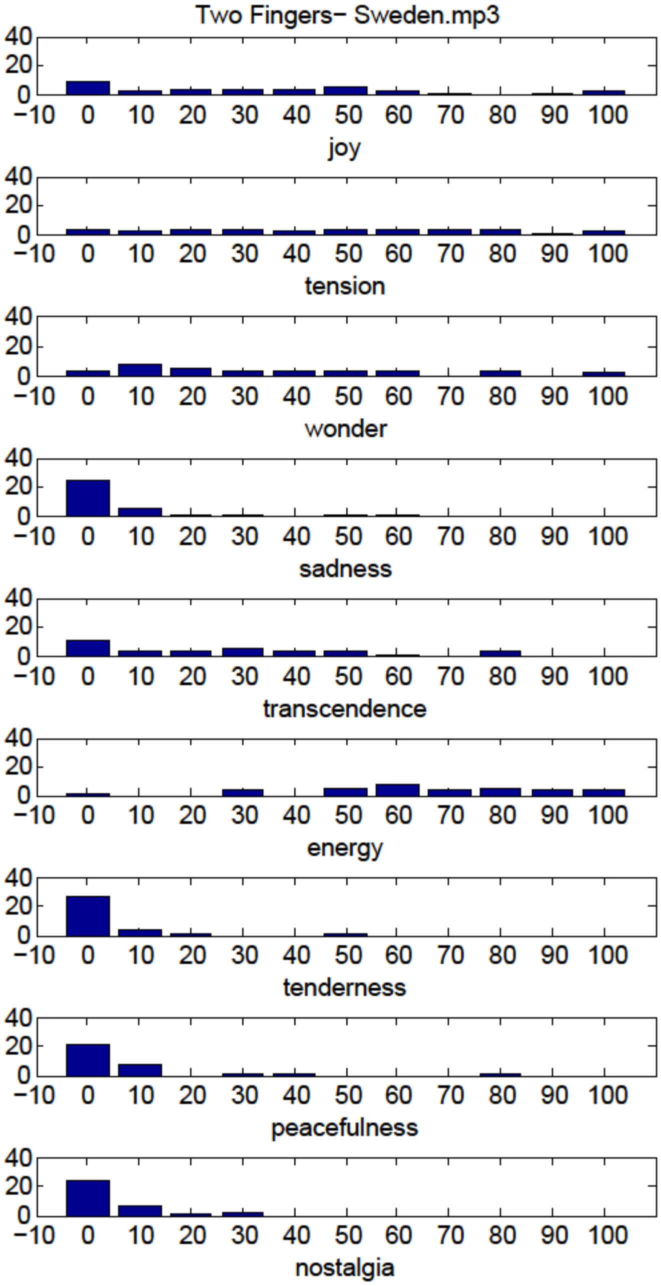


Figure 5.2: This figure shows the distribution of GEMS for *Sweden*.

