Diana Kayser

Do We Feel What We Perceive and Perceive What We Feel?

A Review of Methods in the Study of Emotions in Music

Master's thesis in Musicology

Department of Musicology
University of Oslo
Spring 2016

Do We Feel What We Perceive and Perceive What We Feel?

A Review of Methods in the Study of Emotions in Music

Master's thesis

Diana Kayser

Abstract

The study of emotions in music is a popular topic in the field of music psychology, but methods used to investigate this phenomenon rather focus on how emotions in music are perceived rather than how they are experienced or felt. This problematic leads to the two main objectives of this thesis: In the first part, a review of methods used for studying emotions in music will be given and it will further be evaluated to what extent the results obtained with these methods give insight about the two distinct phenomena referred to as *perceived emotions in music* and *music-evoked emotion.* The methods that have mainly been used in this area are self-report, and physiological measurements or a combination of the two. Self-report is based on introspection and can be classified as *first-person description* of musical experience, whereas methods based on repeatable measurements fall into the category of *third-person descriptions*. Both these standard methods are highly based on interpretation and retrospective evaluation of an experience, which makes it questionable if a subjective and partly subconscious phenomenon like an emotion can reliably be inferred this way. In the second part of this thesis a methodological framework for studying music-evoked emotion will be developed that takes a *second-person description* as its point of departure. The main characteristic of a second-person description is that methods following this framework can give insight to a subjective experience via articulation and expressive behaviour, without requiring introspection and or interpretation. As facial expressions for a specific set of so-called *basic emotions* are suggested to be universal across cultures and provide a reliable indicator of an emotional experience, it will be discussed to what extent these expressions could be used as a form of second-person description to study music-evoked emotions. A pilot-study using this framework has been conducted and the results will be presented.

Foreword

This thesis was written for my Master’s degree in Musicology at the University of Oslo. Writing this thesis has been a very interesting journey from on the start until the finishing line. The first time I have heard about the significance of facial expressions for the study of emotions was in the TV-series ‘Lie to me’ which is based on the studies byPaul Ekman, and was consulted by a number of behavioural psychologists. That was about four years ago, and ever since I was hoping to find out if facial expressions of emotions can be used to study emotions in music. If my findings concerning this relationship have been positive or negative can be found somewhere in this large amount of text that I have produced during the last months.

Nevertheless, writing this thesis would not have been possible without the help of some non-fictional people as well. First and foremost I want to thank Alexander Refsum Jensenius who has been my supervisor for the last months of this thesis. He provided very useful and valuable comments, and pushed me in the right directions at the right time. I also would like to thank Rolf Inge Godøy who has been my supervisor in the first months and who believed in this project right from the start. His lectures and presentations were a great inspiration. The same can be said for Hallgjerd Aksnes who gave me a very motivational feedback in addition to some very helpful literature tips. Another thanks goes to Minho Song who helped me with some of the photographs that can be found in chapter 5. As there are many more people at this institute I would like to thank than there is space here, I would like to give a very special THANK YOU to everyone who joined me for coffee- or lunchbreaks, or did let me take naps in their offices when the exhaustion was too big to be handled with coffee. Words cannot express how grateful I am for being allowed to call you my friends and colleagues.

I would also like to thank Eric Clarke for a fruitful discussion and helpful comments at the SysMus Conference 2015. Last but not least I have to thank my family for all the great support and motivation that I have experienced throughout the last years. I am very grateful for having you! I am also very grateful for my friends Luc and Swea who motivated me from far away, and Thea, who could not have been of bigger help!

Table of Contents

Chapter 1

Introduction

This chapter provides some introductory remarks as well as the motive for this thesis. Further, research questions I am going to address in this framework will be presented. Relevant definitions will be provided, and the methodology used to approach the research questions will be given. The thesis structure is going to be presented in the end of this chapter.

1.1 Motivation

Have you ever been listening to music and had the feeling that some songs just seem to trigger a certain feeling in you? With the effect that you, for example, experience a warm feeling in your chest and feel light and uplifted – or suddenly feel sad or depressed? This phenomenon referred to as music-evoked emotion has amazed me since I can remember. From on early childhood I experienced moments where certain songs would just *feel right*. I remember using music as a ‘tool’ to regulate my emotions from the age of 10 or even earlier. Because my knowledge of the English language was still very limited at the time, I was not able to understand much of the lyrics of the Britpop- and Indie- songs my sister introduced me to, but still I had a feeling of *understanding* them in a way that was much deeper. The music spoke to me in its own language: the language of emotion.

The strongest emotional experience in response to music I remember has occurred to me at a concert of the band Death Cab for Cutie: When they started playing *Transatlanticism* – a song that I have never heard before – I suddenly found myself in tears. And even now, about 10 years later, this particular song still has the power to evoke this feeling of sadness in me - though in a weaker form than in the live concert. I soon learned that other people do experience various emotions in response to music as well. It was about that time that I read the book *Musicophilia* by the neurologist Oliver Sacks (Sacks, 2007). In this book he describes various cases of neurological conditions that lead to partly puzzling experiences with music. With its focus on music and the brain, this book did not only introduce me to the field of music psychology and cognitive neuroscience of music, but also triggered a strong motivation to study musicology myself, and investigate and understand the phenomenon of emotions in music that has accompanied me all my life.

During the last decade I studied musicology, psychology, philosophy of the mind and cognitive neuroscience. For studying emotions in music an interdisciplinary background like that is necessary in order to be able to grasp such a complex phenomenon.

1.2 Main Objectives and Limitations

The main research question that I am going to address is:

> *How can we study music-evoked emotions rather than perceived emotions in music?*

In this thesis I will define *emotions* as affective states that are characterised by their brief duration and their episodic progression. Emotions occur in response to phenomena in the internal or external environments and can be inferred from spontaneous and involuntary changes in physiology and expression. A more detailed discussion about the challenges that have to be dealt with by defining such a complex phenomenon will be given in chapter 2. There is a limited set of emotions that is discussed to be universal cross-culturally. These so-called *basic emotions* share some of the same characteristics and are suggested to be unique in expression (Ekman, 1994).

Perceived emotions in music and *music-evoked emotions* both describe emotions that are somehow experienced in connection with music. The former term describes the perception or recognition of emotional characteristics in music based on prior knowledge, meaning that a melody can be perceived as ‘sad’ without evoking the respective feeling in the listener. The latter term refers to emotions that are experienced or felt in response to music, meaning that one is subjectively experiencing a feeling of ‘sadness’ in reaction to music. Perceived emotions in music and music-evoked emotions can overlap, but do not necessarily need to.

The main research question introduced earlier can be divided into three sub-questions. The first question that needs to be answered is:

i.) *Which methods have been used to study emotions in music?*

As stated previously, emotions in music are not to be seen as one single phenomenon but as either recognized (*perceived*) or felt (*evoked*). Because these are two very different entities, we would expect research methods to differ. The question that becomes relevant after these methods have been identified is:

ii.) *Do those methods give results that are reliable and have ecological validity for studying music-evoked emotions?*

Once methods are described, we have to take a closer look at these and have to identify the capability they provide to obtain results that are both *reliable* and have *ecological validity* for the study of music-evoked emotions. Reliability can be reached simply by actually measuring the data one was aiming to measure (Kelley, 1927:14). Important here is also to avoid *demand characteristics*. Demand characteristics are suggested to be a problematic artefact in experimental research. It is suggested that due to blundering formulation of instructions and/or tasks, participants form an understanding of the research hypothesis and subconsciously adjust their behaviour to please the researcher. Ecological validity can be understood as the urge to study a specific phenomenon by using methods, materials and settings that approximate the real world (Brewer, 2000). The idea behind this is that the results obtained might be more prone to generalization if the conditions of reliability and ecological validity are fulfilled.

After the methods for studying music-evoked emotions have been identified, a third sub-question has to be discussed:

iii.) *Is there a more reliable way in which music-evoked emotions can be studied?*

This question follows the presumption that there are ways to study music-evoked emotions that are more reliable and ecologically valid than previous methods. For this approach the study of *facial expressions of emotion* will be in focus. Facial expressions of emotion can be one means to infer an emotional experience. Evidence from cross-cultural studies suggests that facial expressions that occur together with basic emotions are universal across cultures (Ekman, 1994).

The aim of this thesis is to give a detailed review of the study of music and emotion with the main focus on methods that have been applied as well as the reliability and ecological validity of the results that can be obtained with these methods. Based on the reviewed material that will be presented in the first part, I am going to develop a novel framework that is suited for studying music-evoked emotions, and gives the possibility to better differentiate between perceived emotions in music and music-evoked emotions. A test of this method has been conducted in form of a small pilot-study which will be presented in the second part of this thesis.

1.3 Methodology of Thesis Work

Before I started my research, I formulated the hypothesis that facial expressions of emotion can be used as an indicator to infer an emotional experience in response to music. Searching for studies in different online databases (university library, jstor, google scholar) on this particular topic did not bring any results, which forced me to broaden my focus on methods used on studying emotions in music in general and suggesting a novel framework for studying music-evoked emotions by using facial expressions of emotion. I proceeded by getting an overview about the field of music and emotion. To approach this, I was reading through relevant and influential monographs on the topic. After getting an overview about the theories underlying music and emotion and the methods being used, I formulated my main research question. With that in mind, I searched through the already mentioned online databases using different keywords to find studies on emotions in music. These were excerpted and analysed. Subsequently, I looked for literature on emotions in general by browsing through the shelves of the university library and using the databases described above. I compared the definitions used critically and made an overview about the different theories and approaches that have been used to explain the occurrence and origin of emotions. The same steps were followed for the specific area of facial expressions of emotion and the underlying theories that are discussing the function and meaning of facial displays.

Based on the results of this review study I have been developing a framework that suggests that using facial expressions of emotion can provide more specific data on music-evoked emotion. The method suggested is supposed to be applicable for reliably studying music-evoked emotion cross-culturally in an ecological valid way. To develop this framework I considered different technologies and methods that could be used to measure facial expressions of emotion. All of these methods and technologies have been analysed and the advantages and disadvantages have been discussed. Finally, the most promising method has been picked and implemented into the experimental design. A pilot experiment based on this framework has been conducted and evaluated.

1.4 Structure

In the first part of this thesis I will give an overview of the study of emotions. Chapter 2 starts with a discussion about the challenge of defining the complex phenomenon emotion constitutes. I will present different approaches and models that have been developed with the intent to give an explanation on the occurrence of emotion and suggestions about underlying mechanisms. In chapter 3 I will give an overview about theories and methods that have been devised in the study of facial expressions and discuss the relevance of facial expressions in measuring emotions. The second part of this work provides an overview of the study of music and emotion. A review on theories and methods will be given in chapter 4. This chapter will also include a discussion about the reliability of classical methods that have been used to study music evoked emotions. In chapter 5 I am going to discuss to what extent the methods that have been used to study facial expressions of emotion can be applied to investigate music-evoked emotions. I developed a novel experimental design for studying music-evoked emotions with using facial expressions of emotion, and will present a pilot study and preliminary results in chapter 6.

Chapter 2

Emotion

Before looking at the topic of emotions in music, it is important to get an understanding about what an emotion is, and how it can be studied. The following chapter will provide a detailed discussion about the challenge of finding a definition of emotion. Different approaches about the origin of emotions will be discussed and models attempting to classify emotions will be presented. Subsequently, ways of inferring if an emotion has been experienced will be introduced. Finally, I will give an overview about the influences cognition and emotion have on each other.

2.1 Definition

Even though emotions accompany us in many situations and circumstances and seem to be such a 'normal' thing, psychologists are still struggling in getting a complete grasp of this phenomenon. Already trying to find an agreeable definition seems to be rather challenging, as there is no consensus about which characteristics can be ascribed to emotion, and which features differentiate emotion from other affective states, such as mood and feelings (Zentner & Eerola, 2010). Keltner and Gross (1999) define emotions as

> *episodic, relatively short-term, biologically-based patterns of perception, experience, physiology, action and communication that occur in response to specific physical and social challenges and opportunities* (p.468).

Although including many characteristics that can be found in other definitions of emotion, Keltner and Gross raise more questions than they answer. They are not very clear about if these so-called 'patterns' have to consist of all the mentioned components simultaneously, and if emotional responses are solely reactions to the outside world, or if they might as well occur in reaction to internal processes (e.g. imagery, thought processes).

In their own attempt to define the phenomenon, Niedenthal et al. (2006) state that: "in addition to subjective feeling, […] emotions also contain action tendencies, physiological appraisals, and expressive motor behaviour" (pp.6-8). This definition can be seen as very controversial as well: Even though emotions often result in subjective feelings, emotions can occur without knowing the cause (Johnson-Laird & Oatley, 1989:90) or being consciously experienced as being *felt*. There is a major difference between *emotions* and *feelings,* which will be discussed in the following section of this chapter. Two important points Niedenthal et al. make in their definition is the connection between emotion, expressive motor behaviour, and the occurrence of changes in physiology. These two components are of high significance when it comes to measuring emotions, which will be outlined further down.

In the *Handbook of Music and Emotion*, Sloboda and Juslin (2010) state that emotions are brief in duration and further can be defined as

> *rapidly changing responses to potentially important events (subjective challenges and opportunities) in the external or internal environment, usually of a social nature, which involve a number of subcomponents (cognitive changes, subjective feelings, expressive behaviour, and action tendencies that are more or less 'synchronized')* (pos 1822)[1].

With 'rapidly changing responses' Sloboda and Juslin refer to changes in the autonomic nervous system that lead to measurable physiological changes, e.g. changes of electrical characteristics in the skin (= electrodermal activity) and heart rate. As opposed to Keltner and Gross they clearly involve events in the external and internal environment as being capable of eliciting emotion. Even though this definition is a little more precise and gives us a much better understanding about what an emotion is, it also contains points that require a closer look. For example has the suggested synchronization of subcomponents mentioned in this definition to be regarded with caution. Whereas expressive behaviour and action tendencies (e.g. a fight or flight reaction as a result of *fear*) are spontaneous, automatic, mostly involuntary, and thus most likely in synchrony with an underlying emotion, the same cannot be said for cognitive changes and subjective feeling. It has been suggested that emotions

[1] References taken from the *kindle* version of a publication will be referred to as 'pos'.

occur on the subconscious level (see Zentner & Eerola, 2010), which means that naturally there must be a delay between the subconscious onset of an emotion and the point where one becomes consciously aware of that emotion. Adolphs (2002) criticises that the distinction between the emotional reaction (manifested in the physiological emotional response), and the feeling of the emotion is often overlooked. It can thus be assumed, that emotions can occur without crossing the barrier between subconscious and conscious. This suggests that even if one becomes consciously aware of an emotion, there will always be a delay due to several computational processes - and thus no synchrony. The same counts for subjective feelings. In order to be able to experience a feeling, one has to be aware of the underlying emotion. One does not necessarily have to be aware of experiencing an underlying emotion at all times, so even though there might be a certain overlap between an affective state and a subjective feeling as a result of the underlying emotion, they do not always co-occur and most certainly do not synchronize. Expressive behaviour on the other hand can be seen as automatic, spontaneous behaviour and is thus in synchrony with the underlying emotion. The relationship between emotion and expressive behaviour will be further elaborated in the following chapter.

It can be concluded, that emotions are brief and episodic phenomena in response to events that are perceived in either the internal or external environment. They are spontaneous, involuntary, and automatic and lead to changes in physiology. Emotion does not necessarily have to involve cognition, but can possibly lead to changes in cognition. Vice versa it has been suggested that cognition is also capable of eliciting and influencing emotion (Yiend & Mackintosh, 2005). The relationship between emotion and cognition will be further outlined in the end of this chapter.

2.2 Emotion, Mood, Feeling – Similarities and Differences

In everyday life, the words *emotion*, *feeling*, and *mood* often are used interchangeably to refer to some sort of inner emotional state. Even though the phenomena underlying these terms are somehow related to each other and thus not dissimilar per se, they involve different characteristics we have to be aware of when talking about these concepts in the context of emotion research.

Sloboda and Juslin (2010) suggest that an *emotion* occurs in response to a salient stimulus whereas a less intense state without an immediately apparent cause is considered a *mood* (pos 1853). Further, Davies (2010) emphasizes that emotions are directed towards objects, whereas *moods* can be described as involving "rather general feelings" (pos 593). While emotions can be considered as having an onset, a finite duration, and an offset (Adolphs, 2002:24-25), moods can last for an undefined amount of time and usually cannot be measured in terms of defined phases like emotions. Whereas emotions are considered to feature communication, moods or feelings are not thought of serving this purpose. Zentner and Eerola (2010) define *feeling* as being the "subjectively experienced, consciously accessible part of an emotion" (pos 4572). This strengthens the argument made in the previous section that emotions occur on the subconscious level. Once they are perceived and noticed, they 'turn' into subjective feelings, which are the result of a cognitive interpretation of an emotion.

2.3 What Causes Emotions?

There are different theories and approaches trying to find probable explanations for why we experience emotions and which mechanisms might be involved in the occurrence of emotions.

The *Evolutionary Approach to Emotion* refers to an underlying biologically determined reason. It is believed that specific objects and events trigger emotional responses in a human being (Niedenthal, Krauth-Gruber, & Ric, 2006). This approach is also followed by Paul Ekman (1992) who considers events from our ancestral past to be influential for the development of the emotions we experience today (p.171). Johnson-Laird and Oatley (2002) share this view as well by stating that "each emotion […] prompts us in a direction which in the course of evolution has done better than other solutions in recurring circumstances that are relevant to goals".

Another theory is that emotions occur as a result of evaluation processes. The individual would link events to its goals and needs. In this case emotions would serve as responses of either attaining or maintaining a goal (e.g. happiness), or failing to do so (e.g. sadness) (Stein and Tabasso in Ekman, 1992:171). Niedenthal et al. (2006) refer specifically to events in the environment as being relevant for the evaluative process. This definition needs to be more specified by referring to not just the outer environment but also the inner environment of a human being (thought processes, imagery, etc.), especially because it has been argued elsewhere (Johnson-Laird & Oatley 1989:97) that at least basic emotions can be experienced without knowing the underlying cause. Furthermore, Ekman (1992) points out that emotions also do occur when being alone and without imagining other people (p.171).

As a last approach it has been suggested that the mechanisms that are involved in the occurrence of an emotion are a combination of evolutionary and evaluative processes (Niedenthal, Krauth-Gruber, Ric, 2006).

2.4 Emotion Models and Theories

There are different approaches that are trying to define the appearance of the experience of an emotion and find ways to categorise them. The two main theories in this attempt are presented in this section.

