



Thesis for a Master of Science in the master degree course
Audio Communication and Technology

Masterarbeit für die Prüfung zum Master of Science im Studiengang
Audiokommunikation und Technologie der Technischen Universität Berlin

Martin Keil



Influence Of Music And Lyrics On The Perceived Emotions In Songs

Supervisor: Dr. Jochen Steffens
Prof. Dr. Stefan Weinzierl

Technical University of Berlin, Audio Communication and Technology
Faculty I - Humanities
Fakultät I – Geistes- und Bildungswissenschaften

Contents

| | |
|---|-----------|
| 1. Abstract | 1 |
| 2. Zusammenfassung (Deutsch) | 2 |
| 3. Introduction | 5 |
| 4. State of Research | 7 |
| 4.1. Music and Emotion | 7 |
| 4.2. Music, Vocal Expression, and Language | 13 |
| 4.3. Music and Lyrics, intertwined | 15 |
| 4.4. Mixed Emotions | 19 |
| 4.5. Melancholy & Irony | 21 |
| 4.6. Personality traits and differences | 22 |
| 4.7. About this study | 24 |
| 5. Method - experimental design | 27 |
| 5.1. Participants | 28 |
| 5.2. Material, Stimuli | 29 |
| 5.3. Questionnaire and design | 31 |
| 5.4. Pretest | 32 |
| 5.4.1. Data analysis | 33 |
| 6. Results | 34 |
| 6.1. Basic Emotions | 34 |
| 6.1.1. Music vs. Lyrics - A linear mixed model approach with control variables | 34 |
| 6.1.2. Congruent Combinations - Ratings of intended emotion | 40 |
| 6.1.3. Congruent Combinations - Ratings for other emotions | 44 |
| 6.1.4. Comparing congruent and incongruent combinations | 46 |
| 6.1.5. Interaction with other emotions | 48 |
| 6.1.6. Mixed Emotions | 53 |

Contents

| | |
|---|------------|
| 6.2. Valence and Arousal | 59 |
| 6.2.1. Music vs. Lyrics - A linear mixed model approach with control variables | 59 |
| 6.2.2. Congruent Combinations | 61 |
| 6.2.3. Comparing Congruent and Incongruent Combinations | 67 |
| 6.2.4. Discrete and Dimensional Model | 71 |
| 7. Discussion | 74 |
| References | 82 |
| A. Music and lyrics | 89 |
| A.1. Happy | 89 |
| A.1.1. Gibberish for Happy | 89 |
| A.2. Sad | 90 |
| A.2.1. Gibberish for Sad | 90 |
| A.3. Angry | 90 |
| A.3.1. Gibberish for Angry | 90 |
| A.4. Tender | 91 |
| A.4.1. Gibberish for Tender | 91 |
| B. Distribution of ratings | 92 |
| B.1. Ratings for the four basic emotions in all combined stimuli | 92 |
| B.2. Ratings for congruent combinations and control stimuli | 96 |
| C. STOMP: Median Split Differences | 103 |
| D. Digital Resources | 104 |

List of Figures

| | |
|--|-----|
| 4.1. Circumplex model of affect | 11 |
| 6.1. Q-Q-Plots for the main LMER models | 36 |
| 6.2. ME-MS Score distribution | 38 |
| 6.3. STOMP Score distributions | 39 |
| 6.4. Barplot for intended emotion ratings on all congruent and control stimuli . | 41 |
| 6.5. Barplot for melancholy ratings on all combined stimuli | 54 |
| 6.6. Barplot for irony ratings on all combined stimuli | 55 |
| 6.7. Boxplots for valence ratings on all congruent and control stimuli | 62 |
| 6.8. Boxplots for arousal ratings on all congruent and control stimuli | 65 |
| 6.9. Estimated marginal means for valence model | 69 |
| 6.10. Estimated marginal means for arousal model | 70 |
| 6.11. Mean ratings for all stimuli in the valence and arousal plane | 72 |
| B.1. Ratings for happiness in all combined stimuli | 92 |
| B.2. Ratings for sadness in all combined stimuli | 93 |
| B.3. Ratings for anger in all combined stimuli | 94 |
| B.4. Ratings for tenderness in all combined stimuli | 95 |
| B.5. Ratings for Congruent Combination: Happy | 96 |
| B.6. Ratings for Music Only: Happy | 97 |
| B.7. Ratings for Lyrics Only: Happy | 97 |
| B.8. Ratings for Congruent Combination: Sad | 98 |
| B.9. Ratings for Music Only: Sad | 98 |
| B.10. Ratings for Lyrics Only: Sad | 99 |
| B.11. Ratings for Congruent Combination: Angry | 99 |
| B.12. Ratings for Music Only: Angry | 100 |
| B.13. Ratings for Lyrics Only: Angry | 100 |
| B.14. Ratings for Congruent Combination: Tender | 101 |
| B.15. Ratings for Music Only: Tender | 101 |
| B.16. Ratings for Lyrics Only: Tender | 102 |

List of Tables

| | |
|--|-----|
| 4.1. Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions | 14 |
| 6.1. Variance and SD for random effects of all four main LMER models | 37 |
| 6.2. Values for fixed effects of all four main LMER models | 37 |
| 6.3. Mean ratings for intended emotions | 40 |
| 6.4. F-Statistics for all ANOVA-Tables of the linear mixed effects models for Intended vs. Other Emotions | 46 |
| 6.5. Music with incongruent lyrics and the two main emotions | 52 |
| 6.6. Lyrics with incongruent music and the two main emotions | 52 |
| 6.7. MIN-Statistics (mean) for all incongruent stimuli | 58 |
| 6.8. Variance and SD for random effects of valence & arousal main LMER models | 60 |
| 6.9. Values for fixed effects of valence & arousal main LMER models | 60 |
| 6.10. LMER for comparison of discrete and dimensional model | 73 |
| C.1. Median split of the data set by liking or disliking a genre | 103 |

Eidesstattliche Erklärung

Ist jeder an der TU Berlin verfassten schriftlichen Arbeit eigenhändig unterzeichnet beizufügen!

Hiermit erkläre ich an Eides statt gegenüber der Fakultät I der Technischen Universität Berlin, dass die vorliegende, dieser Erklärung angefügte Arbeit selbstständig und nur unter Zuhilfenahme der im Literaturverzeichnis genannten Quellen und Hilfsmittel angefertigt wurde. Alle Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind kenntlich gemacht. Ich reiche die Arbeit erstmals als Prüfungsleistung ein. Ich versichere, dass diese Arbeit oder wesentliche Teile dieser Arbeit nicht bereits dem Leistungserwerb in einer anderen Lehrveranstaltung zugrunde lagen.

Titel der schriftlichen Arbeit

VerfasserIn/VerfasserInnen*

Name

Vorname

Matr.-Nr.

Betreuende/r DozentIn

Name

Vorname

Mit meiner Unterschrift bestätige ich, dass ich über fachübliche Zitierregeln unterrichtet worden bin und verstanden habe. Die im betroffenen Fachgebiet üblichen Zitiervorschriften sind eingehalten worden.

Eine Überprüfung der Arbeit auf Plagiate mithilfe elektronischer Hilfsmittel darf vorgenommen werden.

Ort, Datum

Unterschrift**

*Bei Gruppenarbeiten sind die Unterschriften aller VerfasserInnen erforderlich.

**Durch die Unterschrift bürgen Sie für den vollumfänglichen Inhalt der Endversion dieser schriftlichen Arbeit.

Words make you think a thought.
Music makes you feel a feeling.
A song makes you feel a thought.

E. Y. HARBURG

1. Abstract

In this master's thesis for the *Audio Communication and Technology* degree program at the Technical University of Berlin the effects of music and lyrics on the perceived emotions in songs were investigated. This was done by doing an online survey including stimuli expressing four different emotions in three media: Spoken word, songs with lyrics in a fictional language, and songs with German lyrics. The emotions expressed by the music and the emotions expressed by the lyrics are presented in congruent and incongruent combinations as songs to examine the effects of mixed emotions in music and text. These emotions were rated by the participants using both discrete emotion terms as well as the circumplex model of affect. Additionally, participants were asked to complete a survey designed to classify them according to the Music Empathizer Music Systemizer Theory (ME-MS) by Kreutz, Schubert, and Mitchell (2008), and for genre preferences (STOMP, Rentfrow and Gosling (2003)) in order to inquire whether these characteristics have an impact on the perception of emotions. Contrary to previous studies (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014; Fiveash & Luck, 2016), the results suggest that music and lyrics are equally important for emotion perception of valence and discrete emotions in songs, but not for arousal; lyrics have little impact on the perception of arousal. Gender, listener types, or genre preferences did not emerge as significant moderators for emotion perception in this study. Cases of mixed emotions in incongruent song combinations were found, and the terms of melancholy and irony were found to not be the best choices in describing mixed emotions in songs.

2. Zusammenfassung (Deutsch)

Musik und Text sind die zwei wesentlichen Bestandteile eines modernen Songs. Die bereits belegten Ähnlichkeiten zwischen Musik und Sprache in Form von z.B. Rhythmus und Tempo, Tonhöhe und Timbre (Arbib, 2013; Patel, 2010) legen die Erforschung der direkten Verbindung dieser beiden Kommunikationsformen in Form von Text bzw. Lyrics in modernen Songs nahe. Diese ist bisher allerdings nicht sehr weit fortgeschritten. Die bis hierhin letzten Arbeiten zu diesem Thema (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014; Brattico et al., 2011; Fiveash & Luck, 2016) sind sich einig darin, dass positive oder negative Lyrics unterschiedliche Wirkungen auf die emotionale Wahrnehmung der Hörer von Musik haben. Diesen Studien zufolge lenkten positive Lyrics von den Emotionen der Musik ab, während negative diese verstärkten. Außerdem wurde festgestellt, dass die Musik bzw. Melodie der für die emotionale Wahrnehmung von Songs wichtigere Teil sei.

Die vorhergehenden Studien hatten dabei jedoch unterschiedliche Schwächen oder Eigenheiten. So wurde entweder kein Kontrollstimulus für die Wahrnehmung des Textes allein genutzt und auch keine Songs, die moderner Populärmusik entsprechen (Ali & Peynircioğlu, 2006), oder es wurden nur wenige Emotionen geprüft und verschiedene Verarbeitungsprozesse gleichzeitig (geschriebener und gesungener Text) genutzt (Mori & Iwanaga, 2014). Darüber hinaus lag der Fokus bisheriger Forschung auf den von den Teilnehmern der Studien empfundenen Emotionen und nicht den wahrgenommenen, was ein wichtiges Unterscheidungsmerkmal für die Emotionsforschung darstellt und somit im Zusammenhang mit Songs noch weiter erforscht werden sollte (Hunter, Schellenberg, & Schimmack, 2010). Auch wurde pro Studie meist nur ein Form der Emotionsmessung angewandt, oft diskrete bzw. Basisemotionen (Ekman, 1992) und andere, wie etwa dimensionale (Russell, 1980), außen vor gelassen.

Die vorliegende Masterarbeit versucht, diese Forschungslücken zu schließen und außerdem weitere Fragen darüber hinaus zu beantworten. Dafür wurde eine Onlinestudie erstellt, die den Teilnehmern vier verschiedene Emotionen (Freude, Traurigkeit, Wut und Zärtlichkeit) jeweils in Form von Musik oder Sprache präsentierte. Daber kamen folgende Präsentationsformen zum Einsatz: Als vorgelesener Text; als Musik

2. Zusammenfassung (Deutsch)

mit Gesang in einer Fantasiessprache ohne semantischen Inhalt; als Musik mit Gesang in deutscher Sprache. Die Stimuli mit deutschem Text und Musik wurden neben emotional passenden Kombinationen auch als emotional inkongruenten Kombinationen erstellt. Dadurch konnten neben der Erforschung des Einflusses von Musik und Lyrics außerdem weitere Informationen zu gemischten Emotionen (Larsen & McGraw, 2014) gesammelt werden. Sämtliche Stimuli wurden speziell für den Zweck der Studie geschrieben bzw. komponiert und aufgenommen, damit eine hohe ökologische und externe Validität der Forschung sichergestellt werden konnte, und waren in der Komposition an moderne Populärmusik angelehnt.

Die wahrgenommenen Emotionen sollten von den Studienteilnehmern sowohl auf unipolaren Likert-Skalen für die vier Basisemotionen bewertet werden, als auch auf der zweidimensionalen Valenz-Erregungs-Skala (*Valence-Arousal*). So sollte ein direkter Vergleich der zwei Skalen zu Emotionen in Songs ermöglicht werden. Neben diesen Emotionen wurden außerdem die musik- und ästhetikbezogene Emotion der Melancholie und der Begriff der Ironie untersucht. Außerdem sollte der Einfluss persönlicher Faktoren der Teilnehmer erforscht werden; dazu wurde die noch recht neue Theorie der *cognitive listening styles* (Kreutz et al., 2008) angewandt, die besagt, dass die Menschen sich als Musik-Empathizer (ME) oder Musik-Systemizer (MS) kategorisieren lassen. Die zur Kategorisierung nötigen Fragen wurden bereits systematisch in Deutsche übersetzt und waren somit bestens für die Studie geeignet (Linnemann, Kreutz, Gollwitzer, & Nater, 2018). Darüber hinaus sollte der Einfluss von Musikstil-Präferenzen (STOMP) bestimmt werden (Rentfrow & Gosling, 2003; Langmeyer, Guglhör-Rudan, & Tarnai, 2012) und Unterschiede der Emotionswahrnehmung durch das Geschlecht der Versuchsteilnehmer untersucht werden. Für die Analysen wurden stets lineare gemischte Modelle mit zufälligen Effekten genutzt, um die Messwiederholungen der Teilnehmer miteinzuberechnen.

Die Studie ergab, dass Musik und Lyrics gleich hohe Anteile an den durch die Teilnehmer wahrgenommenen Emotionen im Bezug auf die Valenz (positiv/negativ) aufwiesen. Dabei wurden kongruente Kombinationen von Musik und Lyrics stets als intensiver in der Hauptemotion wahrgenommen als in den Stimuli ohne Text oder ohne Musik, oder als in den inkongruenten Kombinationen. Dies ist anders als in den vorherigen Forschungen, in denen positive und negative Lyrik unterschiedlichen Einfluss hatte. Die Unterschiede könnten in direktem Zusammenhang mit den verwendeten Stimuli stehen, insbesondere da es sich um populäre Musik handelt und nicht um klassische oder Soundtrack-Musik wie zuvor. Wurden die Emotionen von inkongruenten Stimuli betrachtet, so konnten auch hier zwischen der durch die Musik bestimmten Emotion und der durch die Lyrics bestimmten Emotion keine signifikanten Unterschiede festgestellt werden.

2. Zusammenfassung (Deutsch)

Während dies für die Valenz der Emotionen zutreffend war, verhielt es sich für die emotionale Erregung jedoch anders: Hier dominierte die Musik die Intensität der Wertung deutlich. Dies ist ein Unterschied, der in vorherigen Studien durch Auslassen der Erregungs-Skala nicht festgestellt und übersehen wurde.

Sowohl Melancholie, da deren Intensität fast ausschließlich von der (traurigen oder zärtlichen) Musik abhing, als auch Ironie, da dessen Wertungen stets sehr gering waren, wurden als mögliche Begriffe für gemischte Emotionen durch Musik und Lyrics ausgeschlossen. Gemischte Emotionen in Stimuli mit inkongruenten Kombinationen konnten per MIN-Statistik festgestellt werden (Hunter, Schellenberg, & Schimmack, 2008), nicht nur in Kombinationen aus fröhlichen und traurigen Stimuli.

Gewichtige Geschlechterunterschiede oder Unterschiede durch Genrepräferenzen oder kognitive Hörstile konnten dagegen nicht festgestellt werden werden.

Zärtlichkeit war eine problematische Emotion in dieser Studie, da sie am ehesten ambivalent erschien und oft nicht gut zwischen den Emotionen Zärtlichkeit, Traurigkeit und Melancholie unterschieden werden konnte.

Mit dieser Studie konnten neue Erkenntnisse über die Zusammenwirkung von Musik und Lyrics im Bezug auf wahrgenommene Emotionen gewonnen werden. Außerdem konnten Unterschiede zwischen diskreten und dimensional Emotionsbewertungen festgehalten werden. Die Studie trug darüber hinaus zur Untersuchung von gemischten Emotionen bei und konnte keine relevanten Einflüsse von persönlichen Unterschieden zwischen den Teilnehmern auf die Wahrnehmung von Emotionen in Songs feststellen.

3. Introduction

Listening to popular music is, with the always-available digital resources, a very important leisure activity nowadays (Sloboda, 2010). The effect that music has on people and their emotions has been studied thoroughly in the last years, but the focus of these studies was mainly on classical or instrumental music, as was shown in an in depth literature review by Eerola and Vuoskoski (2013), with 48% using classical stimuli and only 3% pop or rock style music. Music and language and its similarities in melody, dynamics and rhythm has also developed its own scientific research field (Patel, 2003, 2012; Arbib, 2013). The role of music and lyrics combined in popular songs on the perceived emotions is lacking the same attention and has yet to be analyzed in the same depth. There are a few studies that tried to shed light on the effects of lyrics on melodies or music in general (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014; Fiveash & Luck, 2016), but they either still used mainly classical music as the stimuli, mixed stimuli types of written and sung lyrics for control ratings, or exchanged the sung melody with another instrument. In both cases, there are discrepancies to how music and lyrics are usually processed in the real world.

Most of these studies also only used one rating scale, either discrete or dimensional emotion models, for self-reports from the participants, but did not compare different approaches (Eerola & Vuoskoski, 2013). By doing so, the aspect of valence (positive or negative emotions) was of more importance to the studies than that of arousal (high or low arousal).

This Master's thesis and its survey about music and lyrics in contemporary songs aimed to close this gap in research and tried to do so by conducting an online survey including stimuli expressing four different emotions (happy, sad, angry, tender), each corresponding to one of the four segments in the circumplex model of affect (Russell, 1980). These were presented in three different ways: In the form of 1) spoken text in German, 2) songs with a fictional language, and 3) songs with the German text as lyrics. By combining these stimuli in emotionally congruent and incongruent ways, this study tried to show how and in what way lyrics can change the perception of emotions in music. These stimuli were composed and recorded specifically for this

3. Introduction

purpose to have high ecologic validity with good external validity, sounding like real songs as much as possible, but still having high internal validity by not changing the sound and style of the recordings too much from one to another, and thus be comparable. In that way, possible problems of other studies might be reduced and a new view on music and lyrics could be possible.

By incorporating a survey about different cognitive styles of music listening (Kreutz et al., 2008), this study tried to show if these cognitive styles have an influence on the perception of music with lyrics and the expressed emotions and tried to shed further light on the rather new concept of the Empathizer-Systemizer-Theory (Baron-Cohen, 2002) that describes these cognitive styles in general; questions relating to the more conventional STOMP scales (Rentfrow & Gosling, 2003; Langmeyer et al., 2012) about genre preferences were also included, to check for another possible difference in perceiving music, as this has also not been done before in recent studies about music and lyrics in songs. It might also help to better understand the concept of mixed emotions by using emotionally incongruent combinations of music and lyrics. In these ways, this thesis aimed to provide a new view on emotions in songs.

4. State of Research

At the very core of this study is the research about music, language and emotions. Music and spoken word both have the potential to express and also evoke emotions (Scherer & Zentner, 2001; Liebenthal, Silbersweig, & Stern, 2016, and others). Research has shown what the different properties in music can express emotionally (e.g. slow tempo, low volume and minor key for sadness) (P. N. Juslin & Laukka, 2004), and there are studies about it for speech and even non-verbal vocalizations as well (P. N. Juslin & Laukka, 2003; Liebenthal et al., 2016). Various research about the connection of language and music (e.g. rhythm, tonality, etc.) has also been done. (Patel, 2010; Arbib, 2013) The connection between music and lyrics in combination is of main importance here, though. A brief introduction to all these concepts and the state of research will be done in this chapter.

4.1. Music and Emotion

Before moving on any further, it is important to say that neither everyday language nor scientific research has fully agreed upon a terminology for emotions that is valid in every case (Gabrielsson, 2001). To not avoid any misunderstandings, the terms of *emotion* and *feeling* are used in a broader sense in the following thesis.

Music and emotion, and the research of their interactions is as interesting as it is complex. For researchers, music and emotion studies poses several problems: Understanding the differences in induced and perceived emotions in and through music; how the situation of listening and listener itself affect their emotions; how to measure these emotions adequately; and how much of it is of subjective and objective nature (P. Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008; P. N. Juslin & Sloboda, 2011; Fiveash, 2014).

The connection between music and emotion is evidently strong. As seen in various previous studies, music listeners frequently reported that they listened to music for emotional reasons, e.g. to change their emotions or just to enjoy themselves

4. State of Research

(P. N. Juslin & Laukka, 2004; P. N. Juslin & Västfjäll, 2008; Evans & Schubert, 2008, and others)

The emotions music can express and induce are not limited to basic emotions like happiness or sadness. It has been shown that people associate significantly more emotions with music, such as nostalgia or tenderness (Zentner, Grandjean, & Scherer, 2008). The system of classification and interpretation of the different emotions is still under debate though. While it is still hard to say, even for music and emotion researchers, what the term *emotion* in music really means (P. N. Juslin & Västfjäll, 2008), there have been lots of different approaches in recent studies, showing that emotion research within the field of music is of growing popularity and importance (Eerola & Vuoskoski, 2013).

Emotions perceived vs. emotions felt One of the most important things to consider first when doing a music-emotion related study is, what kind of emotions are in question: Emotions *perceived* (expressed by the music) or emotions *felt* (induced by the music). An in-depth review on this topic was done by Gabrielsson (2001). The evaluation of emotions expressed by music “is mainly a perceptual-cognitive process.” (Gabrielsson, 2001, p. 124) Perceiving an expression of, for example, happiness, in the music does not necessarily mean that the listener has to feel happy him- or herself - distinguishing between the listener’s emotional response to the music and his or her perception of the expressed emotional content is crucial. This distinction however can not always be clearly observed. The literature suggests that the lines between these two alternatives can be seen “as somewhat blurred” (Gabrielsson, 2001, p. 124) and hard for the listener to differentiate between the two, even with clear instructions. It has even been suggested that they could just be opposites of the same spectrum (Gabrielsson, 2001). Empirical studies (Eerola & Vuoskoski, 2011; Evans & Schubert, 2008; Hunter et al., 2010; Kallinen & Ravaja, 2006) have found that there might be more similarities between the two than previously thought. Kallinen and Ravaja (2006) for example found that the emotions perceived in a musical stimulus were usually the same as the reported induced emotions. The only differences were in negatively perceived emotions, that aroused weaker negative or even positive felt emotions. Concerning pleasure, they could show that the induced emotions were stronger in this regard than the perceived emotions, but weaker in regards of arousal.

A reason why perceived and felt emotions might be so similar in many regards is the way we communicate emotions verbally. As there are so many individual differences in feeling emotions and the vocabulary to describe it, it is not easy to differentiate

4. State of Research

between what is felt and what is only perceived (Gabrielsson, 2001).

To be certain about what was measured, the distinction should still always be made for different reasons. Depending on the study design, what is actually happening when we perceive or feel an emotion might be different. It is also harder to measure induced emotions than it is to measure perceived emotions, so it is necessary to modify the study design accordingly to take this into account. And finally, the emotions we are *able* to feel when listening to music and the emotions we are able to perceive may in fact be different from each other (P. N. Juslin & Laukka, 2004). As said by Gabrielsson (2001, p. 139) concerning perceived and induced emotions:

The music-person-situation interplay can never be disregarded. It certainly complicates questions, but human response to music is complicated; that is why we still understand so little about it.

To measure the perception of emotions expressed in music is relatively easy as it is a process that does not necessarily need any emotional involvement of the listener and can thus be, in that sense, objective (as much as emotions can be). Especially in a setting where people are asked to evaluate music in a self-report style and not just listen to it like they would in everyday life, it is much easier to detect emotions rather than to feel them. Measuring emotions induced by music is in that regard harder to do, as induced emotions are often connected with personal experiences and memories of events or with (re)listening to a certain piece of music (P. N. Juslin & Laukka, 2004). Evidence showing that all kinds of positive emotions can be both aroused by music and perceived by music, but negative emotions are much more often perceived than felt, is another factor why distinguishing between perception and induction of emotions is so important (Zentner et al., 2008).

For all these reasons, this master's thesis will thus focus on perceived emotions, the emotions that are represented by music and lyrics, and perceived as such by the listener.

Measuring emotions perceived in music The next crucial step in music and emotion research is to decide on how to measure musically expressed emotions. In modern research, the evaluation of these emotions is done with different models. One way of approaching this problem is to use models that were intended for basic, everyday emotions and to transfer them to emotions perceived in music (P. N. Juslin & Västfjäll, 2008). The basis for most of these models are the findings and the theory by Ekman (1992) on the **basic emotions**. The basic emotions theory by Ekman states that there are six basic evolutionary emotions, that are universal for every human being, namely: anger, disgust, fear, happiness, sadness, and surprise.

4. State of Research

These can then be combined to other, more complex emotions. In terms of music, emotions that are harder to express like disgust have sometimes been exchanged for other emotion terms like *tenderness* (Eerola & Vuoskoski, 2011), to account for the findings that it is generally harder to express and induce negative emotions with music, as aesthetic emotions¹ have no severe negative impact on the life of the recipient as actual negative events would have (Zentner et al., 2008), and they are sometimes even perceived as something enjoyable (Vuoskoski & Eerola, 2012). Models based on basic emotions are called *discrete emotion models*, as they are measured on discrete scales, usually on a Likert-type scale. These scales can be unipolar, so that all emotions can coexist, or bipolar, so that some emotions are linked to each other (for example, sadness and happiness can be argued to only exist on a bipolar scale). When used in the context of music, it can be argued that these aesthetic emotions are not sufficiently expressed by basic emotions, hypothesised to initially be linked to survival (P. N. Juslin & Västfjäll, 2008; Zentner et al., 2008). Scherer (2004) suggested that researchers were in need for a new, musical emotion scale, as the scales used so far had many limitations concerning aesthetic emotions.