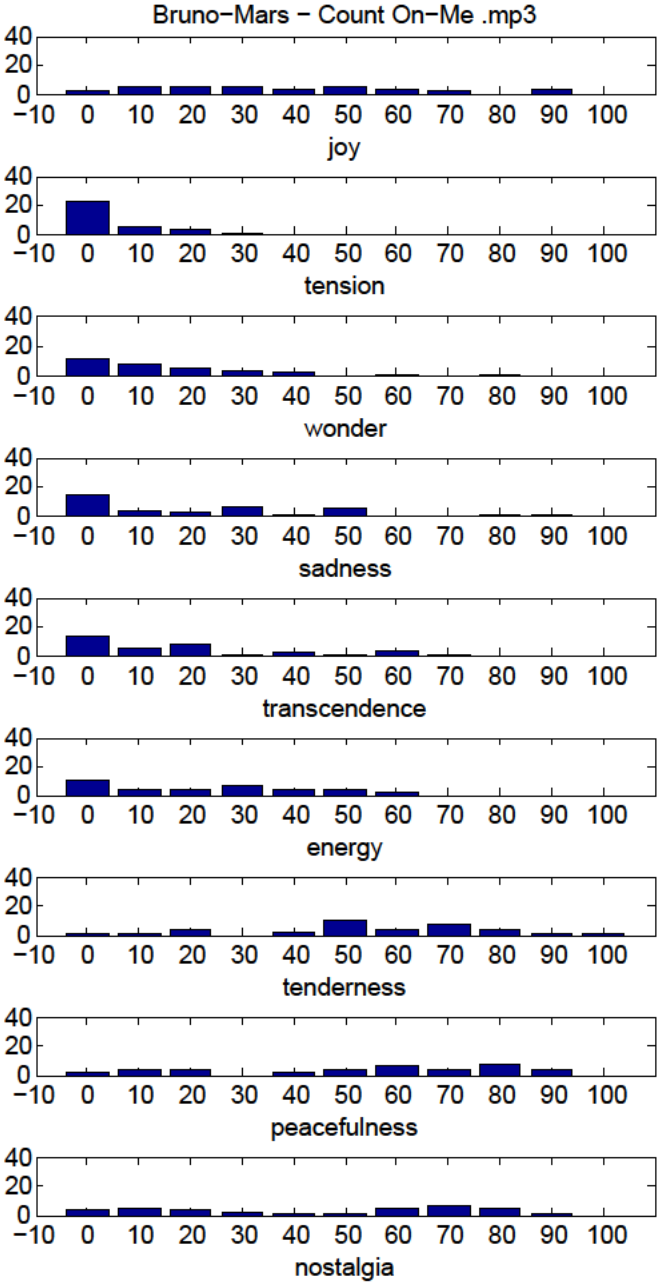


Figure 5.3: This figure shows the distribution of GEMS for *Count on Me*.

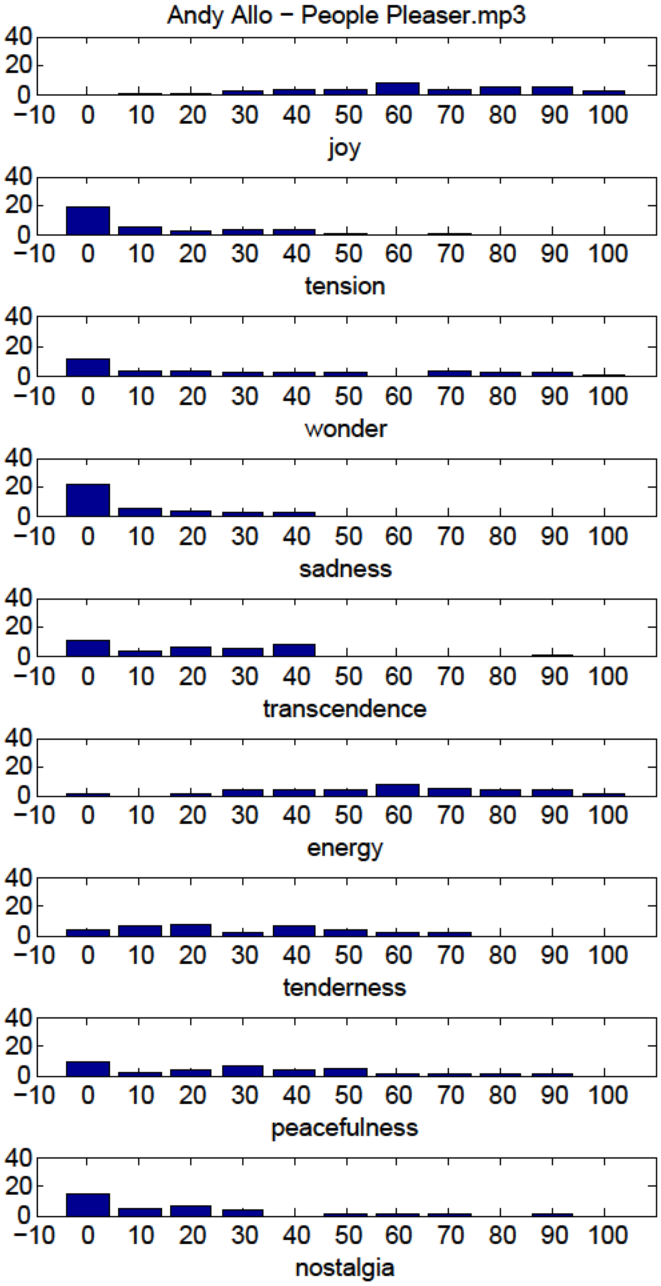


Figure 5.4: This figure shows the distribution of GEMS for *People Pleaser*.