2.4.1 Discrete Emotions Theory

The *Discrete Emotions Theory* (also referred to as Categorical Emotion Theory) suggests that emotions are experienced as distinct categories (Sloboda & Juslin, 2010:1868). A part of this framework is the idea that a limited set of categories forms the so-called *basic emotions*. Basic emotions are considered to be a small set of discrete emotions that form the basis for the immense variety of emotion that can be experienced (Yiend & Mackintosh, 2005). Paul Ekman (1992) suggests that a basic emotion should not be characterised as being a single affective state but should rather be seen as a *family* of related states. The claim is that there is not one single expression for *anger* but several expressions with similar enough characteristics to belong to the *anger*-family. The expressions of the *anger*-family also need to be distinct enough from expressions of other emotion families. In his article, Ekman further suggests that an emotion family consist of what he calls "a *theme* and *variations*" (p.172). The theme can be seen as a core element that is found in all variations of that emotion. Ekman states, that "[v]ariations in the family of anger facial expressions are hypothesised to reflect whether or not the anger is controlled, [and] whether the anger is simulated or spontaneous […]" (p.172). It can thus be differentiated between voluntary (posed) and involuntary (spontaneous) expressions of emotion. The knowledge of this differentiation plays an important role in the experimental design developed for studying music-evoked emotion that will be presented in chapter 5.

There are different opinions amongst psychologists on how many and which emotions exactly are to be included in the set of basic emotions. According to Juslin and Sloboda (2010), basic emotions should:

> *(a) have distinct functions that contribute to individual survival, (b) are found in all cultures, (c) are experienced as unique feelings, (d) appear early in development, (e) involve distinct patterns of physiological changes, (f) can be inferred in other primates and (g) have distinct facial and vocal expression.* (pos 1868)

Paul Ekman (1992) defines nine characteristics that need to be fulfilled by an emotion in order to be classified as 'basic':

1. *Distinctive Universal Signals*

Distinct facial expressions (as well as vocal expressions) have to be observed for the selected emotion. There has to be a high agreement on labelling these expressions with an emotion term across cultures.

2. *Comparable Expression in other Animals*

Other primates generate similar facial expressions as humans.

3. *Emotion Specific Physiology*

Distinctive patterns of changes in the autonomic nervous system have to be evident. The changes in physiology prepare the organism to respond differently to different emotions that can be experienced.

4. *Universal Antecedent Events*

Ekman suggests that common elements in the contexts in which emotions occur can be expected. Evolution and social learning are thought to contribute to the events that lead to emotional experiences and thus the contexts in which they occur may vary within and across cultures. The elements should not be seen as determined but more on an abstract level. For example can physical or psychological harm be seen as an antecedent event for fear, whereas the loss of a loved one would be an element that underlies an antecedent even for sadness.

5. *Coherence in Response Systems*

For emotions that show a distinct pattern in the autonomous nervous system, a systematic relationship between expression and autonomic changes during emotion should be present. Also, Ekman assumes a relationship between facial expressions of emotion and distinct patterns in the activity of the central nervous system. He points out that it is important to only consider spontaneous facial expressions instead of voluntary ones that are exposed in social situations.

6. *Quick Onset*

It is believed that emotions can start so fast that one does not necessarily become aware of it. This might be due to the evolutionary significance of emotion as a response to the event that provoked the emotional response has to require a small amount of time. The occurrence of facial expressions can be observed milliseconds after stimulus presentation, which might be an indicator for the quick onset of an underlying emotion.

7. *Brief Duration*

Emotions are believed to not last longer than several seconds. Ekman argues, that if under certain conditions an emotion lasts longer than a few seconds, a closer look might show that the emotion may not have been evoked only once but repeatedly. That means we do not experience a single emotion for a longer period of time but a chain of the same emotion.

8. *Automatic Appraisal Mechanism*

Ekman proposes two appraisal mechanisms: The automatic appraisal mechanism is assumed to operate without one being aware of it and rather fast. This assumption has been made on behalf of the very short interval between stimulus presentation and emotional response. The extended appraisal mechanism Ekman proposes refers to the responses that are deliberate and include cognition. This mechanism seems to be more focused on reflection rather than reaction, as for the extended appraisal mechanism an emotion can, but does not have to, be underlying.

9. *Unbidden Occurrence*

Emotions often occur automatically, with little awareness (or no awareness at all), have a rapid onset, and involuntary changes in expression and physiology. An emotion is often experienced as something we did not chose to happen to us. Even though they occur without intention, they partly can be brought under volitional control. This can be a challenge, and the success in hiding or masking an emotion is also dependent on factors including individual differences, motivation and intensity of the emotion.

As of today there are five emotions that seem to fulfil all of the nine characteristics defined by Ekman: *anger*, *fear*, *sadness*, *disgust*, and *enjoyment* (Ekman, 1992:170). An analysis of emotion words by Johnson-Laird and Oatley (1989) supports this selection as their list contained the exact same categories. It is suggested that all other emotional states can be derived from the categories that form the basic emotions (Ekman, 1992; Sloboda & Juslin, 2010). Even though psychologists mostly agree on this selection of basic emotions, Ekman does not exclude the possibility of other emotions sharing the same characteristics. As further candidates he suggests *shame* (*embarrassment*), *excitement*, and *contempt*. Due to the lack of empirical evidence, further studies need be conducted in order to investigate if all the nine characteristics that define a basic emotion are present in these emotional states.

2.4.2 Dimensional Model of Affective States

In contrast to the Basic Emotion Theory, models have been developed that represent emotions on one or several different dimensions instead of labelling them as distinct categories. Ever since the Basic Emotion Theory has been developed, theorists like Ortony and Turner (1990) have been asking why there is so much disagreement on which emotions should be classified as *basic* while some contenders (e.g. interest, desire) are in some cases not even considered as being an emotion (in: Yield p. 474). These discrepancies in agreement led to the development of explanations other than the Discrete Emotions Theory. In the *Handbook of Music and Emotion*, Sloboda and Juslin (2010:pos 1896) give a short overview of a selection of *Dimensional Models of Affective States* (also: Dimensional Models of Emotion). Duffy (1941) for example categorizes emotions by using *arousal* as the only dimension. Osgood et al. (1957) on the other hand have developed a three-dimensional model, using *valence*, *activation*, and *power* as classifiers. Among the different models, the most influential so far has been the two-dimensional *circumplex model* by Russel (1980), which I will give an overview of now.

In his *circumplex model* Russel attempts to represent emotions (or rather words that describe cognitive representations of emotion) inside a circular two-dimensional space. In this bipolar space, the horizontal axis represents valence in form of a pleasure-displeasure continuum, whereas the vertical axis represents the activation or deactivation of arousal.

In order to develop his model, Russel (1980) has placed eight affect concepts in a circular model in such a way that opposite concepts form a bipolar dimension (e.g. excitement-depression, arousal-sleepiness)
(see Figure 2.1).
In total, Russel conducted a series of three studies: His first study consisted of two parts: In the first part of the study he instructed subjects to place twenty-eight emotion words into one of eight provided categories: arousal, excitement, pleasure, contentment, sleepiness, depression, misery, and distress. There was no restriction concerning the number of categories a word could be placed into.

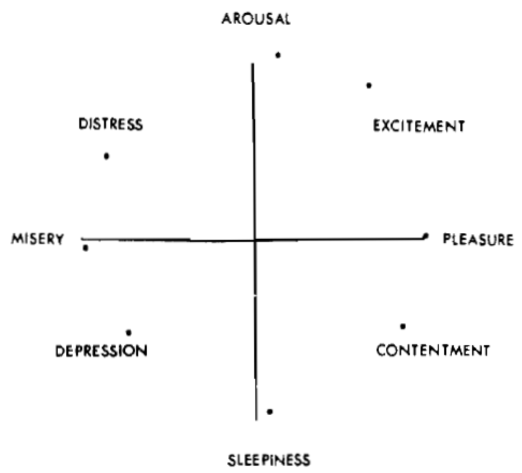

Fig. 2.1: Eight affect concepts a in circular order. (Russel, 1980:1164)

In the second task, another group of participants was asked to place the eight categories introduced in the first study into a circular order, in a way that words opposite to each other would represent opposite feelings, and words that are closer together in the circle would represent more similar feelings (see Figure 2.2).

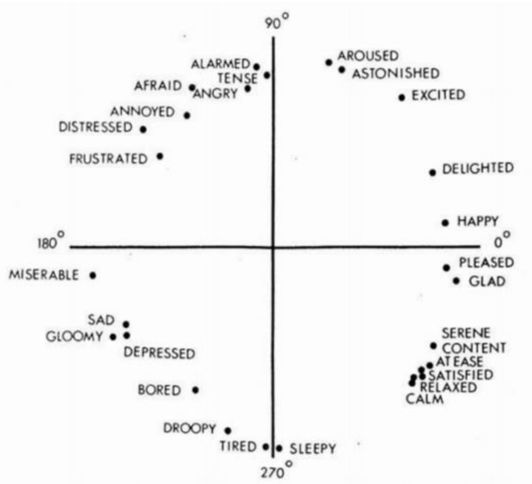

Fig. 2.2: Direct circular scaling coordinates for 28 affect words. (Russel, 1980:1167)

A second study was performed on multidimensional scaling of terms. The participants were instructed to group together the twenty-eight emotional words into 4, 7, 10, and 13 while judging how similar the words are. Words that have been sorted into the same group together did get a higher rating and were mapped closer together in the model than words that have been sorted into separate groups in the different trials (see Figure 2.3).

The third study investigated how participants would place an emotion word into a two-dimensional bipolar space, with a pleasure-displeasure continuum on the horizontal axis and the level of arousal being placed on the vertical axis (see Figure 2.4).

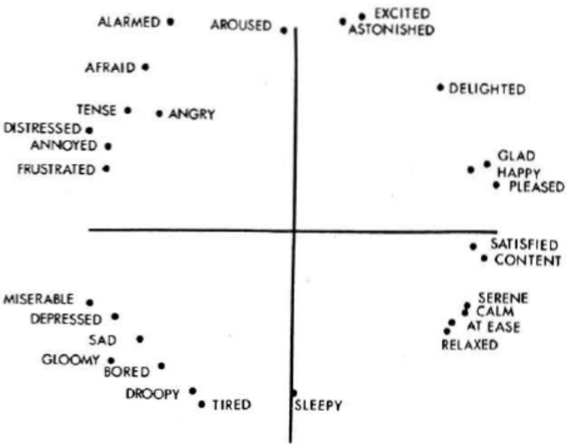

Fig. 2.3: Multidimensional scaling solution for 28 affect words. (Russel, 1980:1168)

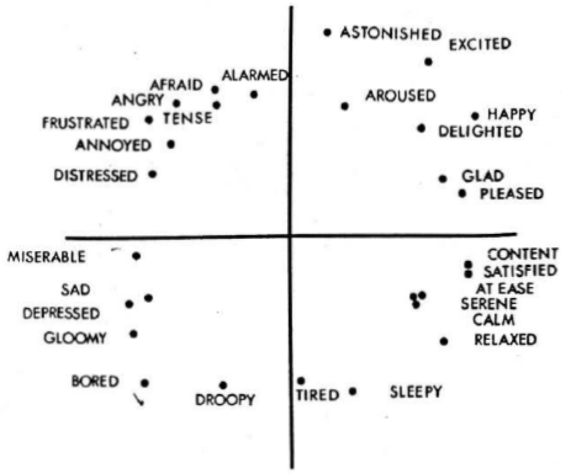

Fig. 2.4: Unidimensional scaling of 28 affect words on pleasure-displeasure (horizontal axis) and degree of arousal (vertical axis). (Russel, 1980:1168)

The *circumplex model* that has been developed based on these studies has been suggested to show the full range of emotional experience on the basis of two key dimensions. The advantage of this model is that it gives an idea about the perceived relationships between emotions and, as Yiend et al. (2005) suggest, “makes it easy to understand how different languages could have developed different words to describe subtly different mixes of emotion experience” (p.474).

The studies by Russel were rather focused on the perceived relations between subjective experienced feelings rather than the underlying emotion. I claim that this model cannot give a full understanding of an emotional experience as it is not taking *intensity* as a third dimension into account. The intensity with which an emotion is experienced should be taken into consideration if this model were used to describe an actual emotional experience and not only the mental representation of the relationships between emotions. I suggest that the intensity with which an emotion is experienced would have a significant impact on where people would categorise the emotion on the arousal axis. In his affect theory, Silvan Tomkins (2009) considers *density*, which he describes as a combination of duration and intensity as an important factor for an affective experience (p.166). As the duration of an emotional experience could also be a relevant factor on how an emotion is experienced and characterised, density could serve as a relevant dimension in a model of affective states. Another problem is, that the model does not explain mixed emotions. For example can *nostalgia* be interpreted as a mixture of joy and sadness, which according to the model are on the bipolar ends of both axes.

It can be argued that the *circumplex model* of affective states does not give any insight about what we define as an emotion and solely takes the cognitive representation, the subjective feeling, into account.

2.5 Measuring Emotions

At least three characteristics have been identified as being embodied when an emotion is experienced: behaviours, bodily responses (in terms of physiology), and subjective feelings (Sloboda & Juslin, 2010; Yiend & Mackintosh, 2005).

Emotions cannot be *measured* in the traditional sense, they can only be inferred from a selection of physiological responses like changes in heart-rate and variation in perspiration. These physiological responses should be interpreted with great care as they can also occur under several other circumstances, like fever in illness (Yiend & Mackintosh, 2005:494). Subjective feelings on the other hand can solely be obtained through self-report and are discussed to be "the most important kind of evidence of emotions" (Sloboda & Juslin, 2010:pos 1835). Even though self-report may be used as one of the main tools to study emotion, I disagree with this statement by Sloboda and Juslin. As will be further elaborated in chapter 4, self-report is already an interpretation of an underlying emotion that has been detected by introspection. If we take a look at the definition provided further up, we can see that at the moment we become aware of an emotion, it turns into a cognitive representation of an affective state that is consciously experienced – thus, a feeling. As we have seen, emotions can also occur without awareness. For that reason, studies only relying on information of an emotional experience gathered through self-report have to be regarded rather controversial as will be discussed in chapter 4.

We can also infer an emotion from how a person behaves by observing their movements, their posture and their facial expressions (Davies, 2010:pos 701).

This behaviour occurs subconsciously and represents a form of embodiment. Nevertheless, it is possible to conceal certain behaviour if one consciously wants to hide their underlying emotion. The behaviour can be altered with the intention of deceiving other people.
Ekman (1992:193) states that

> *[a]n observer can infer that an emotion is likely to be occurring when:*
> - *the response system changes are complex, when it is not just facial, or skeletal, or vocal, or [physiological], ... but a combination;*
> - *the changes are organised, in the sense of being inter-related and distinctive…;*
> - *the changes happen quickly;*
> - *some of the response system changes are ones common to all people; and*
> - *some of the responses are not unique to* homo sapiens.

Ekman further clarifies that these indications do not give a complete guarantee that an emotional experience has occurred, but that it is highly probable. This suggests that even though all the characteristics can be observed, this does not proof an underlying emotional experience, but that there is a high chance of that being the case.

2.6 Emotion and Cognition

Emotion and cognition are usually considered as two sides of the same coin. Even if some definitions of emotion include cognitive appraisal, it is not a necessity, nor are we always consciously aware of that we are experiencing an emotion. But even though these two entities are so opposed to each other, they also impact each other in many ways.

Yiend et al. (2005) suggest that emotions can be divided into *state emotions* and *trait emotions* (p.482). State emotions are the same as mood and affect and refer to the feeling one has ‘right now’. They can change very quickly and are usually measured by self-report. Furthermore, Yiend and colleagues state that “[state emotions] relate directly to the behaviours and physiology […] and can be measured in the same way” (p.482). As the name suggests, trait emotions refer to personality characteristics and their influence on how an emotion is subjectively experienced. Individuals that have the tendency to be worried most of the time are thus prone to experience sadness more often or more intense than people who are optimistic and have a positive view on the world (see Yiend et al., 2005:482). It has furthermore been suggested that state emotion, thus the mood a person is in, and personality traits can influence an emotional experience. Certain emotions can be amplified if the state or trait emotion is similar to the evoked emotion while others might be damped when an opposed state or trait emotion is underlying the experience (see Yiend et al., 2005:482).

Trait emotions should thus also be considered when studying music-evoked emotions as they can be assumed to be a significant variable that would explain differences in results when a between subjects design is applied. Changes in state emotions could on the other hand be responsible for differences in results within subjects. As to my knowledge, trait- and state emotions have not been taken into consideration, the previous statement is pure speculation. Evidence shows that physiological patterns are only distinguishable in some, but not all emotions (Yiend et al., 2005:496). Furthermore it has been shown that various physiological responses can be found in other sensations than emotion, for example fever during illness (see Yiend et al., 2005:494). It has been found that even though physiological arousal can be responsible for the occurrence of an emotion, they were not quite distinguishable from another until cognitive interpretations were applied (Yiend et al., 2005:496). As has been discussed earlier, after cognitive interpretations of an emotion have been made, we refer to the sensation as subjective feeling, and not emotion.

Johnson-Laird & Oatley (1989) point out the need to differentiate between basic emotions and complex emotions. They claim that basic emotions can (but do not have to) be experienced without an apparent reason whereas complex emotions are always dependent on external situations being evaluated on a high cognitive level (p.105). Emotions are not only evaluated in one's own consciousness, but are also communicated to others through the use of emotion labels. This becomes problematic when a particular emotion fails to be interpreted so that it is difficult or impossible to describe the underlying experience with words (Johnson-Laird & Oatley, 1989:89).

2.7 Summary

Emotions have been defined as brief, episodic, spontaneous, involuntary, and automatic manifestations that are caused based on evolutionary influences, evaluation processes, or both.

The two most influential theories that have been developed to explain the phenomenology of emotion have been presented. Whereas the Discrete Emotions Theory emanates the assumption that all emotions can be attributed to a fixed set of emotion categories, the Dimensional Model of Affective States takes one or more dimensions that are assumed to be characteristic for an emotional experience as a point of departure. The most influential model following this approach has been the *circumplex model*.

Emotions can be inferred from changes in physiology and self-reported experience. These two methods can be discussed as being controversial as they do require a high degree of interpretation on different levels and self-report can only reflect the cognitive representation of an emotion which we defined as *feeling*. It has been stated that in addition to the methods described above, emotions can be inferred from embodied parameters such as movements, gestures, and facial expressions. It has further been discussed that even though cognition should be separated from emotion, both phenomena do influence each other in both directions in different ways.

Chapter 3

Facial Expressions of Emotion

Before I am going to discuss how facial expression of emotion could serve as a means to study music-evoked emotions, the reader should get a general understanding what facial expressions of emotions are, and what kind of evidence the research on this topic provides.
This chapter is starting with a brief historical overview about the most influential theories and findings on the topic of facial expressions of emotion, before moving to the theories and studies that have become influential in the second half of the 20th century. Subsequently, a brief overview on the evidence obtained from neuroscience will be given. Finally, three methods that can be used to study facial expressions of emotion will be briefly introduced.

3.1 A Brief Historical Overview

The following section is giving a brief historical overview of the most influential ideas and findings in the study of facial expressions. The first notions on facial expressions date from the Antiques and have been under continuous development during the last centuries.