Following this, Zentner et al. (2008) empirically demonstrated that these primary emotions can be inadequate to describe emotions evoked in a context such as music. They therefore designed another model, especially suited to the emotion terms linked to music; they called it the **GEMS**, Geneva Emotional Music Scale. By compiling a list of music-relevant emotion terms and studying the frequency of both felt and perceived emotions across listeners with distinct music preferences, they could show that emotional responses varied a lot among the musical genres and among induced vs. perceived emotions. For that reason, they focused only on induced emotions in the follow-up research, where they could reduce the list of music-relevant emotion terms from 515 to, in the end, a 9-factorial model of music-induced emotions (Zentner et al., 2008). They were able to show that these were more efficient in describing music-induced emotions than the basic emotion or the dimensional model (see below). Even though there now even exists a validated scale for German emotion terms based on the GEMS (Lykartsis, Pysiewicz, von Coler, & Lepa, 2013), this rather new scale is only linked to music-*induced* emotions, and has also not been properly tested with lyric-based music, such as modern pop songs, as the music used for their experiments were mainly classical stimuli. Thus, this second way of measuring emotions on a discrete, music-related scale is not perfectly suited for the research at hand, that will focus on perceived emotions in songs with lyrics.

¹Emotions linked to the perception of art like music, movies, and paintings

4. State of Research

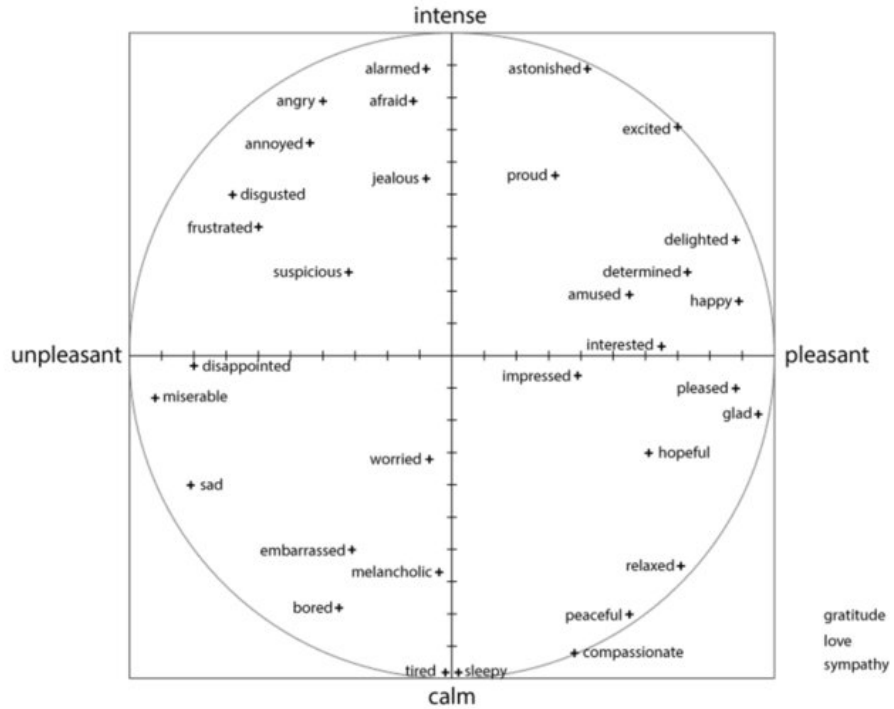


Figure 4.1.: Circumplex model of affect, including discrete emotions mapped onto the space (Haritaipan & Mougenot, 2015)

The third common way of measuring self-reported emotions is to use dimensional models, the most famous being the **circumplex model of affect** (Russell, 1980; Posner, Russell, & Peterson, 2005), that has seen various attempts at improving and changing the first concept to be perfectly suited for music emotions (Eerola & Vuoskoski, 2011). The circumplex model of affect maps all human emotions into a 2-dimensional spectrum of valence and arousal. When using this scale in a study, participants are asked to rate the emotions on how positive or negative (valence, pleasure-displeasure) and how low or high the arousal (activation-deactivation) was. In this way, all emotions can be represented as varying degrees of these two parameters. An alternative to the valence-arousal model was proposed by Thayer (1990), discriminating the emotions only by two arousal-dimensions, namely energetic arousal and tense arousal, with Vieillard et al. (2008) using this theory for music related emotion research. The dimensional approach for music emotions has even been extended to a three-dimensional approach, consisting of the three spaces pleasure displeasure, arousal calmness, and tension relaxation. However Eerola and Vuoskoski (2011) were able to show, that for music related emotions, a third dimension was mostly redundant and the system could be reduced to two without significantly reducing the quality of the analysis. Even though further testing and researching would help to find out more about the dimensional models for music emotion research, it should be save to assume for now that using a two-dimensional

4. State of Research

model of valence and arousal is appropriate. As the study by Eerola and Vuoskoski (2011) also focused on perceived emotions, using this approach for the study at hand was deemed useful.

In the mentioned study by Eerola and Vuoskoski (2011), they also compared discrete models like the basic emotion model, optimized for music related emotions including tenderness, to the dimensional models. They were able to show that discrete and dimensional models are often similar in results, so that discrete, basic emotions can be mapped into the valence and arousal space (Eerola & Vuoskoski, 2011). Referencing other, neurological studies (Gosselin et al., 2005) there seems to be evidence that discrete and dimensional approaches to emotions have different underlying processes in the brain, implicating that the choice of model has an effect on the outcome of the study; even though the discrete emotion terms might be, to some degree, mappable into the dimensional emotion space, the terms are processed differently and are not completely interchangeable - nevertheless they are highly compatible. Eerola and Vuoskoski also decided to focus on film music as the stimuli to be evaluated, so it can not be said for certain how these ratings change for other genres. Moreover, it was shown that listeners can experience both sad and happy feelings at the same time when listening to a stimulus with mixed emotional cues (Hunter et al., 2008, see section 4.4 of this thesis on mixed emotions, p. 19) Therefore this study will not only use basic emotions on unipolar scales, focusing on the emotions anger, tenderness, happiness, and sadness, like in the study by Eerola and Vuoskoski (2011), but also a valence-arousal two-dimensional approach to see if music containing lyrics are validated differently on these two models.

Another way of measuring musical emotions is to use physiological measures, such as brain scans or facial activity while listening to music (Brattico et al., 2011). But these measurements were outside the scope of the study at hand.

Summing up, these are the most important methods in recent studies to measure emotions:

- **the basic emotions theory** (Ekman, 1992) discrete emotion model
- **musical emotions scales**, discrete emotion models especially designed for the evaluation of emotions in music (e.g. the GEMS (Zentner et al., 2008))
- **the circumplex model of affect** (Russell, 1980) dimensional emotion model
- **physiological measures**, such as MRI studies (Brattico et al., 2011)

A note on music stimuli in emotion research: As stated before, most recent studies focused on classical music as the stimuli (Eerola & Vuoskoski, 2013) and have often

been chosen by the judgement of the researchers. Another problem lies in the choice of known versus unknown samples. Well-known music examples as stimuli might lead to different results because participants may already be familiar with them (Eerola & Vuoskoski, 2011), leading to associations with previous situations where the stimulus was encountered. As an alternative, researchers can opt to use synthetic examples, where all musical features can be manipulated as needed. Yet these do have the tendency to sound just like that, synthetic, and might lead to problems of external validity. It is possible to create own musical stimuli, that still rely on instruments and structures inherent to the musical genre in question, but they need to be composed and recorded first, specifically for the purpose of the research questions. This was done in this master's thesis.

4.2. Music, Vocal Expression, and Language

Both music and language are two of the most human traits of this species. There are many things that we have in common with other animals, but using meaningful sound sequences as either spoken word or music is a unique trait of human beings (Patel, 2010). They both draw on lots of similar patterns to create meaning, like pitch contrasts, timbral contrasts, or rhythmic contrasts. For that reason, it comes as no surprise that the research of music and language as a comparative field has drawn a lot of attention. There is literature about the development of language and music in various research fields (Arbib, 2013; Patel, 2010). Music and language seem to not only rely on the same resources when they are being created, like syntax and structure (Lerdahl, 2013), but also when perceived and in emotional terms (Scherer, 2013).

Neuroscientific studies show that the processing of music and language is in some ways done in the same parts of the brain (Koelsch et al., 2002). When it comes to semantics of music and language - the meaning they try to convey - Koelsch et al. (2004) could show that music can influence the processing of words and also “prime representations of meaningful concepts” (p. 303). They did this by measuring EEG-feedback from the participants when visually presenting words that were shown after hearing a spoken sentence or a musical stimulus. It shows that music can transfer a lot of semantic information (even though they are not necessarily the same semantics as in language) and that both language and music have a lot in common when it comes to processing emotional information.

In the book *Music, Language, and the Brain* by Patel (2010), which includes various recent research in the field, the author summarizes the main aspects of music and

4. State of Research

language processing as follows (p. 417):

1. As cognitive and neural systems, music and language are closely related.
2. Comparing music and language provides a powerful way to study mechanisms that the mind uses to make sense out of sound.

Concerning the ability of the voice alone to transport emotions, a systematic “review of 104 studies of vocal expression and 41 studies of music performance” was done by P. N. Juslin and Laukka (2003, p. 770). The results suggested that there are many similarities between the two channels of communication, for one the accuracy of being able to communicate discrete emotions, and for the other the acoustic cues used to communicate these emotions. The cues for both music and vocal expression can be found in Tab. 4.1. These results were cross-culturally valid. They also found that, in terms of basic emotions, “Anger and sadness were generally better communicated than fear, happiness, and tenderness” (P. N. Juslin & Laukka, 2003, p. 797).

| Emotion | Acoustic cues (vocal expression/music performance) |
|------------|--|
| Anger | Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity |
| Fear | Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity |
| Happiness | Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity |
| Sadness | Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and microstructural irregularity |
| Tenderness | Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and microstructural regularity |

Note: F0 = fundamental frequency

Table 4.1.: Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions (P. N. Juslin & Laukka, 2003, p. 802)

All of these findings have implications for the current study, because a) both music and lyrics may be processed differently depending on the semantics and emotional

meaning of each one when combined into one stimulus, b) using vocal expressions without semantics of language can also transport emotions and they cannot be separated completely.

Most of these studies have been conducted using English as the main language, so using a different language, like in this case German, adds to the reasons why this study was conducted.

4.3. Music and Lyrics, intertwined

Although there have been various research projects on music and emotions as well as language and emotions, the research on the combination of music, language and emotion in the form of songs with lyrics has been surprisingly sparse. We still don't know too much about the way lyrics may affect our perception of emotions in music (Stratton & Zalanowski, 1994). As this is the case, the following will try to summarize every major study on this topic so far.

One of the earliest studies in this field was done by Galizio and Hendrick (1972), who had subjects listen to folk songs, that had or did not have guitar accompaniment, and whose lyrics were either sung or spoken. These lyrics included political information. They found that, independent of whether the lyrics were sung or spoken, the guitar accompaniment always increased positive emotional arousal in the participants. They also stated that in the cases with a guitar accompaniment, the semantics of the lyrics were more persuasive, so they concluded that the music creates positive arousal and in turn leads to higher acceptance of persuasion.

In another study, Coffman, Gfeller, and Eckert (1995) compared the effect of atonal music and a negative poem, either alone or combined, on both musicians and non-musicians. The authors had previously conducted a study (Gfeller & Coffman, 1991) where they investigated the contributions of music and text to affective response in trained musicians only, using tonal and atonal music alone, a recited poem alone or a combination of the poem and each music style. In both cases, ratings for the text only conditions were more positive than with atonal music. They could also recreate the finding that text sung or text spoken did not make a significant difference. Coffman et al. (1995) hypothesized that the more negative ratings were due to an optimal level of complexity for music/text stimuli being surpassed. When this happens, emotional responses are more negative, which was the case for atonal music. They concluded that text and music contribute different kinds of information to the listener, with different emotional responses. They also state that the music vs. non-music condition was the most important factor, and not the level of

4. State of Research

musical expertise of the listener, at least in their later study, where both musicians and non-musicians participated (Coffman et al., 1995). They did use a poem for this research though, and not a specific song lyric.

Contrary to the findings above, where music always played a significant role, three experiments conducted by Stratton and Zalanowski (1994) resulted in the conclusion that lyrics have a better ability to direct mood change than music alone, and that lyrics can play a crucial role in songs. In the first experiment, they used the song and the lyrics of a sad composition by Oscar Hammerstein and Jerome Kern and played only the melody on a piano, lyrics recited without music, or lyrics sung accompanied by the piano. When they compared the music alone with the combination, the music alone was rated as a lot more pleasant and positive than the one including depressing lyrics. For the second experiment, they changed the lyrics to a positive one, and added an up-beat tempo music, and cross-combined them with the stimulus from the first experiment. This again showed an importance of lyrics, where even the happy song was changed to be more depressing with negative lyrics. To study the lasting effects of lyrics on melody, they did a third experiment, where the melody of the song was first presented including the negative lyrics and then, a week later, without the lyrics. These ratings were then compared to the (positively rated) melody only condition of the first experiment. They could show that this melody was rated significantly less pleasant after the participants knew that they belonged to a negative lyric, even when they did not hear the text again. They also hypothesized that in case of incongruent combinations of emotions in different stimuli, the lyrics will overtake the emotional evaluation. Criticism about these experiments included that the original melody was emotionally too ambiguous and that the more positive, upbeat version did not show statistical significance in the change of ratings for positive affect, therefore these findings may only be appropriate for ambiguous stimuli.

Sousou (1997) built on the findings of Stratton and Zalanowski, but tried to use more obviously sad or happy musical stimuli, and they did not change the melody to be sung, but rather to be read while listening to the music. They used happy and sad lyrics, and the conditions *no music*, *happy music*, and *sad music*. In this study, the music appeared to be the dominant factor for emotions in songs. Combining their findings with their critique on the research by Stratton and Zalanowski (1994) they argue that “music is more likely to influence mood when the music is less ambiguous, but lyrics may influence mood when the music is ambiguous.” (Sousou, 1997, p. 39)

As the studies to that point lead to conflicting results, further research was necessary. All the above mentioned research also focused only on emotions induced and

4. State of Research

not on emotions perceived.

In a more recent, major study, Ali and Peynircioğlu (2006) found that the addition of negative lyrics enhanced the perception of negative emotions for listeners, while the addition of positive lyrics detracted from the perception of positive emotions. In their experiment, they asked participants to rate four different emotions, each corresponding to one part of the circumplex model of affect (positive: happy & calm, negative: sad & angry) in music stimuli that were combined with different lyrics. These combinations were done congruently and incongruently concerning the emotion quality of music and emotion. They also found that in all combinations the melodies were more important for the ratings than the lyrics. An effect of lyrics that depends on the emotional content of the music is theorized, without providing further evidence. Interestingly, it is not completely clear whether the participants rated perceived or felt emotions; they were asked to rate “their emotional judgments” (Ali & Peynircioğlu, 2006, p. 516) or the emotions “conveyed” by the music (p. 514), so it is not clear if the participants might have mixed these up. Ali and Peynircioğlu said that they wanted to measure induced emotions (p. 512), but the wording *emotions conveyed* suggests that the emotions rated could also be only perceived and not felt. The music stimuli used were again mainly classical or from soundtracks, sometimes jazz, and they did not incorporate a lyric-only condition to control for the effectiveness of the text itself.

In line with the theory by Ali and Peynircioğlu (2006), that the effect of the lyrics for songs depends on the type of music that it is accompanied by, Brattico et al. (2011) conducted an fMRI study that came to the result that “lyrics are more important for inducing sad emotions by music, but that instrumental cues have a greater significance in inducing happy emotions through music.” (p. 12)

Another empirical study by Mori and Iwanaga (2014) used music representing happiness with sad lyrics sung in a foreign language so that the participants could not understand it at first. After that, a textual translation of the lyrics was given to them to read, and then a combination of the song with the translated music readable at the same time was presented. The participants were asked to rate the pleasantness of the emotions that were induced by the music as well as their perception of the emotions in the material. Mori and Iwanaga found evidence suggesting that the listener has pleasant feelings generated by the perception of sadness from sad lyrics when the music has a happy quality. The happy music with translated sad lyrics was rated as equally happy and sad in perception, but induced as much pleasantness in the listener as the untranslated music. In a previous study, Mori (2009) could also show that happy lyrics in happy music did not change the perception of the happiness in the song, but when combined with sad lyrics, the perceived happiness

4. *State of Research*

was lowered and the perceived sadness increased. For happy music, it seems that the lyrics cannot enhance the happy feelings, but it can be lowered by using sad lyrics, again supporting the idea that the effect of lyrics changes depending on the context.

The latest study in the field for music, lyrics, and emotion was an experiment by Fiveash and Luck (2016) on error detection in lyrics while listening to happy and sad music. They presented happy and sad songs to the participants and asked them to detect possible errors in the (either sad or happy) lyrics. Aiming for mood induction in the listener, they only presented congruent combinations of music and lyrics, and played them in blocks of only happy and then only sad songs (or vice versa). The participants were asked to rate their own feelings after listening to each excerpt. Fiveash and Luck found the previous studies to be true about the effect, that lyrics are processed differently depending on the music they are paired with. The study showed that negatively valenced music leads to increased detection of errors in lyrics. They connect this to the idea that negatively valenced music induces a negative mood, and therefore activates a more “detail-oriented processing style” (Fiveash & Luck, 2016, p. 1346).

All in all, the literature to date on the topic of music, lyrics, and emotions comes to the conclusion that the type of music is the most important factor in processing lyrics, especially the valence of the music. There is conflicting evidence on the matter whether music or lyrics are more important for the listener’s emotions (Stratton & Zalanowski, 1994; Sousou, 1997; Ali & Peynircioğlu, 2006), but the majority of the studies see the music as more important. Yet with the exception of the research done by Mori and Iwanaga (2014), all mentioned studies focused on induced emotions. The current study will try to explore the side of perceived emotions in songs with lyrics in more detail and add to the research done so far.

A note on research done on the processing of lyrics and music, that was not related to emotions: Bonnel, Faita, Peretz, and Besson (2001) showed in a study using operatic songs that music and lyrics can be processed independently from one another. In contrast, concerning memory representation of songs, a study by Serafine, Davidson, Crowder, and Repp (1986) suggested that music and lyrics are integrated, attributed to phonology or prosody of the lyrics, as nonsense lyrics had the same effects as the real lyrics. For this study, folk songs were used, which might be a reason for the difference between the results of two experiments. It is also very much possible that lyrics and music can be processed independently when heard, but are stored in an integrated way in the memory.

4.4. Mixed Emotions

Research on mixed emotions is a rather old question, that is still not answered conclusively. For the case of pleasure and displeasure, Schimmack (2001) shows that these emotions are not mutually exclusive, but are also not completely independent from each other; they interact and can seemingly not be represented fully by two separate, unipolar dimensions - but a bipolar dimension would not be sufficient either. He was able to recreate the findings in another study (Schimmack, 2005), that focused on the validity of reports, leading to further evidence of mixed emotions. He proposes to use a different method of measuring feelings like pleasure and displeasure, that do not behave strictly negatively correlated with each other (Schimmack, 2001). This measurement is called the MIN-Statistic and is used in research on attitude as well. It can describe the intensity of mixed feelings by looking at the weaker rating of the two. The higher the rating for the weaker feeling, the higher the index. Before that, the way of comparing these kinds of emotions was most often done by looking at the Pearson correlation, but Schimmack criticizes this. He proposes this different approach of comparing emotions for the following reason (Schimmack, 2001, p. 91):

Pearson correlations test the presence of linear dependencies, that is, whether changes in one variable are reflected in changes in the other variable. MIN values test whether values greater than zero on one variable are accompanied by zero values on the other variable.

So these two approaches measure very different things, and Schimmack was able to show that the Pearson correlation is strongest (negative) when MIN values are highest; by looking at the correlation, one would have missed that there were in fact two different strong emotions induced in the listener.

In a literature review called *The Case for Mixed Emotions*, Larsen and McGraw (2014) provide various examples of evident mixed emotions: For example, in the case of a highly ambivalent movie called *Life is Beautiful* (Benigni, 1997), that features a high degree of conflicting emotions of happiness and sadness, it was shown that moviegoers were reporting both emotions at the same time (Larsen, McGraw, & Cacioppo, 2001). Nearly half of all of those moviegoers reported to feel happy and sad at the same time, which was four times more often than when they asked these people the same questions before watching the movie. They repeated these findings in controlled laboratory settings (Larsen & McGraw, 2011), where they also checked for any alternative interpretations and possible mistakes, like acquiescence of the moviegoers to get away from the survey as soon as possible, or whether par-

4. State of Research

ticipants might have answered in the way what they thought the researches might like to hear; but none of this was found to be true. When they asked participants to press buttons when feeling sad or happy, it was shown that they would press them both at the same time when watching scenes of mixed emotions, again indicating the actual presence of mixed induced emotions (Larsen & Green, 2013).

Music with conflicting cues for happiness and sadness also elicited more mixed emotions than songs with consistent cues in a study by Hunter et al. (2008). They mixed different cues for affect induction from music (e.g., slow temp for sadness with major mode for happiness) and used congruent combinations (e.g. fast tempo, major mode) as control settings. Participants were asked to rate their emotions on independent happiness and sadness scales. Hunter et al. based their findings on the MIN-Statistic that Schimmack (2001) proposed to use for these kinds of studies. These excerpts were all rather short (30s), but they still reliably and predictably evoked cases of mixed emotions. A question left unanswered by this was whether people could actually feel these emotions at the same time or just switched between them, as they were asked after listening to the stimuli. To tackle this, Larsen and Stastny (2011) did a similar study, but included the possibility to report feelings of happiness and/or sadness *while* listening to the music by pressing buttons corresponding to these emotions. They were able to show that mixed emotions occurred significantly more often when music with incongruent cues was played, concluding that “positivity and negativity are separable in experience” (Larsen & Stastny, 2011, p. 1469).

Beyond the emotions of happiness and sadness, it was shown that people who enjoy horror-movies feel pleasure by watching these movies while they can also experience displeasure, such as fear (Andrade & Cohen, 2007).

A meta-analysis by Berrios, Totterdell, and Kellett (2015) shows that mixed emotions are a “robust, measureable, and non-artifactual experience.” (p. 1). Of the 63 studies that were of interest for this analysis, only 6 used music as a way of inducing mixed emotions, so further research in this area is of interest. As this current study will feature both congruent and incongruent song stimuli concerning emotions in music and lyrics, this is a good opportunity to check for mixed emotions.

Another interesting point about aesthetic emotions is that negatively perceived emotions, such as fear, are often not felt as such, and in some cases even induce pleasure (Kallinen & Ravaja, 2006). This is not necessarily a mixed emotion, but it might be the case that especially in these cases the definitions of perceived and felt emotions (see Section 4.1) might get mixed up by the listener and, when reported, may look like a mixed emotion.

4.5. Melancholy & Irony

Melancholy Melancholy is an emotion that has, in the past, often only been used to describe clinical conditions, equated primarily with a mental illness (Brady & Haapala, 2003). Taking a step back from definitions in older history, melancholy can be seen as a lot more. As it is said in the jazz classic 'It might as well be spring' by artists like Frank Sinatra or Nina Simone, "I feel so gay in a melancholy way", (Hammerstein & Rodgers, 1945) melancholy is usually seen as a mixture of sad and happy feelings, a sense of longing, that is not only connected to negative, but also positive emotions. In the circumplex model of affect, it sits more or less in a neutral position of the valence space, with low arousal (see Fig. 4.1), indicating that it is an emotion able to represent both positive and negative emotions. In their article about melancholy in the context of aesthetic emotions, Brady and Haapala (2003, p. 2) say: "But the complexity of melancholy, the fact that it is fascinating in itself, suggests a further thought - that melancholy might be considered as an aesthetic emotion per se." They categorize it as a complex emotion, with both displeasurable (loneliness, sadness) and pleasurable sides (happy memories and fantasies, through reflection) to it. In that sense, melancholy can be seen as the perfect option to choose for this study as an emotion category, because it is an example of mixed emotions that is of particular interest for the arts, and it might lead to new insights into mixed emotions. Brady and Haapala suggest that, in music, melancholy can arise through slow tempi with moments of faster tempi to create a sense of reflection through contrast, and can thus not be classified so easily. It is also a part of the larger GEMS-Scale for music induced emotions (Zentner et al., 2008), as an adjective for the factor of 'Tender Longing', so using melancholy as an emotion term in this study also incorporates an emotion directly related to music and aesthetics.

Irony Irony is in itself not an emotion, but rather a way of creating distance to a subject or topic without requiring distance to the emotions associated with it (Steele, 2010). By contrasting things that seemingly don't go together a disassociation is created that can lead to a distance. In language, it functions as a tool to communicate the opposite of the word-by-word interpretation of the things, and it can also enhance the emotional impact of a message (Filik, Hunter, & Leuthold, 2015). But just as there is little research to be found on the implications of irony in language, there is even less to be found concerning musical and aesthetic emotions. For this reason, irony was introduced into this study in order to have the possibility to rate conflicting cues of emotion in music and lyric, not as a term for emotion quality, but as a tool to find out more about conflicting cues. These ratings may

be helpful in further analyses to find hypotheses about mixed emotions and irony in music with lyrics.