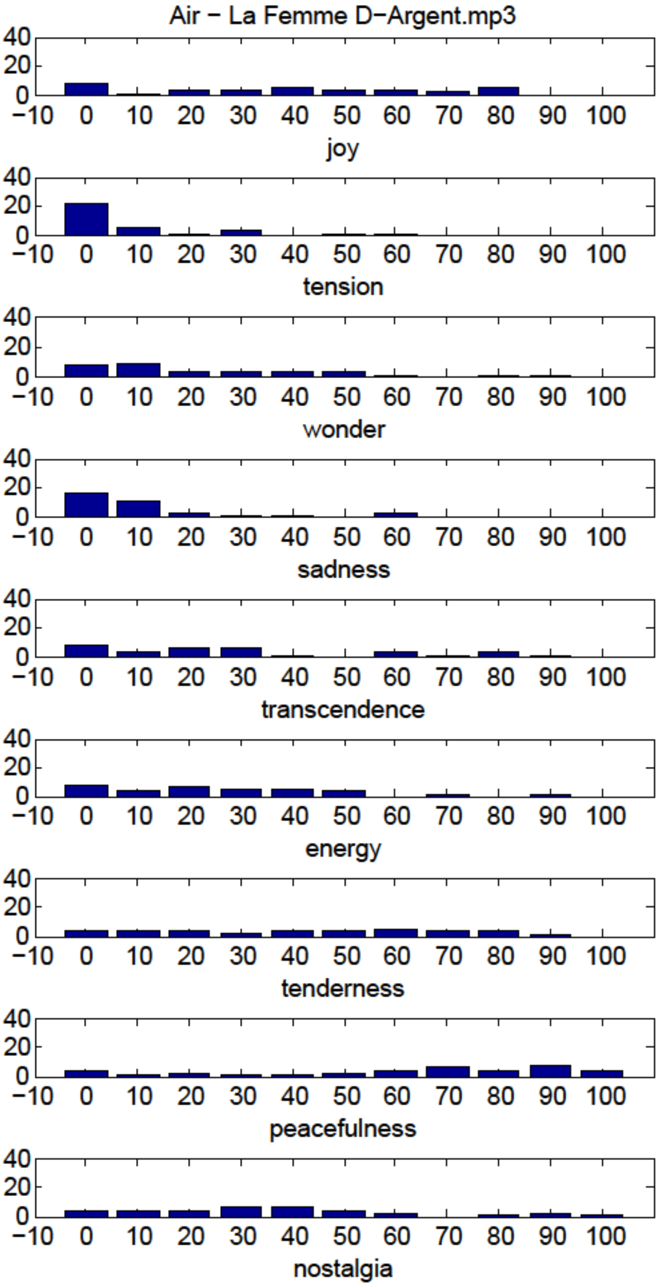


Figure 5.5: This figure shows the distribution of GEMS for *La Femme d'Argent*.

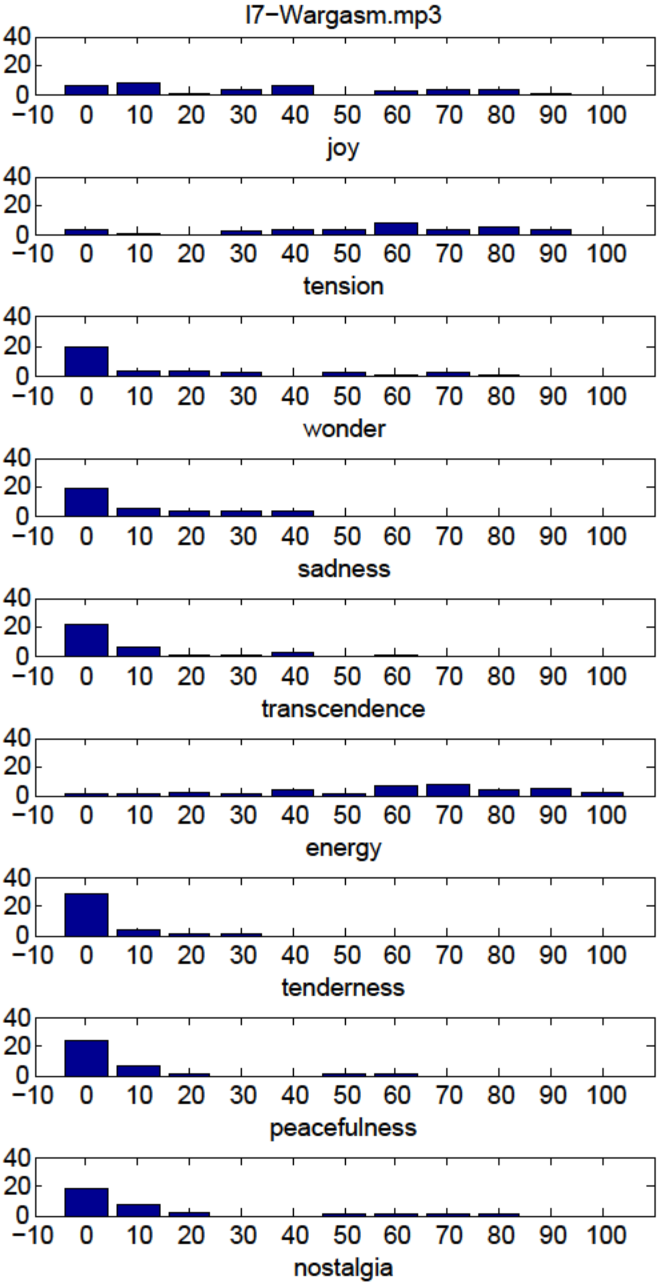


Figure 5.6: This figure shows the distribution of GEMS for *Wargasm*.