3.1.1 Antiques

The assumption that expressions of the face can reveal inner states and emotional experiences has already been made by Aristotle. He stated that "[t]here are characteristic facial expressions which are observed to accompany anger, fear, erotic excitement, and all the other passions" (cited in Russell&Fernández-Dols, 1997:3). Even though Aristotle acknowledged the existence of facial expressions *per se,* he considered them as being too short to be informative (Fridlund, 1994:3). It is also known that in the tragedy and comedy performances in ancient Greece and Rome masks were used which usually exaggerated facial expressions, so that the audience could see and interpret them from far away.

3.1.2 Medieval Times

During the Medieval Times much attention was drawn to the face, but rather to its shape and features. It was believed that different characteristics found in the face were seen in connection to personal traits rather than a possible display of an underlying emotion. In the Introduction to Darwin's 3rd edition to *The Expression of the Emotions in Man and Animals* (1998), Paul Ekman specifies that prior to Darwin's notions, facial expressions had solely gotten attention from physiognomists, "who maintained that character or personality was revealed by the static appearance of the face, the size and shape of the features, and their proportions" (p.xxvii). The belief that personality traits could be inferred from these characteristics was very popular only a few decades before Darwin's influential studies and thoughts on the expression of emotion (Ekman, in: Darwin, 1998:10).

3.1.3 After 1800

In 1862 the physiologist Duchenne de Boulogne conducted some controversial research using electricity to determine which muscles are involved in facial expressions of emotion. He stimulated facial muscles by using galvanisms (Fridlund, 1994:11). When using the galvanism method, facial muscles are stimulated with electrodes or metal probes that release electrical current (Wells, 1859). The changes in electrical current activate the underlying muscle. Today galvanism is referred to as electrophysiology. One of the major findings Duchenne made in his experiments is that not only the muscles around the mouth (*zygomaticus major*) are involved in a genuine smile, but also the muscles around the eyes (*orbicularis oculi*) (Fridlund, 1994:11) (see an overview of the muscles of the face in figure 3.1). Duchenne's explanation of facial expressions of emotion was from a creationist perspective, much like the one of Christoph Bell (Fridlund, 1994:11). Like Bell, Duchenne believed that facial expressions are universal all over the world and that they were created for the purpose of communicating emotion.

About a decade after Duchenne's findings, Charles Darwin re-discovered the area of facial expressions and took his findings as a point of departure for his own investigation on facial expressions (Fridlund, 1994:11, Darwin, 1998). Darwin's aim was to explain facial actions from an evolutionary perspective. His motivation has been to challenge and find an alternative to Bell's creationist view about the universality of facial expressions (see Ekman in Darwin, 1998:8). It has been suggested that Darwin's *Expressions of the Emotions in Man and Animals* (1998) has been the first attempt to give an evolutionary account of human facial behaviour. In his approach, Darwin does not generally deny that certain muscles form expressions that can be used to communicate emotions, but sees it as problematic that the *only* purpose of facial expressions should be the expression of emotions. He argues that they can also serve as a means of communication in contexts where emotions are not involved.

In general it can be said that Darwin suggested that the face is a display of inner states of the mind, without specifying what these inner states might be. This leaves room for interpretation. Nevertheless, he considered facial expressions to be universal across cultures and sent letters containing questionnaires to correspondents worldwide to investigate on his hypothesis (Fridlund, 1994:12). Darwin did not consider facial expressions to be innate, but argues that previous consciously performed actions might over time have developed into what he calls *reflex actions* through habit and association. Due to this automation, he states, that "the sensory-nerve cells excite the motor cells, without first communicating with those cells on which our consciousness and volition depend" (Darwin,1998:40; Fridlund,1994:22). This way expressive habits became non volitional and automatic and were inherited by future generations. Darwin thought of those involuntary actions as being 'of the least use' in terms of communication but he considered that mimicking these so-called habits wilfully as being rather useful and as containing a signal value (Darwin, 1998:356; Fridlund, 1994:24-25).

It can be summarized that Darwin did not consider facial displays to *express* emotion, only as *accompanying* them. For him, only intentional expressions, including those that mimicked the habitual ones, could be used for the purpose of communication (Fridlund, 1994:25).

3.2 Modern Theories

The following section will present an overview about modern theories on the function of facial expressions. Whereas the *Emotions View of Faces* is mainly a further development based on the theories and findings by Duchenne du Boulogne and Charles Darwin, the

Behavioural Ecology View sees facial expressions rather as social signals than expressions of emotion.

3.2.1 Emotions View of Faces

The Emotions View of Faces (or Facial Expression Theory) regards facial expressions as readouts of inner states (Buck, 1994). Facial expressions of emotion occur involuntarily and spontaneously with little to none volitional control (Ekman, 1991). They can also be produced voluntarily if one has the intention to hide his or her emotional state and thus wants to deceive in a social situation. If the underlying emotion one is attempting to conceal is rather strong it can 'leak' through the masking expression for a very brief moment. These 'leakages' form the so-called *micro-expressions* which are revealed for no longer than 1/15 to 1/25 of a second and can only be spotted with the help of technology and/or experienced observers. Evidence suggests that the 'leaked' expression reveals the emotion one has been trying to conceal (Ekman, 1991).

The Emotions View of Faces takes the assumption that facial expressions of emotion are innate and recognized universally across cultures as its starting point. The most influential evidence on the universality hypothesis has been collected by Paul Ekman and colleagues in field studies conducted cross-culturally amongst both literate and illiterate populations. Ekman and colleagues conducted cross-cultural studies with the intention to investigate if facial expressions of emotion were judged the same way in different cultures. It has been found that for the so-called basic emotions, facial expressions of emotion are recognized similarly across cultures (Ekman, 1994). Although recognition scores across studies were generally higher than chance (Russell, 1994:106), the scores obtained by individuals from non-Western cultures have been lower than the scores gathered by subjects from Western cultures (Fridlund, 1994).

Not all languages provide words for the emotion categories that are conceptualized in the English language. For this reason, different methods had to be developed in order to investigate how facial expressions of emotion are judged cross-culturally. In studies by Ekman and Friesen (1971) photographs of faces depicting various facial expressions of emotion that were understood to be prototypical were shown to individuals from different cultures. In cases where there was no direct translation of the English words for the basic emotions, participants were told a story that was supposed to transport information about a target emotion. The subjects were asked to match the picture that was showing the facial expression with the emotion that they associated with the story they have been told before. Although these studies revealed important insights on the probable universality of a discrete set of emotions, this method contains several pitfalls: First of all, the pictures shown to the participants were posed expressions that were supposed to express the displays that are considered being prototypical for the basic emotions. These posed expressions can be seen as exaggerated versions of a pre-defined prototype that is less likely to occur in real-life situations. This critique has also been made by Adolphs (2006). In his article he also pointed out the importance of individual differences and context.

The lack of context can be considered as one of the most important problems in these studies. It goes without saying that a smile can not only mean that someone is expressing *joy*, as smiling behaviour is also to be found in other situations. As it has been discussed in the previous chapter, one characteristic of an emotion is that it evolves in time. Taking this into consideration, using only pictures of posed expressions shows another problem: the photographs shown to the subjects are static. This means that the highly relevant information about onset, apex, and offset is missing for the receiver. These experiments go against the nature of facial expressions as being involuntary and spontaneous and thus lack ecological validity.

Last but not least, these studies by Ekman and colleagues were focused on concepts and terminologies that are used in the Western culture in general, and in the English speaking culture in particular. The evidence shows that not all cultures do provide translations for the emotion words the English language uses. In addition, researchers did not take into account that different cultures might perceive and conceptualize phenomena in different ways. This would be an explanation for differences in the judgement of facial expressions of emotion in different cultures. Judgements in non-Western cultures have still been above chance but significantly lower than in Western populations. It can be suggested that the concepts used in other cultures might be more fine- or coarse-grained than the concepts we are used to. Facial expressions of emotion may thus be perceived differently. Consequently, labels that are used to define expressions vary based on the cognitive representations of the expression perceived.[2]

[2] Another example from a different research field would be the rather recent study by Roberson et al. (2006) on colour categorization in the Himba where it was shown that the Himba tribe from Namibia does not have a word for the colour blue but differentiates between different shades of green. Additionally the Himba do not differentiate between green and blue in general and fail to spot a blue square amongst green ones. A task that has

Russell (1994) states that the labels used to describe facial expressions of emotion vary widely across individuals even in Western cultures (p.104).

It can be summarized that the focus in the study of facial expressions of emotion has mainly been on perception and judgement rather than production. Even though differences in the judgement of facial expressions across cultures can leave some doubt concerning the universality hypothesis, it can be suggested that the results underlie something Russell (1995) calls ‘minimal universality’, meaning that there is a small amount of agreement on how facial expressions of emotions can be interpreted across cultures (p.383). Even though judgements of facial expressions of the basic emotions slightly vary across cultures, the fact that all these cultures display facial expressions remains to be undisputed. To my knowledge no culture has been found so far that does not display facial expressions that are associated with the underlying basic emotions, which leads to the conclusion that the faculty to form these facial displays in response to an underlying basic emotion can be considered innate and universal.

3.2.2 Behavioural Ecology View

In the *Behavioural Ecology View*, facial expressions are rather seen as social signals which evolved to serve as a pre-language form of communication (Fridlund, 1994:139). Rather than communicating or expressing emotions, facial displays are considered to communicate intentions in social situations. The concept of basic emotions does not play a role in the Behavioural Ecology View. Rather than studying facial expressions themselves, the Behavioural Ecology View finds its arguments in theories and in experiments that did *not* show enough evidence for the theory that emotions can be expressed in the face. It has been suggested that facial expressions are understood and have an impact upon others’ behaviour because the signals and the vigilance for, and the faculty to understand these signals evolved parallel with each other (Fridlund, 1994:128). As opposed to the Emotions View, it is suggested that facial displays are not readouts but can be seen as “social tools that simplify social encounters.“ (Fridlund, 1994:129).

The Emotions View and the Behavioural Ecology View do not have to be seen as mutually exclusive (Ekman, 1992; Ekman in Darwin, 1998:xxx). While the latter strictly neglects the idea that facial displays can express an underlying basic emotion, the Emotions

---

been easy for the English participants, who on the other side had problems differentiating between different shades of green, even spotting the different square was hardly possible.

View solely states that *some* expressions communicate emotional states without excluding that some facial displays exist that are used for *other* forms of social communication. In agreement with Adolphs (2002), I argue that in the moment facial expressions are shaped, both emotional and social aspects interact, and are highlighted by the circumstances in which they occur (p.23). That means in detail that basic emotional reactions are accompanied by involuntary expressions, whereas voluntary expressions can be modulated by culturally or socially shaped display rules.

3.3 Evidence from Neuroscience

In the last decades, the field of neuroscience has become more and more influential in different research areas. Also in the study of facial expressions and specifically facial expressions of emotion, looking at the activation of specific brain areas can give further insight on how facial expressions are perceived and produced.

A study by Hopf, Muller-Forell, & Hopf (1992) found that voluntary facial expressions could not be produced when damages in the motor cortex were found, but that these lesions would not impair the ability to form spontaneous facial expressions. Lesions in the insula, basal ganglia, or pons resulted in the impairment of forming involuntary facial expressions but left the ability to form voluntary ones (see Adolphs, 2002:23). These results highly suggest that involuntary and voluntary facial expressions are operated in different brain areas. Another suggestion that has been made based on results of findings neuroscience provided is, that the subcortical route may be a fast and also automatized way in the case of face perception. It is suggested to operate parallel with the cortical route (Adolphs, 2002:31). In addition, strong evidence for the activation of the amygdala in the perception of facial expressions has been shown.[3] This supports the assumption that facial expressions of emotion are perceived subconsciously. Whalen and colleagues (1998) found that activity in the amygdala was increased even in the cases where facial expressions where shown for a time span that is too short to be consciously perceived. Right after the expression-stimulus, a neutral expression was presented for a longer period of time. This method is called backward

[3] Going into detail about the brain regions involved in perceiving and recognizing facial expressions of emotion would extend the frame of this Thesis. The interested reader can find a very detailed overview on this specific topic in: Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1(1), 21-62.

masking. Participants reported to have only seen the neutral (masking) expression, whereas increased brain activity in the amygdala shows strong evidence that the masked expression was perceived.

3.4 Other Evidence

Studies using methods that are not connected to neuroscience *per se* also found that the subconscious plays a major role in the perception of facial expressions (Adolphs, 2002:30). It has been found that observing the facial expressions of another person results in activation of one's own muscles or muscle groups. Even though this is not visible, it can be measured with electromyography (EMG, see next section). Those reactions have also been observed when facial expressions have been presented subliminally (Dimberg, Thunberg, & Elmehed, 2000).

3.5. How can Facial Expressions of Emotion be Studied?

In this section I am going to present three methods that can be, and have been, used to study facial expressions of emotion in different contexts. The aim of this section is solely to give an overview about methods being used, and will not refer to specific studies or experiments. The methods now presented will become relevant in chapter 5 when the development of an experimental design for using facial expressions of emotion as an indicator of an experienced emotion in response to music will be discussed.

3.5.1 Electromyography

Facial expressions of emotion can be measured by using an electromyograph. An electromyograph is a device that measures electrical activity in the muscles. When a muscle cell is activated a so-called *action potential* is generated, which can be measured by the device. In order to measure an action potential, specific electrodes have to be placed on the skin just above the centre of the muscle of interest. For measuring facial muscles, surface electrodes are preferred. Before the electrodes can be attached, the skin needs to be cleaned with alcohol pads, to assure that no dust or dead skin is between the electrode and the skin. If necessary, hair has to be removed. When surface electrodes are used, we refer to the method as surface EMG (sEMG).

The muscles that can be assessed with sEMG are the muscles directly under the skin. As the voltage difference between two distinct electrodes needs to be measured, at least one pair of electrodes needs to be attached. The raw signal generated needs to undergo several steps of post-processing before the data can be used for further analysis.

3.5.2 Facial Action Coding System

The Facial Action Coding System (FACS) developed by Ekman and Friesen (1978) is an observer based system of facial expression measurement (Cohn et al., 2007:203). Ekman and Friesen have studied the muscles of the face extensively and have identified more than 40 so-called *action units* (AU) that are associated with the muscles being involved in any human facial expression. The FACS can thus be regarded as a catalogue that simplifies the identification of facial displays and their categorisation. The most recent version of the FACS is FACS 2002. According to this version, 9 AU are located in the upper, and 18 in the lower face. The remaining AUs define for example head- and eye positions. The intensity with which an AU is activated is indicated with letters from A-E, where A indicates a trace of an activation and E the maximal intensity (Cohn et al., 2007:206). So far, more than 7000 different combinations of AUs have been observed (Scherer & Ekman, 1982).
If one wants to use the FACS for stusying facial expressions of emotion, a videorecording of the subject one intents to analyse has to be made. Based on that recording, researchers can code different events they observe in the face and the AUs that are involved. Before any researcher is able to apply the FACS in research, one has to undergo a specialised training that requires approximately three months (Cohn et al., 2007:215). Not just the training requires a big amount of time, also the application of the FACS is rather work- and time intensive as it is assumed that it takes about 100 minutes to code a video with the length of one minute. The reason is, that videos usually are analysed frame-by-frame, based on changes in behaviour observed on the surface of the face and the duration of certain movements. Depending on the framerate chosen for the recording, the time that is needed for the analysis can vary.

Two approaches can be followed when using the FACS. Whereas the *judgement-based* approach would rate a certain combination of action units (AU 6 + AU 12) as the facial expression of ‘happiness’, the *sign-based* approach would solely code “the face as having an

Basic emotion	AUs	Muscles (and AU description)
Happiness	6 12	Orbicularis oculi, pars orbitalii (Cheek Raiser) Zygomaticus major (Lip Corner Puller)
Sadness	1 4 15	Frontalis, pars medialis (Inner Brow Raiser) Corrugator supercilii, Depressor supercilii (Brow Lowerer) Depressor anguli oris (Lip Corner Depressor)

Table 3.1 Action Units and facial muscles that are associated with facial expressions of 'happiness' and 'sadness' (Cohn et al. 2007:207-210).

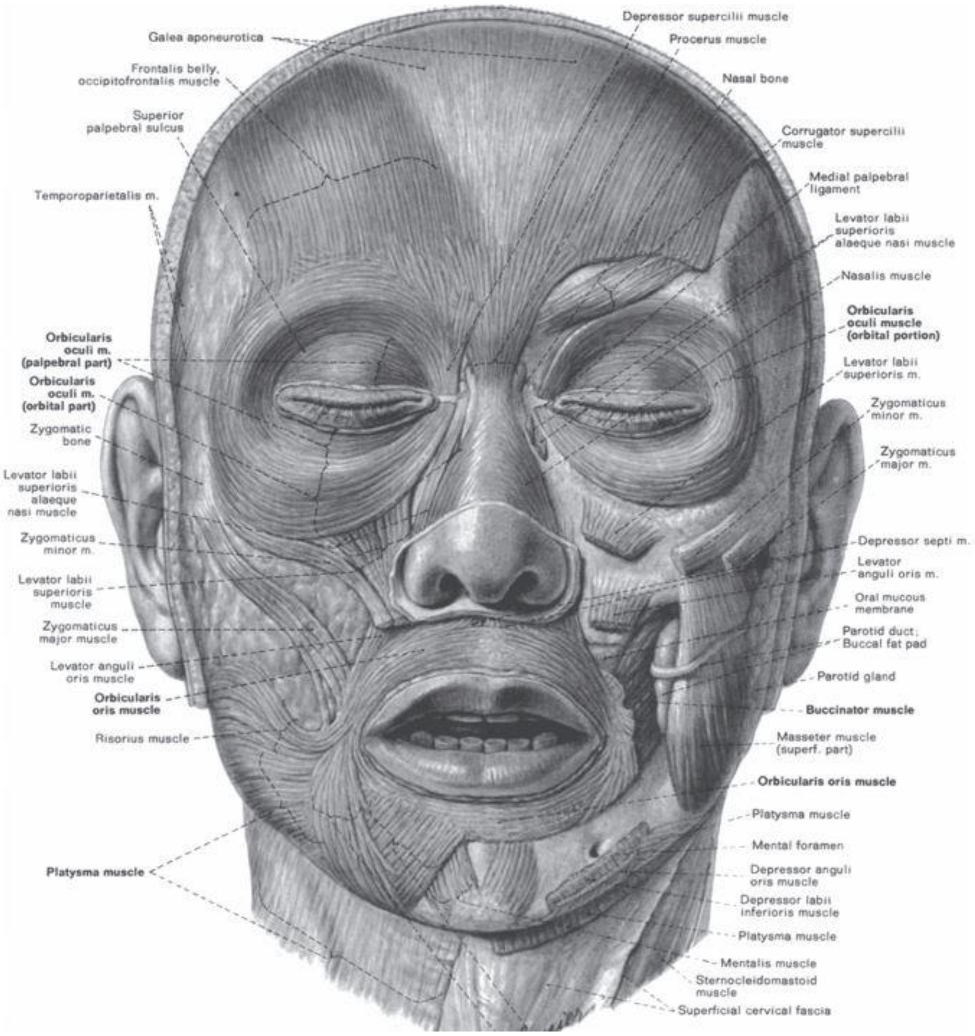

Fig 3.1 Muscles of the face (Clemente, 1997 in Cohn et al. 2007:207-210).

upward, oblique movement of the lip corners" (Cohn et al., 2007:204). The AUs that are most relevant for the study of emotion are the ones that are involved in the facial expressions that accompany the basic emotions. As 'happiness' and 'sadness' are emotions that are experienced in response to music more than the other basic emotions, the AUs that are categorised as combining to facial expressions of 'happiness' and 'sadness' (see table 3.1) can be put selectively into focus when studying facial expressions of emotion in response to music-evoked emotions. A visualisation of the AU-combinations for 'happiness' and 'sadness', respectively, can be found in figures 5.8. and 5.9 in chapter 5.