4.6. Personality traits and differences

Personality traits and differences between participants can sometimes lead to differences in rating emotions. It can be the difference of usually listening to music analytically or emotionally (Kreutz et al., 2008), just having different music genre preferences (Rentfrow & Gosling, 2003), or simply the gender of the listener. Concerning the gender, various differences in emotional everyday life have been reported by social psychology (A. Fischer, 2000), and in music a study by Kamenetsky, Hill, and Trehub (1997) for example showed that women rated music excerpts as “more emotionally expressive and more likeable than did men” (p. 149). In the experiments by Coffman et al. (1995), women reported higher ratings on negative emotions and men on positive emotions. But all in all, reported differences in emotion perception, especially in music, are rare and recent literature suggests that these are minor, or not at all present (Ali & Peynircioğlu, 2006; Robazza, Macaluso, & D’urso, 1994; A. H. Fischer, Kret, & Broekens, 2018). Even though Ali and Peynircioğlu (2006) were not able to show any differences for gender in rating emotions in their study on songs with mixed emotions in music and lyrics, it is nonetheless interesting to test whether these findings can be recreated for the stimuli and the setting used in this study. For other personality traits and how they might affect self-report for emotions, the following two approaches are considered:

Empathizer-Systemizer-Theory A question that is under debate in literature is that of the listener type, of how the listener actually listens to music (Gabrielsson, 2001). Adding to this debate is the Empathizer-Systemizer-Theory, a rather new theory, based on findings by Baron-Cohen, Knickmeyer, and Belmonte (2005). It states, that there are two general cognitive styles, namely *Empathizing* and *Systemizing*. They define these styles this way (Baron-Cohen et al., 2005, p. 819):

Empathizing is the capacity to predict and to respond to the behavior of agents (usually people) by inferring their mental states and responding to these with an appropriate emotion. Systemizing is the capacity to predict and to respond to the behavior of nonagentive deterministic systems by analyzing input-operation-output relations and inferring the rules that govern such systems.

4. State of Research

In other words, empathizing is the ability to change one's own perception of emotions of other people or stimuli by changing the own emotions appropriately and in this way understanding and predicting the possible results of this emotion. Systemizing on the other hand is an analytic way of tackling the question of prediction without emotional involvement.

An empathizing (Baron-Cohen & Wheelwright, 2004) and a systemizing quotient (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003) have been developed to test individuals for these different cognitive styles and have been positively tested as useful instruments for measuring these cognitive styles (Wakabayashi et al., 2006). The concept also states that there are individual differences in empathizing and systemizing, for example gender differences, where systemizing tends to appear more often in male participants and empathizing in female participants (Baron-Cohen et al., 2005).

The concepts of this theory have been transferred to the music domain by Kreutz et al. (2008). In their study, the validity of the E-S-theory has been shown for music listening styles by building a new inventory of questions about music empathizing (ME) and music systemizing (MS). Since the development of this questionnaire, it has been shown that there is a positive relationship between enjoyment of evoked negative emotion in response to music with music empathy, using the ME-MS (Garrido & Schubert, 2011). The inventory for questionnaires on ME-MS has very recently been translated to German and validated as effective to be used in further studies (Linnemann et al., 2018).

Without using this ME-MS-Inventory, but rather the standard E-S-Inventory, it has also been shown that musical preferences of genre and psychological attributes of the music are related to the cognitive brain type (Greenberg, Baron-Cohen, Stillwell, Kosinski, & Rentfrow, 2015).

As the ME-MS theory has never before been used in a context of songs or music with lyrics, and emotions, it is of high interest for this study whether these personality traits can have an influence on the ratings for perceived emotions.

STOMP and music preferences It has been empirically shown that preferences for musical styles and genres are linked to personality traits of the listener (Rentfrow & Gosling, 2003). The research done by Rentfrow and Gosling (2003) resulted in a 14-item scale to assess preferences for music genres, the STOMP (Short Test Of Music Preferences). The 14 items, namely Classical, Blues, Country, Dance/Electronica, Folk, Rap/hip-hop, Soul/funk, Religious, Alternative, Jazz, Rock, Pop, Heavy Metal, and Soundtracks/theme songs can then be reduced to four music preference dimensions: Reflective & Complex; Intense & Rebellious; Upbeat & Conventional; Ener-

getic & Rhythmic.² The STOMP has been updated to a revised version, called the STOMPR³, including 23 musical genres and a fifth dimension. For use in German-speaking countries, Langmeyer et al. (2012) did a cross-cultural replication study of the original STOMP. The genres folk and country were there exchanged with “New German Wave” (*Neue Deutsche Welle*) and “Popular German Music” (*Populäre Volksmusik*) and all others were confirmed to work as well. This version was chosen to be used in the present study to look for any personal differences related to genre preferences in rating emotions perceived in music with lyrics.

4.7. About this study

As outlined above, there are various reasons for conducting this research about emotions in songs with music and lyrics. There is still little known about emotions in modern popular music genres, as most studies focused on classical or soundtrack music (Eerola & Vuoskoski, 2013). The current study built on many of the design choices made by Ali and Peynircioğlu (2006) and Mori and Iwanaga (2014), but in comparison more meaningful control stimuli with only spoken word for the lyrics and with a sung, fictional language in the music-only stimulus were used to control the semantics of both language and music. A different emotion, *tenderness*, was used for the negative arousal, positive valence stimulus. Stimuli specifically composed for the purpose of the survey were incorporated to have a good internal, as well as external validity, using non-artificial sounding music in the style of popular songs, where the different variables (singer, instruments, tempo and mode, etc.) can be controlled. This way it is also possible to compare the whole song against the lyric, and not just a melody. To control for possible influences of the listener’s personal preferences and listening styles on the ratings, the ME-MS score (Kreutz et al., 2008; Linnemann et al., 2018) and the STOMP (Rentfrow & Gosling, 2003; Langmeyer et al., 2012) were added, and gender differences were checked. Another main difference to other studies was the inclusion of two different emotion scales, using both a discrete (basic emotions) as well as a dimensional (valence-arousal) model to compare possible rating differences. The focus was on perceived emotions, as previous studies have focused on induced emotions.

²The way how to group the 14 items to these four categories and the norms can be found here:
<http://www.midss.org/content/short-test-music-preferences-stomp>

³<https://sites.google.com/view/pjasonrentfrow/measures>

Hypotheses Mainly based on recent research by Ali and Peynircioğlu (2006), Mori and Iwanaga (2014), and Fiveash and Luck (2016), the main hypotheses and concepts of interest in this study are as follows:

- The valence of the music changes the effects of the lyrics (Fiveash & Luck, 2016) in the following ways:
 - In emotionally congruent combinations of music and lyrics, the lyrics strengthen negative emotions (sadness, anger, negative valence) and weaken positive emotions (happiness, tenderness, positive valence) (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014)
 - Emotionally incongruent combinations of lyrics and music lead to higher ratings for the emotion conveyed by the music than for the emotion conveyed by the lyric; the music is more important for emotion in songs (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014)
- There are mixed emotions perceived in stimuli with incongruent combinations of music and lyrics (Hunter et al., 2008; Larsen & McGraw, 2014)
- There are differences in rating emotions in songs, depending on the listening style of the participants (ME-MS) (Kreutz et al., 2008; Linnemann et al., 2018)
- People who favour reflective and complex music change their ratings more depending on the lyrics than people who favour upbeat and conventional music (STOMP) (Rentfrow & Gosling, 2003; Langmeyer et al., 2012)
- There are no gender differences in rating perceived emotions in songs (Ali & Peynircioğlu, 2006)
- Mixed emotions can happen when listening to songs with music and lyrics, in the following ways:
 - Emotionally incongruent combinations of lyric and music leads to the perception of irony when happy and angry stimuli are combined
 - Emotionally incongruent combinations of text and music leads to an emotion of melancholy when sad and happy stimuli are combined
 - Emotionally incongruent combinations of music and lyrics lead to perception of strong mixed emotions (Larsen & Stastny, 2011; Larsen & McGraw, 2014)

Apart from the above hypotheses, it was of interest to the study how effective the survey stimuli were in communicating emotions to validate that they were

4. State of Research

effective in communicating the intended emotions. Thus, they were checked and compared for intensity in the intended as well as the other rated emotions. Moreover, the different models used for measuring emotions (discrete and dimensional) were put into contrast to see which model is best fit to capture (mixed) emotions in songs with music and lyrics.

5. Method - experimental design

The aim of this study is to investigate the influence of music and lyrics on the perceived emotions in modern songs and whether listening styles or the gender of individuals had an impact on the perception of these emotions. For this purpose, the participants listened to 24 stimuli composed and written solely for this study. The stimuli presented were as follows: German lyrics alone in the form of spoken text; songs with a fictional language as lyrics; and music and German lyrics combined to a song. The emotions in question for this study were chosen by their position in the circumplex model of affect by Russell (1980) - one for each main sector in the model. Happy for positive valence and arousal; sad for negative valence and arousal; tender for positive valence and negative arousal; and angry for negative valence and positive arousal. Text and music were combined congruently and incongruently concerning the respective emotional content, to look for reinforcing, attenuating or other effects on the perception of emotions by adding lyrics or music to the respective stimulus.

An online study was conducted to achieve this. The questionnaire was designed with the online tools provided by limesurvey¹ and then distributed via social media, e-mails, as well as forums and survey platforms. In this way a quantitative, experimental study was conducted with randomized participants in a within-subjects-design. It was a cross-sectional study with only one survey at one point in time. An online survey was chosen in favour of a laboratory test to allow for possibly more and a higher variety in the participants, aiming for higher external validity by having a listening setting that is closer to a real life situation.

After being presented with one stimulus, the participants were asked to rate the emotional content of it on different scales: a) On scales for valence and arousal, each ranging from -10 to +10. b) unipolar Likert scales for the terms 'happiness', 'sadness', 'anger', 'tenderness' and, to account for possible mixed emotion effects, 'melancholy' and the describing term 'irony' ranging from 1 = 'not at all' to 5 = 'very much'. All stimuli were presented in a randomized order to every participant. They

¹Limesurvey: www.limesurvey.com

were also asked to find another emotion term that they perceived in the music, if possible. As only 0.61% of all possible cases reported any additional emotion terms, these will be omitted for now.

5.1. Participants

Overall, 252 people participated on this survey, with 142 participants completing it. Using the median of the time needed to complete the survey ($M_{Interviewtime} = 1735.84$ secs), every person who was faster than roughly half of the median (900 secs) or slower than more than double of the median (4000 secs) were excluded as well (2 and 4 participants, respectively). By visually analyzing the answers of every participant, one person was found that rated every stimulus the same on every scale, always choosing the lowest possible option, and was excluded as well. Of these remaining participants, 2 rated their own german language skills as equal or lower than 4 on a scale from 1 = “Very low / I don’t speak German” to 6 = “Very High / Like a native speaker”, and were thus excluded from the following analyses, to make sure that everyone was able to fully understand the instructions as well as the lyrics of the songs. By filtering with a check question on whether the participants did understand the instructions, another 11 were excluded, as they rated the test song ‘Happy’ by Pharell Williams as negative on the valence and / or the arousal scale. This test song was, as the title of it suggests, chosen as a rather clear example of a stimulus using happy-sounding music and lyrics with positive valence and arousal.

This left 122 participants’ (76 female, 46 male, $M_{age} = 31.43$ years, age range: 19 - 85) complete data sets for further analyses. 69.67% of these people were redirected to the survey via the survey platform *surveycircle.de*. When asked to rate their own proficiency in music, 49 participants identified as ‘Hobby-Musicians’ (40.1%), 67 as ‘Non-Musicians’ (54.9%) and 6 as ‘Professional’ (4.9%).

The participants were asked to listen to the stimuli via headphones. 42.6% of them did so, while 43,4% used the speakers of their laptops for playback. The final 14% used either hifi speakers or another playback device.

5.2. Material, Stimuli

The survey included 24 different stimuli that the participants were asked to rate for the perceived intensity of emotions in the songs. For each emotion representing one sector of the circumplex model of affect and the various congruent and incongruent combinations at least one stimulus was needed. These were put together under aspects of ecological validity. To achieve this, four different song sections were composed which were all similar in instrumentation, but expressed different emotional content. The piano was always used as the main instrument because it is considered as emotionally neutral (see Chau, Wu, and Horner (2014)) and was combined with a simple rhythm and bass accompaniment to make it feel more like a modern production and an actual song. The compositions are based on study results by P. N. Juslin and Laukka (2004) about how emotions can be evoked by different musical attributes (e.g. fast tempo and major key for happy, slow tempo and minor key for sad music).² They were composed and recorded to fit to a modern song setting, rather than a classical composition, as most studies on the subject to date have focused primarily on classical music and the effects of music and lyrics on popular music haven't been studied as well so far (Eerola & Vuoskoski, 2013). The length of the songs was limited to between 23 and 42 secs with a fade-in and -out in order to keep the duration of the experiment under control, keep the stimuli comparable and also get enough completed polls (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005). Original songs were composed and chosen to minimise any effects of familiarity in the listener. By recording the same instrument and the same vocalist for songs composed by the same composer, variability across the stimuli was minimised as well. A few possible songs were composed for each emotion, but only the songs that were considered to be the most effective in conveying the respective emotions were chosen by experts who are musicology graduates. These stimuli included a short music-only intro with a fade-in of 5 to 7 seconds; after it, the vocalist entered the song. The melodies of the song always consisted of 10 to 12 notes per phrase and about 60 notes per song, so that it was possible to match the emotionally congruent and incongruent lyrics to every song without altering the melody.

For the control condition of music-only, gibberish lyrics were created to convey emotions without any semantics from the text. This fictional language was created with an online generator for fantasy languages to produce recognizable phrases and an apparent grammar (Scriboly, 2019). A placebo question in the survey was asked

²Basic emotions of happiness, sadness, anger, fear and tenderness are coverable with it

5. Method - experimental design

about whether the listener was able to recognize the language of the song, even though it was impossible to do so. This way, it was possible to create convincing stimuli including only music as the emotion transporting element, while at the same time keeping the vocalist and not replacing him with another instrument to keep the melody, and thus ensuring that the difference between songs with and without lyrics was minimised to the actual text and its semantic content. As there were no differences as a function of gender of the vocalist in the experiment conducted by Ali and Peynircioğlu (2006), this study used only one male vocalist for every stimulus.

The lyrics were selected by aspects of the same positive and negative valence as the songs and were mostly self-written or in small parts borrowed from unknown songs. Inspiration was taken from google searches for poems about the respective emotion, as well as social media pages like Twitter and Tumblr to see how German people expressed these emotions, as it was done by Fiveash and Luck (2016) in their study on error detection in lyrics. Every phrase was matched in length to the melody of the lyrics, thus containing 10 to 12 syllables per phrase and an average of 60 syllables in total, again ensuring comparability.

For the control condition lyrics-only, these texts were then individually spoken and recorded, whereby the speaker was to not reflect the emotional content of the text in the expression of the voice without making it feel too disconnected, to make sure that the emotion conveyed by these stimuli came mainly from the text itself and not the voice and expression of the narrator. Literature shows that the voice alone can transport emotions, even without meaningful semantics from the language (P. N. Juslin & Laukka, 2003, see Tab. 4.1).

When sung, they were then combined with emotionally congruent and incongruent music. For the incongruent combinations of music and lyrics, the vocalist had to sing the lyrics in the same manner as he did for the respective congruent song, to ensure that the only differences between the congruent and the incongruent matchings of songs and lyrics was again only the semantic content of the text and not the expression by the vocalist or changes in melody. For the combination of music and gibberish language without content, the text was also sung as close to the stimuli with real lyrics as possible by the singer, expression-wise.

In total, 24 stimuli were created in the ways described above. One for each sector from the circumplex model with a congruent lyric (4) and one for each possible combination of song and lyrics in an incongruent way (12). Furthermore, the 4 spoken lyrics and the 4 songs with gibberish lyrics as control conditions were recorded

as well. Everything was recorded in an acoustically optimised room with professional equipment: A TLM102 microphone by Neumann into an Apollo Twin interface by Universal Audio connected via Thunderbolt to a Macbook Pro running Logic Pro X for vocal recordings and high-quality sampled drums and pianos from Logic Pro X, recorded via MIDI. The vocals were slightly tuned with the pitch-correction software Melodyne after recording, to ensure that the intonation of the vocals were the same for every note of the matching song combinations. Minimal processing was added to the recordings, including equalizers and slight compression and reverb, to enhance the feeling of a conclusive, modern recording and of the spoken text. The songs and texts were then mastered to the same loudness (LU), so that there was no important difference between the stimuli and the participants did not have to adjust the volume throughout the survey.

5.3. Questionnaire and design

As an introduction to the questionnaire, the participants were informed about the goals of this study, to examine the influences of lyrics and music on the perception of emotions in songs and that they were about to listen to some audio examples after they answered a few personal questions. They were allowed to take a break if needed, as the survey was rather long, by using the functions provided by Limesurvey to save progress. Participation was voluntary and confidential, and it was possible to withdraw from the study at any point. They were informed that there was no right or wrong answer and that they could answer the questions spontaneously. It was possible to win 2x20 EUR amazon.com-vouchers and a recording for a song in the studio of the creator of the survey.

Participants were asked to wear headphones while doing the survey, that they were in a quiet surrounding and that they could attend the survey without being interrupted. They were then asked to play a sound clip of a person asking to put a slider that was visible on the website to the value of +10 to ensure their audio equipment was working, that the volume was appropriate, and that they were paying attention to the audio, before entering the experiment.

Participants were then specifically asked to perceive the emotional content of the music and what the composer and writer might have intended, not their own emotions during the pieces. They were also asked to rate the overall impact of the song and not separate parts of it. To capture these emotions, the circumplex model of affect by John Russel (1980) was used and queried on two sliders, one for valence

5. Method - experimental design

and one for the arousal of the song (Russell, 1980), in a range from -10 (low arousal / negative valence) to +10 (high arousal / positive valence). In addition, four categorical emotion terms (happy / *fröhlich*, sad / *traurig*, angry / *wütend*, tender / *zärtlich*) and two different terms, which could arise from the incongruent combination of stimuli, namely melancholy / *melancholisch* (Brady & Haapala, 2003) and irony / *ironisch*, were inquired on unipolar Likert scales from 1 ('not at all' / *Trifft überhaupt nicht zu*) to 5 ('very much' / *Trifft vollkommen zu*) asking about how the audio material was perceived concerning emotions. For the gibberish language example, another question was added to make the allusion that the language was real by asking the participants if they knew the language. The questions were answerable while the stimulus was playing, skipping to the next one before at least ten seconds of it were played was not possible.

A questionnaire about gender, mother tongue, education, and about the proficiency in music was recorded at the beginning of the study.

Another important aspect of the study was the question about the listening style of the participants. Based on the theory of Empathizers and Systemizers by Baron-Cohen et al. (2005) and the adaption for music (Kreutz et al., 2008), it was of interest whether the participants were more likely to listen to music systematically or emotionally. This was done via a questionnaire by Linnemann et al. (2018) for Music Systemizers and Music Empathizers (9 per group, 18 items in total) in German to later be able to determine if there were any moderating effects (music systematizer or music empathizer) and if they had an impact on the evaluation of the perceived emotions in the stimuli. To find out whether there are any other moderating effects depending on the participant's preferences for music styles, a STOMP-Questionnaire consisting of 20 different musical genres and a question for each about how much they liked the respective genre, was added as well (Rentfrow & Gosling, 2003; Langmeyer et al., 2012).

5.4. Pretest

Before presenting the online questionnaire to the public, it was tested by 5 people (3 males, 2 females, age between 27 and 59 years) to check for any (technical) problems that might arise while filling out the survey. For 2 of them, it was done by talking to the experimenter while filling out the survey one after the other about questions they might have. The other three completed the survey online on different devices and sent a detailed feedback to the experimenter afterwards. The results of this pretest lead to slight changes in the 'Tender' stimulus, as it was rated too high on

5. Method - experimental design

the arousal scale and thus changes in the melody and the lyrics had to be made. The melody was changed to have less high intervals and the text was changed to use less arousing words.

Furthermore, the explanations and instructions on what valence and arousal are and how the participants should rate them were changed to be more detailed, to lessen any confusions about the terms. The basic emotion terms seemed to be easily understandable, as they were used in everyday life, but the terms of valence and arousal appeared to be less common to the pretest participants.

Apart from these changes, the survey seemed to work very well, even though it took every participant a much longer time to finish it than for the people who participated in the final survey, as they all tried to be very thorough and detailed in their feedback.

After the examination of these results and incorporating these changes, the final survey went online for 5 weeks.

5.4.1. Data analysis

All analyses were performed using *R* and *RStudio* (R Core Team, 2018). Most of the main analyses were calculated using linear mixed effects models with random effects to account for repeated measures. These models were created with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), and p-values were obtained with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). Comparison of the model's performances were done using R's *anova()*-function. The models were fitted by maximum likelihood. All reported R^2 values were calculated using the *r.squaredGLMM()*-function from the *MuMIn* package (Bartoń, 2018) and always reports the marginal R^2 of the linear mixed model, which is in this case the variance explained by the fixed effects only.

No data points in the ratings for emotions were missing, as every participant had to answer all questions before being able to proceed to the next question; this is true for all analyses.

6. Results

To compare how the audio files performed on the different scales, all analyses for the basic emotions are summarized first, followed by the analyses for valence and arousal, and the results will then be put in contrast.

6.1. Basic Emotions

6.1.1. Music vs. Lyrics - A linear mixed model approach with control variables

In order to answer the hypotheses on the influence of music and lyrics on the perception of emotions, and to also be able to discuss possible differences between the study participants in terms of listening styles (ME-MS), genre preferences (STOMP) and gender, separate linear mixed effects models with random intercepts (LMER) for each basic emotion were created. The ratings on the four basic emotions in question for all combined stimuli for these models can be found in the Appendix, B.1.

The basic models tried to predict the ratings for the emotion of the combined stimuli via the two fixed effects of the ratings for the control stimuli of music-only and lyrics-only. This was done using all possible combined stimuli with all emotions, congruent and incongruent. As an example, in this way the model tried to predict not only the ratings of happiness in the stimuli containing either happy music or happy lyrics, but also the ratings of happiness in stimuli not containing happiness by intention, e.g. in sad music with sad lyrics compared to happiness ratings of the sad control stimuli and thus looking for occurrences of mixed emotions as well and how they can be predicted by the participant's ratings for the control stimuli.

As a random effect, the participant's ID was incorporated into the model to account for the possibility that participants are all affected differently by the experimental manipulations (Barr, Levy, Scheepers, & Tily, 2013). To further define these more complex models, random slopes for the control stimuli ratings were also created and

6. Results

then compared to the basic models, to find the maximal fit.

In the next step, the participant's ME-MS scores and the genre preference scores of interest, Upbeat & Conventional or Reflective & Complex, were added separately as fixed effects to new models and then compared to the more basic models as well, to see whether there are interaction effects with the participant's ratings.

Assumptions of linear mixed models were checked visually by using Q-Q plots on the distribution of the residuals of the models.

To complete the search for the model with maximal fit, the final model is compared with a simple linear regression model, to see whether incorporating the random effects structure actually makes a significant difference.

After following these steps, the best fitting models for all emotions incorporate uncorrelated intercept and slopes without interactions. Their implementation in R being:

$$\text{Rating}_{\text{Emotion, Combined}} \sim \text{Rating}_{\text{Emotion, Lyrics}} + \text{Rating}_{\text{Emotion, Music}} + (1|\text{ID}) + (0 + \text{Rating}_{\text{Emotion, Lyrics}} + \text{Rating}_{\text{Emotion, Music}} | \text{ID})$$

The important values for estimating the models can be found in Tables 6.1 and 6.2. By looking at the results in the mentioned tables, and judging by the estimate for both music and lyric effects, it can be said that they are more or less equal for each emotion. The estimates for music and lyrics differ only by small margins of 0.09 max., meaning that the participant's ratings for both the music and the lyric-only conditions are of more or less equal importance in estimating the combined stimulus ratings. In these small differences, the music is a little stronger than the lyric for the happiness, sadness, and tenderness ratings, and it only changes for the anger ratings, where the lyrics estimate is a little higher than the music estimate. These findings differ from previous studies by Ali and Peynircioğlu (2006) or Mori and Iwanaga (2014), where they found that lyrics weakened positive emotions and helped communicate negative emotions. In the case of this study, it seems that both are of equal importance. To investigate this more closely, more detailed analyses are done in the following sections.

Looking closer at the models, it can be seen that the variance and standard deviation of all random effects are rather low compared to the residual, so it can be said that there are only small differences in the individual participants and their ratings of basic emotions. Judging by the R^2 , the models seem to do quite well in explaining the variation in the ratings.