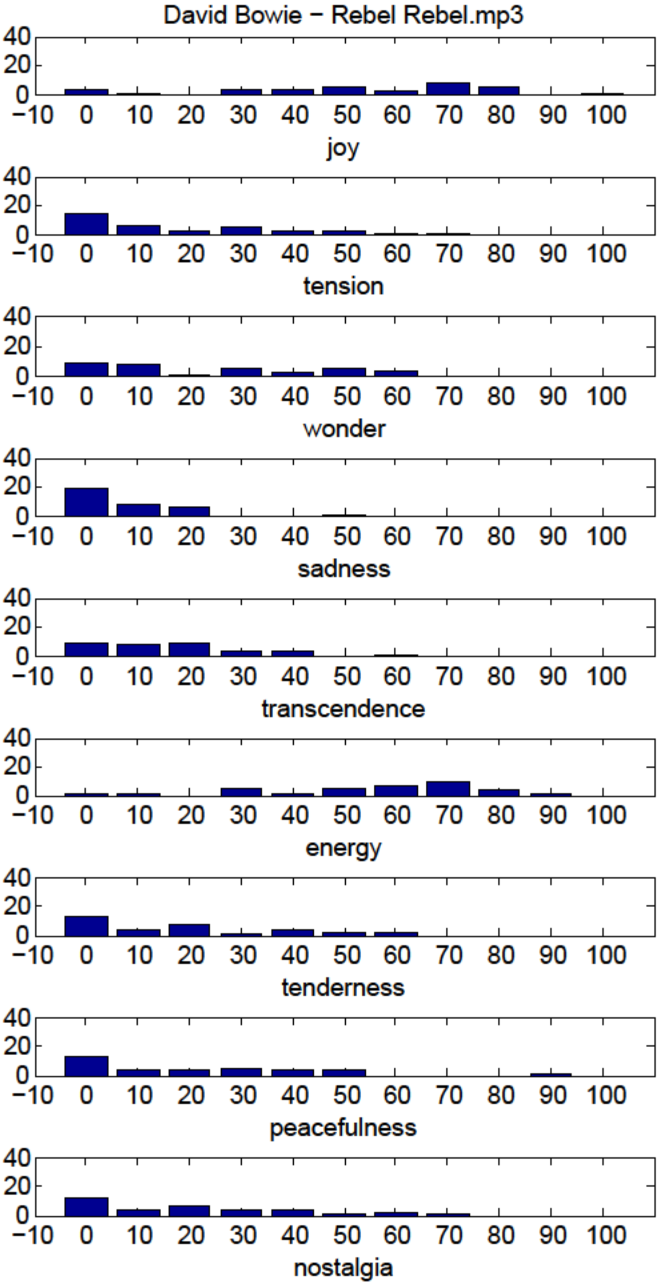


Figure 5.7: This figure shows the distribution of GEMS for *Rebel Rebel*.

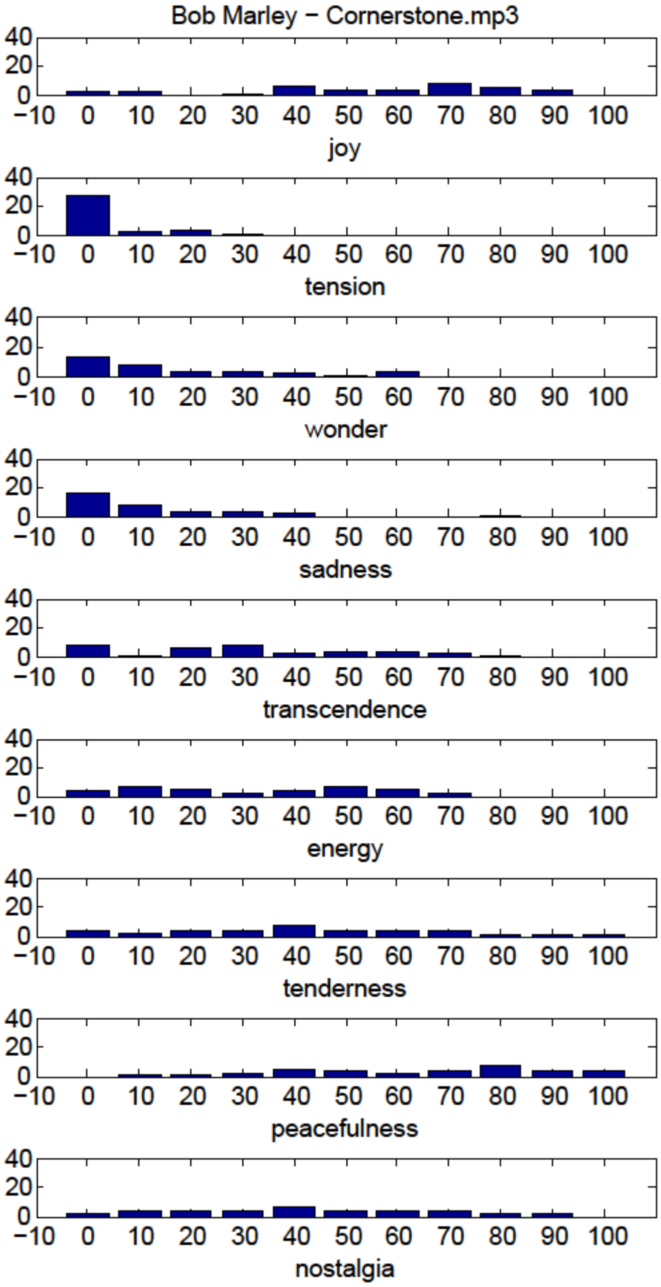


Figure 5.8: This figure shows the distribution of GEMS for *Corner Stone*.