3.5.3 Automatized Face Analysis

During the last few years, face recognition software has been invented and developed. This technique has become more and more popular in different areas. In various computer games this technology can be used to model one's own face into a computer game character. Another example is the social app *Snapchat* that has a feature that allows face recognition and lets the users apply different objects (e.g. dog ears and dog mouth) to their own face for entertainment.

Face recognition software cannot only be used for the sake of entertainment but also to study human behaviour. One commercial software that can be used to analyse facial expressions of emotion is the FaceReader6 developed by NOLDUS Information Technology (2015). The software detects facial expressions of emotion in three processing steps. First, the face is detected, subsequently the face is modelled with 500 facial keypoints. Finally, the facial expression is classified. According to their website, more than 10.000 annotated images were used to train the software and determine the keypoints. FaceReader6 is trained based on the Facial Action Coding system and detects six basic emotions (sadness, happiness, surprise, fear, anger, disgust). In addition contempt and neutral expressions can be coded.

When it comes to analysing the data, one can choose between using the *circumplex model* of affect (Russel, 1980) and a selection of 20 action units of the Facial Action Coding System (Ekman & Friesen, 1978). The NOLDUS FaceReader has been used in different research areas so far, mainly for studying consumer behaviour, market research, educational research and in psychology ("Noldus Information Technology," 2015). Only one study (Weth, Raab, &

Carbon, 2015) has applied this technology to study facial expressions of emotion in response to music. The study by Weth and colleagues will be presented in more detail in chapter 5.

3.6 Summary

Facial expressions of emotion have not always been seen in connection with emotion. Even modern theories neglect the evidence that the expressions for some of the basic emotions can be considered universal to some extent. In cross-cultural studies carried out on facial expressions of emotion, the focus has been mainly on the perception and categorisation rather than the production. Findings from studies in Western and non-Western cultures suggest that the faculty to perceive a facial expression as being the display of an underlying emotion is to some extent universal. Evidence from neuroscience lead to the conclusion that facial expressions are processed in the subconscious and that one does not necessarily have to be aware of showing or seeing them. Methods that can be used to study facial expressions of emotion are the measurement of muscle activity with the help of EMG, observing the changes of facial expressions and code the activated muscles with the FACS, or doing an automated face analysis by using a specific software. How these methods can be applied in the study of music-evoked emotion will be discussed in chapter 5.

Chapter 4

Emotions in Music

This chapter is providing an overview about the study of emotions in music. After providing definitions for the two distinct phenomena of perceived and evoked emotion in music, I will present different theories that are attempting to explain why we are able to associate emotions to music. Subsequently, the underlying mechanisms that are discussed to be involved in this process will be presented. Finally, methods that have been used study emotions in music are described and their reliability of the results obtained will be discussed.

4.1 Definition

The term *emotion in music* can be used and interpreted in various different ways. In the introductory chapter of the *Handbook of Music and Emotion* (Patrick N. Juslin & Sloboda, 2010), one of the standard works in this research area, Juslin and Sloboda state that 'emotions in music' is used as an abbreviated version of the term for 'emotions that were somehow induced by music'(pos 357-368) This definition is rather restricted as it is solely focusing on induced (or evoked) emotions and is lacking to include the phenomenon of perceived emotions in music.

In this section I am going to provide definitions for both, perceived emotions in music and music-evoked emotions. Why this differentiation is so important when studying emotions in music will become more obvious in the scope of this chapter.

4.1.1 Perceived Emotions in Music

Perceiving emotions in music means that based on certain features in the music's structure and dynamic, the expression of a distinct emotion can be *recognized*. This recognition is based on previous experiences we have had with emotion in everyday life and in different contexts (e.g. prosody in speech, body language). That does neither implicate that an emotion can be perceived in any piece of music, nor that anyone is able to perceive emotion in a certain piece of music. As the term *recognition* suggests, this process is based on previous experiences and prior knowledge, and dependent on someone's individual and cultural background. Although we can say that a certain piece of music *sounds* sad to us, it does not mean that we believe the music to *be* sad, like a person would be (Davies, 2010: pos 487). On a more abstract level, we can still perceive features in the music that let us make this judgement. It can be assumed that we somehow compare the musical features with what we already know. As a basis we would take the abstract knowledge we have about emotional cues (e.g. in speech prosody or body language) that we apply to the musical features. The question if the faculty of perceiving emotion in music is innate or learned is still to be solved. I myself would argue that, as with speech perception, the faculty to recognise emotion in music is innate, but that the 'rules' by which we perceive and label emotion in music is learned and varies cross-culturally.

4.1.2 Music-Evoked Emotions

The emotion that is perceived in music and the emotion that is induced by the music do not necessarily have to correspond.

We generally speak of evoked (or induced) emotion in music if emotions are subjectively experienced, or *felt* in response to music. Usually we only refer to the consciously experienced emotion in the context of evoked emotions, but as we have seen in the previous chapters, there is evidence for that an emotion can be *experienced* without the conscious being involved. It has also been shown that changes in physiology, heart rate and electro-dermal activity can occur in response to music (Koelsch, 2010:217). Some of these changes might be too weak to cross the bridge between the subconscious reaction and the conscious experience. It can thus be speculated that music evokes emotions more often than we are aware of. It has been suggested elsewhere (Peretz, 2010: pos 2412) that music can evoke a variety of emotions, ranging from basic emotions (e.g. happiness, sadness, fear), to more mixed emotions (e.g. nostalgia). In cases where music evokes emotion (whether they reach the stage of consciousness or not) it can be speculated that changes in facial expressions of emotion should be present as well, even though they would be very brief. To my knowledge, only one study (Weth et al., 2015) has looked at facial expressions in response to self-selected sad music and found activation of the facial muscles that are underlying the facial expression of sadness (see details in chapter 5).

There is still an ongoing discussion about the question if emotions in music are different to 'real life' emotions (Koelsch, 2012; Koelsch, Siebel, & Fritz, 2010). If we can find evidence for the assumption that, in addition to physiological changes, facial expressions of emotion can be detected in response to music, these findings would probably bring more light into this discussion, as facial expressions of emotion are a more specific indicator of an emotional experience than changes in physiology, heart rate and electro-dermal activity.

4.2 How can we 'Understand' Emotions in Music?

There are different theories that try to explain why and how emotions can be detected (or perceived) in music and why and how emotions can be induced by music. The most prominent theories are the *expression theory*, the *arousal theory*, and the *contour theory* (Davies, 2010). Each theory does only cover a certain aspect and cannot be considered as a general explanation for all phenomena that are known so far in connection with music and emotion.

4.2.1 Expression Theory

In the *Handbook of Music and Emotion* Stephen Davies (2010) gives a good and brief definition of the *expression theory*. He states that "[t]he *expression theory* analyses the music's expressiveness as depending on the composer expressing his or her occurrent emotion through the act of composition" (pos 716). It can be speculated that this does not necessarily mean that the composer has to have the intention to express an emotion, or that the 'encoding'-process is happening consciously. The same counts for the listener. The process of 'decoding' an emotion can be assumed to happen automatically without conscious awareness.

Davies also points out the role the dynamic structure of the music plays in the resemblance to human expression (Davies, 2010:pos 701). Further, Davies compares the way we register the expressiveness of music being rather like seeing a person who feels the emotion rather than reading the emotion's label (e.g., 'sad') or a description of the emotion (Davies, 2010:pos 654). Both of these statements are very important when we study perceived and evoked emotion in music, since both, music and emotion can be considered dynamic phenomena that evolve in time and cannot (and should not) be studied without the context they are embedded in. Often only small parts of a whole musical piece are used as stimuli in the study of music and emotion. This is somehow equivalent to the practice in the study on facial expressions of emotion, where expressions are shown on a photograph, which represents a fixed point in time and does not give any information about the context this (often posed and exaggerated) expression was produced in.

4.2.2 Arousal Theory

The *arousal theory* states that music evokes emotion in the listener and considers this being the reason for why we consider music as being expressive (Davies, 2010:pos 732). Davies suggests that we project the emotion experienced in response to the music on the composer's intention to express the emotion in question (Davies, 2010:pos 747). He further suggests that the person we attach the emotion to, as well as the emotion itself, are imagined.

In other words, the arousal theory states that music is considered expressive because of an imagined emotional experience and not a subjective felt emotional experience of our own. No emotion is experienced or felt, we are not affected by the music ourselves, but attribute the experience to an imagined, external person. The expression we attribute to the music is not linked to our own emotional experience but to the imagination of someone else's emotional expression.

4.2.3 Contour Theory

The final theory Davies presents is the *contour theory*. As point of departure this theory takes the assumption that certain characteristics in the expression of emotion are universal and can be recognized cross-culturally (Davies, 2010:pos 833). Davies states that music is experienced as expressive because it shares the same characteristic appearances that are associated with the expression of emotion in other modalities, for example speech, gait, and physiognomy. The core of the *contour theory* is, that music solely copies a shape we are familiar with, without being filled with the same (or any) content. Music, in Davies' view, does not express emotion, but is nevertheless judged as being expressive due to association with emotional characteristics. As an example he uses the phenomenon, that we rate objects (e.g. willows) or animals (e.g. dogs) as looking 'happy' or 'sad' even though we *know* that objects and creatures do not experience emotions like humans do. We associate their appearance with characteristics we usually see in connection with emotional expression in the environment that we are familiar with.

In a meta-study by Juslin and Laukka (2003) 104 studies of emotional speech were analyzed. The results showed that the same features that lead listeners to judge speech as 'happy' also have been the underlying characteristics in the judgements for 'happy' music. The researchers concluded that "the same features that underlie the speech-emotion code are also responsible for listener's judgements in musical experiments" (Sloboda & Juslin, 2010:pos 1961). With 'speech-emotion code' it is referred to significant characteristics in speech that make it possible to infer an underlying emotion. For example will a fast tempo, higher pitch, and dynamic speech be judged as sounding 'happy' whereas a slow tempo, lower pitch, and monotonous speech will rather be rated as sounding 'sad'.

None of the presented theories does satisfactorily explain why or how we perceive music as expressive and they do not at all explain why and how some music evokes emotion to a listener. In my view, especially the *arousal theory* seems rather far-fetched as it assumes that an affective state of a composer is imagined. This explanation would come with lot of cognitive load since one would have to actively imagine an external figure the expression can be projected on.

To sum up, it can be said that the theories presented by Davies only take the perceived emotion in music into account. None of these approaches explain how music can lead to experiences of subjective feelings in response to music. To get a further understanding about the phenomenon of music-evoked emotions, some of the underlying mechanisms will be presented in the next section.

4.3 Underlying Mechanisms for ‘Understanding’ Emotions in Music

In the following section I am going to present mechanisms that have been suggested to be involved in the phenomenon of emotions that are evoked by music. As these concepts are very complex themselves, only a brief overview will be given. My aim is to outline the general idea of these mechanisms and the role they play in ‘understanding’ and experiencing emotions in music.

4.3.1 Theory of Mind

The *Theory of Mind* (ToM) describes the faculty to relate mental states such as wishes, intentions, knowledge, beliefs etc. to one self and to others (Premack & Woodruff, 1978). This involves our understanding that other peoples’ internal representations can vary from our own wishes, intentions, knowledge, etc. We use this faculty in our everyday life in order to understand the mental processes of other humans and to predict their behaviour based on this understanding. Johannson (2005) makes the assumption that there might be a connection between ToM and language as children acquire both around the age of four or five years (p.149). The ToM seems to be an innate faculty in humans that needs a few years of experience to fully develop. Further, Johannson suggests that all humans form a ToM from the day they were born until the age of five. Social interactions and environment play an important role in the development of this faculty. It can be suggested that every human possesses a more or less strong form of ToM.

The ToM can only be a *theory*, as these mental processes cannot be observed directly (Johannson, 2005:148). One mechanism that might bring more insight in how this process of understanding and ‘mind-reading’ could work is the Mirror Neuron System.

4.3.2 Mirror Neuron System

The neural basis for recognising mental states of another individual could be provided by the so-called mirror neurons. Mirror neurons have first been discovered by Rizzolatti and colleagues in the 1990ies (see Johansson, 2005:152). It is suggested that the mirror neuron system (MNS) consists of different populations of motor neurons that discharge both when a goal-directed action is executed as well as when the same action is observed being executed by another individual (Gallese & Goldman, 1998:493; Rizzolatti, 2007:318). In detail that means that while observing another individual, the same motor trajectories that are activated in the brain of the protagonist are also activated in the brain of the passive observer, solely that the last part of the action – the execution - is inhibited in the latter, so that the action is mentally simulated. Kohler et al. (2002) have shown that this procedure of inner simulation does also occur in a specific group of motor neurons when a monkey *hears* the sound of a meaningful action without any visual information about the source. This group of neurons was not activated when other interesting sounds were perceived, but it was shown that they could discriminate between sounds of different actions. First these neurons were found in the brains of macaque monkeys, but it has been shown that motor neurons found in the analogous brain region in humans react similar. The discovery of the MNS provided neuro-psychological explanations for a number of social phenomena that play an important role in terms of communication, for example action and intention understanding, mind reading, empathy, and language (see Rizzolatti, 2007:329). Since emotion can also be considered as a goal-directed communication system, communicating emotional states to the self and to others, it can be suggested that the MNS plays an important role in understanding emotions perceived in and evoked by music.

4.3.3 Cerebral Intentionality and Corporeal Intentionality

Intentionality describes the faculty to form representations of objects inside of the mind, without them having to be present in the outside world (Jacob, 2003). Brentano suggested that intentionality refers to content that is physically non-existent, meaning that mental attitudes such as thinking or believing only exist in the mind (see Leman, 2007:78). Whereas objects can be created by thinking, the process of thinking does not automatically create those objects in the real world. Intentionality is defined by an imaginary reality and is not to be confused with *intention*. Music can be considered as being a product based on intentionality, which has been transformed into an object.

As an underlying concept for this process Leman suggests *corporeal intentionality*, which suggests that mental states, for example ideas, intentions, or mental representations can be turned into physical energy. This physical energy can then be turned into ideas, intentions, or mental representations again (Leman, 2007:84). In other words, cerebral intentionality represented by thoughts, ideas, intentions etc. is transformed into moving sonic forms by corporeal intentionality, which is represented by the motor patterns that are underlying the production of actions that lead to creating these moving sonic forms. Corporeal intentionality is translated into those corporeal articulations like, for example, playing on an instrument. If we want to take emotions as an example, this means that emotions can get encoded into these moving sonic forms by our mental representation of that emotion. Emotions can be defined as a unique communication system and thus represent goal-directed actions. The mental representation of emotion is formed by knowledge from different sources we bring in connection with an experienced emotion, like characteristic changes in speech parameters that are very similar to the characteristics we associate with emotions in music. If one succeeds encoding this mental representation of an emotion into the music, the mirror neuron system in the listener gets activated when the moving sonic forms are perceived. Similar motor trajectories that were involved in the music producing action by the interpreter are activated in the listener. The listener thus comes to associate the moving sonic forms with his or her own actions. These actions are simulated in the listener's brain and can be translated to the corporeal expression and subsequently the experience of the emotion the musician (consciously or subconsciously) embedded into the music. With the help of these processes we have the impression to *understand* emotion in music and also *feel* it. This process is also known as empathy. Because we cannot observe these processes, this order of potential events underlying an emotional experience in music is merely speculative.

4.4 Approaches for Studying Emotions in Music

In his book *Embodied Music Cognition and Mediation Technology* (2007), Marc Leman introduces three ways in which musical experience can be described: *first-person description*, *third-person description*, and *second-person description*. In the following chapter I am going to give an overview of these approaches and will give examples of methods that have been used in the study of music and emotion that can be assigned the respective description method. Additionally, it will be discussed to what extent these methods can give insight about emotions experienced in connection with music. Table 4.1 provides a first attempt to give a schematic overview about the methods and approaches and about which emotional experience can be inferred from these. The aim of this chapter is not to give a very detailed overview about every single method that can be used to study emotions in music as this has been done in great detail by other scholars (e.g. Zentner & Eerola, 2010).

In addition to giving this overview, I am going to discuss these to what extent obtained results can be considered reliable and ecological valid for studying music-evoked emotion, and if these methods are useful for cross-cultural application.

4.4.1 First Person Description of Musical Experience

Leman (2007) defines the *first-person description* as being a language-based description of the world, resulting from a subjective interpretation (p.79). In relation to musical experience, music can be defined as physical energy that is taking the shape of moving sonic forms. These are perceived as objects in action. This action-oriented bias makes it possible for the listening subject to assign intentional qualities to music and attribute intentions and interpretations based on the information that is picked up by the physical energy. First-person descriptions can solely be obtained by introspection and cannot be observed from the outside. We can thus refer to it as cerebral intentionality (Leman, 2007:79).

When studying emotions in music, first person description is mainly obtained by using *self-report.* Self-report can be obtained in many different ways: with forced choice categories, free description methods, in a controlled setting in a natural listening environment, verbally, nonverbally, post-stimulus, continuously, and so on. When doing research on emotions in music, the most popular methods being used are going to be presented briefly in this section. In the *Handbook of Music and Emotion*, Zentner and Eerola (2010) are providing a very detailed overview about forms of self-report that can be applied to the study of emotions in music (pos 4588).

One form of obtaining self-report is through judgement studies. Judgement studies can take different shapes and forms, the most relevant in this field of research being *forced choice*, *free descriptions*, and *continuous rating*s. In the forced choice method, self-report is often based on the discrete emotions theory or the dimensional models presented in chapter 2. In studies using the discrete emotions theory as theoretical background, emotion categories are provided by the researcher (e.g. the Differential Emotions Scale, Izard et al., 2003). Studies using fixed emotion categories often treat these categories as mutually exclusive. More often than not only one of the categories provided can be selected. The judgements are often made after stimulus presentation and solely provide an overall judgement of the musical stimulus rather than providing information about which part or parts of the music provided the basis for this particular judgement. This method thus does not give a valid presentation of the underlying musical experience, as music and emotion both evolve in time and it can thus be assumed that several emotional experiences can occur in the scope of a single musical piece, even though one might not always be aware of this experience. As I have discussed in chapter 2, emotions occur subconsciously and might have to cross a certain threshold in order to reach conscious awareness.