As can be seen in Fig. 6.1, the Q-Q-Plots of the residuals show that models are generally not as good in predicting the highest and lowest values; this is probably

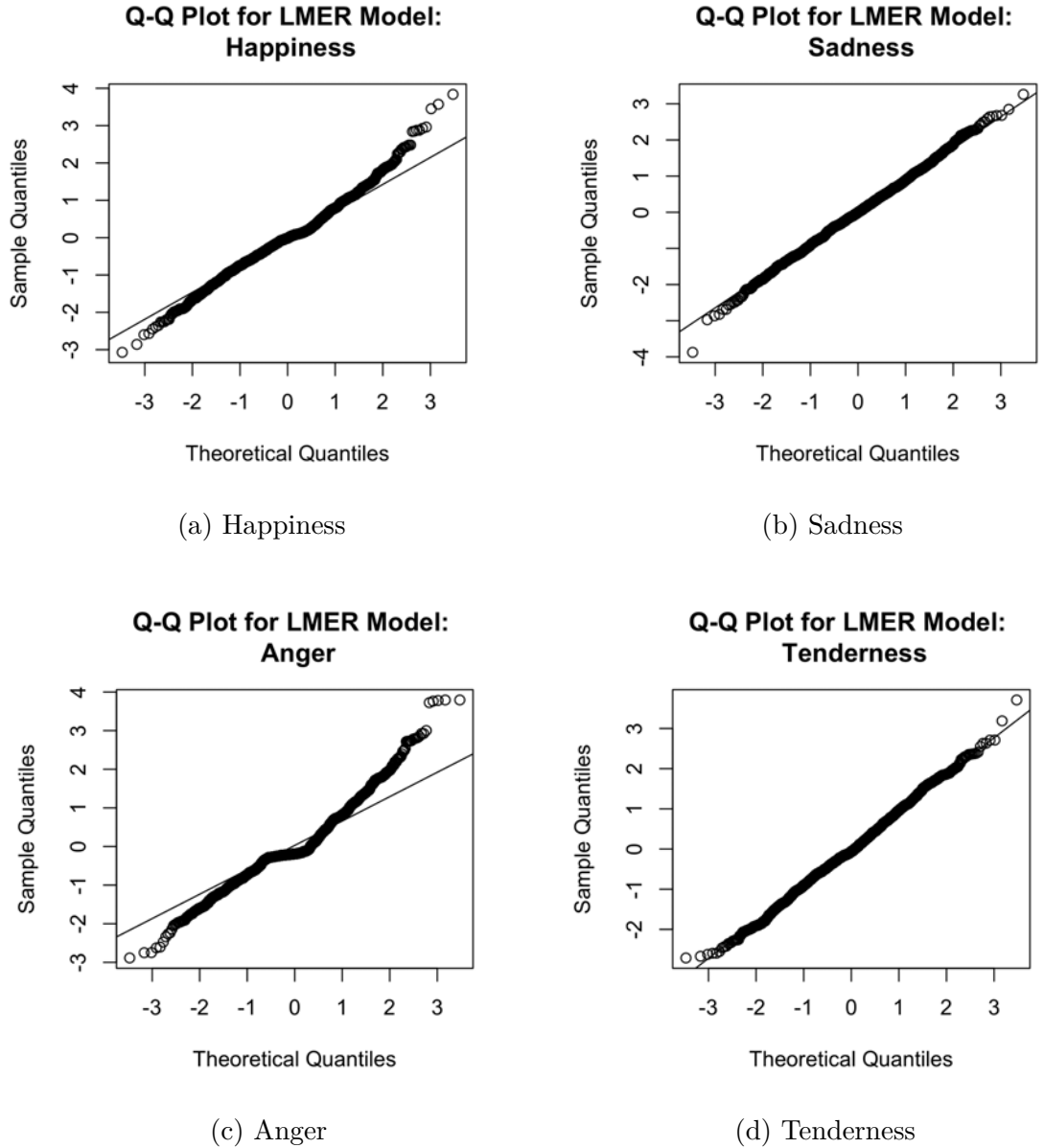


Figure 6.1.: Q-Q-Plots for the main LMER models

due to ceiling effects caused by the rather small range of 1-5 in the Likert scales and the high intensity of perceived emotion in many stimuli. Especially the model for anger should be judged with caution.

ME-MS and STOMP

In the following step, the ME-MS and STOMP scores of interest were added. Looking at the data for the ME-MS scores first, it can be said that the scores are normally distributed among the participants, with most participants being close to a 0, making them balanced listeners (see 6.2a and 6.2b) Only few participants rated themselves

6. Results

| Emotion | Parameter | Variance | SD |
|-----------------------------|-----------|----------|--------|
| Happiness (ICC = 0.026) | ID | 0.0205 | 0.143 |
| | Music | 0.0361 | 0.190 |
| | Lyric | 0.0388 | 0.197 |
| | Residual | 0.7426 | 0.862 |
| Sadness (ICC = 0.056) | ID | 0.0547 | 0.234 |
| | Music | 0.0296 | 0.172 |
| | Lyric | 0.0321 | 0.179 |
| | Residual | 0.8981 | 0.948 |
| Anger (ICC < 0.001) | ID | <0.001 | <0.001 |
| | Music | 0.0456 | 0.214 |
| | Lyric | 0.0359 | 0.189 |
| | Residual | 0.795 | 0.892 |
| Tenderness (ICC = 0.017) | ID | 0.0164 | 0.128 |
| | Music | 0.0348 | 0.186 |
| | Lyric | 0.0291 | 0.170 |
| | Residual | 0.9374 | 0.968 |

Table 6.1.: Variance and SD for random effects of all four main LMER models

| Emotion | Parameter | Estimate | 2.5% | 97.5% | df | p |
|--------------------------------|-----------|----------|--------|-------|------|--------|
| Happiness ($R^2 = 0.54$) | Intercept | -0.079 | -0.186 | 0.027 | 436 | 0.145 |
| | Music | 0.545 | 0.501 | 0.590 | 175 | <0.001 |
| | Lyric | 0.485 | 0.439 | 0.531 | 139 | <0.001 |
| Sadness ($R^2 = 0.48$) | Intercept | 0.291 | 0.148 | 0.433 | 1084 | <0.001 |
| | Music | 0.526 | 0.482 | 0.571 | 191 | <0.001 |
| | Lyric | 0.430 | 0.386 | 0.473 | 172 | <0.001 |
| Anger ($R^2 = 0.55$) | Intercept | 0.112 | 0.016 | 0.207 | 1840 | 0.022 |
| | Music | 0.501 | 0.448 | 0.554 | 157 | <0.001 |
| | Lyric | 0.587 | 0.542 | 0.632 | 139 | <0.001 |
| Tenderness ($R^2 = 0.42$) | Intercept | 0.279 | 0.141 | 0.417 | 379 | <0.001 |
| | Music | 0.481 | 0.431 | 0.531 | 183 | <0.001 |
| | Lyric | 0.406 | 0.361 | 0.452 | 158 | <0.001 |

Table 6.2.: Values for fixed effects of all four main LMER models

as strong music empathizers (score > 1) or music systemizers (score < 1). Just like it was in the study by Linnemann et al. (2018) for the German ME-MS, female participants were a little more on the music empathizer side (Median for males = 0, females = 0.33).

To investigate whether the scores have any impact on the ratings for the emotions, they were incorporated into the linear mixed models as fixed effects, with possible interactions with both the lyric-only or the music-only rating. However, for no model were the AIC ratings ever lower than the previous best fitting models without these interactions, and a further check on the ANOVA-Tables of the models showed that

6. Results

there were no main or interaction effects for the ME-MS score on any of these models (all $ps > 0.05$ ¹). On the model for the happiness ratings, the ME-MS score only had an estimate of 0.009, $SE = 0.083$ as a main effect, and -0.013 and 0.008 as interaction with Music and Lyric condition ($SE = 0.035, 0.037$, respectively). The ME-MS score thus does not seem to make a real difference in the participant's ratings of perceived emotions in songs. One explanation for this might be that this is due to the task of evaluating the perceived emotions in the stimuli and not the own felt emotions while listening. This is also reflected in the low variance in the participant's ID as a random effect in the models, so it seems there really are no substantial differences in-between the different raters.

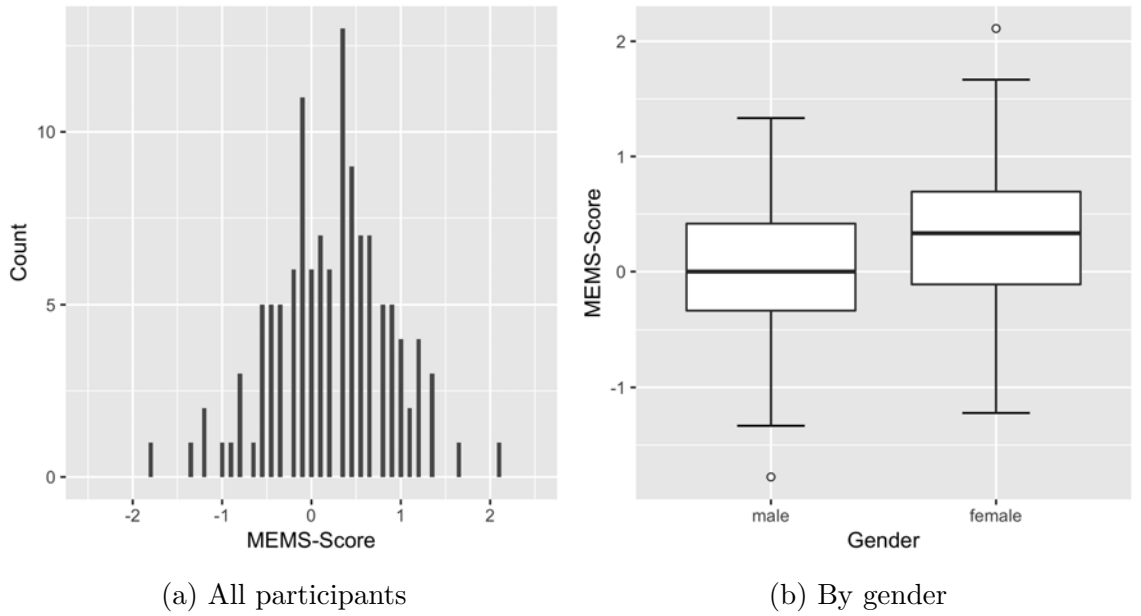


Figure 6.2.: ME-MS Score distribution

Looking at the STOMP scores for the genres of interest for the hypotheses (Upbeat & Conventional and Reflective & Complex), this impression is reinforced. For each model, the two scores of interest were added as fixed effects with interactions with the ratings for the lyrics and music control condition. This again lead to no better model fit compared to the original models (for happiness: $AIC_{\text{Original}} = 5224$, $AIC_{\text{STOMP}} = 5233$, $p = 0.86$).

To be certain that there really are no effects concerning the STOMP scores in these models, a median split on the data was performed, splitting it by liking or disliking Upbeat & Conventional or Reflective & Complex music. By splitting at the median, it might be possible to see differences due to these personality traits more clearly.

¹As an example, for the model of happiness the F-Statistics for the two interactions with Music and Lyrics were $F(1,171) = 0.16$, $MSE = 0$, $p = 0.69$; and $F(1, 147) = 0.05$, $MSE = 0$, $p = 0.82$, respectively. All other models had similar values.

6. Results

Comparing the final models of the emotions using the full data set or the two split data sets by looking at the values of the estimates of the linear models, it can again be said that there are no major differences between these groups. The differences between the groups and the full data set is marginal and might just be due to some normal variability between the participants.

A distribution of the STOMP scores of the participants can be seen in Fig. 6.3. The estimates for the models with the median split can be found in the Appendix in Table C.1.

Summing up, there seem to be no significant differences in between the participants when rating the perceived emotions in the stimuli, neither by cognitive listening style (ME-MS) nor by music genre preferences (STOMP). As a last check, models incorporating possible interactions with the participant's gender as a fixed effect were done for all four emotions and compared with the original models. Again, neither were there any models with a better fit comparing AIC values, nor were there any significant main or interaction effects for the gender condition (all $ps > 0.05$). So it seems that, when asking people to rate perceived emotions of music and lyrics on unipolar scales, there are no significant differences between the participants. Another factor might be the rather low variability in the Likert scales of 1-5 used in this survey, reducing any possible differences in ratings between the raters.

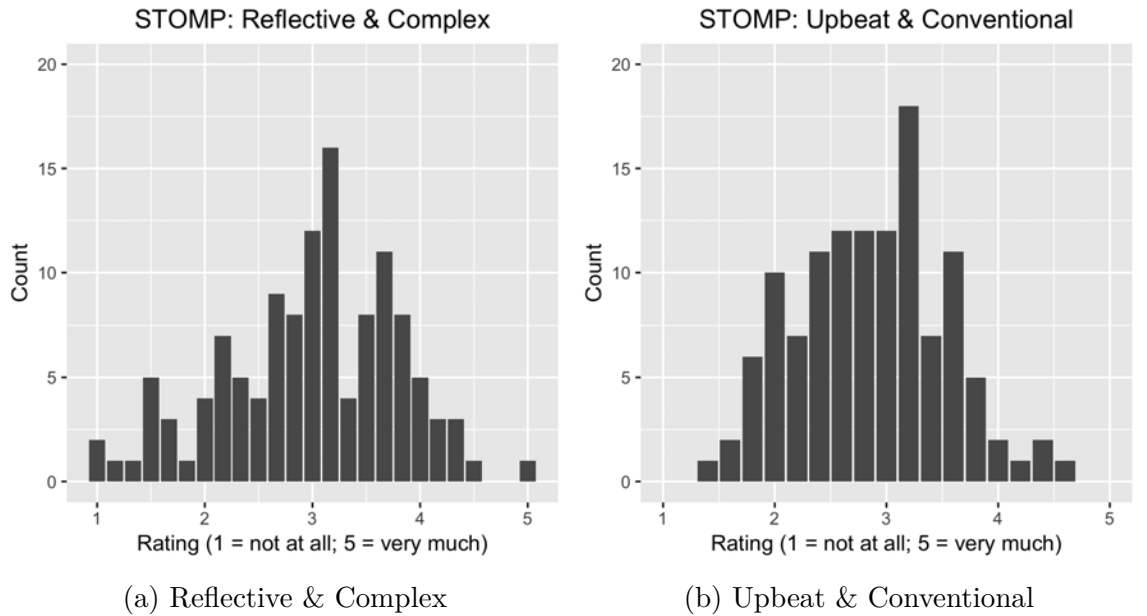


Figure 6.3.: STOMP Score distributions

6.1.2. Congruent Combinations - Ratings of intended emotion

To look for similarities and differences between the results of the study by Ali and Peynircioğlu (2006), who used mainly classical and jazz music and additional lyrics recorded on top of them as their stimuli, the next interesting analysis was that of the difference in intensity of emotional responses by intended emotion of the stimuli, with congruent combinations only, by the stimulus condition (lyrics only, music-only, combination) and any possible gender effects.

As can be seen in table 6.3, the overall mean intensity for the congruent combinations is always higher than both the music or the lyric on its own, while differences between male and female ratings seem to be negligible. For the first analysis of these ratings, a linear mixed model with associated analysis of variance was conducted. The model was built with the Ratings as the dependent variable and interacting fixed effects of the Stimulus and the Emotion conditions. To look for gender differences, a separate model including interactions of the gender condition with the fixed effects was created and compared to the model without it using R's built-in `anova()` function. This showed that the gender did not help to increase model fit ($AIC_{\text{without gender}} = 3832.6$ vs $AIC_{\text{with gender}} = 3840.4$, $p = 0.646$) and detailed analysis of the larger model did not show any main or interaction effects for the gender condition (all $p > 0.3$) and was thus excluded from the following analysis. No random slopes were added, as models with it could not converge, only the participant's ID as a random intercept was incorporated.

| | Lyrics | | | Music | | | Combined | | |
|---------|--------|-------|---------|-------|-------|---------|----------|-------|---------|
| Emotion | Women | Men | Overall | Women | Men | Overall | Women | Men | Overall |
| Happy | 3.697 | 3.870 | 3.762 | 4.250 | 4.391 | 4.303 | 4.526 | 4.674 | 4.582 |
| Sad | 4.579 | 4.457 | 4.533 | 4.434 | 4.304 | 4.385 | 4.632 | 4.761 | 4.680 |
| Angry | 4.092 | 3.913 | 4.025 | 3.382 | 3.435 | 3.402 | 4.539 | 4.544 | 4.541 |
| Tender | 3.961 | 4.087 | 4.008 | 3.513 | 3.565 | 3.533 | 4.263 | 4.326 | 4.287 |
| Overall | 4.082 | 4.082 | 4.082 | 3.895 | 3.924 | 3.906 | 4.490 | 4.576 | 4.522 |

Table 6.3.: Mean ratings for intended emotions

The model without gender condition showed a main effect for intended emotion $F(3, 1342) = 38.350$, $MSE = 26.615$, $R^2 = 0.156$, $p < 0.001$. Post-hoc comparisons with emmeans and Tukey showed that, over all stimuli in the model, ratings for happiness were higher than those for anger and tenderness; sadness ratings were higher than those for happiness, anger and tenderness, all $ps < 0.001$. Anger and tenderness ratings did not differ significantly, $p = 0.876$. When combined to positive (happy and tender) and negative (sad and angry) emotions, it could be seen that negative emotions were rated as slightly higher in intensity than the positive emotions, $F(1, 1342) = 15.406$, $MSE = 12.083$, $R^2 = 0.078$, $p < 0.001$.

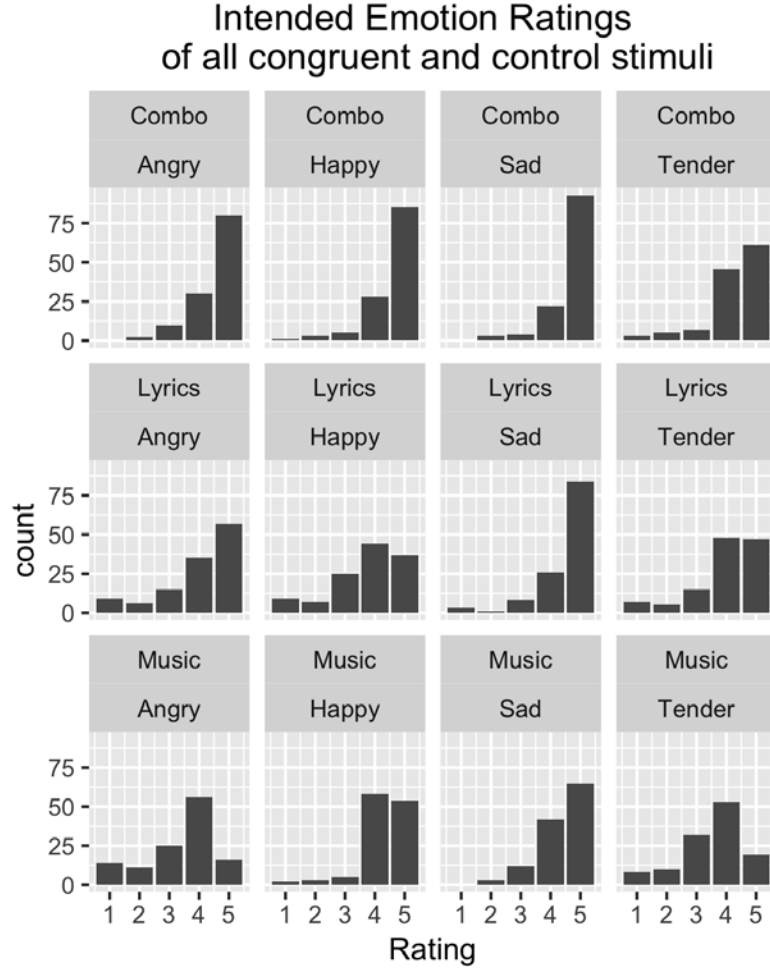


Figure 6.4.: Barplot for intended emotion ratings on all congruent and control stimuli

These results differ from the findings by Ali and Peynircioğlu (2006) and Mori and Iwanaga (2014), who found that positive emotions were conveyed more effectively than negative emotions by music and lyrics. They suspected this difference might also be due to the effectiveness of the representing stimuli of the selected emotion, and not be an effect by the music and lyrics, which might be the case here as well, as they are not in line with the findings of the previous studies. Another explanation might be however that for actual songs and more popular music styles, the perception and importance of music and lyrics changes for the listeners. It could be the case that positive emotions can be communicated slightly easier by music and lyrics than negative emotions, and that the basic emotions of happiness and sadness are easier to convey than anger and tenderness.

Another main effect was found for the stimulus condition, meaning there was an effect when comparing lyrics only, music-only and their combined stimuli, $F(2, 1342) = 70.976$, $MSE = 49.265$, $p < 0.001$. The rating for the music-only as well

6. Results

as the lyrics only were both lower than those for the congruent combinations ($p < 0.001$ for both); the lyric ratings were lower in the mean by -0.441, while the music ratings were lower by -0.617. This shows that the congruent combination of the two stimuli was adding to both separate stimuli in intensity of emotions, making the combination more effective, and adding music to a lyric, or vice versa, enhances the perceived emotions. Comparing the lyrics only and the music-only, the mean ratings were significantly higher for the lyrics than for the music, even though only by a small margin ($diff = -0.176$, $p = 0.003$), so it can be said that the music composed for the study were a little weaker in conveying the intended emotions than the lyrics written for it.

There was a significant interaction between the stimulus and the emotion condition when comparing all four emotions separately, $F(6, 1342) = 15.388$, $MSE = 10.679$, $p < 0.001$, as well as when comparing only positive and negative emotions, $F(2, 1342) = 6.799$, $MSE = 5.333$, $p = 0.001$.

For both positive and negative emotions, adding music to the lyrics making it a combined stimulus, the intensity of intended emotion gets stronger, by a difference of 0.549 and 0.332, respectively (both $p < 0.001$). Adding lyrics to music makes it stronger in intensity of intended emotion by 0.516 for positive emotions and 0.717 for negative emotions, both $p < 0.001$. Looking at the lyrics only and the respective music-only condition comparing the effectiveness of the stimuli, participants rated the music as less intense in intended emotion than the lyrics for negative emotions ($p < 0.001$), while the positively valenced music and lyrics were not significantly different in intensity from each other ($p = 0.999$).

Comparing the negative and positive stimuli in their effectiveness in communicating intended emotions, no significant differences between the intensity of the combined stimuli as well as between the music-only stimuli could be found, while the negative lyrics were rated as significantly higher in intensity by the participants ($p < 0.001$, $diff = 0.393$). Overall, looking at positive and negative emotions only, the survey stimuli seem to have been created to be comparable in intensity, as judged by the participants.

Looking at the four emotions separately, the participants rated the lyrics for the happy and angry emotion as less intense than the respective combined stimulus (all $p < 0.001$), but the sad and tender lyrics were not significantly weaker before adding music ($p = 0.968$, and 0.279 , respectively). The ratings were significantly higher for the combined stimulus for the angry and tender emotion condition when compared to the respective music-only condition ($p < 0.001$ for both, respectively), while the happy and sad music did not differ significantly in intensity after adding lyrics to it ($p = 0.279$ and 0.200 , respectively). These ratings suggest that the sad emotions

6. Results

cannot be easily enhanced in intended emotion intensity by adding either a fitting lyric or music, the happy music cannot be enhanced by adding lyrics to it, and the tender lyrics not by adding music, while all other emotions were rated as stronger in the respective combined stimulus. This could be due to the effectiveness of the three stimuli trying to communicate sadness, but it might also be that for sadness, lyrics and music can be equally good in communicating the emotion, and that it does not make a lot of difference when adding music or lyrics to either one. For happiness, it can be speculated from these results that it does not matter much for a happy song whether it has understandable happy lyrics or not. It can communicate happiness just as good on its own as it does with fitting lyrics. If it matters at all what kind of lyrics are added to the happy music will be checked further when comparing with non-congruent combinations. The tender stimuli are a bit ambiguous, as their mean ratings are never as high as those of the comparable stimuli of the other emotions (s. Table 6.3) and in this case the lyrics seem to be of higher importance for emotion perception than the other stimuli.

Checking the comparability of the stimuli in detail, ratings for happy lyrics did not differ significantly in intensity from those for angry and tender lyrics ($p = 0.374$ and 0.479 , respectively), while the sad lyrics were rated as more intense than the happy, tender and the angry lyrics ($p < 0.001$, $\text{diff} = 0.770$; $p < 0.001$, $\text{diff} = 0.525$; $p < 0.001$, $\text{diff} = 0.508$); angry and tender lyrics did not differ significantly in intensity, $p = 1$. For the music condition, the happy music was rated as more intense in intended emotion than the angry and tender music, both $p < 0.001$, but did not differ significantly from the ratings for the sad music, $p = 0.999$. The sad music was rated as more intense than the angry and tender music, both $p < 0.001$. Angry and tender music were not significantly different from each other in rated emotion intensity, $p = 0.987$. These comparisons suggest that the sad and happy music composed for the survey were better at communicating the intended emotion than the angry and tender music. This might just be the case for these stimuli, but it could also be possible that it is not as easy for music to convey more complex emotions as it can with happiness and sadness. For the lyrics, it seems that the sad lyrics were the best at communicating the intended emotion, while the others performed equally well and all above average, looking at Table 6.3.

As mentioned above, no main effect was found for the gender of the participants, recreating the findings by Ali and Peynircioğlu (2006), that there were no gender differences in judgement of the conveyed emotions, $F(1, 122) = 0.158$, $MSE = 0.110$, $p = 0.691$.

And just as there was no main effect for the gender condition, there were also no

6. Results

significant interactions with the emotion ($F(3, 1342) = 1.1436$, $MSE = 0.791$, $p = 0.330$) or the stimulus condition ($F(2, 1342) = 0.321$, $MSE = 0.221$, $p = 0.725$) with the gender condition, so it seems that there were no differences for the ratings of intended emotion between genders.

For the ME-MS score, no significant effect was found as well $F(1, 122) = 3.26$, $MSE = 2.3$, $p = 0.074$, even though in this case the score was close to the significance level of 0.05. Both STOMP scores of interest (Reflective & Complex, Upbeat & Conventional) showed no significance either. (Reflective: $F(1, 122) = 1.70$, $MSE = 1.2$, $p = 0.20$; Upbeat: $F(1, 122) = 1.38$, $MSE = 1.0$, $p = 0.24$) It seems that personal traits of the participants do not matter for rating emotions from songs with lyrics .