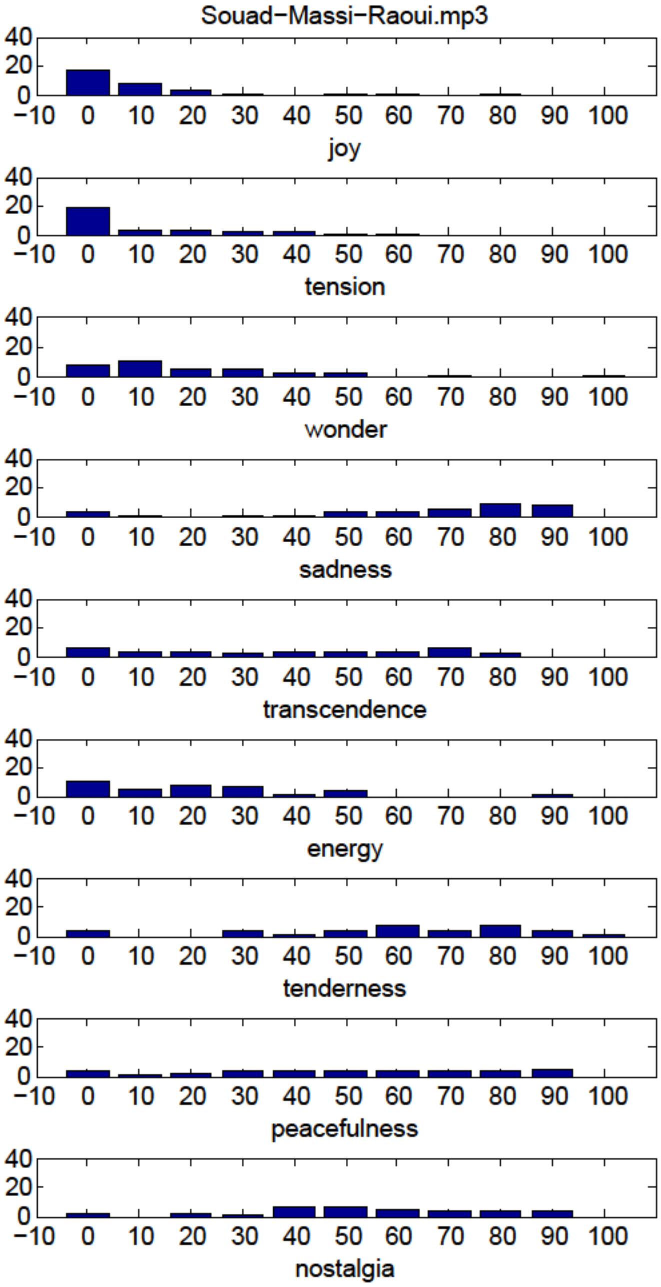


Figure 5.9: This figure shows the distribution of GEMS for *Raoui*.