Studies using the dimensional model are more interested in obtaining information about the valence (positive-negative) and the level of arousal (high-low) of the emotional experience as data obtained with theoretical framework is believed to be better comparable with data obtained through, for example, physiological measurements.

The data gathered through the forced choice method is by definition qualitative, but as most of the questionnaires work with nominal scales (= fixed emotion categories) and / or ordinal scales (= number as representing the degree to which the respective emotion is felt or perceived), the data can easily be quantified. This process does not work that easily when free description methods are used. Free descriptions are often also referred to as narrative approach to study emotions in music (Zentner & Eerola, 2010:4663). In this method, subjects are encouraged to describe an emotional experience in their own words. This naturally leads to an enormous amount of description words, which makes it more difficult to quantify the collected data. One strategy here is to sort the individual labels into broader categories and so obtain a more limited number of categories. The problematic that arises here is that the modified data does not represent the answers given by the subjects and the reliability of the results can be questioned.

Both of the methods just presented above have important limitations and pitfalls. Usually the judgement has to be made after the stimulus has been presented. The results obtained thus do not represent the full scope of an emotion experienced in reaction to music. Data collected with these methods have to be regarded with caution as it cannot be said with absolute certainty if subjects report an evoked emotion, a perceived emotion, or their own feelings they experience independently from the presented music. This critique has also been made by Zentner and Eerola (2010). It can be speculated that with these methods rather perceived

approach	description			method		results	emotions in music measured	
	1st	2nd	3rd	qualitative	quantitative		evoked	perceived
self report								
forced choice	x			x	x	ambiguous, possible confusion about the task	(x)	x
free descriptions	x			x	(x)		(x*)	
continuous self-report	x				x	cognitive load involved in the results, more likely perceived emotions are reported	(x*)	x
physiological measurements			x		x	changes in physiology in response to music suggest that an emotion is evoked	x	
articulations		x		x			x	
facial expressions of emotion		x		x		involuntary, spontaneous facial expressions indicate an emotional experience	x	
body movements		x		x		focus on action-sound couplings, relationship between musical gestures and sound		x**

Table 4.1: Schematic overview about methods for studying emotions in music

x* This method is more suited for the study of perceived emotions, but can also be used to report evoked emotions, even though the reliability has be questioned.
x** When body movements or articulations are used to infer emotions, we can only articulate and move to what we are perceiving

emotions than evoked emotions can be studied, as a certain amount of introspection and cognitive load are involved prior to the judgement and limited awareness of one's own emotion can become problematic in this process. The judgement methods discussed here are also vulnerable to demand characteristics (see also Zentner & Eerola, 2010:pos 5009).

It can also become difficult to compare studies, as labels used for the emotion categories tend to vary amongst researchers. This represents a form of researcher bias as he or she shows a preference for some labels over others, which might influence the judgements made by the participant.

Emotion categories are often represented by the list of basic emotions by Ekman or variations of it. As I have stated in chapter 2, not all languages provide translations for the emotion categories that are used in the English language. This makes a cross-cultural application of this method problematic. Because it has been criticised that the labels used for basic emotions to not represent the full range of emotional experiences in connection to music, Zentner et al. (2008) have developed the Geneva Emotional Music Scales (GEMS) which have been reported to be preferred over the discrete emotion model and the dimensional model (pos 4985). Nevertheless also here we have to be careful since there cannot be a one-to-one application of this scale in different cultures for the same problems as mentioned earlier.

The free description method comes with problems as well as one has to be aware of that participants can have difficulties with finding the right words of the musical experience because they lack the vocabulary or, as has been stated elsewhere, music sometimes evokes feelings that cannot be described with words (Zentner & Eerola, 2010:pos 5077). Even though one experiences some kind of sensation, introspection does not lead to any kind of verbalization, meaning that an interpretation of the sensation is not fully accessible or that it cannot be described with the available vocabulary. This suggests that at least for the study of evoked emotions language with its arbitrary and ambiguous nature does, in my view, not serve as a useful tool to gain insight into the sensation of music evoked emotion.

The method of continuous self-report provides an approach that is intended to capture not solely an overall impression of the emotional character that is associated with a musical stimulus but the development of emotional experiences over the entire scope of the presented music. The idea behind this approach is that emotion needs time to unfold and also information about several experiences in reaction to one musical stimulus can be collected. Continuous self-report can for example be obtained by asking a subject to press a button whenever an emotion is experienced or detected. Even though this method might be able to gather information about several emotional experiences in reaction to a single stimulus, the results do not really tell us anything about the nature of the emotion. This means that a form of verbal self-report has to be obtained in addition to find out which emotion has been experienced at what point and to what degree. Another method that has been used to gather continuous ratings is through a dial slider that, for example, represents one or two dimensions (e.g. valence and arousal) deviated from the dimensional emotion models.

Even though these continuous ratings can be used to obtain information about the time points at which an emotional experience has been identified, there naturally must be a slight delay between the time point of the onset of the emotion and time point of the rating. This so-called response delay can be suggested to be caused by two events: becoming consciously aware of the experience and the reaction time it takes to press a button or move a slider. In addition, continuous rating demands continuous introspection and awareness of one's inner state and thus causes a high amount of cognitive load. This makes me question if the subjects get the chance to engage in the listening experience if they have to constantly monitor their inner states. Another pitfall is that, like in the methods presented previously, the evoked emotion or the perceived emotion is reported.

As I have discussed in chapter 2, emotions occur on the subconscious level, which means that music listeners are not necessarily consciously aware of experiencing them in response to the music. With 'experiencing', I mean that one might show physiological indicators that are underlying an emotion, but the barrier between those physical characteristics and consciousness has not yet been crossed. In some cases an emotion might just be too brief or too weak to become aware of. For this reason it can be claimed that the method of self-report used in numerous studies on music and emotion is not very useful for several reasons: First and foremost, it requires introspection and interpretation. This leads to cognitive load interrupting the musical experience. Another issue is that participants are prone to confuse perceived emotions in music with their own experiences (Zentner & Eerola, 2010:5009). The data gathered using this method have a rather small level of reliability. Here often demand characteristics play a role, meaning that in some cases participants can consciously or subconsciously infer the underlying hypothesis. This possible knowledge of the purpose of a study leads subjects to alter their answer instead of reporting their honest experience.

Self-report may be still a reliable method to study perceived emotions in music as the process of recognition requires cognitive processes by definition. Due to language constraints, this method is not suited for cross-cultural application. For the study of evoked emotions, a method that avoids cognitive load and is more immediate without causing any response delay has to be applied. How such a method could look like will be described later in this thesis.

4.4.2 Third Person Description of Musical Experience

The *third person description* is based on *repeatable measurements* (Leman, 2007:79). As we have seen in chapter 2, physiological changes occur when an emotion is experienced. These changes in electro-dermal activity, heart rate, perspiration, etc. can be measured. As a result we get curves that represent the changes that have been measured in the respective physiological parameter. Even though this method uses objective measurements, this does not mean that the method itself is more objective or more specific than the first-person description. We need to interpret these changes that have been measured and these interpretations may be taken with caution. As we have seen previously, changes in physiology do not solely happen in the presence of an emotion, but can also occur, for example, when a high fever is experienced. Also circumstances from the outer environment lead to these changes. Thus, we can only speculate if these changes represent an emotional response to a musical stimulus (Leman, 2010:81). Interpretations of the measurements can thus be considered as arbitrary. This makes the third-person description, in my view, only applicable in combination with a form of first-person description.

4.4.3 Second Person Description of Musical Experience

As the first- and third person description of musical experience both require a high amount of interpretation of inner states and changes in physiology, respectively, Leman (2007) suggests a third form of description that involves intentionality but does not require any, or only little, interpretation (p.81). *Second-person descriptions* are formed by the expression and articulation of a subjective experience from one person to another. If two subjects are capable of expressing and articulating music, they are suggested to both be able to understand a musical experience due to a shared underlying basis of that expression and articulation. Leman further proposes that corporeal articulations could serve as a form of second-person description of musical experience.

Leman points out that the most relevant difference between the first-person and second-person descriptions is that the first-person description reflects the (cerebral) interpretation of an intended act whereas the second-person description is characterised by (corporeal) articulation of an intended act (p.77). Articulations can be considered as being more spontaneous, and more body based.

Because of the spontaneity corporeal articulations provide, I would suggest that using a second-person description approach would provide a more universal way of signalling an underlying emotional experience and can help to develop a methodological framework that could be applied cross-culturally without losing its reliability.

One way to apply the second person description of musical experience in general (not specified to emotion) is the study of *body movements* or *body motion* in connection with music. Here, several approaches have been applied, but to my knowledge the focus has not been on emotional experiences in connection with music. Research on body motions and music have, for example, covered a very wide range in the investigation of action-sound couplings, like music- and sound-related body movements of different stages of the musical experience (see e.g. Godøy & Leman, 2010; Jensenius, 2007; Nymoen, 2013).[4] Body movements related to music have, for example been studied by using optical motion capture (see chapter 5).

When it comes to the study of emotions on music, not many researchers have turned to the option of using the second-person description of musical experience. Why that is can only be speculated, but I understand that it is a challenge to create a framework that studies a very subjective phenomenon like emotions without having to deal with introspection or interpretation.

My suggestion is that facial expression of emotion in response to music could be included in such a methodological framework, since they serve as nonverbal, immediate, subconscious articulations of an emotional experience. Due to the suggested universality of facial expressions in connection with basic emotions, we can avoid interpretations and can infer an emotional experience, or several emotional experiences, in response to music.

How facial expressions of emotion can be used to study emotions in music and how an experimental design could look like, will be presented in the following chapter.

---

[4] The book *Musical Gestures* (Godøy & Leman (eds.), 2010) gives a very broad and interesting overview about this topic.

Chapter 5

Methods for studying music-evoked emotion by analysing facial expressions of emotion

In the following chapter I am going to discuss the usefulness of using facial expressions of emotion as an indicator of music-evoked emotions. Two studies that have used three different methods to study facial expressions in connection with music will be presented: optical motion capture, electromyography, and automatized face recognition software. Subsequently to each method I am going to briefly discuss these approaches in terms of reliability, ecological validity, and probable underlying problems for studying music-evoked emotions.

5.1 Facial Expressions of Emotion as a Tool for Studying Music-Evoked Emotions

As I have presented in the previous chapters, a number of studies investigating emotions in music (perceived and evoked) use self-report as their primary method. This is problematic for several reasons: in many cases, self-report is gathered *after* the presentation of a musical stimulus. This means that a dynamic experience (such as emotion) in response to a dynamic phenomenon (such as music) is reported post-stimulus as a single, overall rating. Using such a static method neglects the fact that there may be several emotional experiences in response to one single piece of music. Besides the retrospective approach, several forms of continuous measurement can be used to obtain self-report. This approach is also problematic as it requires cognitive load and introspection throughout the whole listening experience. In general, reporting an emotional experience requires introspection and interpretation of a subconscious phenomenon. Thus, what is reported is not the phenomenon itself but its cognitive representation. Another problem with these approaches is that it is not always clear if subjects report the emotion *induced* by the music or the emotion they *recognized* in the music (e.g. Darrow et al., 1987). Also, it is not always possible to capture an emotional experience with the arbitrary selection of words language offers. The use of emotional terms throughout the field is inconsistent. This makes the comparison of results from different studies difficult, if not even impossible. Even if the emotional terms would be used in a more systematic manner, there is not always an equivalent in other languages, which makes any approach using language unsuited for cross-cultural application. The measurement of physiological responses in relation to music also seems to provide little reliability, as there seem to not be distinguishable patterns of physiological changes for every basic emotion and changes in physiology can also be caused by other means.

As we have seen in chapter 3, facial expressions of emotion are a more specific indicator of an experienced emotion than self-report, and due to the avoidance of self-reported, verbal emotion ratings, this approach is perfectly suited for cross-cultural application. Juslin and Sloboda (2010) request a way to study responses to music that leads to more valid conclusions about experienced states in response to music (pos 287) and also Leman (2007) advocates the need for a method that uses less interpretation than required by the first-person and third-person description. Analysing facial expressions of emotion in response to emotion evoked by music provides a method that is neither dependent on introspection, nor interpretation.

In the following sections, I will give an overview about methods that can be, and partially also have been, applied to study facial expressions in connection with music.

5.2 Optical Motion Capture

Motion capture technology can be applied to record motion of objects and people. It finds utilisation in a big variety of areas: entertainment, animation in movies and computer games, sports and last but not least it is used in dance and movement studies in musicology (e.g. Nymoen, 2013; Danielsen et al., 2015).

Optical motion capture is only one of many forms of motion capture. If one wants to track motion, devices that are used in everyday life can already serve as motion capture systems. For example can smartphones be used as they have a built-in accelerometer that uses three parameters (x, y, and z-coordinate) to determine the change of position in a 3-dimensional space. The optical motion capture system that is going to be described here uses a set of cameras that send out infrared-light from an area close to the camera lens. The light is then reflected by markers coated with reflective material. These markers are available in different sizes, starting at 3-4mm. The reflected light is captured by the cameras. As the light is not sent out by the markers but only reflected, we call this a *passive system*. The data that is collected can be used to e.g. calculate a three-dimensional representation of a subject and to determine the positon in the two-dimensional or three-dimensional space. In addition it translates the data captured by the sensor into information about the motion of the captured subject (Nymoen, 2013:14). The body motions of the subject can be tracked quite accurately with this method.

Due to a high resolution of the cameras, even rather small movements, like facial expressions, can be captured. This has been mainly done to create animations for movies and computer games, but only one study has been found where facial expressions have been captured in connection with music. The study has been conducted by Livingstone, Thompson, Russo (2009). The researchers investigated facial expressions in relation to perception and production of emotional singing. The researchers conducted two experiments, one using a three dimensional passive optical motion capture system and one using Electromyography (EMG; see next section). From a professional singer emotional singing for six phrases in three different intentions (sad, happy, neutral) were recorded on video. The verbal content throughout all the emotion conditions stayed identical so that in total 18 stimuli were produced.

For each individual stimulus participants were asked to observe the facial expressions of the singer in order to be able to imitate the emotional singing as accurate as possible. In order to capture the motion of participant's facial expressions, in total thirteen reflective markers were placed on the forehead, inner and middle of each eyebrow, the nose bridge as well as the nose tip. Additional markers were placed on the left and right side of the head, upper and lower lip, and the corners of the lips. Ten of the thirteen markers were 3mm in diameter, the remaining three markers were 9mm in diameter. Four cameras were used to collect the data. In their paper, Livingstone and colleagues did not refer to the exact positions of the markers. Figure 5.1 and 5.2 suggest one way of how the markers could have been placed during the study.

The experiment was divided into four phases: During the presentation of the stimuli, participants were requested to pay close attention (visually and aurally) to emotion being expressed by the singer (phase one). Phase two was intended for the preparation of the

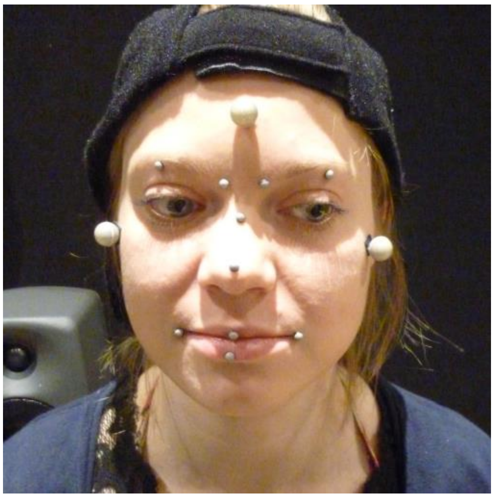

Fig. 5.1 This figure shows the attempt to replicate the marker positions (front view) from Livingstone et al.'s (2009) setup. As they only reported the size of the markers and the positions, but not which markers are placed on which position, it could only be assumed that the bigger markers (9mm) are placed on the forehead and the sides of the face and the smaller ones (3-4mm) on the more central positions.

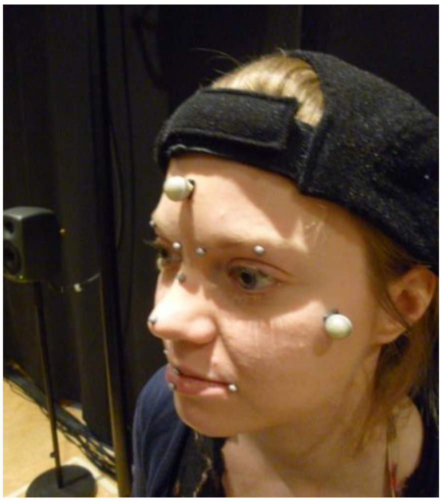

Fig. 5.2 The same setup attempting to replicate the marker positions reported by Livingstone et al. (2009),
side view. The markers are attached directly to the face with a special sticky material.

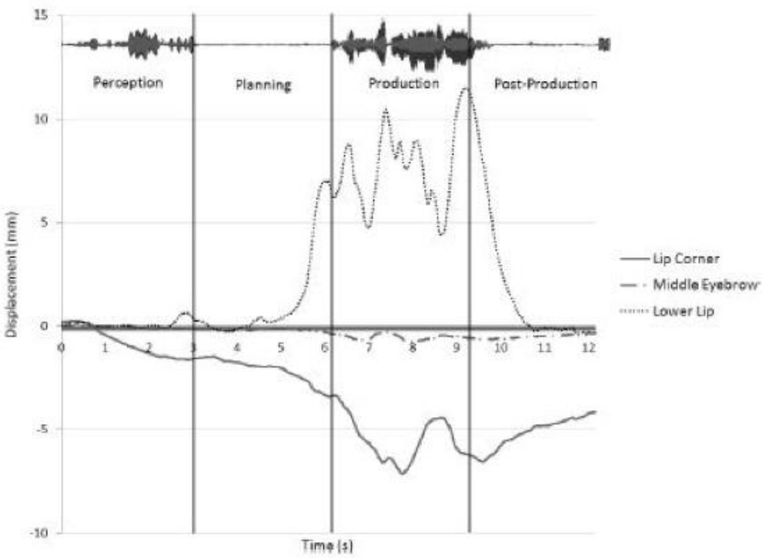

Fig. 5.3 Time course of feature displacement. In the happy condition, averaged across captures and participants. Time course is divided up into four distinct phases or epochs: perception, planning, production, and post-production. Vocalization is shown in the acoustic signal. (Livingstone, Thompson, Russo, 2009:478)

imitation. In the third phase participants had to express and imitate the emotional singing as accurately as possible. Phase four was referred to as the post-production phase, which had no specific task to it and directly followed the imitation. Livingstone and colleagues predicted to find facial expressions of emotion in all four phases. Using the motion capture system, it was possible to analyse facial movements that have been associated with emotional singing.

Only preliminary results of the data are presented in the paper. The data suggests that the eyebrows and the lip corners were significantly relevant across conditions, which matches with the theories that the eyebrows and lip corners play a significant role in the production and perception of facial expressions of emotion.