6.1.3. Congruent Combinations - Ratings for other emotions

In this survey, the participants were asked to always rate the perceived emotions on all available scales and not just the ones with emotions that were intended to be in the stimuli. In that way, they could answer on multiple unipolar scales and thus it was possible to look for how well the stimuli performed concerning the congruent emotions, but also whether there might be cases of mixed emotions.

To check whether the intended and the perceived emotions aligned for the congruent stimuli and the control stimuli of music and lyric only, they were analyzed on every possible rating (happiness, sadness, anger, tenderness, melancholy, irony) and compared to the intended emotion of the respective stimulus. To achieve this, separate linear mixed effects models with the participant's ID as a random intercept were created to show how the rating changed for every emotion, with the respective intended emotion as the reference level. As fixed effects, the different emotions were used and maximum likelihood was the estimation method. Post-hoc comparisons of the ratings were again performed with the least-square means and the p-values were adjusted using Tukey. In this case, no random slopes could be calculated, as the number of observations would be as high as the number of random effects for these terms. The participant as a random effect accounted only for low variance in all of the models (below 0.1, Std. Dev. < 0.4), suggesting that there are no large differences in the ratings of the perceived emotions between all the participants. Assumptions of linear mixed models were analyzed by visual inspection of residual plots.

The distributions of the ratings for all the above-mentioned stimuli can be found in detail in the appendix, chapter B.

6. Results

The happiness ratings for the congruent combination of the happy music and lyric, the happy music only and the happy lyrics only were all significantly greater than all the other possible ratings for the respective stimulus (all $ps < 0.001$), as expected. The same goes for all participants' ratings of anger for the congruent angry combination, the angry music only and the angry lyrics only (all $ps < 0.001$). For the sad lyrics and the sadness rating this is equally true ($ps < 0.05$). Interestingly, the ratings for sadness for the sad music only were not significantly higher than the ratings for melancholy ($Estimate = -0.04$, $p = 0.998$), while it was true for all other comparisons ($ps < 0.001$). As they both have very high mean ratings ($M_{Sadness} = 4.385$ vs. $M_{Melancholy} = 4.426$) with the mean for melancholy even being slightly higher, it seems that a differentiation between the two is difficult to do when listening only to music. In the combination of sad music and sad lyrics, the same can be said, as the estimate in contrasts is just 0.213 ($SE = 0.106$), $p = 0.337$. They again both have very high mean ratings ($M_{Sadness} = 4.68$ vs. $M_{Melancholy} = 4.467$), but this time the sadness ratings are a little higher.

The participants' ratings for tenderness for the respective tender stimuli were also not always higher than the other emotions; for the tender music, ratings for tenderness were not significantly higher than for sadness ($Estimate = -0.131$, $p = 0.863$) as well as for melancholy ($Estimate = -0.328$, $p = 0.05$). In fact, the mean ratings for both melancholy ($M = 3.861$) and sadness ($M = 3.664$) were both higher than those for tenderness ($M = 3.533$). Anger, happiness, and irony were not rated as higher in intensity than tenderness. This might be a problem of the tender music stimulus, making it hard for the participants to differentiate between the categories of sadness, tenderness and melancholy. They were all rated equally high in intensity and none of them were higher than 4 in the mean. Tenderness in music might be difficult to communicate, even though many different cues for making a song sound tender (slow tempo, major mode, small variability of sound level, consonance, lowered singer's formant, etc.) (P. N. Juslin & Laukka, 2004) have been incorporated into the song. It might be the case that it is either a) always hard to differentiate between the mentioned emotions; b) they in fact are all perceived in the music; c) the cues for tenderness in music by P. N. Juslin and Laukka (2004) are meant for classical music only and cannot simply be reproduced by popular music styles.

These conclusions get further support when looking at the ratings for the tender lyrics-only condition. As it was the case for the congruent combination of the two, the ratings for tenderness were all significantly higher than the others (all $ps < 0.001$), making just the tender music stimulus a rather ambiguous one in terms of emotions, and it only gets fully convincing as a tender song when combined congruently with tender lyrics.

| | df1 | df2 | F | MSE | p | Marginal R^2 |
|-----------------|-----|-----|--------|--------|--------|----------------|
| Happy Combined | 5 | 610 | 301.96 | 203.01 | <0.001 | 0.656 |
| Happy Music | 5 | 610 | 248.7 | 163.21 | <0.001 | 0.615 |
| Happy Lyrics | 5 | 610 | 103.41 | 97.471 | <0.001 | 0.398 |
| Sad Combined | 5 | 610 | 452.01 | 306.77 | <0.001 | 0.753 |
| Sad Music | 5 | 610 | 456.48 | 270.45 | <0.001 | 0.747 |
| Sad Lyrics | 5 | 610 | 297.81 | 252.69 | <0.001 | 0.648 |
| Angry Combined | 5 | 610 | 201.58 | 173.66 | <0.001 | 0.569 |
| Angry Music | 5 | 610 | 44.251 | 47.023 | <0.001 | 0.226 |
| Angry Lyrics | 5 | 610 | 202.22 | 175.78 | <0.001 | 0.543 |
| Tender Combined | 5 | 610 | 206.79 | 179.19 | <0.001 | 0.580 |
| Tender Music | 5 | 732 | 208.52 | 165.87 | <0.001 | 0.588 |
| Tender Lyrics | 5 | 610 | 147.29 | 138.03 | <0.001 | 0.481 |

Table 6.4.: F-Statistics for all ANOVA-Tables of the linear mixed effects models for Intended vs. Other Emotions

So apart from the tender music stimulus and the sad music and combination, all stimuli performed as expected. Those that did not suffered from a lack of good distinguishability between the low arousing emotions and the high amount of melancholy in the sad music.

6.1.4. Comparing congruent and incongruent combinations

Another analysis of interest was that how congruent and incongruent combinations of lyrics and music differ in detail, to possibly find out about the effectiveness of music and lyrics in communicating emotions. To achieve this, the factors of Music Emotion and Lyric Emotion of the stimuli were combined to a factor of congruency or incongruency in emotion, meaning if the music and lyric of the combined stimulus were of the same emotion type, the stimulus was a congruent one; if music and lyric emotion did not align, it was an incongruent combination. Choosing either the music or the lyric as the intended emotion to compare between them, the factor thus had the possibility of being congruent to music (incongruent to lyrics), congruent to lyrics (incongruent to music) or congruent to both, just as it was done in the previous study by Ali and Peynircioğlu (2006).

The linear mixed model for this again had the participant's ID as the random effect, and the congruency condition and the type of intended emotion condition as fixed effects, calculated with maximum likelihood. Another model with the gender as a fixed effect was compared against the simpler model and again neither a better fit was found ($AIC_{\text{without gender}} = 10759$ vs. $AIC_{\text{with gender}} = 10766$, $p = 0.62$) nor

6. Results

were there any main ($F(1, 150) = 0.62$, $MSE = 0.8$, $p = 0.43$) or interaction effects with the gender condition ($F_{\text{Congruency:Gender}}(2, 3294) = 0.29$, $MSE = 0.4$, $p = 0.75$; $F_{\text{Intended Emotion:Gender}}(3, 3294) = 1.11$, $MSE = 1.5$, $p = 0.34$), showing no signs of any gender differences in rating the stimuli.

The main model shows a significant effect for the Congruency condition, $F(2, 3294) = 242.8$, $MSE = 317$, $p < 0.001$. Post-hoc comparisons with emmeans and Tukey-adjustment show a significant difference between the congruent stimuli and the incongruent songs, $p < 0.001$. The ratings for the congruent stimuli on the intended emotion were higher than the stimuli with music not fitting to the intended emotion, by 1.197.

The congruent stimuli were also rated as higher in intensity than the stimuli with incongruent lyrics, $p < 0.001$, by 1.259. Just as expected, the congruent combinations are seen as higher in intensity of intended emotion than the incongruent combinations. No significant difference between the stimuli with incongruent music and the ones with incongruent lyrics were found, $p = 0.305$; difference of 0.062, which is different to the results from Ali and Peynircioğlu (2006), where the melodies played a more dominant role in eliciting the intended emotion. It is as it was in previous analyses of this study, that lyrics and music seem to contribute equally as much to the perception of emotion in music. The difference to previous studies can again be due to different stimuli, different genres of music and the question of perceived vs. induced emotion.

There was also a main effect for the intended emotion $F(3, 3294) = 42.9$, $MSE = 56$, $p < 0.001$. The sad ratings were higher than the happy, angry and tender ratings (all $p < 0.001$), but the other emotion categories did not differ significantly in intensity of perceived emotions ($ps > 0.8$). For this comparison, lyrics with mismatched music and music with mismatched lyrics, as well as the all matching stimuli were put into this one category of intended emotion, but apart from the sadness ratings being higher, no significant differences were found, so it can be said that the participants did not really perceive positive or negative emotions differently; the difference in sadness ratings might again be because of the effectiveness of the sad stimuli.

The interaction of type of intended emotion and the congruency condition showed an effect as well, $F(6, 3294) = 21.6$, $MSE = 28$, $p < 0.001$. The congruent combination of both music and lyric was always rated higher than the stimuli with incongruent lyrics or those with incongruent music, concerning the intended emotion. Comparing congruent lyrics and congruent music, the lyrics were rated as being weaker than the music for happiness ($p < 0.001$), but as stronger for the anger and tender stimuli ($ps < 0.001$). No significant difference was found between the congruent sad lyrics and congruent sad music. That the congruently combined stimuli were always stronger

than those where either music or lyric was mismatched comes as no surprise, showing that both music and lyric help in conveying the desired emotion. It is interesting that for the more specialized emotions of anger and tender, the ratings for the stimuli with fitting lyrics were seen as more intense than those with fitting music and mismatched lyrics, meaning that in these cases the lyrics can better transport an emotion than music, showing that music might not be as good in communicating emotions like tenderness and anger compared to sadness and happiness.

6.1.5. Interaction with other emotions

In the next step, it was of importance to see how exactly the different emotions interact with each other when music and lyrics are matched incongruently. This was done by looking at the participant's ratings for both the emotion of the music as main intended emotion as well as the lyric as the main intended emotion and comparing which combinations are more intense in the intended emotion of the music or the lyric. Again, a linear mixed model approach was deemed fit for this task. The model should depict how intended emotion of the music interacted with the emotions of the mismatched lyrics and vice versa. As a random effect again only the ID was used into the model. Everything was calculated using the maximum likelihood method.

Music

Looking at the ratings for the intended emotion of the music and how they changed depending on the mismatched lyrics, the model showed an interaction effect of the music emotion and the lyric emotion, $F(9, 1830) = 115.93$, $MSE = 112.6$, $p < 0.001$, $R^2 = 0.36$. Comparing the emotion ratings in detail using emmeans with Tukey adjustment, it showed the following:

The happy music matched with

- sad lyrics was not rated as happier than with angry lyrics ($p = 0.984$);
- sad lyrics was rated as happier than with tender lyrics ($p < 0.001$);
- angry lyrics was rated as happier than with tender lyrics ($p < 0.001$);

Interestingly, the combinations with negatively valenced lyrics (happy, angry) are seen as happier than the one with tender lyrics. It's also interesting that the congruent combination of happy music and lyrics is not rated as significantly happier than

6. Results

the incongruent combination with tender lyrics ($p = 0.9516$), so it does not seem to matter much of what kind the positive lyrics are, as long as they are positive.

The sad music matched with

- happy lyrics was rated as less sad than with angry lyrics ($p < 0.001$);
- happy lyrics was not rated as sadder than with tender lyrics ($p = 0.857$);
- angry lyrics was rated as sadder than with tender lyrics ($p < 0.001$);

For the sad music, the valence of the lyrics is of main importance; the combination with negative lyrics (angry) are rated as sadder than happy lyrics. Again, the congruent combination of sad music and sad lyric is not seen as significantly sadder than the incongruent match with the other negative emotion, anger, $p = 0.951$.

The angry music matched with

- happy lyrics was rated as less angry than with sad lyrics ($p < 0.001$);
- happy lyrics was not rated as angrier than with tender lyrics ($p = 1$);
- sad lyrics was rated as angrier than with tender lyrics ($p < 0.001$);

Just as it is for the sadness ratings, the negative lyrics (sad) communicate anger better than happy or tender lyrics. Here all congruent combinations are stronger in intended emotion than the incongruent combinations ($ps < 0.001$).

The tender music matched with

- happy lyrics was rated as less tender than with sad lyrics ($p = 0.01$);
- happy lyrics was not rated as more tender than with tender lyrics ($p = 0.952$);
- sad lyrics was rated as more tender than with angry lyrics ($p < 0.001$);

For tenderness, sadness seems to fit best for the intended emotion ratings, so that here the valence is not as important, but rather the arousal.

These results seem to draw a very differentiated picture. The positively valenced music seems to be affected by mismatched lyrics in a less intuitive way than the negatively valenced ones. For example it is a bit surprising that the happy song is conceived as happier with a sad lyric than with a tender lyric. The tender music also seems to better interact with the sad lyrics than with the same valenced one (happiness). For tenderness, it seems that the arousal of the mismatched emotion is more important than the valence, which will be checked further down below in the section for valence and arousal. Generally speaking, it seems that all lyric types

have different effects on different music types and thus again seem to be of great matter for evaluating the emotion of a song with lyrics.

Lyrics

Taking a look at the ratings for the intended emotion of the lyrics and how they changed depending on the mismatched music, the model again showed an interaction effect of the lyric emotion and the music emotion, $F(9, 1815) = 114.47$, $MSE = 96.5$, $p < 0.001$, $R^2 = 0.33$. The results were:

The happy lyrics matched with

- sad music was rated as less happy than with the angry music ($p < 0.001$);
- sad music was rated as less happy than with the tender music ($p < 0.001$);
- angry music was rated as happier than with the tender music, but not significantly ($p < 0.052$);

Sad music with happy lyrics seems to be a lot less happy than angry and tender songs with happy lyrics. Angry and tender music influence the happiness ratings of the lyrics equally as much. The congruent combinations are all seen as happier than the incongruent ones ($ps < 0.001$)

The sad lyrics matched with the

- happy music was rated as less sad than with the angry music ($p = 0.007$);
- happy music was rated as less sad than with the tender music ($p < 0.001$);
- angry music was rated as less sad than with the tender music ($p < 0.001$);

Sad lyrics and tender music are the strongest in sad emotions when matched together. Interestingly, for sadness the congruent combination is not seen as sadder than the incongruent match with the tender music ($p = 0.8840$), possibly due to a similar arousal of the two songs.

The angry lyrics matched with

- happy music was not rated as angrier than with the sad music ($p = 1$);
- happy music was not rated as angrier than with the tender music ($p = 0.319$);
- sad music was rated as a bit angrier than with the tender music, but not significantly ($p < 0.0650$);

6. Results

For anger, it does not matter much which type of emotion is contained in mismatched lyrics, the music seems to take over here. For anger, all congruent combinations are stronger in intended emotion than the incongruent ones.

The tender lyrics matched with

- happy music was rated as less tender than with the sad music ($p = 0.016$);
- happy music was rated as more tender than with the angry music ($p < 0.001$);
- sad music was rated as more tender than with the angry music ($p < 0.001$);

Just like for the sad lyrics, the other way around seems to work as well. Tender lyrics are most affected by the sad music, getting more tender than with happy or angry lyrics. Again, the congruent combination of tenderness is not seen as significantly different from the incongruent match of sad music and tender lyrics, in terms of tenderness ($p = 0.675$).

Summing up, it can again be said, that music and lyrics seem to both be of a high matter when evaluating emotions. Depending on the type of intended emotion of song and lyric, the different incongruent combinations have different effects, and no one conclusion can be drawn from this, they need a differentiated look.

Comparing both emotions of incongruent combinations

One of the last comparisons of interest were those between the ratings for both emotions contained in the incongruently combined stimuli, eg. if sad music and a happy lyric are combined, how do the ratings for sadness and happiness on this stimulus compare? This can give another insight on what the listener might attend more to in these incongruent stimuli, the music or the lyric. To do this, various t-tests were performed for each stimulus. The results of the comparisons can be found in tables 6.5 and 6.6.

These comparisons again show that both music and lyric play an important role in affecting the perception of emotions in songs. The ratings for the songs with happy music take over the feel of the mismatched lyric emotion, when the lyrics are negative, and stay positive when combined with tender lyrics. It is interesting to note that, when listening to happy music with tender lyrics, the participants do not rate the happiness lower than the tenderness, while for sad lyrics, the sadness is stronger. So they in fact do not seem to have problems with differentiating between sadness and tenderness in lyrics. As can be seen in sad music: Here, the song appears to be more tender than sad when combined with tender lyrics. Tenderness has a

6. Results

| Music Emotion | Rating (Mean) | Lyric Emotion | Rating (Mean) | t | p |
|---------------|---------------|---------------|---------------|-----------------|-----------|
| Happiness | 2.9 | Sadness | 3.3 | $t(200) = -3$ | 0.002 |
| | 2.7 | Anger | 3.4 | $t(200) = -5$ | < 0.001 |
| | 4.4 | Tenderness | 3.6 | $t(200) = 6$ | < 0.001 |
| Sadness | 3.7 | Happiness | 2.2 | $t(200) = 10$ | < 0.001 |
| | 4.5 | Anger | 3.5 | $t(200) = 7$ | < 0.001 |
| | 3.5 | Tenderness | 4.0 | $t(200) = -4$ | < 0.001 |
| Anger | 2.5 | Happiness | 3.2 | $t(200) = -4$ | < 0.001 |
| | 3.7 | Sadness | 3.8 | $t(200) = -0.9$ | 0.4 |
| | 2.5 | Tenderness | 2.5 | $t(200) = 0.1$ | 0.9 |
| Tenderness | 2.9 | Happiness | 2.8 | $t(200) = 0.7$ | 0.5 |
| | 3.4 | Sadness | 4.5 | $t(200) = -9$ | < 0.001 |
| | 2.7 | Anger | 3.1 | $t(200) = -3$ | 0.002 |

Table 6.5.: Music with incongruent lyrics:
Comparison of ratings for both emotions in the stimuli

| Lyric Emotion | Rating (Mean) | Music Emotion | Rating (Mean) | t | p |
|---------------|---------------|---------------|---------------|-----------------|-----------|
| Happiness | 2.2 | Sadness | 3.7 | $t(200) = -10$ | < 0.001 |
| | 3.2 | Anger | 2.5 | $t(200) = 4$ | < 0.001 |
| | 2.8 | Tenderness | 2.9 | $t(200) = -0.7$ | 0.5 |
| Sadness | 3.3 | Happiness | 2.9 | $t(200) = 3$ | 0.002 |
| | 3.8 | Anger | 3.7 | $t(200) = 0.9$ | 0.4 |
| | 4.5 | Tenderness | 3.4 | $t(200) = 9$ | < 0.001 |
| Anger | 3.4 | Happiness | 2.7 | $t(200) = 5$ | < 0.001 |
| | 3.5 | Sadness | 4.5 | $t(200) = -7$ | < 0.001 |
| | 3.1 | Tenderness | 2.7 | $t(200) = 3$ | 0.002 |
| Tenderness | 3.6 | Happiness | 4.4 | $t(200) = -6$ | < 0.001 |
| | 4.0 | Sadness | 3.5 | $t(200) = 4$ | < 0.001 |
| | 2.5 | Anger | 2.5 | $t(200) = -0.1$ | 0.9 |

Table 6.6.: Lyrics with incongruent music:
Comparison of ratings for both emotions in the stimuli

special role compared to the other three emotions, seemingly taking over positive and negative valence, depending on the underlying second stimulus. Angry and tender music does not seem to be perceived as especially angry or tender anymore when matched with different lyrics; mixed emotions seem to be in action here.

When taking the lyrics as the main component and adding incongruent music to it, sadness seems to be the strongest emotion again, as the music and its emotion are never rated as more intense. For all other emotions, it's again a mix and it gives a hint that conveying mixed emotions seems to be possible.

6.1.6. Mixed Emotions

Another goal of this study was to provide new findings on mixed emotions. This was done by allowing the participants to rate the songs, especially the incongruent stimuli, with the emotion of melancholy and the expression of irony. The emotion of melancholy is an expression of both positively and negatively valenced feelings and could, in the hypothesis, be produced by incongruent stimuli. Although sarcasm can not be considered an emotion, it may still be a good indication of when the study participants perceived something as emotionally mixed or inappropriate.

As a further remedy, the MIN-Statistic, as used by Hunter et al. (2008) in a mixed emotion study with music, was applied to the scores to determine in which stimuli mixed emotions are most likely to encounter. The MIN-Statistic states that two (or more) different scores are compared with each other, and the lowest of these scores, the minimum, is selected as representative to see what the value of the minimum is. So when the averages of two emotions and their ratings are compared and both have a high mean, so too is the MIN-Statistic high which is an indication of mixed emotions.

Melancholy and Irony

To study melancholy and irony, linear mixed models were created just like in section 6.1.1, using the ratings of the control stimuli of music-only and lyrics-only, this time with the ratings for melancholy or irony, and with this predicting the melancholy/irony ratings in all stimuli. Again, the ID of the participants were used as a random effect, with uncorrelated random slopes for the control variables.

Before going into detail, a quick look at the distributions of the ratings for melancholy and irony in Fig. 6.5 and 6.6 provides interesting insights: While the ratings vary quite a lot for melancholy, depending on the underlying music emotion, and can be quite high, they are overall pretty low for all irony ratings. It seems that melancholy is transported especially by the music emotion, as ratings for all sad and tender songs appear to have high melancholy ratings as well. For irony, it seems that if the participants felt there was irony expressed in the songs, then it was usually with an incongruent combination of positive and negative music or lyric, like happy music and sad lyric or vice versa. To confirm these assumptions, the first two LMER models are created in the way described above.

While for all four basic emotions the same models showed that music and lyrics seem to be equally important, the model for melancholy shows something new, and

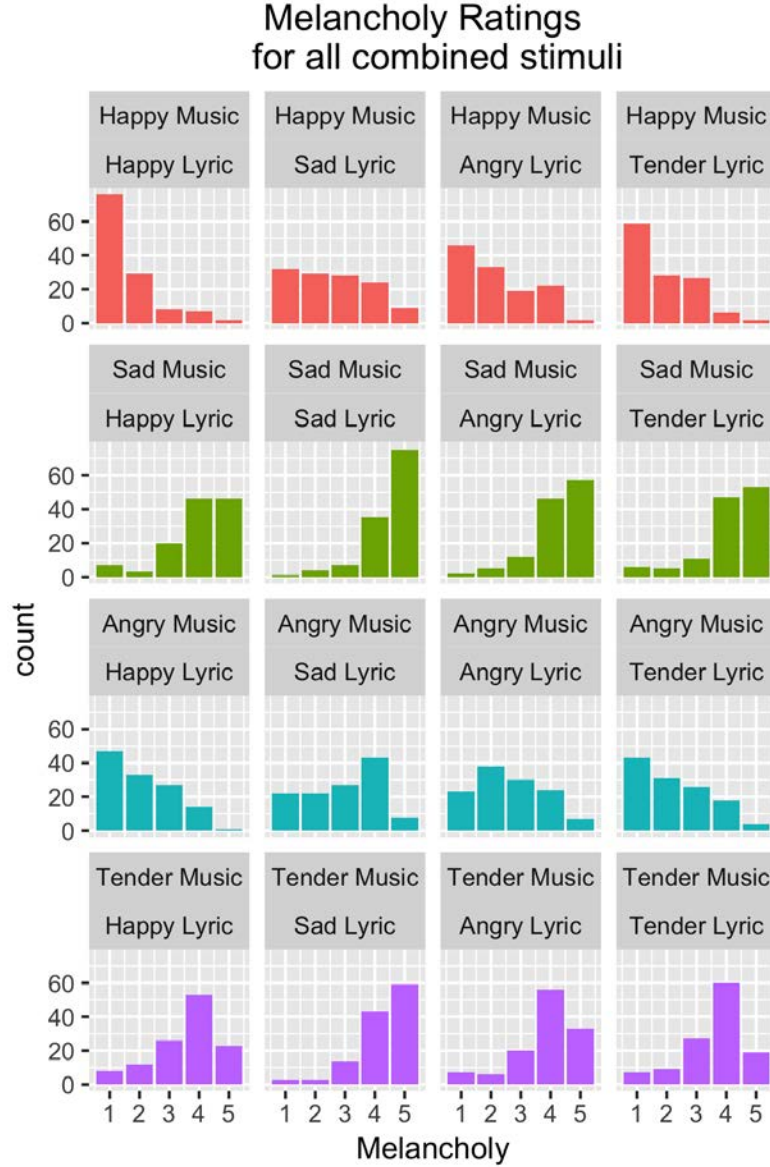


Figure 6.5.: Barplot for melancholy ratings on all combined stimuli

what was assumed after looking at the distributions. The intercept has an estimate of 0.45, the estimate for the music-only condition is 0.64, while the estimate for the lyrics only is just 0.24, all $p < 0.001$. In this case, the music seems to have a stronger influence on whether or not the participants can recognize melancholy in the stimuli. The random effects and slopes again have a low variance below 0.07, with an ICC of 0.08 for the ID, $R^2 = 0.5$.