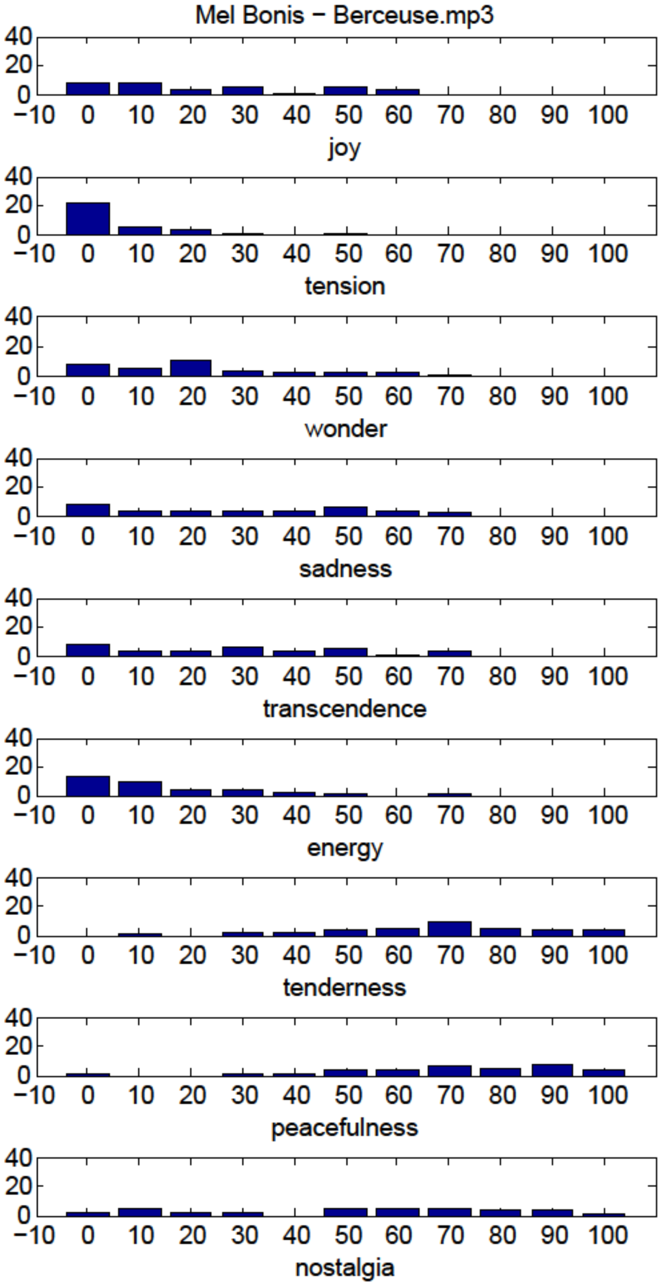


Figure 5.10: This figure shows the distribution of GEMS for *Berceuse*.

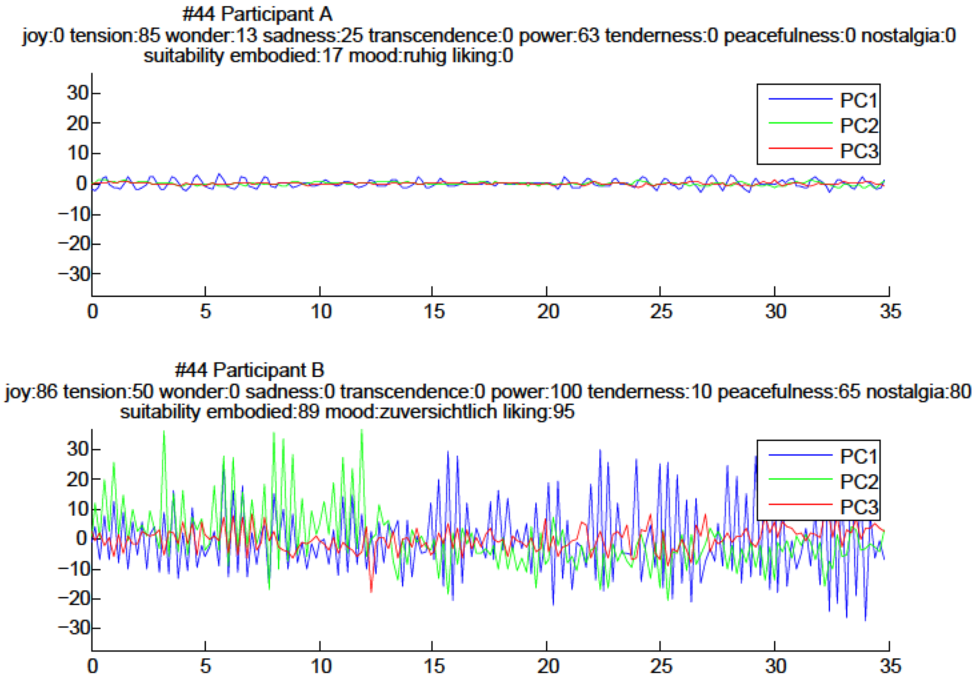


Figure 5.11: The figure shows the corporeal and verbal articulations of two participants for *Wargasm*.