The study by Livingstone, Thompson, Russo (2009) shows that optical motion capture can be used as one method to study facial expressions of emotion in response to music. As the study has shown, the eyebrow and lip corner movement can be detected in the perception phase already. Due to the nature of the task, it cannot be appraised if the movement occurred in reaction to the emotional singing or if it has been an imitation of the expressions the participants had observed in the singer.

In my view, using optical motion capture in studying facial expressions in response to perceived emotion in music might be a somewhat useful, although it would not be my first choice of method and I would argue that for studying emotions evoked by music it is not a suited at all for several reasons: In order to get a good representation of the face and facial expressions, ideally the whole face or at least a substantial part of it would need to be covered with markers in order to be able to get a good representation of the face and to track the changes in the different regions that are involved in facial expressions of emotion. This can lead to a feeling of unpleasantness in the participant. The reflective markers would need to be placed on the face, which draws the participant's attention to the area of interest. This awareness may lead to voluntary and involuntary movements of the face and impact the data. In a test-trial I have also discovered that cameras need to be placed rather close to the participant in order to be able to capture the reflection of the small face-markers. Markers used have been 3-4 millimetre in diameter. Using bigger markers would not be recommended. During a test trial where I tried to replicate the marker positions reported in the study by Livingstone and colleagues it was discovered that markers would fall off the face eventually. Another disadvantage is that using optical motion capture is lacking flexibility. It takes a long time to set up a system like that, so having access to a lab with a fixed system would be beneficial. In terms of ecological validity, the setting is rather unnatural and might even be intimidating for some participants. Researchers are dependent on participants coming to the lab, as the lab cannot be moved to the participants.

5.3 Electromyography

Facial expressions of emotion can also be measured by using an electromyograph. As it has been described in chapter 3, electrical activity in the muscles is measured through electrodes that are attached to the skin on the surface of the muscle of interest.

When a muscle cell is activated a so-called action potential is generated, which can be measured by the device. An action potential can be measured prior to the firing of a neuron and thus indicates activation of the respective neuron. In order to measure an action potential, specific electrodes have to be placed on the skin just above the centre of the muscle of interest. For measuring facial muscles, electrodes that are attached to the surface of the skin are preferred, as the alternative needle electrodes used in intramuscular EMG are invasive and should not be applied by unexperienced researchers. Due to their invasiveness it is less likely that a method using those electrodes by musicologists would fulfil ethical research standards. In addition the muscles involved in facial expressions of emotion are to be found on the surface rather than in depth. Before the electrodes can be attached, several preparation steps have to be performed (see chapter 3).

In their article referred to in the previous section, Livingstone, Thompson, Russo (2009) conducted a second experiment using sEMG. The setup of the experiment was exactly the same as for the optical motion capture experiment, the only difference was, that instead if reflective markers, surface electrodes were attached on the skin above the zygomaticus major muscle that is involved in smiling, and the corrugator that is involved in frowning. Figure 3.1 in chapter 3 shows where the respective muscles are positioned. Figures 5.4 and 5.5 show an attempt to visualise how the setup would have looked like, although Livingstone et al. probably used smaller electrodes in their study.

As predicted the activation of zygomaticus major was related to positive states whereas corrugater activity was related to negative states. Muscle activity was already measured in the perception phase. Also in this experiment it is not sure if the predicted muscle activity was observed due to emotion or because of conscious or subconscious imitation of the singer.

Fig 5.4 This figure shows the attempt to replicate the electrode positions (side view) from Livingstone et al.'s (2009) setup. The electrodes used in the study were probably smaller than the ones that were available to me.

Fig 5.5: The same setup attempting to replicate the electrode positions reported by Livingstone et al. (2009), front view. Ideally electrodes would need to be placed on both sides of the face, which was not possible due to space constraints.

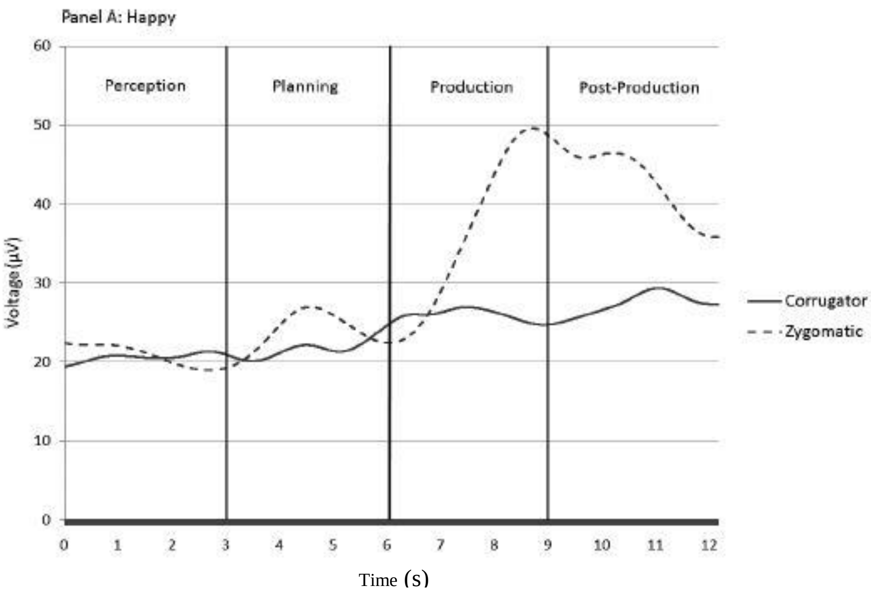

Fig 5.6 EMG activity of zygomatic and corrugator muscles in happy condition. (Livingstone, Thompson, Russo, 2009:482)

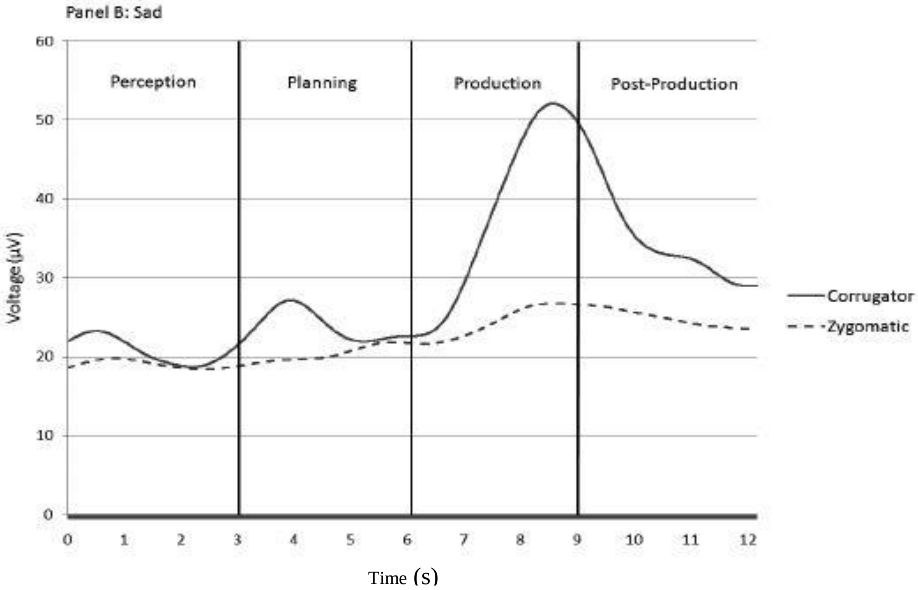

Fig 5.7 EMG activity of zygomatic and corrugator muscles in sad condition (Livingstone, Thompson, Russo, 2009:482).

Although electromyography seems to serve as a convenient method to study facial expressions in emotional singing, I do not see any advantage using this method for studying emotions evoked by music. Even though the device is way more flexible than the optical motion capture and get rather specific data on muscle activation, the disadvantages overweigh. The electrodes need to be placed on the face, which draws attention to it.

Electrodes attached to the face can feel very irritating due to the texture of the gel the electrodes are filled with and the weight it puts on the face. Only a certain amount of electrodes can be applied to the limited space. Another problem is, that muscle sizes and the exact position of a muscle vary from person to person, which makes precise positioning of the electrodes rather challenging.

5.4 Facial Action Coding System

One method that can be used to study facial expression of emotion without drawing attention to the face is the Facial Action Coding System (Ekman & Friesen, 1978) introduced in chapter 3. An advantage FACS has in comparison to EMG and optical motion capture is that nothing needs to be attached to the participant's face which makes it far more convenient for both the researcher and the participant. As mentioned in chapter 3, the participant is recorded on video during the whole experiment. The coding of changes in the face is carried out on a frame-by-frame basis. The higher the frame rate that has been chosen for the recording, the longer it takes to code all the events occurring, but also a more detailed coding is possible when a higher frame rate is used. To avoid that the participant's attention gets drawn to the face, a cover story can be invented to avoid demand characteristics. For example can physiological measurements be used not just to serve as a cover but also to record additional data that can be used for comparison and validation of the results obtained with the FACS.

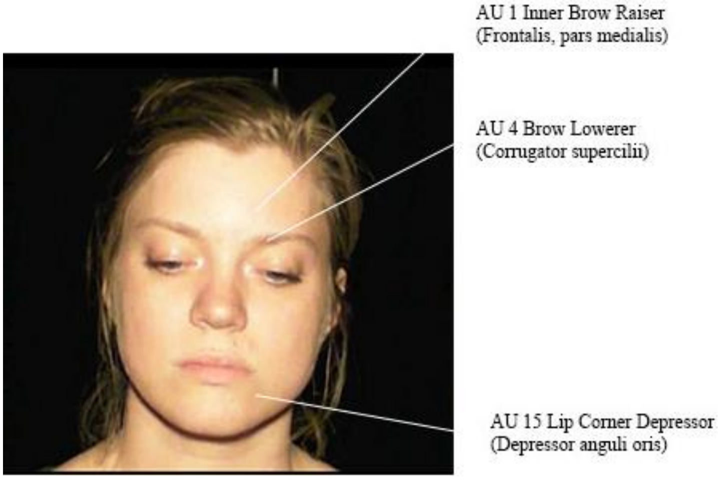

Fig 5.8 This figure shows the positions of AU 1, AU 4, and AU 15 and the associated muscles. In FACS this combination of AUs would be judged as a facial expression of 'sadness'.

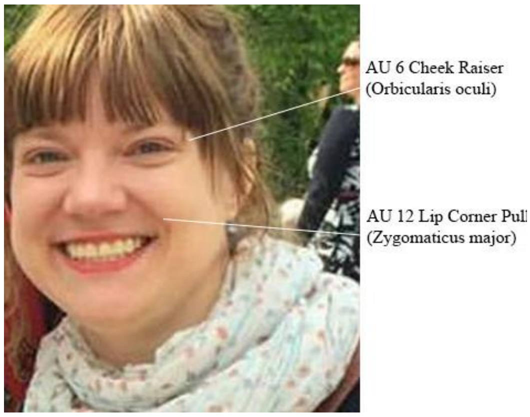

Fig 5.9 This figure shows the positions of AU 6 and AU 12 and the associated muscles. In FACS this combination of AUs would be judged as a facial expression of 'happines'.

Although the FACS has been around for several decades and been applied to a wide range of studies on facial expression of emotion, no study has applied this method to study facial expression of emotion in reaction to music-evoked emotion. The emotional responses that are most likely to expect in reaction to music can be assumed to be 'sadness' and 'happiness'. Figures 5.8 and 5.9 show the positions of the AUs that in combination are suggested to form the facial expressions of 'sadness' and 'happiness', respectively. Currently, there are no studies that have attempted to apply the FACS in connection with music.

5.5 Automatized Face Analysis

During the last few years, face recognition software has been invented and developed, and has become more and more popular in different areas. In various computer games this technology can be used to model one's own face into a computer game character. Another example is the social app Snapchat that has a face recognition feature that which lets its users apply different objects (e.g. dog ears and dog mouth) to their own face for entertainment. Face recognition software cannot only be used for the sake of entertainment but also to study human behaviour. One commercial software that can be used to analyse facial expressions of emotion is the FaceReader developed by NOLDUS Information Technology (NOLDUS, 2015).

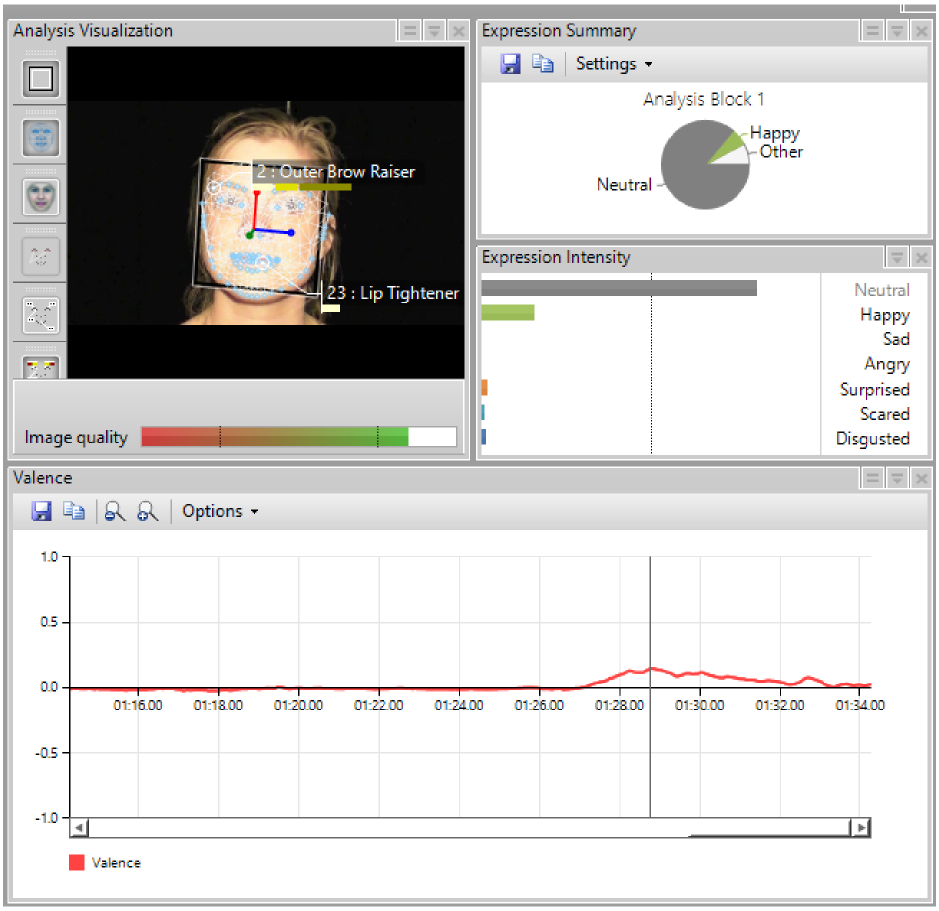

Fig 5.10 This figure shows a screenshot taken from the layout of the FaceReader 6 developed by Noldus Information Technology. The upper left frame shows the input video with the keypoins and the detected AUs. The upper right frame shows an overall summary of the expressions detected throughout the whole analysis and the intensity of the expression rated for each frame of the input file. The lower window shows the valence (positive or negative) of an expression.

The software detects facial expressions of emotion in three processing steps. First, the face is detected, subsequently the face is modelled with 500 facial keypoints. Finally, the facial expression is classified. According to their website, more than 10.000 annotated images were used to train the software and determine the keypoints. FaceReader is trained based on the Facial Action Coding System and detects six basic emotions (sadness, happiness, surprise, fear, anger, disgust). In addition contempt and neutral expressions can be detected.

When it comes to analysing the data, one can choose between using the circumplex model of affect developed by Russel (see chapter 2) and a selection of 20 action units of the Facial Action Coding System. An screenshot of the layout of the FaceReader6 software by NOLDUS taken during the analysis process is shown in figure 5.10. The user can either perform an analysis in real-time or after the experiment. The frames that are displayed show the input file, a summary about the most dominant expressions, the intensity of the expression for every single frame, and if the valence of the expression is positive or negative. Different features can be added to the input video, for example the keypoints that are applied for modelling the face, or the information about the AUs detected by the software during the analysis.

The NOLDUS FaceReader has been used in different research areas so far, one of them being in different branches of psychology, excluding music psychology. Only one study so far (Weth, Raab, Carbon, 2015) has applied the software to conduct a study with the intention to investigate facial expressions of emotion in response to self-selected sad music. Weth and colleagues used a multi-method approach using self-report and automated facial analysis. Self-report included questionnaires about familiarity, liking and emotion experienced (= felt) for self-selected sad music, unfamiliar happy music and unfamiliar sad music. The participants were asked to bring music that they themselves experience as sad and others would at least perceive as sad. The unfamiliar stimuli were provided by the researchers. Unfamiliarity of the stimuli was validated by a questionnaire. Before the experiment started, the participants were informed about being filmed, but not about the purpose of the filming.

Videos were recorded individually for each participant and the stimuli were presented in a counterbalanced method. Before the participants were released, they have been debriefed about the purpose of the filming. None of the participants retracted after being informed about the purpose of the study. For the automated face analysis the Noldus FaceReader was used.

The outcome of the study confirms all of the main hypotheses of the researchers:

Results show that self-selected music compared to both unfamiliar sad and happy music would generate i.) more sadness; ii.) more happiness than unfamiliar happy music; iii.) stronger overall emotions; and iv.) stronger mixed emotions.

The automatized face analysis has several advantages and seems to solve some of the problems that have been discovered with the alternative methods that have been presented. When using the FaceReader-software, nothing has to be attached to the face, which makes an experiment more convenient for the researcher and the participant. Almost no preparation time and material is needed, which makes this method also attractive from an economic point of view. As only a video camera is required during the experiment, the equipment is very flexible and can be taken anywhere. As opposed to the FACS where facial movements are coded manually, the software does this automatically. In addition, facial expressions of emotions are identified in only one step. Because of the flexible and timesaving nature of this method, a higher number of participants can be studied and analysed in a shorter amount of time than the other methods allow for. The only disadvantages that could be found are, that a cover story needs to be invented (e.g. by measuring physiological data in addition), and that the software is very costly.

All in all it can be stated, that the automatized face analysis is a method that can be used to study music-evoked emotions in an ecological valid and reliable way. The method is ecological valid because the equipment can be taken anywhere and thus the listener can be studied in a natural listening environment. Reliability can be reached because the whether an emotion is experienced in response to the music is displayed on the face and does not require interpretation or introspection.