The ratings for irony are more in line with the original four models, in the way that the estimate for the intercept is 1.16, for the music-only condition 0.24 and for the lyrics only 0.27, $ps < 0.001$, so they are more or less equal and also not as good in predicting the ratings for the other stimuli, $R^2 = 0.082$ (random effects variance below 0.1, residual 1.155, ICC = 0.07), and are overall very low. After that, a few

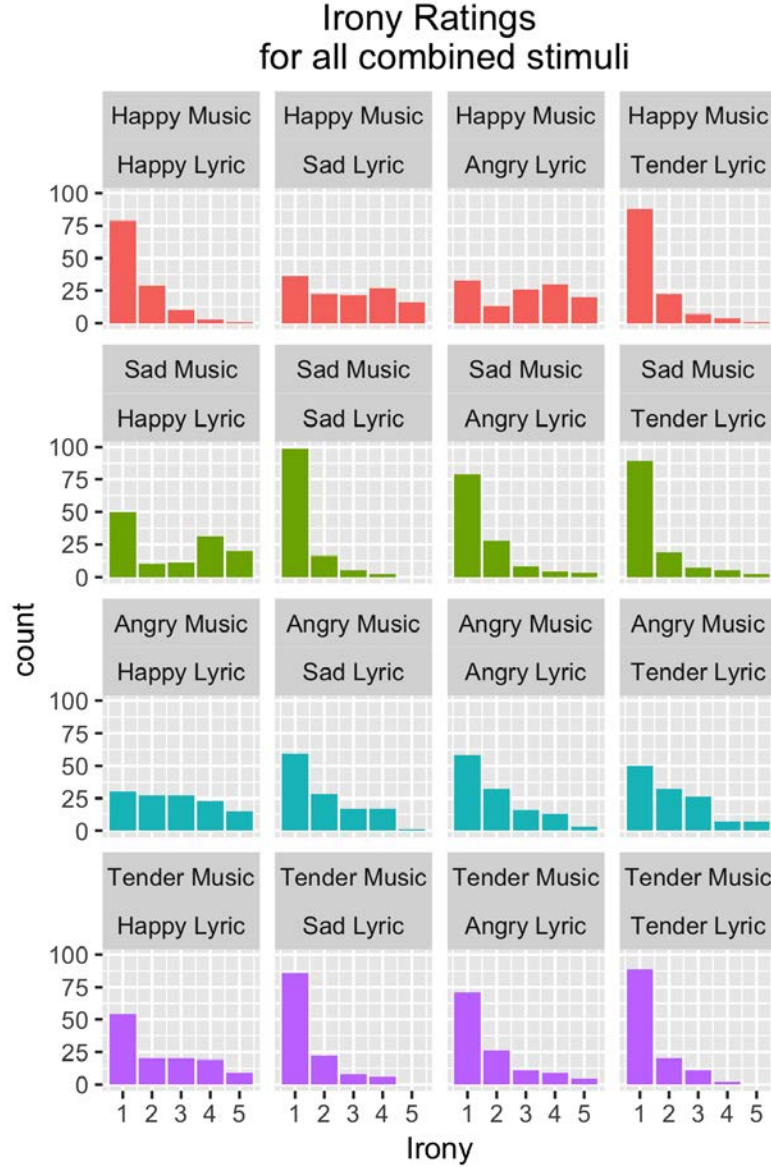


Figure 6.6.: Barplot for irony ratings on all combined stimuli

more detailed analyses were done, for both irony as well as for melancholy.

To find out how the ratings changed depending on how music was matched with a lyric and vice versa, another two linear mixed models were created, that use the factors of the music emotion and the lyric emotion of the combined stimuli as predictors for the melancholy ratings, to see if there are significant interactions between them. Surprisingly, the ANOVA for the melancholy model showed only significant main effects for the music emotion ($F(3, 1815) = 572.92$, $MSE = 518$, $p < 0.001$) and the lyric emotion ($F(3, 1815) = 56.90$, $MSE = 51$, $p < 0.001$), but not for their interaction ($F(9, 1815) = 1.51$, $MSE = 1$, $p < 0.14$), $R^2 = 0.43$. It seems, that the participants actually did not perceive any important interaction between

6. Results

music and lyric and perceived the stimuli always as melancholic if the specific music or lyric emotion on itself was included that creates melancholic feelings. Post-hoc analysis of the main effects showed that the two main emotions to evoke melancholy in songs were sadness and tenderness, as they were always the strongest in the ratings (all respective $ps < 0.001$), with sadness being even stronger than tenderness ($p < 0.001$). This is probably due to the similarities in the music of low arousal and mainly being different in the way that one is in major and one is in minor mode. For lyrics, post-hoc comparisons showed that it is again the sadness that mainly evokes melancholy and is stronger than the other three emotions (all $ps < 0.001$). The other strong emotion for melancholy in the lyrics surprisingly is anger, being rated as more intense in melancholy than happiness and tenderness, but weaker than sadness ($ps < 0.001$). Happiness and tenderness in lyrics are not significantly different in intensity ratings for melancholy, $p = 0.06$). All in all, melancholy does not seem to be the best way of measuring mixed emotions in songs, as it seems to mainly be attributed to the music's feeling and there especially to sad and tender music. A mixed model that predicted melancholy ratings by using the ratings for the four basic emotions in all stimuli was created to check this in detail, with uncorrelated random slopes with the ID for these ratings. The estimates of the model were: Intercept = 1.8604, Happiness = -0.2281, Sadness = 0.4696, Anger = -0.1396, Tenderness = 0.2436 (all $p < 0.001$). This model further underlines that sadness and tenderness are the most important factors for melancholy in songs, while the other emotions detract from it.

The model for irony and its ANOVA did show significant interactions between music emotion and lyric emotion ($F(3, 1815) = 42.9$, $MSE = 39.2$, $p < 0.001$) as well as for music condition ($F(3, 1815) = 33.1$, $MSE = 30.2$, $p < 0.001$) and lyric condition only ($F(3, 1815) = 47.1$, $MSE = 43.1$, $p < 0.001$), $R^2 = 0.19$. So in this case, it can be said that both music and lyrics help evoke irony. If irony was evoked, then it was easiest with happy or angry music, as they were rated as more intense in irony than the sadness and tenderness ($ps < 0.001$), but not significantly different from each other ($p = 0.941$). Looking at the lyrics, happiness was the strongest factor to communicate irony, as all stimuli containing happy lyrics had higher irony ratings than the other three emotions (all $ps < 0.001$). The tender lyrics were the weakest in irony and thus might again be the emotion to take over emotions of the music rather than elicit mixed emotions or irony. In general the ratings for irony are rather low, with a mean over all stimuli of $M = 1.9$, which is below average, and thus it is best to say that irony does not seem to be a good way of measuring mixed emotions in songs as well.

6. Results

For a final look on the hypotheses that a happy song with sad lyrics (and vice versa) creates melancholy and a happy song with angry lyrics (and vice versa) creates irony, new models were created containing the combined stimuli of only one music emotion (happiness or anger) or one lyric emotion and its (in)congruent combinations, with the intended emotion as a reference level and the ratings for melancholy or irony as the dependent variable. Using this method it can be seen that the intercept of melancholy ratings for the model with a happy song is 1.61, and it gets increased by 0.98 with sad lyrics, which is the highest compared to 0.58 with angry lyrics and 0.28 with tender lyrics (all $p < 0.001$). For the model with happy lyrics, the effect is even stronger, with the intercept at 1.61, and an estimate of 2.39 with a sad song, again higher than 0.48 with an angry song and 1.98 with a tender song ($ps < 0.001$). Based on this, the hypothesis that melancholy is highest with an incongruent combination of sad lyrics and happy music seems to be true. Yet the other way around draws a different picture; having the sad music as the baseline, with an intercept of 4.47 for melancholy, all other lyrics weaken the melancholy ratings by -0.48, -0.23, and -0.35 for happiness, anger and tenderness (all $p < 0.03$). With the sad lyric as the baseline the results are not much different: With an intercept of 4.47, the estimate with happy music is at -1.89 ($p < 0.001$), -1.52 with angry music ($p < 0.001$) and -0.22 with tender music ($p = 0.06$). So the hypothesis, that happy and sad stimuli evoke melancholy when combined incongruently is only true for happy music with sad lyrics and happy lyrics with a sad song, when compared to the congruent combination. With sad stimuli, the melancholy is always rated as strongest in the congruent combination and thus does not get more intense when they get mismatched with other emotions.

Looking at the irony models for happiness and anger, the second hypothesis seems to be true as well. While the intercept with happy music and happy lyrics is at 1.51, the estimate with angry lyrics is the highest with 1.42 compared to 1.20 for sadness and -0.08 for tenderness. Using the lyric as the main emotion and then comparing to the different matching music, the intercept is at 1.51 and the estimate for anger is again the highest at 1.21 compared to 1.17 for sadness and 0.75 for tenderness. The same goes for the angry music with an intercept of 1.94 for irony with angry lyrics, and an estimate of 0.78 with happy lyrics ($p < 0.001$), compared to 0.02 and 0.15 for sadness and tenderness, which aren't even significant ($p = 0.9$ and 0.25 respectively). The angry lyrics as the baseline have an intercept of 1.94 for irony and an estimate 0.98 ($p < 0.001$), while the irony even gets lower at an estimate of -0.39 for sadness ($p < 0.001$) and -0.16 for tenderness ($p = 0.21$). So it can be said that the hypotheses about melancholy interacting with sadness and happiness and irony with happiness and anger is, at least partly, true.

Mixed Emotion Scores

As a last step in analysing the basic emotion ratings, the MIN-Statistic (Hunter et al., 2008) to compare different unipolar ratings for all incongruent was calculated. For example, in the case of happy music with sad lyrics, the ratings of one person on happiness and sadness were compared, and the lower rating was chosen as the MIN value for this stimulus. On all these MIN values, a mean was calculated, to compare where the highest cases of mixed emotions can be found. Table 6.7 shows that the highest scores in the mean are happening when happy music and tender lyrics are combined (3.5) and when sad music and angry lyrics are combined (3.4). Five of the 12 combinations show a MIN-statistic of more than 3, so it seems very much to be possible to evoke mixed emotions with mismatched music and lyrics. A linear

| Music Emotion | Lyric Emotion | MIN-Statistic (mean) |
|---------------|---------------|----------------------|
| Happiness | Sadness | 2.3 |
| | Anger | 2.1 |
| | Tenderness | 3.5 |
| Sadness | Happiness | 1.9 |
| | Anger | 3.4 |
| | Tenderness | 3.0 |
| Anger | Happiness | 1.9 |
| | Sadness | 3.2 |
| | Tenderness | 1.7 |
| Tenderness | Happiness | 2.2 |
| | Sadness | 3.3 |
| | Anger | 2.2 |

Table 6.7.: MIN-Statistics (mean) for all incongruent stimuli

mixed model with the MIN-Stats as the dependent variable, the music and the lyric emotion as interacting fixed effects, and the participant's ID as a random effect, showed main effects for both music emotion ($F(3, 1331) = 32.8$, $MSE = 28.8$, $p < 0.001$), and lyric emotion ($F(3, 1331) = 80.8$, $MSE = 71.1$, $p < 0.001$), as well as an interaction of them ($F(5, 1331) = 71.5$, $MSE = 62.9$, $p < 0.001$), $R^2 = 0.28$, showing that both music and lyric have an influence on the perception of mixed emotions in terms of the MIN-Statistic. The overall results are in line with the previous analyses that tenderness plays a special role within the four emotions in question, as it shows high MIN-values for both happy and sad music when combined with tender lyrics, but not with its counterpart, anger. Interestingly, it is not about a mix of positively and negatively valenced emotions concerning this statistic, as for both happy and sad music the combination with the other respective positive / negative emotion is the highest in the MIN-Statistic, as well as for angry music.

6.2. Valence and Arousal

The participants were asked to rate the perceived emotion of the stimuli not only on the basic emotion scales, but also on the valence and arousal plane, on scales from -10 to +10. To compare these ratings to the basic emotion ratings, similar analyses were performed. First a linear mixed model approach including all stimuli, then a comparison of the control stimuli with the congruent combination of music and lyrics, an analysis of how the ratings change when the stimuli are mismatched, and finally a vector approach, mapping all mean ratings on the valence/arousal plane like a vector and trying to predict the rating by using the ratings for the control stimuli.

6.2.1. Music vs. Lyrics - A linear mixed model approach with control variables

The approach for finding the model with maximal fit was the same as described for the basic emotions in section 6.1.1. Null models without the fixed effects were created, and then compared to the models with the ratings for the control stimuli of music and lyrics only as the fixed effects, both with just the ID as a random effect. As expected the models with the fixed effects had a lower AIC rating ($AIC_{null,v} = 12560$ vs. $AIC_{fixed,v} = 11185$, $p < 0.001$ for valence, and $AIC_{null,ar} = 12576$ vs. $AIC_{fixed,ar} = 10507$, $p < 0.001$ for arousal), and were then compared to a model with the same uncorrelated slope of music + lyrics ratings. Just as it was for the basic emotions, this was the best model fit ($AIC_{slope,v} = 10956$, $AIC_{slope,ar} = 10487$), models with correlated slopes to the ID did not converge. Adding the gender as a fixed effect with interactions and checking the gender conditions in detail showed that there were no main effects or interaction effects in the valence model and could thus be left out of the model again. But, while the gender seemed to have no influence whatsoever in all other models, the model for arousal shows a slight main effect for the gender condition ($F(1, 117) = 5.12$, $MSE = 58$, $p < 0.025$), with ratings from female participants being a little lower than those from male participants (diff = 0.59). The AIC comparisons of the models with and without the gender did not suggest a significant difference between the models ($p = 0.082$) though, so the overall effect is only low. The model was reduced to a fixed effect of gender, without interactions with the control stimuli ratings, and this showed a significant difference to the model without it ($AIC_{gender,ar} = 10483$, $p = 0.014$), so only for arousal the gender seems to make a difference in judging the perceived emotions in songs, and

6. Results

is the model for further analysis.

As it was one of the main interests of this study, the ME-MS score was again added to the models as possible fixed effects with interactions (see 6.1.1). For valence, a comparison of the models again did not show any better fit with the ME-MS ($AIC_{ME-MS, val} = 10962$, $p = 0.93$) and also did not show any main or interaction effects in an analysis with the anova-function. The same goes for the arousal ratings ($AIC_{ME-MS, ar} = 10487$, $p = 0.63$), so it seems that the ME-MS score does not affect the perception of emotion in songs, neither on basic emotion scales nor in the valence and arousal plane.

As a last step, the chosen models were compared to the simple linear model without random effects and random slopes, to check that linear mixed models were a good approach, which proved to be the case ($ps < 0.001$) for both valence and arousal.

| | Parameter | Variance | SD |
|--------------------------|-----------|----------|-------|
| Valence (ICC = 0.029) | ID | 0.422 | 0.65 |
| | Music | 0.029 | 0.17 |
| | Lyric | 0.073 | 0.27 |
| | Residual | 14.001 | 3.74 |
| Arousal (ICC = 0.062) | ID | 0.759 | 0.871 |
| | Music | 0.017 | 0.129 |
| | Lyric | 0.009 | 0.097 |
| | Residual | 11.370 | 3.372 |

Table 6.8.: Variance and SD for random effects of valence & arousal main LMER models

| | Parameter | Estimate | 2.5% | 97.5% | df | p |
|-----------------------------|----------------|----------|--------|--------|-----|--------|
| Valence ($R^2 = 0.49$) | Intercept | -0.527 | -0.741 | -0.313 | 111 | <0.001 |
| | Music | 0.518 | 0.473 | 0.563 | 106 | <0.001 |
| | Lyric | 0.463 | 0.406 | 0.521 | 114 | <0.001 |
| Arousal ($R^2 = 0.65$) | Intercept | 0.533 | 0.161 | 0.905 | 116 | 0.006 |
| | Music | 0.807 | 0.771 | 0.843 | 86 | <0.001 |
| | Lyric | 0.072 | 0.034 | 0.110 | 109 | <0.001 |
| | Gender(female) | -0.591 | -1.055 | -0.128 | 113 | 0.014 |

Table 6.9.: Values for fixed effects of valence & arousal main LMER models

The main parameters of the model can be found in Table 6.8 and 6.9. These show that again the random effects and especially the random slopes do not play a major role, so it seems that, as the ME-MS score and gender were both of no importance as well, the differences in between participants are minor. The most important thing to note is that while the estimates of the fixed effects for the valence suggest that

music and lyric are again of more or less equal importance for the perception of the valence of emotion in songs, the fixed effects coefficients of the arousal model however differ a lot from each other, with the estimate of the lyric condition being close to zero. This suggests that arousal ratings do vary a lot more depending on the music that is heard than on the lyrics. This is an important finding that also sheds light on the importance of music and lyrics, in comparison to the ratings on basic emotions, where this was not suggested by the analyses. Whereas in previous studies (Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2014) the important difference in emotion ratings for music and lyrics seemed to lie within the valence of positive or negative music or lyrics, in this case using the valence and arousal scale shows that the most important difference seems to lie within the arousal that is better communicated by music than by lyrics. Another look on the congruent, and on the incongruent combinations is necessary to further determine these influences.

6.2.2. Congruent Combinations

To see how the congruent stimuli performed compared to the control stimuli on the valence and arousal plane, eight linear mixed models (to account for repeated measures) for each main intended emotion and the respective ratings on the valence and arousal scales were performed, as a function of the stimulus condition and the gender of the participants. These models all had the ratings as the dependent variable, and the Stimulus condition as the fixed effect. For further analyses of possible gender differences, interactions between the stimulus condition and the gender were added to the model; as a random effect, the participant's ID was used. Another linear mixed model comparing all different emotions with each other was created as well to see how much they differ in valence or arousal; this tried to predict the ratings based on the fixed effects of Emotion and Stimulus condition as well as their interactions, again with possible interactions with the participants gender, and the ID as a random effect. The models were then analyzed by using a Type III Analysis of Variance to look for significant effects and least-square means with Tukey adjustment for post-hoc analysis of differences.

Just as it was for the comparison of the intended with the other emotions in the section on basic emotions, again comparisons of models showed that the AIC for the models without the gender as a fixed effect were lower than the ones with it or there was no significant difference ($ps > 0.3$), apart from the tender stimulus in valence, which will be analyzed in more detail below. Separate analyses of the models for the other three emotions including the gender also did not show a main or interaction effect for it in any model, again suggesting that the gender does not have an effect

there, and was thus excluded from the following analyses.

Valence

It can be seen in the boxplots in Fig. 6.7, that only for sadness and happiness the highest/lowest absolute ratings (positively or negatively valenced) in all stimuli were those for the combined stimulus condition, while the angry and tender stimuli have the highest ratings in the lyric condition and the song being the lowest rated (absolute). It can also be seen that these ratings all have a rather high spread of the values, even though the data was filtered beforehand for people seemingly not understanding the instructions on the test songs; there are still quite a few outliers and in some cases a very high interquartile range. This could be due to people not being so familiar with the nomenclature of valence and arousal compared to an easier to understand system like the basic emotions.

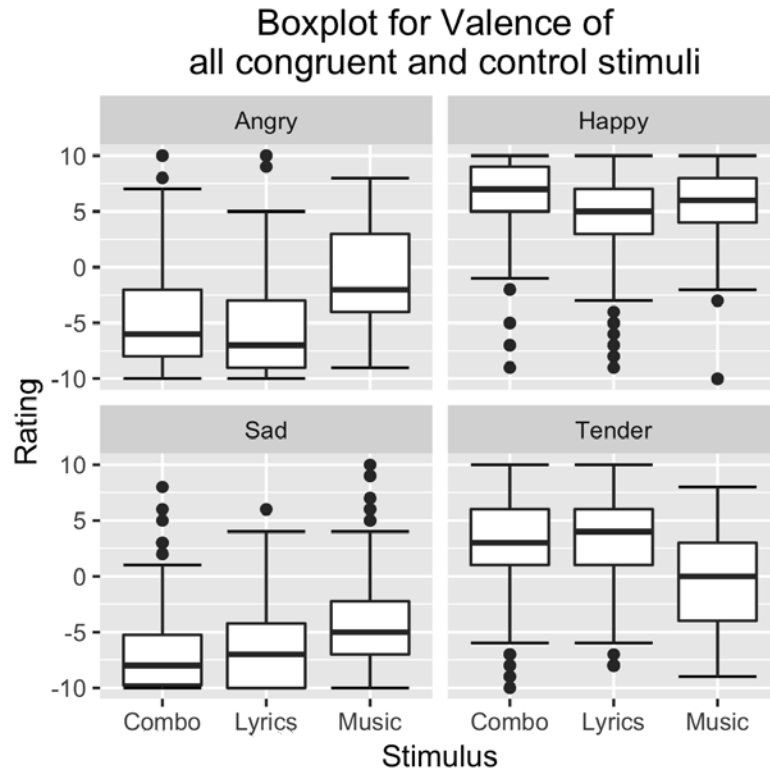


Figure 6.7.: Boxplots for valence ratings on all congruent and control stimuli

The linear mixed model and the corresponding ANOVA for the happy song, happy lyrics and their combination showed a main effect for stimulus, $F(2, 244) = 12.903$, $MSE = 115.95$, $R^2 = 0.045$, $p < 0.001$, and post-hoc comparisons showed that the ratings for the combined stimulus were significantly higher in valence than those for the lyric condition, $p < 0.001$ and $\text{diff} = 1.9$. The ratings for the song only and

6. Results

the combined stimulus did not differ significantly, $p = 0.250$, meaning there was no significant increase or decrease in valence when adding happy lyrics to a happy song. This is in line with the findings for the comparison of the basic emotions, where the ratings for the happy song and the congruent combination of happy song and happy lyrics did not differ in intensity either (see p. 43). The song condition was rated as higher in valence than the lyrics, $p = 0.003$.

The linear mixed model and the corresponding ANOVA for the sad song, sad lyrics and their combination showed a main effect for stimulus as well, $F(2, 242) = 31.525$, $MSE = 248.37$, $R^2 = 0.077$, $p < 0.001$. In this case, the ratings for the song and the combined stimulus differed significantly ($p < 0.001$), with a difference of 2.647, which in this case means that the song is seen as more positive on the valence scale, and thus adding lyrics to it makes the combination seem more negative. The ratings for the lyric-only condition and the combination did not differ significantly, $p = 0.504$, making the sad lyrics about as negative as the sad combination in perceived valence. The song was rated as more positive than the lyrics, $p < 0.001$, $\text{diff} = 2.246$.

For the angry stimuli, the linear mixed model and ANOVA again showed a main effect for stimulus, $F(2, 242) = 52.643$, $MSE = 773.6$, $R^2 = 0.163$, $p < 0.001$. In this case, the combination was perceived as less positive than the song ($p < 0.001$, $\text{diff} = -3.615$), but the combined stimulus was rated as significantly more positive than the lyrics ($p = 0.034$, $\text{diff} = 1.229$), and the song was also seen as more positive than the lyrics ($p < 0.001$, $\text{diff} = 4.844$). This is interesting, as both songs, that were intended to be negative (angry and sad), were rated as more negative when lyrics were added. In the case of anger, the lyrics were rated as even less positive than the combined stimulus, so the song actually lowers the intensity of perceived negative emotions when added to the lyrics.

In the case of tenderness, the comparison of models with and without gender interaction effects showed a significant difference and a slightly better fit of the model for the one with gender interactions for the case of the tender emotions, ($\text{AIC}_{\text{without gender}} = 2119.0$ vs ($\text{AIC}_{\text{with gender}} = 2116.1$, $p = 0.030$), so in this case the model with gender interactions was used for further analyses. This model and an analysis of variances showed a main effect for stimulus as well, $F(2, 244) = 29.245$, $MSE = 407.15$, $p < 0.001$, $R^2 = 0.137$. Post-hoc comparisons showed that the combination was seen as more positive than the music condition ($p < 0.001$, $\text{diff} = 3.130$) and the lyrics were seen as more positive than the music ($p < 0.001$, $\text{diff} = 3.384$). For the tender lyrics condition and the the combined condition, no significant differences could be found ($p = 0.867$). The model also showed an interaction

6. Results

effect for the Stimulus * Gender condition, $F(2, 244) = 3.9442$, $MSE = 54.91$, $R^2 = 0.137$, $p = 0.02$, but post-hoc comparisons showed no significant effects for the relevant cases (Lyrics:Male - Lyrics:Female, Music:Male-Music:Female, Combo:Male - Combo:Female, all $ps > 0.2$), again suggesting no real difference in gender.

These results suggest that, apart from the happy stimulus, the lyrics are always seen as just as positive or negative as the combined stimulus with added music, thus making a bigger difference for the rating than the music. Only for the happy stimuli, the music seemed to take over, being seen as just as positive as the music with matching lyrics. This could, as always, be due to the choice and composition of the stimuli, with the lyrics usually being intrinsically more effective. Comparing the congruent stimuli with the control stimuli suggests that the music is not as good in communicating the tendency of valence of the stimulus, whereas for all discrete emotions, as seen in the previous section, they seem to be of equal importance. It must be added that the R^2 for all models are rather low, so it might be the case that there are other effects that were not considered in the models or the study itself concerning valence.

Arousal

Taking a look at the boxplots for the arousal ratings of the congruent and control stimuli in Fig. 6.8, it can be seen that for the stimuli where the arousal ratings were intended to be positive (angry and happy), the ratings for the lyrics condition seem to be significantly lower than for the other two respective stimuli. For the low aroused sad and tender stimuli, the combo, song, and lyrics seem to not differ much in median rating from each other.

Just as it was for the valence ratings, the range of the values for each stimulus are rather high with several outliers despite previous filtering.