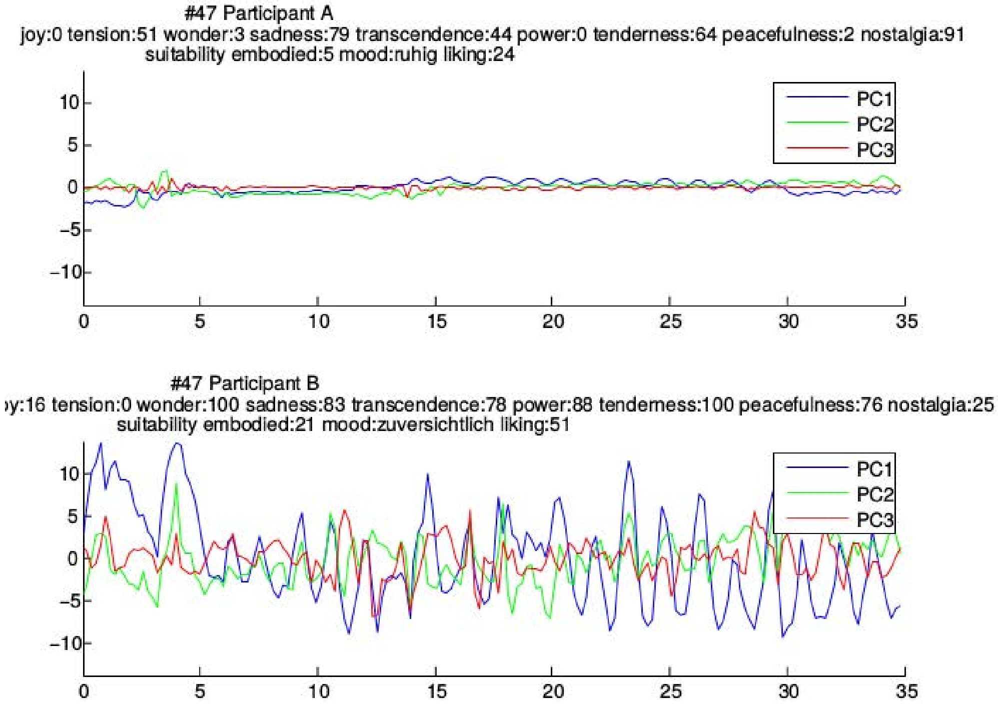
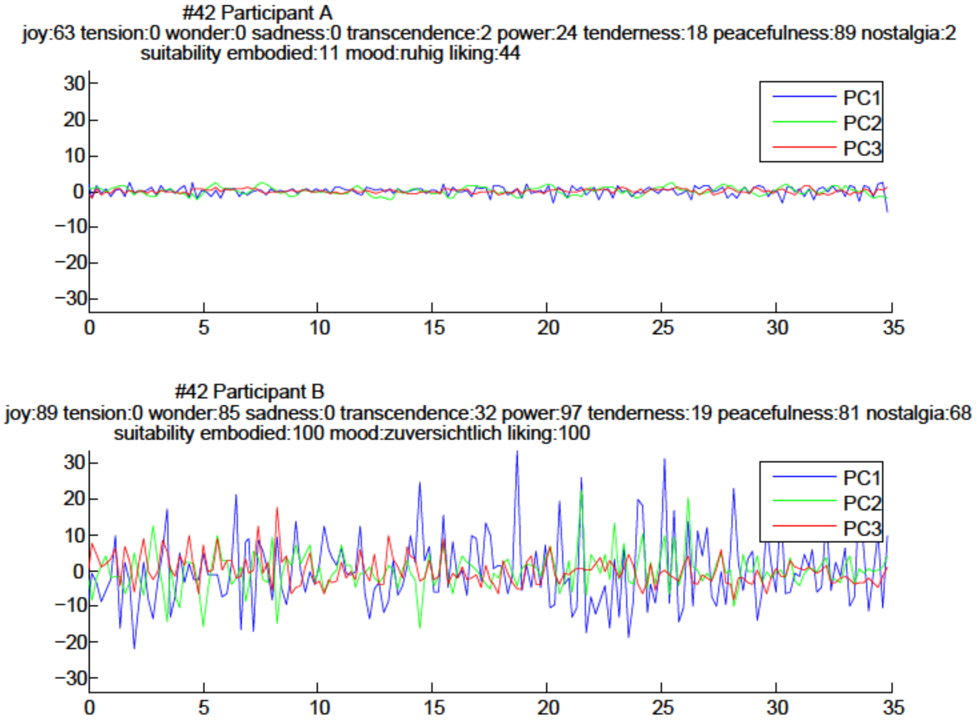


Figure 5.12: The figure shows the corporeal and verbal articulations of two participants for *Raoui*.



#42 Participant A
joy:63 tension:0 wonder:0 sadness:0 transcendence:2 power:24 tenderness:18 peacefulness:89 nostalgia:2
suitability embodied:11 mood:ruhig liking:44

#42 Participant B
joy:89 tension:0 wonder:85 sadness:0 transcendence:32 power:97 tenderness:19 peacefulness:81 nostalgia:68
suitability embodied:100 mood:zuversichtlich liking:100

Figure 5.13: The figure shows the corporeal and verbal articulations of two participants for *People Pleaser*.