5.6 Summary

Even though not much research has been done to investigate how facial expressions of emotion can serve as a reliable indicator for an emotional experience in connection with music, the studies presented above indicate that this approach is promising. Not only can we avoid the confusion between perceived and evoked emotions since the results are rather unambiguous, the facial expression approach has the potential to be applied in cross-cultural research. Out of all the presented methods, it can be suggested that the automated face analysis provides the most reliable method. The markers used in optical motion capture and

	motion capture	electromyography	FACS	automatized face analysis
advantages	• data collected represents the motion of the surface of the face in a 3-D space • if one chooses to use enough markers to cover the face this would lead to a good representation of facial features • markers are wireless • objective measurement of motion	• precise measurement of muscle activation • objective measurement of muscle activation • does not necessarily need a fixed setting	• flexible – coding and analysis are performed on video recordings of the subjects • nothing needs to be attached to the face	• flexible - analysis is performed on video recordings of the subjects • nothing attached to the face • software automatically identifies AUs, their intensity, valence • automatically judges the facial expression of emotion • takes into account that more than one emotion can be felt and displayed • timesaving
disadvantages	• markers have to be attached to the face • markers can fall off during the experiment • attention to the face can lead to demand characteristics • the more markers are used, the longer post-processing and analysis take • experiments have to take place in a fixed setting (e.g. lab)	• wires • electrodes have to be attached to the face • limited space in the face, only a selection of muscles can be measured • attention to the face can lead to demand characteristics	• training and coding are time intensive • coders need to be blind to the hypothesis • a cover story needs to be invented and probably additional physiological measurements need to be added so that no attention is drawn to the face	• the software is costly • a cover story needs to be invented and probably additional physiological measurements need to be added so that no attention is drawn to the face

Table 5.1 Overview over advantages and disadvantages in the different methods on measuring facial expressions of emotion

the electrodes used for EMG need to be attached to the face and thus draw attention to the area of interest. In addition, having something attached to one's face is rather unpleasant. These factors could influence the results. The face recognition software on the other hand can not only be used in disguise, but also can be applied outside of the laboratory which makes it much more flexible and convenient for both participant and researcher. The participant can be studied in his or her natural listening environment which adds ecological validity to this approach. The researcher does not have to rely on resources such as a laboratory that has a motion capture system installed. An overview about the advantages and disadvantages of each method can be seen in table 5.1.

In the following chapter I am going to present a pilot study I myself have planned and conducted in order to test the software that has also been used by Weth et al. (2015).

Chapter 6

Using Automatized Face Analysis for Studying Music-Evoked Emotion – A Pilot Study

This chapter presents a pilot study that I have conducted with a trial version of FaceReader6 prior to the publishing of the study by Weth et al. (2015). The aim of this experiment was to see if the FaceReader6 would be useful for the study of emotions evoked by music can be rather seen as a test of method.

6.1 The Software

The version of the software used in the pilot experiment was a 30-day free-trial version of the FaceReader 6, which comes with various limitations: It is only possible to analyse videos with a length of a maximum of 120 seconds. For each project only five videos can be analysed. In addition, some functions of the program (e.g. *gaze-direction*) were deactivated. All in all, the remaining features offered in the trial version were still satisfactory for this test of method.

6.2 Aims and Questions

The aim of this experiment was to test, if the FaceReader6 is able to detect facial expressions of emotion in response to self-selected happy and sad music and if the detected expressions match the reported emotional experience. Questions that have been addressed concerning the usefulness of the tested software to study emotions in music are:

i. How accurate can action units (AU) be identified by the software?
ii. Do the AU detected and the emotion selected by the software match?

iii. Does the detected facial expression match the expected outcome?
    a. Do we see a facial expression of *happiness* when the participant is listening to the *happy* music?
    b. Do we see a facial expression of *sadness* when the participant is listening to the *sad* music?
iv. Are there other AU involved in the detected facial expressions?
v. What does it mean when AU are detected that are usually not present in the detected facial expression of emotion?
vi. Do these AU belong to another facial expression of emotion?

From these questions, the following hypotheses have been developed:

i. A person listening to music rated as *sad* shows activation in the action units that are involved in the facial expression of 'sadness'.
ii. A person listening to music rated as *happy* shows activation in the action units that are involved in the facial expression of 'happiness'.
iii. Music that is familiar to the listener evokes stronger emotional reactions (and therefore more likely a facial expression of emotion) than music that is unfamiliar to the listener.
iv. The data will show different emotional experiences (peaks) in response to individual stimuli.

6.3 Participants

The purpose of this project was to investigate whether the software would pick up facial expressions of emotion while listening to self-selected happy and sad music. For this reason, it was sufficient to test the method on only two persons. Participants were two female students of the age of 24 and 27.

6.4. Stimuli and Materials

To assure good image quality, shadows in the face of the participant have to be avoided. This was accomplished by placing the light source right behind and slightly above the camera. The music was presented through two loudspeakers in stereo. The participants were recorded continuously throughout the experiment using a Sanyo handycam with a framerate of 25 frames per second.

Prior to the experiments the participants were asked to bring their own music. Participant 1 brought one song for each condition (‘happy’ vs. ‘sad’) and was presented with each stimulus twice. White noise was played for the duration of 30 seconds between the stimuli to avoid carry-over effects. Participant 2 brought two stimuli for each condition and was presented with each stimulus once. Due to the aim of the experiment only being a test of method, this imbalance of the number of stimuli is not important.

A detailed overview about the order of the stimuli, their duration, and their condition is given in table 6.1 and table 6.2.

Stimulus	**Duration**	**Condition**
White Noise	30 seconds	‘neutral’
Our Lady Peace – “4am”	1.30 minutes	‘sad’
White Noise	30 seconds	‘neutral’
Los Campesinos – “You! Me! Dancing!”	1.30 minutes	‘happy’
White Noise	30 seconds	‘neutral’
Los Campesinos – “You! Me! Dancing!”	1.30 minutes	‘happy’
White Noise	30 seconds	‘neutral’
Our Lady Peace – “4am”	1.30 minutes	‘sad’

Table 6.1 Stimuli, duration, and reported emotion categories for participant 1.

Stimulus	Duration	Condition
White Noise	30 seconds	‘neutral’
Damien Rice – “9 Crimes”	1.30 minutes	‘sad’
White Noise	30 seconds	‘neutral’
Travis – “Happy”	1.30 minutes	‘happy’
White Noise	30 seconds	‘neutral’
Frank Turner – “Try this at home”	1.30 minutes	‘happy’
White Noise	30 seconds	‘neutral’
Death Cab for Cutie – “Transatlanticism”	1.30 minutes	‘sad’

Table 6.2 Stimuli, duration, and reported emotion categories for participant 2.

6.5 Procedure

Participants were asked to sit in a chair which was positioned in front of the camera and were instructed to listen to the music and relax. The presence of the camera has been explained as being for documentation and later reference. Stimulus representation was counterbalanced between the ‘happy’ stimuli and the ‘sad stimuli’. Prior to each stimulus, thirty seconds of white noise were presented. The duration of the experiment for each participant was about twenty minutes. After the experiment, participants where debriefed about the presence of the camera.

6.6 Analysis

Due to the limitations of the trial-version of the FaceReader 6 software, the raw video data were cut with iMovie 7.1.4 so that the final length of each stimulus did not exceed two minutes. The resulting videos were converted from iMovie-files to .mov files and subsequently imported to FaceReader 6. Different projects needed to be created since the test version only allows the analysis of five videos à 2 minutes.

After the videos were imported into the FaceReader 6, the automatic analysis has been started for several projects simultaneously. Only a selection of the preliminary results for participant 2 will be presented here. Prior to the analysis, the feature that visualizes the activated action units was chosen.

6.6.1 Condition 'happy'

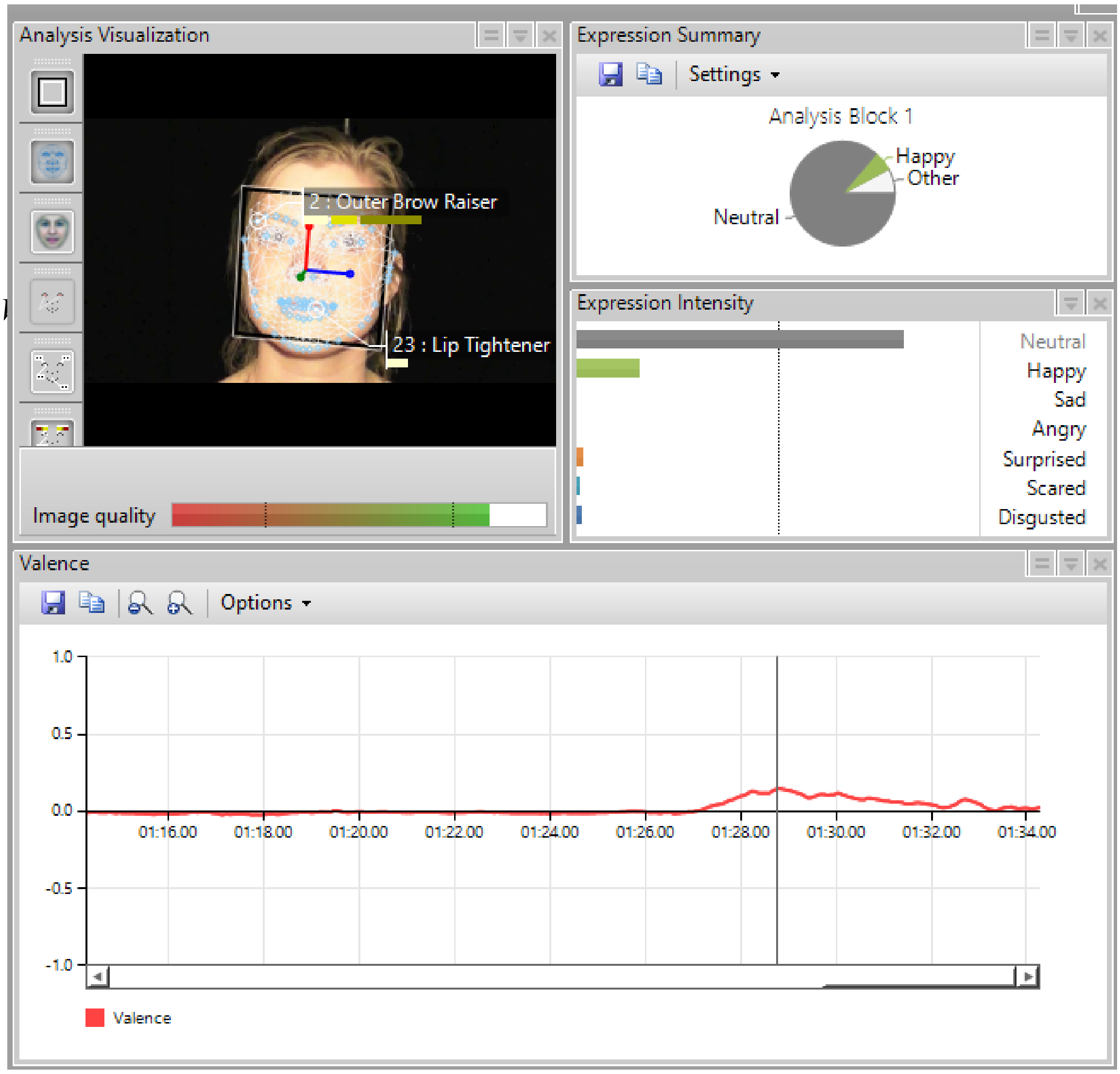

Fig 6.1: Participant 2; condition 'happy'; stimulus: "*Try this at home*"
Time Code 01:14:00-1:34:00

The visualization in figure 1 shows the image of the participant and the detected Action Units. The graph at the bottom of the image shows clearly that the valence (lower graph) for the stimulus is in the positive domain. The expression summary (pie chart) shows that the software rated the dominating facial expression for the stimulus of emotion as ‘happy’. The expression intensity shows that ‘happy’ was (next to neutral) more prominent than other expressions. Action Units that are activated match this interpretation.

6.6.2 Condition ‘sad’

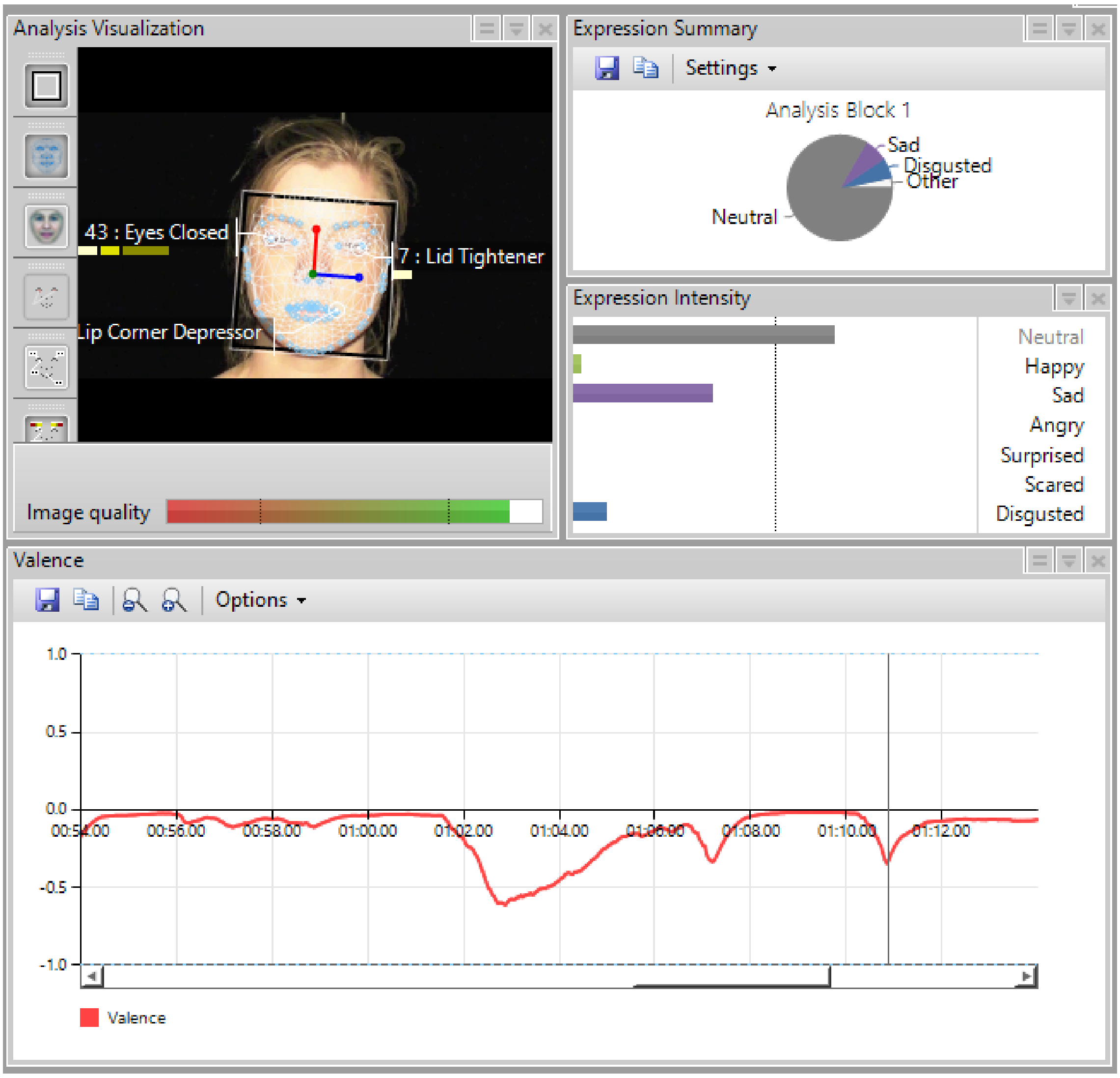

Fig 6.2: Participant 2; condition ‘sad’; stimulus “*Transatlanticism*”;
Time Code 00:54:00-1:14:00

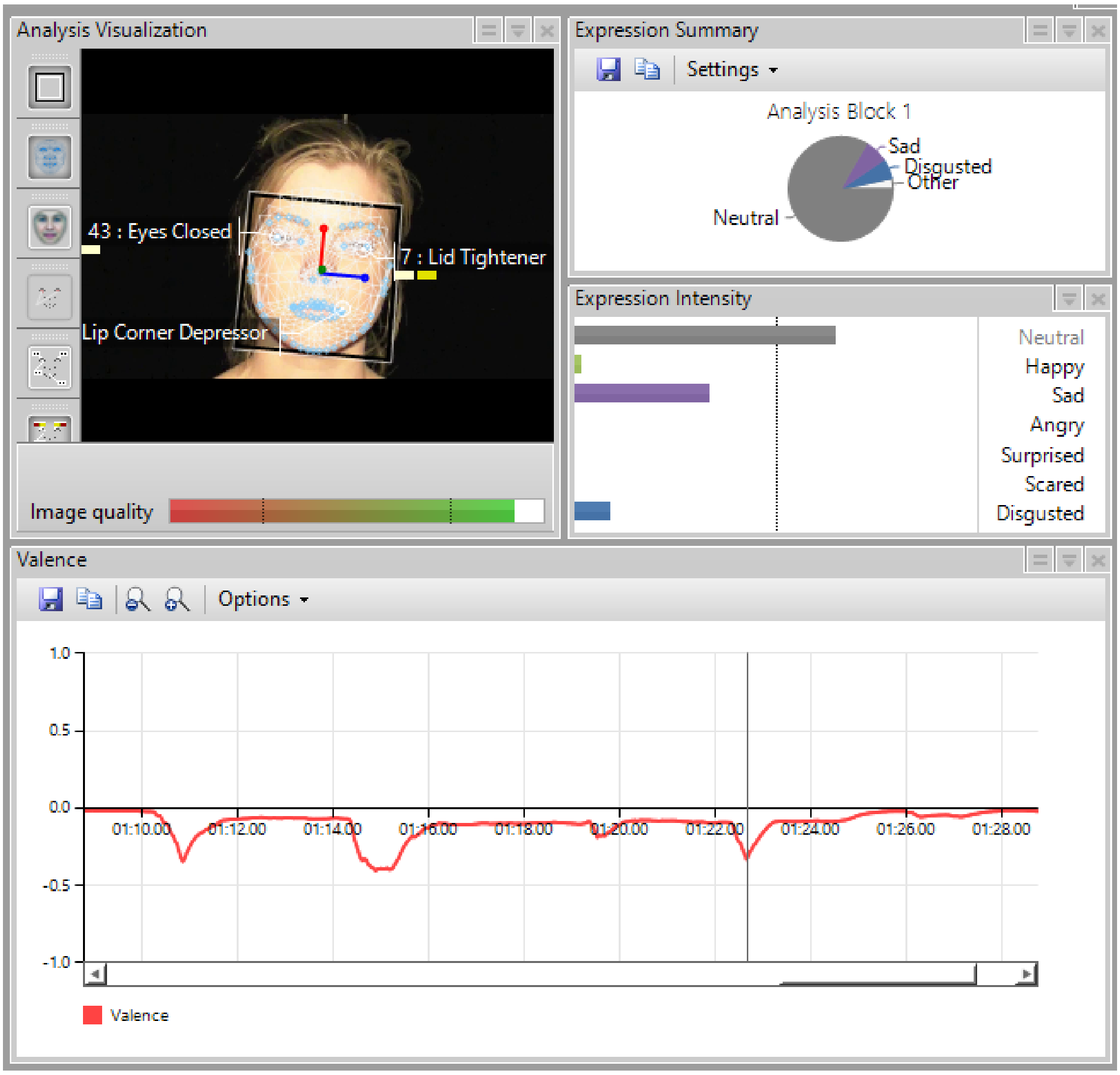

Fig 6.3: Participant 2; condition 'sad'; stimulus **"***Transatlanticism";*
Time Code 01:08:00 – 01:28:00

The visualizations in the figures 6.2 and 6.3 show the images of the participant and the detected Action Units during the presentation of the stimulus "Transatlanticism".