When comparing all arousal ratings for all congruent stimuli and their control stimuli, it is possible to do a more refined model. This includes a non-correlated random slope for Emotion and Stimulus grouped by the ID condition, leading to a model of the following form:

$$\text{Rating} \sim \text{Stimulus} * \text{Emotion} + (1|\text{ID}) + (0 + \text{Stimulus}|\text{ID}) + (0 + \text{Emotion}|\text{ID})$$

This model was compared to a model incorporating possible interactions of the gender condition with stimulus or emotion condition, but it showed it that the model without the interactions did not have a significantly better fit ($\text{AIC}_{\text{without gender}} = 8129.9$ vs ($\text{AIC}_{\text{with gender}} = 8132.0$, $p = 0.207$). The model with random slopes was

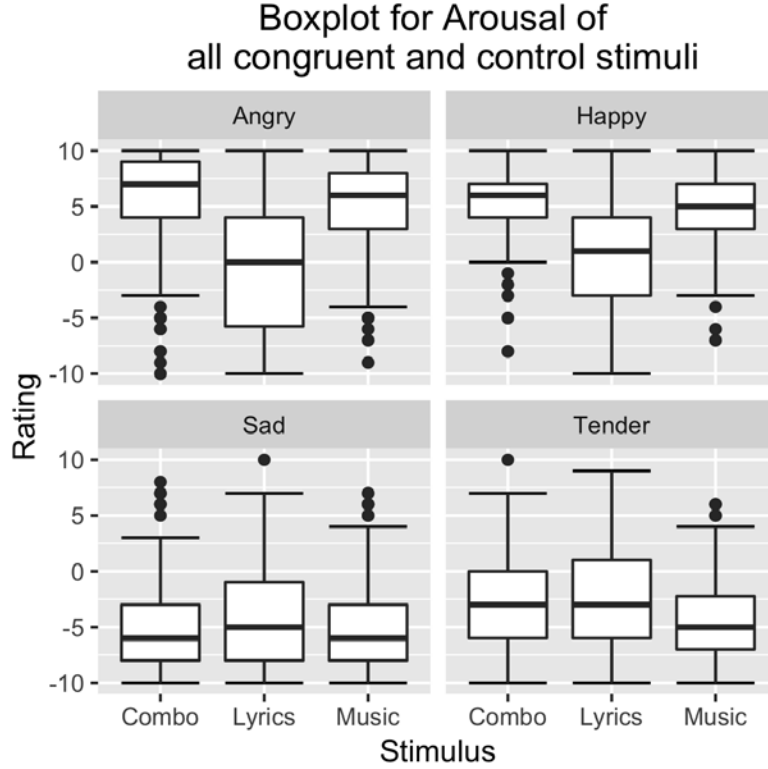


Figure 6.8.: Boxplots for arousal ratings on all congruent and control stimuli

deemed better than the one without them ($AIC_{\text{no slopes}} = 8361.2$ vs ($AIC_{\text{slopes}} = 8129.9$, $p < 0.001$), so the one without gender interactions, but with random slopes was used for further analyses.

The ANOVA for this model showed a main effect for Stimulus, $F(2, 181.41) = 24.958$, $MSE = 241.96$, $R^2 = 0.027$, $p < 0.001$, and all post-hoc comparisons were significant ($ps < 0.05$). While the participants rated the combined stimuli as higher in arousal than both lyrics and songs, the lyrics were rated as less aroused than the song. Collapsed across all stimuli, the lyrics seem to transport less arousal than the song. It must be said that this comparison is hard to draw conclusions from, because it's hard to say if this is because the positively aroused stimuli behave differently than the negatively aroused stimuli, or because the lyrics are just not as good in transporting arousal. They are however in line with the previous model on all congruent and incongruent stimuli with discrete emotions that suggested that lyrics are not as good in transporting arousal. The Emotion condition shows a main effect as well ($F(3, 122.05) = 152.375$, $MSE = 1477.27$, $R^2 = 0.027$, $p < 0.001$) and interestingly, all ratings for arousal grouped by the intended emotions differ significantly from each other ($ps < 0.001$) apart from those for anger and happiness ($p = 0.923$), so the participants rated the stimuli for these two emotions as equally arousing.

6. Results

For more detailed analyses of the differences for the stimuli, separate models divided by the intended emotion were created.

For arousal, the linear mixed model and the corresponding ANOVA for the happy song, happy lyrics and their combination showed a main effect for stimulus, $F(2, 244) = 73.933$, $MSE = 897.06$, $R^2 = 0.244$, $p < 0.001$, and post-hoc comparisons showed that the ratings for the combined stimulus were significantly higher in arousal than those for the lyric condition, $p < 0.001$ and $\text{diff} = 4.983$. Just as it was the case for valence, the ratings for the song only and the combined stimulus did not differ significantly, $p = 0.328$. The song condition was rated as higher in valence than the lyrics, $p < 0.001$, $\text{diff} = 4.344$. This suggests that the lyrics do not play as much of a role in the participant's ratings for arousal.

The linear mixed model and ANOVA for the sad stimuli showed a main effect for stimulus as well, $F(2, 242) = 4.014$, $MSE = 40.404$, $R^2 = 0.011$, $p = 0.019$. Post-hoc comparisons here show only a significant difference between the combined and the lyrics condition, $p = 0.028$, $\text{diff} = 1.049$. Combo and song, as well as song and lyrics did not differ significantly from each other, again supporting the idea that for the case of arousal, the lyrics do not play a significant role, as the ratings for the song and the combined stimulus did not differ much, while the ratings for lyrics and combined stimulus did. The difference, even though not significant, between the song and the combined stimulus was estimated to be -0.115, so that the combo was rated as ever so slightly less aroused.

Comparing the angry stimuli, the linear mixed model and ANOVA again showed a main effect for stimulus, $F(2, 242) = 95.044$, $MSE = 1399.3$, $R^2 = 0.250$, $p < 0.001$. Post-hoc comparison showed that the participants rated the combination as more aroused than the lyrics ($p < 0.001$, $\text{diff} = 6.025$), the lyrics as significantly less aroused than the song ($p < 0.001$, $\text{diff} = -5.693$), while the ratings for song and combination did not differ significantly ($p = 0.780$).

Post-hoc comparisons for the tender stimulus, after checking for a significant main effect in stimulus ($F(2, 242) = 7.6039$, $MSE = 94.557$, $R^2 = 0.028$, $p < 0.001$), showed that both combined stimulus and lyrics were rated as higher in arousal than the song alone ($p = 0.029$ and 0.001 , respectively), while lyrics and combo were not perceived as significantly different in communicating arousal ($p = 0.609$). As the tender song was intended to be negatively aroused, it is interesting to see that adding lyrics to the song made it seem more aroused, just as much as the lyrics on its own, while for the other three emotions song and combined stimulus were not much different in this aspect. Only for tenderness, the lyrics seem to be of higher

importance for the participants.

Overall, there were again no signs of gender differences in the ratings for the congruent stimuli. Comparing the ratings for valence and arousal with those for the basic emotions, it is interesting that differences in lyrics and songs are more pronounced when people are asked to rate on the valence/arousal scale, in the way that lyrics are not as important when it comes to transporting arousal ratings, while they do play a role for the valence rating and the songs are not as strong in this regard.

6.2.3. Comparing Congruent and Incongruent Combinations

Linear mixed models

To understand how valence and arousal ratings are affected by music or lyrics, a comparison of congruent and incongruent stimuli was necessary. To do this, the ratings on valence and arousal of all combined stimuli were used for another two linear mixed models that have either valence ratings or arousal ratings as the dependent variable and the music emotion and the lyric emotion as interacting fixed effects; as a random effect, the participant's ID was used as always. The participant's gender was added in another model to check again for any possible gender effects, but in both models, not significant main effects ($F(1, 121) = 1.04$ and 2.25 , $MSE = 20$ and 32 , $p = 0.308$ and 0.136 , for valence and arousal respectively) ($F(1, 121) = 1.04$ and 2.25 , $MSE = 20$ and 32 , $p = 0.308$ and 0.136 , for valence or arousal) for gender were found and model fit wasn't improved either. So it seems that for neither basic emotions nor valence and arousal ratings it seems to make a difference what gender the participant has.

Both models for valence and arousal without the gender effect have a good marginal R^2 of 0.44 , and 0.53 , respectively.

Using the Type III ANOVA for the valence model, there was a main effect for Music Emotion, $F(3, 1815) = 220.2$, $MSE = 4167$, $p < 0.001$. So collapsed over all different combinations for each music with all four lyrics, the happy music has higher valence ratings than all other music stimuli, all $p < 0.001$, meaning even combined with other lyrics, the happy music is always rated as more positive than the other ones. The sad music stimuli are all rated as more negative than the angry and tender ones, both $p < 0.001$. Interestingly, the angry and tender music did not differ significantly in valence ratings, collapsed over all combinations for each music emotion, $p = 0.944$. The reasons for this are not clear, as the angry music

6. Results

was expected to be rated as a lot more negative than the positively valenced tender music; it might be though, that for these songs, the lyrics play a higher role in changing the perception of valence, so that they both lead to a more mixed response than the sad and happy songs.

The Lyric Emotion condition in the valence model showed a main effect as well, $F(3, 1815) = 319.3$, $MSE = 6042$, $p < 0.001$. Collapsed over all different combinations for each lyric with the four music types, the happy lyrics are more positively valenced than the angry and sad lyrics (both $p < 0.001$), but slightly less positive than the tender lyrics ($p = 0.024$). The sad lyrics are rated as less positive than the tender lyrics ($p < 0.001$), but not significantly different than the angry lyrics ($p = 0.914$). The tender lyrics are rated as more positive than the angry lyrics ($p < 0.001$), on all song combinations. For the lyrics, it seems that the overall valence of the lyric seems to play an important role, and they change the perception of emotion in songs in the same way if they are positive (difference for happy and tender is -0.79) or negative (difference for sad and angry is -0.18), but they do differ a lot between positive and negative (5.76 between happy and sad lyrics, 5.58 between happy and angry, -6.55 between sad and tender, and -6.37 between angry and tender). Overall, both music and lyric seem to be important for the perception of valence in songs.

An interaction effect between Music and Lyric emotion was found in the valence model, $F(9, 1815) = 2.7$, $MSE = 51$, $p = 0.004$. Fig. 6.9 shows the estimated marginal means predicted by the emmeans-Function in R. These are the means of the interaction of music and lyric emotion for valence and the blue bars around them show the 95% confidence intervals. If the bars of two emotion pairs would overlap when placed on the same horizontal line, or be very close to each other, this would mean that they are not statistically significantly different from each other in a contrast, using a Tukey adjustment. This shows what the main effects already suggested: Changing the lyric in each song always makes a significant difference comparing a positive and a negative lyric emotion (all $ps < 0.05$); comparing the negative emotions angry and sad or the positive emotions happy and tender with each other, it doesn't show a great difference in all songs even in the sad song, where happy and tender lyrics might have made a difference ($p = 0.069$). It is interesting to see that the ratings for the angry music with the lyric combinations seem very similar to those for the tender music and its lyrics, both having a lot less extreme ratings to either side of the valence spectrum than sad and happy songs; this was reflected in the main effect analysis as well, where angry and tender music did not differ significantly in valence.

For the model on arousal ratings, the ANOVA showed a main effect for music emotion, $F(3, 1815) = 891.04$, $MSE = 12666$, $p < 0.001$. Post-hoc comparisons

6. Results

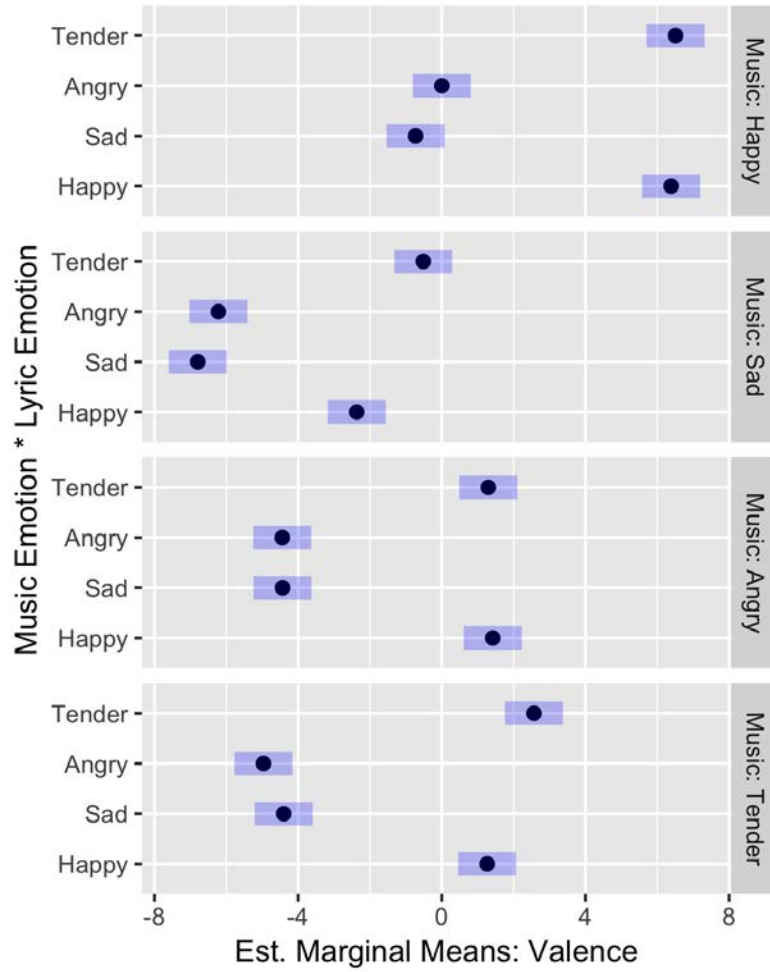


Figure 6.9.: Estimated marginal means for valence model. Blue bars show 95% confidence intervals

with emmeans and Tukey-adjustment showed that the ratings for all music stimuli are significantly different from each other ($ps < 0.001$), apart from angry and happy music, the music stimuli with high arousal ($p = 1$). For the other comparisons, the predicted differences are also pretty high, within 8-9 points; only the sad and tender music stimuli, the low arousal ones, have a lower difference of about 1. This again shows that the music is very important for the survey participants when rating arousal in songs.

The ANOVA also showed a main effect for the lyric condition, $F(3, 1815) = 7.60$, $MSE = 108$, $p < 0.001$. Post-hoc comparisons show that, although significant, the differences between the lyrics are a lot less pronounced: Only two comparisons show significance, between sadness and anger ($p = 0.0003$), and sadness and tenderness ($p = 0.0003$). The estimated marginal means are not significantly different for the comparisons between happy lyrics and sad, angry, and tender lyrics. ($ps > 0.1$), and angry and tender lyrics show even less difference ($p = 1$, estimate = -0.002).

6. Results

Even for the significant comparisons, the differences are very low, with an estimate of about 1. It seems that, even though they can make a small difference, especially with sadness, the most effective lyric stimulus in this survey, lyrics seem to have a lot less impact on arousal ratings than the music. An interaction effect between music and lyrics emotion was found in the arousal model as well, $F(9, 1815) = 2.71$, $MSE = 39$, $p = 0.004$. Fig. 6.10 shows the differences in interaction quite well.

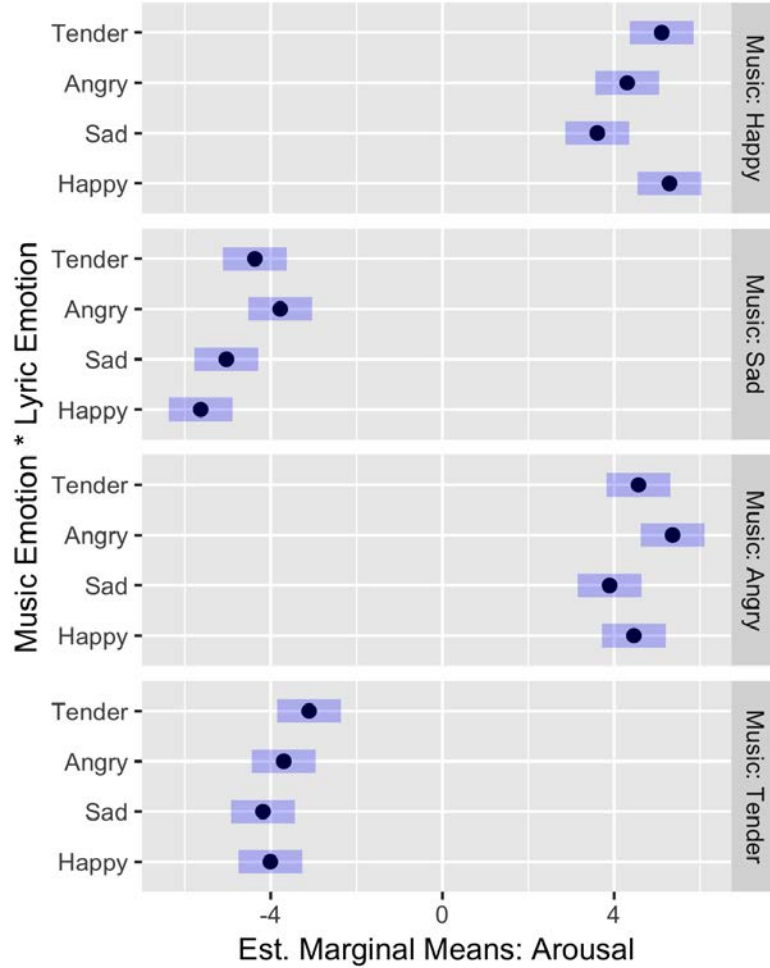


Figure 6.10.: Estimated marginal means for arousal model. Blue bars show 95% confidence intervals

Comparing only the differences in ratings for lyrics for one music emotion, there are basically no changes due to the them. The main effect mainly comes from the comparisons between different music, as the differences between the different music and their lyrics are so pronounced (e.g. comparing the sad music, happy lyric to the happy music, sad lyric). This again shows how important the music is for rating arousal compared to the lyrics.

This is also a finding that is not visible in the analyses for basic emotions, so using the valence & arousal plane as a rating scale, another detail can be found out about

the differences of music and lyric on the perception of emotion in songs.

Vector approach

As the last step in visualizing and analysing the response data on valence and arousal, the mean ratings for every stimulus were plotted on a valence / arousal grid, as seen in the Fig. 6.11. Incorporated into these plots are the ratings on the music-only and lyrics only stimuli; the ratings for the combined stimuli, in all possible combinations; and, in a means of visualizing the predictability of each combined rating, the estimated values when summing up the mean ratings of the two respective control stimuli of each emotion. This was done, apart from showing the data in a more eye-pleasing way, to show how the participants tried to evaluate a mixed emotion stimulus and if that was matching to just mentally summing up the ratings for the music-only and lyric-only stimuli.

As can be seen, the biggest differences in valence for the actual ratings for the combined stimuli and the respective estimates happen with the sad music (mean difference of -1.3). The other music stimuli are pretty similar and can be estimated rather well with a simple addition of vectors. (Mean difference between the actual ratings and estimates: $M_{\text{Happy}} = -0.4$; $M_{\text{Angry}} = -0.42$; $M_{\text{Tender}} = -0.4$). This makes for an absolute mean difference of 0.63 over all stimuli. It seems that ratings for valence can be estimated quite well by a simple addition of vectors of the control ratings.

The arousal ratings are a little different. The mean differences between rating and estimate are: $M_{\text{Happy}} = -2$; $M_{\text{Sad}} = -1.9$; $M_{\text{Angry}} = -1.2$; $M_{\text{Tender}} = -2$. The estimated ratings are always lower than the participant's ratings on arousal, for both high and low arousal ratings, with an absolute mean difference of 1.8. This is most certainly due to the fact that the lyric-only ratings are all around 0 or below for all four emotions and thus an addition of vectors always leads to rather low ratings in arousal; but as can be seen in the section before, the music was shown to be more important for estimating the arousal rating of a song, so a non-weighted, simple addition cannot be as accurate for arousal as it is for valence.

6.2.4. Discrete and Dimensional Model

As a last statistical analysis, the predictability of the discrete emotions terms using the valence and arousal ratings, as well as the other way around was checked. This was done using linear mixed models for each of the four main emotions tested plus melancholy, as well as one for valence and one for arousal. The ID as a random

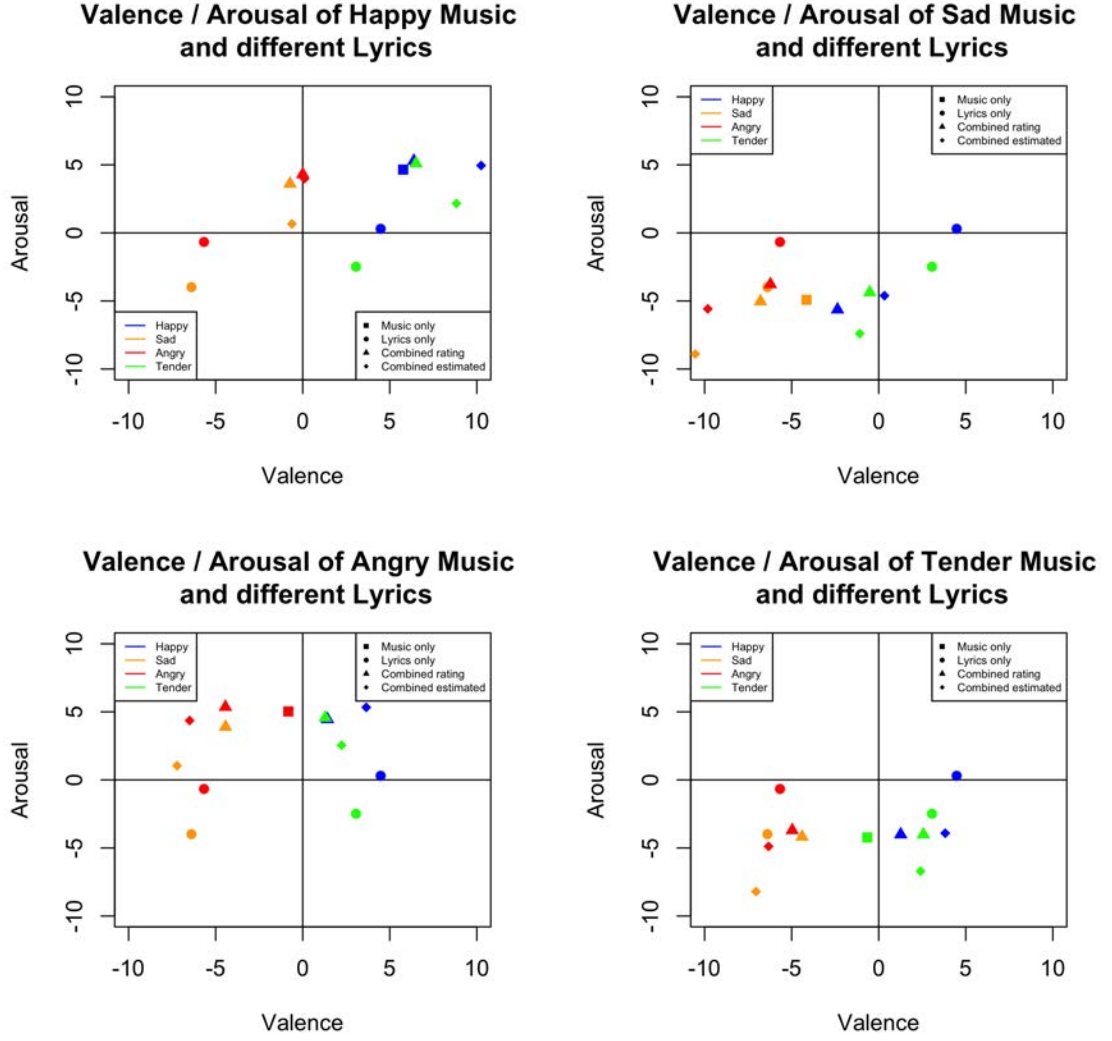


Figure 6.11.: Mean ratings for all stimuli in the valence and arousal plane

intercept and uncorrelated random slopes were incorporated and everything was fit using restricted maximum likelihood.

The results of these models can be found in Tab. 6.10. It is important to note that the values for valence and arousal were recorded using a scale from -10 to 10, whereas the discrete emotions were recorded from 1 to 5 and thus have a lower resolution.

It is hard to say which model is better at predicting the other. While the discrete emotions do well in predicting the valence ratings, they are not so good at arousal, looking at the R^2 values. For the discrete emotions, happiness is the emotion which can be predicted best by valence and arousal, with sadness also being rather high, while anger and melancholy are a little lower and prediction of tenderness ratings by valence and arousal is rather poor. This again shows that happiness and sadness

6. Results

are the simplest emotions to be determined on both scales, and that tenderness is the most ambiguous emotion. Interesting to note as well is the importance of melancholy for the prediction of the arousal ratings: The higher the melancholy ratings, the lower the arousal, even more so than the effects of sadness and tenderness. On the other hand, in the valence model, melancholy wasn't even significant, $p = 0.22$) (using R's lmerTest package to calculate p-values for linear mixed models), whereas in all other models every emotion/dimension was highly significant, all $ps < 0.001$. Looking at the prediction of melancholy by valence and arousal, it can also be seen that arousal is a lot more important for high melancholy ratings than the valence is. This can mean that melancholy can both be positive and negative in valence (seen as slightly more negative), but low arousal is more important for melancholy and leads to higher ratings.

This is different from the findings by Eerola and Vuoskoski (2011), where the dimensional models were better estimated by the discrete emotions than the other way around. In this survey, we cannot conclude hard and fast rules about this, as it differs between dimensions and between discrete emotions.

| | | Marginal R^2 | Fixed effects estimates |
|-------------|------------|----------------|--|
| Dimensional | Valence | 0.59 | Intercept = -2.8357; H = 2.1951; S = -0.9587; A = -0.4414; T = 0.3778; M = -0.1067 |
| | Arousal | 0.39 | I = 2.4606; H = 1.0818; S = -0.5749; A = 1.1802; T = -0.5930; M = -1.3370 |
| Discrete | Happiness | 0.57 | I = 2.55410; V = 0.16795; A = 0.03099 |
| | Sadness | 0.47 | I = 3.05004; V = -0.12913; A = 0.06927 |
| | Anger | 0.31 | I = 2.12317; V = -0.12804; A = 0.09323 |
| | Tenderness | 0.19 | I = 2.79423; V = 0.06521; A = -0.09773 |
| | Melancholy | 0.35 | I = 3.12168; V = -0.05379; A = -0.11538 |

Table 6.10.: LMER for comparison of discrete and dimensional model

Further comparisons of the discrete and dimensional model by the results of the separate analyses are done in the Discussion below.