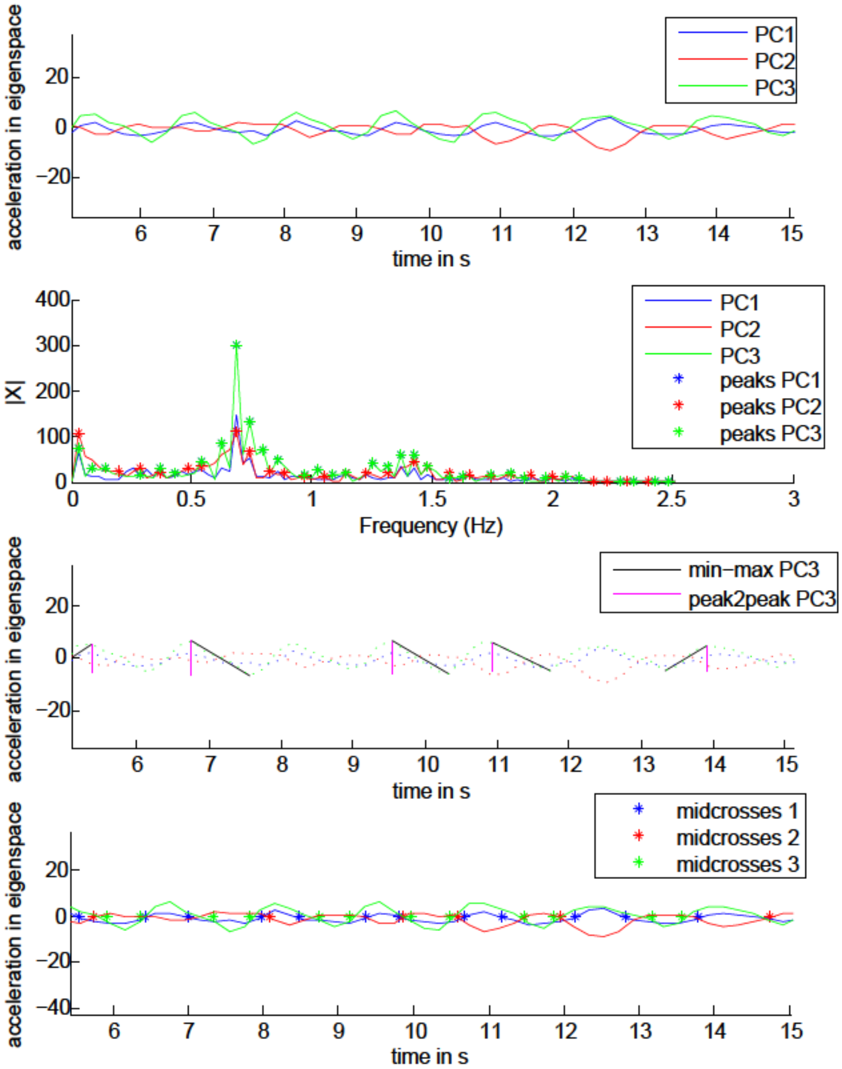


Figure 5.14: The figure shows the features for *Raoui* from one participant.

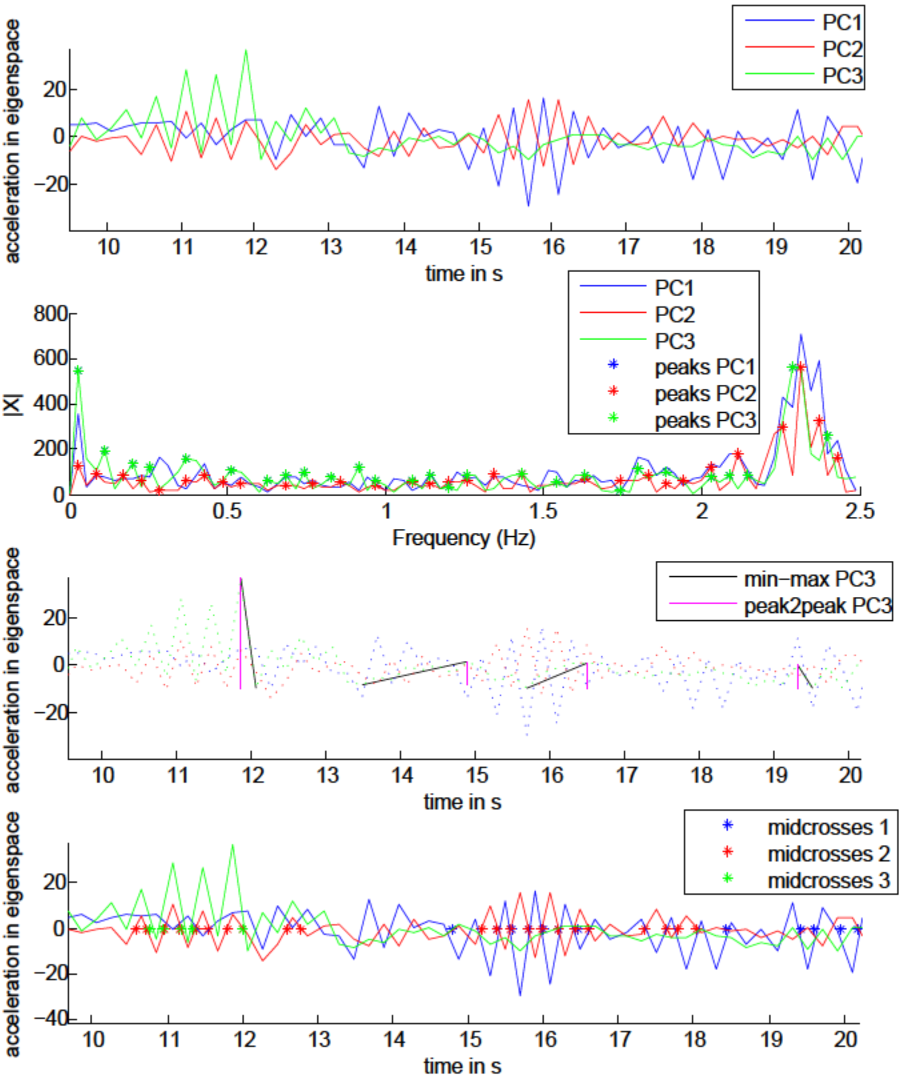


Figure 5.15: The figure shows the features for *Wargasm* from one participant.