The graph at the bottom of the image shows clearly that the valence (lower graph) for the stimulus is having several peaks in the negative domain. The expression summary (pie chart) shows that the software rated the dominating facial expression for the stimulus of emotion as 'sad' and 'disgust'. The expression intensity shows that 'sad' was (next to neutral) more prominent than other expressions. Action Units that are activated match this interpretation.

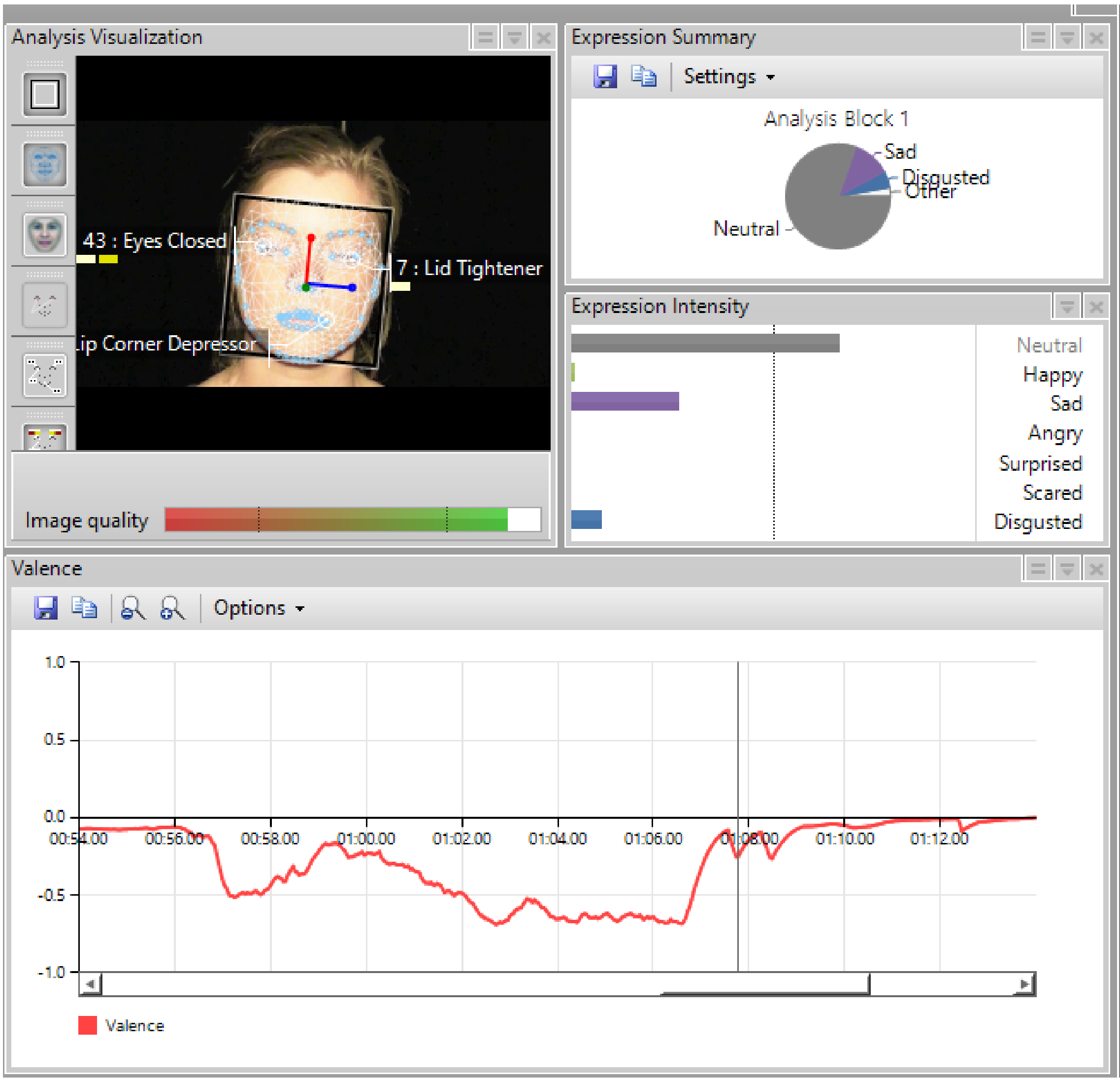

Fig 6.4: Participant 2; condition 'sad'; stimulus "*9crimes*"
Time Code 00:54:00-1:14:00

This visualization shows the image of the participant and the detected Action Units.

The graph at the bottom of the image shows clearly that the valence (lower graph) for the stimulus is having several peaks in the negative domain. The expression summary (pie chart) shows that the software rated the dominating facial expression for the stimulus of emotion as 'sad' and 'disgust'. The expression intensity shows that 'sad' was (next to neutral) more prominent than other expressions. Action Units that are activated match this interpretation.

6.7 Preliminary Results

From the data presented in the previous section we can derive some preliminary results. Hypothesis i.) stated that a person listening to music rated as *sad* shows activation in the action units that are involved in the facial expression of 'sadness' can be confirmed. The software detected activation of AU 7 (orbicularis oculii; lid tightener) and AU 15 (depressor anguli oris; lip corner depressor). Although only AU 15 is involved in the facial expression of sadness, the result from the software seems pretty unambiguous. Hypothesis ii.) stated that a person listening to music rated as *happy* shows activation in the action units that are involved in the facial expression of 'happiness'. In the 'happy' condition we can see that AU 2 (frontalis pars lateralis; outer brow raiser) and AU 23 (orbicularis oris; lip tightener) were activated in the example chosen. Even if none of these action units is involved in the facial expression of 'happiness', the software inferred that a facial expression of happiness with a low intensity was underlying. Therefore, hypothesis ii.) could partly be confirmed. Hypothesis iii.) could not be tested in this experiment. The peaks in the data imply several emotional experiences during listening to individual stimuli which confirms hypothesis iv.).

6.8 Summary

As it has been presented in the previous chapter, several methods can be used to measure facial expressions of emotion in response to emotion in music. Some of them are more useful than others. Not many studies exist that focus on emotions evoked by music and therefore this area of research is lacking methods and routines. It can be argued that the automatized face analysis software method is the most useful out of the presented ones. Both the study by Weth et al. (2015) and my own pilot experiment have shown that the software picks up the predicted facial expression of emotion in response to music. Compared with using EMG and motion capture, using the FaceReader has the advantage that it does not have to be used in a fixed setting which makes it more flexible in general. The software is non-invasive and non-obtrusive. It does not draw the subject's attention to the face as neither markers nor electrodes need to be attached to the participant's face.

Since a video camera and perhaps a good light source is sufficient to conduct the experiment, not much equipment is needed which makes it possible to conduct studies in different places all over the world. The data collection is rather uncomplicated and does not need a lot of time on setting up and calibrating the system. Not much preparation time for the participants is needed, which saves a lot of time during the experiment. This leads to the advantage that more participants can be studied, which can lead to a higher significance of the results. The analysis of the data is very simple and intuitive as well. Together with a set of questionnaires that already are used in the research on music and emotion, some correlations between different parameters can be investigated.

Chapter 7

Summary

This thesis presented a review on the state of the art in the study of music and emotion with particular attention to the reliability and ecological validity of research methods that have been applied to study music-evoked emotion.

7.1 Main Findings and Discussion

The first subquestion stated in the beginning of this thesis was concerned with the methods that have been used to study emotions in music. The methods that have been used to study perceived as well as evoked emotions in music are mainly focused on first-person and/or third person description of musical experience. First-person description methods such as self-report require introspection and interpretation of a plain subjective experience, like an emotion. Emotions are not always consciously accessible and as subjects need to constantly monitor their feelings, methods like this require a high amount of cognitive load. Subjects are not able to experience the music as they would normally do. Physiological measurements which belong to the third-person description are meant to be more objective than the highly subjective self-report. But also this methods requires a substantial amount of interpretation as the data collected solely shows the changes in physiology over time and does not give any information about the emotion being experienced, as only very few emotions show distinct patterns in physiology. Changes in physiology can also be observed without an emotion being experienced.

The second subquestion dealt with the degree of reliability these methods contain in the study of music-evoked emotions in particular. It has been found that the methods used for studying emotions in music in general do not give reliable results for music-evoked emotion in particular. It is not always clear if the experiences that are reported in the self-report method are based on emotional characteristics perceived in the music, the subjects' own emotions independent from the music, or emotions that have been evoked by the music. Demand characteristics are also problematic, as especially in self-report the underlying hypothesis can be inferred from the instructions given prior to the experiment. Also research bias can be an issue as the emotion labels chosen may show certain preferences. Self-report is most of the time obtained as one overall rating in the end of an experiment, which neglects that more than one emotion can be experienced during one musical stimulus. When continuous rating is used as a method, cognitive load and response delay play a role. Also for physiological measurements it is not sure if the changes in the data are based on an underlying emotion or if they are based on something that is independent from the music.

It can thus be stated that the methods used traditionally to study emotions in music are not reliable for studying music-evoked emotion as we cannot be sure about what is actually measured in most of the cases. Concerning ecological validity it can be stated that even though these methods can be applied in a natural listening environment, subjects have to monitor and interpret their emotional experiences in response to the music which leads to a large amount of cognitive load and thus a disruption of the musical experience.

To answer the third subquestion it has been investigated if there is a way to study music-evoked emotions that is more reliable and ecological valid than the traditional methods that are being used. With this question in mind it has been discussed to what extent a method underlying the second-person description of musical experience can be used to study music-evoked emotions. It has been shown that the study of facial expressions in response to music constitutes a good alternative to the methods that have been used in the last decades. Facial expressions of emotion are immediate and spontaneous displays of an underlying emotion and thus response delay and cognitive load can be avoided. It has been shown that the most promising method to study facial expressions in response to music is by using a software which performs an automatized analysis of the facial expressions that are displayed in response to music, detects action units, and categorises the displays detected into emotion categories. Since the emotion labels will be the same for every participant, researcher bias can be avoided. It can thus be concluded, that analysing facial expressions of emotions in response to music gives insight about the emotional experiences that are evoked by music. As this method infers evoked rather than perceive emotions, it is more reliable than the traditional methods. As the equipment is very flexible, the method can be used anywhere and thus subjects can be studied in their natural listening environment. As neither introspection, nor interpretation are required, cognitive load can be avoided and the listening experience is less likely to be disrupted than it is the case in self-report.

The answers to the last subquestion also provide the answers to the main research question. It has been shown that traditional methods do give ambiguous results and that a method that approaches the phenomenon of emotions in music in from the point of view of a second-person description can give results for music-evoked emotions rather than perceived emotions in music.

7.2 Future Work

It can be concluded that the proposed experimental framework can be applied to the study of music-evoked emotion. The method is focused on the study of music-evoked emotions rather than perceived emotions in music and thus makes it much more convenient to get the results one is looking for.

As the method has only been tested in a very simple pilot experiment so far, it has to be tested on a higher number of participants in order to gain knowledge about usability. In addition, ways of analysing the data in a meaningful way have to be defined. Future experiments could first test the method with self-selected music, as it is proposed in the experimental design, in order to obtain a baseline on the frequency and intensity of the expressions and to compare these findings with characteristics found in the music. A second step would be to use the same musical stimuli for different participants and identify differences in expressiveness in individual subjects and investigate which parameters (e.g. EQ, cultural background, musical preference, musical sophistication) might influence these probable differences. In addition, this framework should be combined with a number of questionnaires to get a better understanding about to what extent the expressiveness of the subject correlates with factors such as musical preference, empathy, and musical sophistication. The Short Test on Musical Preference (STOMP) (Rentfrow & Gosling, 2003) can be used to study if and how expressiveness and liking influence emotional experience and expressiveness. The Goldsmith Musical Sophistication Index (Gold-MSI) (Müllensiefen, Gingras, Stewart, & Musil, 2014) can be used to look at differences of emotional experience and expressiveness in musicians versus non-musicians. Another variable that can be useful to study in connection with emotional expressiveness is the Empathy Quotient (EQ) (Baron-Cohen & Weelwright, 2004) as it can be speculated that subjects who score higher on the EQ are more likely to experience emotions in music and might also be more likely to express them. Also it would be interesting to see if personality has an impact on the emotional expressiveness of a subject.

In addition, also physiological measures like electro-dermal activity can be used to provide additional data for comparison and support. Using a device for measuring galvanic skin response can also serve as a cover story to get the attention off of the camera. Musical parameters can then also be compared with the data obtained through facial analysis and physiological mesurements.

Bibliography

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 21-62.

Adolphs, R. (2006). Perception and Emotion: How We Recognize Facial Expressions

Current Directions in Psychological Science, 15(5), 222-226.

Baron-Cohen, S., & Weelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163-175.

Brewer, M. (2000). Research Design and Issues of Validity. In H. Reis & C. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: Cambridge University Press.

Buck, R. (1994). Social and emotional functions in facial expression and communication: the readout hypothesis. *Biological Psychology, 38*, 95-115.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-Based Measurement of Facial Expression With the Facial Action Coding System. In J. A. Coan & J. B. Allen (Eds.), *The Handbook of Emotion Elicitation and Assessment* (pp. 203-221). New York: Oxford University Press.

Darrow, A., Haack, P., & Kuribayashi, F. (1987). Descriptors and preferences for Eastern and Western musics by Japanese and American nonmusic majors. *Journal of Research in Music Education, 35*(237-248).

Darwin, C. (1998). *The Expression of the Emotions in Man and Animals* (3 ed.). New York

Oxford: Oxford University Press.

Davies, S. (2010). Emotions expressed and aroused by music: philosophical perspectives. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion (Kindle Edition)*. Oxford: Oxford University Press.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science, 11*(1), 86-89.

Ekman, P. (1991). *Telling lies*: W. W. Norton and Company, Inc.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion, 6*(3), 169-200.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russel's mistaken critique. *Psychological Bulletin, 115*, 268-287.

Ekman, P., & Friesen, W. V. (1971). Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.

Fridlund, A. J. (1994). *Human Facial Expression - An evolutionary view*. San Diego, California: Academic Press.

Gallese, V. (2007). Mirror neurons and the social nature oflanguage: The neural exploitation hypothesis. *Social Neuroscience, 3*(3), 317-333.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science, 2*(12).

Hopf, H. C., Muller-Forell, W., & Hopf, N. J. (1992). Localization of emotional and volitional facial paresis. *Neurology, 42*, 1918-1923.

Izard, C. E., Libero, D., Putnam, P., & Haynes, O. (1993). Stability of emotion experiences and their relations to personality traits. *Journal of Personality and Social Psychology, 64*, 847-860.

Jacob, P. (2003, 15.10.2014). Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/entries/intentionality/

Jensenius, A. R. (2007). *Action-Sound - Developing Methods and Tools to Study Music-Related Body Movement.* (Ph. D.), University of Oslo, Oslo.

Johansson, S. (2005). Language, Mind, and Self *Origins of Language. Constraints on hypotheses.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion, 3*(2), 81-123.

Johnson-Laird, P. N., & Oatley, K. (2002). Basic emotions, rationality, and folk theory. *Cognition and Emotion, 6*(3-4), 201-203.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin, 5*, 770-814.

Juslin, P. N., & Sloboda, J. A. (2010). Introduction. Aims, Organization and Terminology. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion (Kindle Edition)*. Oxford: Oxford University Press.

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. Great Britain: World Book Company.

Koelsch, S. (2012). *Brain & Music*: John Wiley & Sons, Ltd.

Koelsch, S., Siebel, W. A., & Fritz, T. (2010). Functional Neuroimaging. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion (Kindle Edition)*. Oxford: Oxford University Press.

Kohler, E., Keysers, C., Ultimá, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*: MIT Press.

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. (2014). *The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v1.0.* Retrieved from London:

Niedenthal, P. M., Krauth-Gruber, S., & Ric, F. (2006). *Psychology of Emotion: Interpersonal, Experimental and Cognitive Approaches*. New York: Psychology Press.

Noldus Information Technology. (2015). Retrieved from www.noldus.com/

Nymoen, K. (2013). *Methods and Technologies for Analysing Links Between Musical Sound and Body Motion.* (Ph.D.), University of Oslo, Oslo.

Peretz, I. (2010). Towards a Neurobiology of Musical Emotions. In P. A. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion (Kindle Edition)*. Oxford: Oxford University Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The behavioral and brain sciences, 4*, 515-526.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*, 1236-1256.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. A. (2006). Colour Categories and Category Acquisition in Himba and English. In N. Pitchford & C. P. Biggam (Eds.), *Progress in Colour Studies* (Vol. 2). Amsterdam/Philadelphia: John Benjamins B.V.

Russell, J. A. (1995). Facial Expressions of Emotion: What Lies Beyond Minimal Universality. *Psychological Bulletin, 118*(3), 379-391.

Russell, J. A., & Fernández-Dols, J. M. (1997). What does a facial expression mean? In J. A. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 3-30): Cambridge University Press.

Sacks, O. (2007). *Musicophilia*. New York: Random House.

Scherer, K., & Ekman, P. (1982). *Handbook of Methods in Nonverbal Behavior Research*. Cambridge: Cambridge University Press.

Sloboda, J. A., & Juslin, P. N. (2010). At the interface between the inner and outer world: psychological perspectives. In J. A. Sloboda & P. N. Juslin (Eds.), *Handbook of Music and Emotion (Kindle Edition)*. Oxford Oxford University Press.

Technology, N. I. (Producer). (2015, 8.3.2016). FaceReader. Retrieved from http://www.noldus.com/human-behavior-research/products/facereader

Tomkins, S. S. (2009). Affect Theory. In K. Scherer & P. Ekman (Eds.), *Approaches to emotion*: Psychology Press.

Wells, D. A. (1859). *The science of common things: a familiar explanation of the first principles of physical science. For schools, families, and young students.*

Weth, K., Raab, M. H., & Carbon, C.-C. (2015). Investigating emotional responses to self-selected sad music via self-report and automated facial analysis. *Musicae Scientae, 19*(4), 412-432.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge. *The Journal of Neuroscience, 18*(1), 411-418.

Yiend, J., & Mackintosh, B. (2005). Cognition and Emotion. In N. Braisby & A. Gellatly (Eds.), *Cognitive Psychology*. Oxford: Oxford University Press.

Zentner, M., & Eerola, T. (2010). Self-report measures and models. In J. A. Sloboda & P. N. Juslin (Eds.), *Handbook of Music and Emotion (Kindle Edition)*. Oxford: Oxford University Press.

Zentner, M. R., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*(494-521).