7. Discussion

This study aimed to close a gap in research concerning emotions in songs with lyrics, and how music and text interact to convey emotions. It also tried to address the differences in discrete and dimensional emotion models concerning modern songs, if and how mixed emotions appeared in emotionally ambiguous song stimuli, and whether the listener type or gender had an influence on these perceptions. It tried to do this with songs that were composed and whose lyrics were written solely for this research, and how different combinations of these music and lyric types of the four basic emotions happiness, sadness, anger, and tenderness changed the perception of them.

Discrete and Dimensional Emotions in Songs - Music and Lyrics In this study, music and lyrics actually were equal partners, speaking in general. Both in emotionally congruent and incongruent matchings of music and lyrics, they each have a more or less equal share in defining the emotions of the song.

Also, the combined congruent stimuli were perceived as more intense in emotion than just the lyrics or the music on its own, or than incongruent matches, again leading to the conclusion that both lyrics and music play an important part concerning perceived emotions in songs. While the studies by Ali and Peynircioğlu (2006) suggested that lyrics detracted from positively valenced music and enhanced negatively valenced songs, this cannot be said in this study; both in discrete and dimensional models, the congruent combinations were seen as more intense than the music stimuli, no matter the valence of the song. Fiveash and Luck (2016) also stated that positive music detracts from the detection of errors in lyrics, while negative music enhances it, and in that way also finding an effect of music depending on the valence of it, this was also not the case here, at least not for emotion perception.

Furthermore, negative emotions were rated as slightly more intense in perceived intended emotion, which is again different to the findings of Ali and Peynircioğlu (2006), where the positive emotions emerged as more dominant. It could just be, that there are differences in the stimuli between this study and the previous one,

7. Discussion

or that this is a change due to actual songs in the type of popular music being used.

Interestingly, the dimensional approach created a new insight: The arousal in songs was clearly dominated by the music, the lyrics effectively made little to no difference in arousal in all song types. It is reasonable to assume that music is better at communicating arousal than lyrics are. The valence was best perceived in happy and sad songs, whereas angry and tender lyrics were stronger than the combined song, showing that there are slight differences depending on the emotion. For all emotions, lyrics performed better at communicating the intended valence than the music did, and they were often just as good as or even better than the congruent song stimuli. Both music and lyrics are important for valence in songs, and it can even be argued that lyrics dominate the valence dimension. It might just be the effectiveness of the composed stimuli that lead to these findings, which can never be fully ruled out. Looking at the dimensional model together, it can again be argued that music and lyrics are equally important for determining the emotion of a song, maybe with a slight favor for music, as the arousal dimension is so clearly dominated by it.

The congruency to intended emotion of a song also helped to show that music and lyrics are equal partners: When comparing positive and negative emotions, congruent songs were always better at transporting the intended emotion than both music or lyric. Concerning the intended emotion of music in incongruent stimuli and the intended emotion of lyrics in incongruent stimuli, no significant difference in intensity could be found, again making a case for equal importance of song and lyrics. Split up to all emotions separately in incongruent songs, music dominated in happiness, lyrics dominated for the more complex emotions of anger and tenderness, and in sadness they were seen as equally strong. This underlines the significance of differentiating between the emotions as the importance of music or lyric changes depending on them.

For the happy, angry, and sad music, the other lyric of the same valence (happy-tender, sad-angry) led to no difference in intensity of emotion of the music, only tenderness of the music was better communicated with sadness (negative arousal) than with other emotions. It can again be concluded that lyrics have different effects on the music depending on both communicated emotions, so a well differentiated analysis is necessary.

Differences to previous studies might be because of the control stimuli: In this study, the music-only control stimulus did not change the melody instrument from a singer's voice to something else like a piano (Stratton & Zalanowski, 1994; Ali &

7. Discussion

Peynircioğlu, 2006) and did not carry any actual semantic meaning, not even from a foreign language (Mori & Iwanaga, 2014).

In all stimuli concerning tenderness or sadness, the participants reported similarities between the intensities of those emotions. This might be due to a semantic problem, as the difficulty to distinguish tenderness and sadness have been shown by Eerola and Vuoskoski (2011) as well. Tenderness also emerged as a special emotion compared to the others by taking over the valence of the underlying second stimulus in incongruent matches.

Discrete vs. Dimensional Emotion Categories As already outlined above, both rating possibilities can help find out different things about the interaction of music and lyrics in songs. Especially the notable difference of importance to arousal ratings shows that using only discrete emotions can lead to missing important information about emotions in songs. But using discrete, unipolar rating scales helps in determining which actual emotion term was perceived and to investigate cases of mixed emotions, that would be harder to find using dimensional models only. Many of the emotion terms, especially happiness and sadness, could be represented very well by valence and arousal ratings however, with the exception of tenderness, the most ambiguous emotion term in this study. This is in line with the findings by Eerola and Vuoskoski (2011), that discrete and dimensional models have a lot in common.

Mixed Emotions As expected, combinations of incongruent emotions of music and lyrics lead to the perception of mixed emotions in songs. The strongest cases of mixed emotions happen when sad music and angry lyrics or sad lyrics and angry music are combined, as well as when tender music gets combined with sad lyrics or tender lyrics with happy music. It can be speculated that this is also because of the uncertainty of definitions for tenderness, which might seem both happy and sad to some, especially in music. In cases of incongruent combinations with contrasting emotions like happiness and sadness, which has been studied before (Larsen & McGraw, 2014), mixed emotions still appear, even though not as prominent. Both music and lyrics are important for the perception of mixed emotions. The MIN-statistic applied to discrete emotions helps to find out about these, as the dimensional model cannot show these findings as easily. A problem of the way this study was conducted is that the emotions were not measured while listening, so it is hard to say if it only was a change of perception over the time of listening and all these emotions were then reported, or if it was in fact a mixed emotion at the same

7. Discussion

time; the study by Larsen and Stastny (2011) did show though that, at least when induced, mixed emotions in music for happiness and sadness actually happen.

The aesthetic emotion of melancholy was used to look for special cases of mixed emotions (Brady & Haapala, 2003). It was hypothesized that the combination of happy and sad emotions lead to the highest cases of melancholy. This was found to be true for the case of happy music with sad lyrics, and sad music with happy lyrics compared to the combination with other music types. For sad music compared to all combinations with lyrics on the other hand, melancholy ratings were the highest for the congruent combination, so here positive lyrics did not increase melancholy. It seems that sad songs alone can already communicate high degrees of melancholy, so the effectiveness of melancholy as a variable for measuring mixed emotions is limited when it comes to sad music types. Ratings on melancholy also varied mostly by music and not by lyric type, so melancholy should not be considered a good way of measuring mixed emotions in songs.

Using irony as an indicator for high mismatches of emotions, the theory that anger and happiness combined lead to the highest perception of irony was found to be true. In all other cases, irony was a rather ineffective way of measuring mixed emotions. Even though both music and lyrics seem to have an important influence on the perception of irony, reported perception of irony was all in all very low with a mean below the average.

Listener Types and Gender Differences No significant differences between listener types of music empathizers or music systemizers, between preferences for genres, or gender differences were found, apart from a small main effect for arousal, where females rated arousal in all stimuli as lower than males. Differences that were represented by the user's ID in the mixed models were always very low, so it can be said that these traits are not important when evaluating perceived emotions in songs. This might all be because of the very clear instructions to just rate the perceived emotions intended by composer and lyricist, so that these emotions were rated as "objectively" as possible by the participants and in that way eliminating any personal difference effects. Using these stimuli for another survey concerning induced emotions, the effects might be very different. These results mirror the findings by Ali and Peynircioğlu (2006) and A. H. Fischer et al. (2018) that there are little to no gender differences when it comes to emotion ratings.

On the stimuli created for the survey The music and lyrics composed for this study were overall a good fit for the task. The analysis for intended emotions showed

7. Discussion

that, when comparing the positive and negative control stimuli of music-only, lyric-only and then their congruent combination, all combinations were equally good in communicating the intended emotion, the music stimuli were equally good, and only the lyrics were better in transporting negative emotions. The lyrics were a little better in communicating the emotions than the music was, but that may also be due to the possibility that it is easier to communicate emotions like anger and tenderness in speech compared to music.

This could be seen when looking at the emotions independently: sad and happy music were perceived as more intense in intended emotion than angry and tender music. For lyrics, three of the emotions were equally good at communicating the emotions, only sadness was better than the rest.

When comparing the ratings of the intended emotions to the other emotions of each stimulus, to see whether some stimuli might have been ambiguous in emotions or even better at communicating another one, the stimuli for happiness and sadness always performed best at intended emotions, in sadness the music was equal in ratings to melancholy, and the tender music-only stimulus was actually perceived as sadder and more melancholic than tender. These two problems are probably due to the difficulty of communicating these emotions precisely with music; participants in the study of Eerola and Vuoskoski (2011) have had the same problem of differentiating between sadness and tenderness. These emotion terms are a bit unclear when it comes to music. It might also be that the cues for tenderness in music proposed by P. N. Juslin and Laukka (2004) are not sufficient enough or were not incorporated well enough into the compositions at hand to communicate this emotion. The lyric-only stimuli did not show these problems for tenderness; in fact, the ratings for tenderness in lyrics and the congruent combination were better than for the music-only stimulus, and unambiguous in the combination, meaning that lyrics help tenderness to dominate and help the music to convey tenderness.

A bit of a problem might have been the combination of having only 5 Likert points to choose from when reporting perceived discrete emotions and the rather high effectiveness of most of the control stimuli for music-only and lyrics only, especially in the case of sadness. These were often rated close to the highest possible intensity with not much headroom for a combination. This might have lead to a ceiling effect and any effects might have been lowered in the discrete emotion scales. Still, because of a rather high number of participants, the analyses could still find significant effects, so the problem shouldn't be too large.

Considerations and Limitations, Conclusion While this research tried to combine the best approaches of previous studies on emotions in songs while avoiding their limitations, this study itself still leaves things open for consideration.

One of those things was the use of tenderness as an emotion. It was shown before (P. N. Juslin & Laukka, 2004) that tenderness is an emotion which is hard to differentiate from sadness and happiness. The same was found to be true in this study. Even though it is argued that tenderness is a better term of basic emotions in terms of music (Eerola & Vuoskoski, 2011) compared to something like disgust, it still does not seem to be perfect and choosing another emotion representing low-arousal positive-valence might be helpful for other studies. The next factor was the choice of irony as a term to rate. In the analyses, it might be considered as a non-emotion, but participants might have still changed their other emotion ratings because they could rate the stimuli on the term of irony. It is unclear which effects this might have. As irony ratings were overall pretty low, it can be argued that the effect might be low as well.

Choosing a different range for valence and arousal ratings between -10 and 10 compared to 1-5 with the discrete emotions is another thing to consider, but literature suggests that the effects of using finer scales are rather low for direct comparison (Matell & Jacoby, 1971; Dawes, 2008) and only have an impact on slightly higher mean ratings when using lower (5-7 Likert points) resolutions.

Doing an online survey with repeated measures for every participant and stimulus might also not be the perfect way; it does tackle the problem of differences between raters and make it comparable in this way and also lead to high statistical power, but this form might have led to order effects. The participants might have changed their ratings depending on whether they have heard the control stimuli or congruent matches before hearing the combination of incongruent emotions. The randomized design tried to avoid this pitfall and linear mixed models might also have help to reduce this effect, but it is still noteworthy. The external conditions of an online study are, as always, harder to control, but they do have the positive side of creating a more valid surrounding in which people would actually listen to music.

Participants were asked to only rate the perception of emotion, not felt emotion. For that reason it is not possible to tell, especially in the case of mixed emotions, which emotions were actually induced. Previous research has already focused on induced emotions (Fiveash & Luck, 2016; Hunter et al., 2008; Mori & Iwanaga, 2014), so it was important to find out more about perception of emotions in songs, even though it might be hard for the listener to differentiate between perceived and felt emotions. Furthermore it has been reported that the difference between

7. Discussion

perceived and induced emotions is mainly one of intensity and not valence (Hunter et al., 2010), and that the differences overall are not too high (Eerola & Vuoskoski, 2011; Evans & Schubert, 2008; Kallinen & Ravaja, 2006), so this shouldn't be too much of a problem in that regard. In connection with the personal traits of the participants that were found like ME-MS (Linnemann et al., 2018), STOMP score (Langmeyer et al., 2012), this might have had an unwanted effect though. Differences between the raters might have been minimized by phrasing the question about perceived emotions this way. In further studies on this topic it could be a better idea to let participants rate induced emotions when the ME-MS score is of importance.

Using only 4 songs and 4 lyrics as possible combinations is not perfect. To be able to better generalize the results, further stimuli and combinations are needed. In addition, the control stimuli used were already extremely good in conveying the specific emotions, especially in the case of sadness. This is on one hand a good thing, so most of these stimuli were not ambiguous and comparison of emotions by matching music and lyrics was meaningful. On the other hand, this effectiveness might have lead to ceiling effects, in the way that combinations of highly effective music-only and lyric-only stimuli might not have led to as much difference in the combined stimuli ratings, as they couldn't effectively get much higher. Significance in statistical tests was still found in all stimuli that had this issue, so the problem is marginal. The randomized design tried to avoid this a bit as well.

The tender stimuli, as stated before, were the most ambiguous. This might be due not just because of tenderness as an ambiguous emotion itself, but that the stimuli itself for this emotion were not written well enough. This problem was found in the pre-tests already and changes to melody and backing had been made, but potentially, these changes were not sufficient and further research should either use different emotion terms or write new stimuli for this emotion.

The study at hand did well at using control variables for both music and lyric, that were less artificial than in previous studies. The control stimulus for music still had a singer's voice, but it was made completely certain that noone can understand anything by using a random language generator. Making the lyric control stimulus also with a human sound and not having a mix of reading and listening made sure that no form of processing other than listening was involved (Mori & Iwanaga, 2014). The control stimuli and combined stimuli were also all rated by all participants and were thus comparable in terms of emotions, as they could be rated in intensity in comparison to the previously heard stimuli.

In future research in the field of music, lyrics and emotions, using an emotion

7. Discussion

scale that was intended for music emotions, such as the GEMS (Zentner et al., 2008; Lykartsis et al., 2013) might help avoid drawbacks of this study. It might also help investigate the effectiveness of the GEMS with modern pop music and lyrics. Incorporating more stimuli for each emotion is also advisable; possible fatigue of the participants should be avoided, as this study was already rather long with an average interview time of 28,9 min.

To directly build up on the findings of this study, incorporating a different, non-native language like English, that many people can still understand to some degree, would also be interesting. It can be argued that the possibility of understanding the lyrics on a native speakers level increases the importance of lyrics. This would also reflect the fact that the main part of modern popular music in Germany is written with English lyrics. It is also possible to ask for induced emotions and not perceived emotions, to find out more about listening styles and emotions in songs.

Another interesting investigation would be the interaction of lyrics and music in the genre of Jazz, as this genre is defined by standards that can and are interpreted both with singing voices or instruments and is in that way a perfect choice for research on music and lyrics.

As the results of this study that music and lyrics are mostly equal partners differed from previous research, it is important to undertake further investigations based on the approaches taken here. The new-found equal importance of lyrics on the valence, and the higher importance of music on the arousal changes the understanding of the interactions between music, lyrics and emotions in the context of song, which might have implications for other fields like emotion estimation in music databases, or the overall understanding of music and language. Although there are many similarities, this thesis also showed that the differences between discrete and dimensional emotion models should not be forgotten and that they can both contribute to our understanding of emotions in songs. The case of mixed emotions, as well as the quality of melancholy in songs make a case for further research on this topic.

References

- Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: are lyrics and melodies equal partners? *Psychology of Music*, 34(4), 511–534.
- Andrade, E. B., & Cohen, J. B. (2007). On the consumption of negative feelings. *Journal of Consumer Research*, 34(3), 283–300.
- Arbib, M. A. (Ed.). (2013). *Language, music, and the brain: A mysterious relationship*. Cambridge, Massachusetts: The MIT Press.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in cognitive sciences*, 6(6), 248–254.
- Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: implications for explaining autism. *Science*, 310(5749), 819–823.
- Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358(1430), 361–374.
- Baron-Cohen, S., & Wheelwright, 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.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bartoń, K. (2018). MuMIn: Multi-Model Inference [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MuMIn> (R package version 1.42.1)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Benigni, R. (1997). *La vita è bella (Life is Beautiful)*. Motion Picture. Italy: Miramax.
- Berrios, R., Totterdell, P., & Kellett, S. (2015, April). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in Psychology*, 6, 428.
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8),

- 1113 1139.
- Bonnell, A.-M., Fäy, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, 63(7), 1201–1213.
- Brady, E., & Haapala, A. (2003). Melancholy as an aesthetic emotion. *Contemporary Aesthetics*, 1(1).
- Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S. K., & Tervaniemi, M. (2011, December). A functional MRI study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology*, 2, 308.
- Chau, C.-j., Wu, B., & Horner, A. (2014). Timbre features and music emotion in plucked string, mallet percussion, and keyboard tones. In *ICMC* (pp. 982–989). Athens, Greece.
- Coffman, D. D., Gfeller, K., & Eckert, M. (1995). Effect of textual setting, training, and gender on emotional response to verbal and musical information. *Psychomusicology: A Journal of Research in Music Cognition*, 14(1-2), 117.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61–104.
- Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1), 18–49.
- Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, 30(3), 307–340.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4), 169–200.
- Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12(1), 75–99.
- Filik, R., Hunter, C. M., & Leuthold, H. (2015). When language gets emotional: Irony and the embodiment of affect in discourse. *Acta Psychologica*, 156, 114–125.
- Fischer, A. (2000). *Gender and emotion: Social psychological perspectives*. Cambridge University Press.
- Fischer, A. H., Kret, M. E., & Broekens, J. (2018). Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis. *PloS ONE*, 13(1), e0190712.
- Fiveash, A. (2014). *Effects of musical valence on the cognitive processing of lyrics* (Unpublished master’s thesis). University of Jyväskylä, Jyväskylä.

References

- Fiveash, A., & Luck, G. (2016). Effects of musical valence on the cognitive processing of lyrics. *Psychology of Music*, 44(6), 1346–1360.
- Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, 5(1), 123–147.
- Galizio, M., & Hendrick, C. (1972). Effect of musical accompaniment on attitude: The guitar as a prop for persuasion. *Journal of Applied Social Psychology*, 2(4), 350–359.
- Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception: An Interdisciplinary Journal*, 28(3), 279–296.
- Gfeller, K., & Coffman, D. D. (1991). An investigation of emotional response of trained musicians to verbal and music information. *Psychomusicology: A Journal of Research in Music Cognition*, 10(1), 31–48.
- Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, 128(3), 628–640.
- Greenberg, D. M., Baron-Cohen, S., Stillwell, D. J., Kosinski, M., & Rentfrow, P. J. (2015). Musical preferences are linked to cognitive styles. *PloS ONE*, 10(7), e0131151.
- Hammerstein, O., & Rodgers, R. (1945). *It might as well be spring*. Song.
- Haritaipan, L., & Mougenot, C. (2015, 11). An experimental study of tactile interactions in long-distance emotional communication. In *IASDR 2015* (pp. 511–534). Brisbane.
- Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, 22(2), 327–352.
- Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 4(1), 47.
- Juslin, P., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008, 11). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, 8, 668–683. doi: 10.1037/a0013505
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814.
- Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3), 217–238.

References

- Juslin, P. N., & Sloboda, J. (2011). *Handbook of music and emotion: Theory, research, applications* (1st ed.). Oxford University Press.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5), 559–575.
- Kallinen, K., & Ravaja, N. (2006). Emotion perceived and emotion felt: Same and different. *Musicae Scientiae*, 10(2), 191–213.
- Kamenetsky, S. B., Hill, D. S., & Trehub, S. E. (1997). Effect of tempo and dynamics on the perception of emotion in music. *Psychology of Music*, 25(2), 149–160.
- Koelsch, S., Gunter, T. C., Cramon, D. Y. v., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical “language-network” serves the processing of music. *Neuroimage*, 17(2), 956–966.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307.
- Kreutz, G., Schubert, E., & Mitchell, L. A. (2008). Cognitive styles of music listening. *Music Perception: An Interdisciplinary Journal*, 26(1), 57–73.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. doi: 10.18637/jss.v082.i13
- Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What do music preferences reveal about personality? *Journal of Individual Differences*, 30(2), 119–130. doi: 10.1027/1614-0001/a000082
- Larsen, J. T., & Green, J. D. (2013). Evidence for mixed feelings of happiness and sadness from brief moments in time. *Cognition & Emotion*, 27(8), 1469–1477.
- Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology*, 100(6), 1095–1110.
- Larsen, J. T., & McGraw, A. P. (2014). The case for mixed emotions. *Social and Personality Psychology Compass*, 8(6), 263–274.
- Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, 81(4), 684–696.
- Larsen, J. T., & Stastny, B. J. (2011). It’s a bittersweet symphony: Simultaneously mixed emotional responses to music with conflicting cues. *Emotion*, 11(6), 1469–1473.
- Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. *Language, Music, and the Brain: A Mysterious Relationship*, 257–272.
- Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The language, tone and

- prosody of emotions: neural substrates and dynamics of spoken-word emotion perception. *Frontiers in Neuroscience*, 10, 506. doi: 10.3389/fnins.2016.00506
- Linnemann, A., Kreutz, G., Gollwitzer, M., & Nater, U. M. (2018). Validation of the german version of the music-empathizing-music-systemizing (mems) inventory (short version). *Frontiers in Behavioral Neuroscience*, 12, 153. doi: 10.3389/fnbeh.2018.00153
- Lykartsis, A., Pysiewicz, A., von Coler, H., & Lepa, S. (2013). The emotionality of sonic events: Testing the Geneva Emotional Music Scale (GEMS) for popular and electroacoustic music. In G. Luck & O. Brabant (Eds.), *The 3rd international conference on music & emotion (icme3)*. Jyväskylä, Finland,.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31(3), 657-674.
- Mori, K. (2009). The influence of the meaning of lyrics on the expressed emotion of music valence. In L. Naveda (Ed.), *Proceedings of the second international conference of students of systematic musicology* (pp. 53-58). Ghent, Belgium.
- Mori, K., & Iwanaga, M. (2014). Pleasure generated by sadness: Effect of sad lyrics on the emotions induced by happy music. *Psychology of Music*, 42(5), 643-652.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674.
- Patel, A. D. (2010). *Music, language, and the brain*. New York: Oxford university press.
- Patel, A. D. (2012). Language, music, and the brain: a resource-sharing framework. *Language and Music as Cognitive Systems*, 204-223.
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715-734.
- R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- 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(6), 1236-1256. doi: 10.1037/0022-3514.84.6.1236
- Robazza, C., Macaluso, C., & D'urso, V. (1994). Emotional reactions to music by gender, age, and expertise. *Perceptual and Motor Skills*, 79(2), 939-944.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. doi: 10.1037/h0077714

References

- Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, 33(3), 239–251.
- Scherer, K. R. (2013). Emotion in action, interaction, music, and speech. In M. A. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship* (pp. 107–139). The MIT Press Cambridge; London.
- Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (p. 361–392). Oxford: Oxford University Press.
- Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition & Emotion*, 15(1), 81–97.
- Schimmack, U. (2005). Response latencies of pleasure and displeasure ratings: Further evidence for mixed feelings. *Cognition & Emotion*, 19(5), 671–691. doi: 10.1080/02699930541000020
- Scriboly. (2019). *Scriboly - fake language generator*. Retrieved 20.02.2019, from <http://www.scriboly.com/Generators/Language>
- Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, 25(2), 123–135.
- Sloboda, J. A. (2010). Music in everyday life: The role of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 493–515). Oxford: Oxford University Press.
- Sousou, S. D. (1997). Effects of melody and lyrics on mood and memory. *Perceptual and Motor Skills*, 85(1), 31–40.
- Steele, B. J. (2010). Irony, emotions and critical distance. *Millennium*, 39(1), 89–107.
- Stratton, V. N., & Zalanowski, A. H. (1994). Affective impact of music vs. lyrics. *Empirical Studies of the Arts*, 12(2), 173–184.
- Thayer, R. E. (1990). *The biopsychology of mood and arousal*. New York: Oxford University Press.
- Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, 22(4), 720–752.
- Vuoskoski, J. K., & Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), 204. doi: 10.1037/a0026937
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J.,

References

- Fine, D., . . . Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences*, 41(5), 929–940.
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494–521. doi: 10.1037/1528-3542.8.4.494

A. Music and lyrics

A.1. Happy

Berg und Täler fangen an zu blühen
auch die Sonne hat heut keine Mühen
niemals fühlte ich mich so wie heute
voller Lust und voller Lebensfreude

Ich möchte lachen, tanzen und singen
wie ein irrer durch die Gegend springen

A.1.1. Gibberish for Happy

Gu guyu Deiyugu lola lo gu
megure Fayudei magu yute
redei guyu logu logu la re
yagu Gufa lagu Ya lukaku

tsolo fayo lorelotei
yote fatu ketei loguyago

A.2. Sad

In meinem Herzen ist ein Grab voll Schmerzen
es behütet und bedeckt dich nun
der graue Regen löscht die letzten Kerzen
In meinen Tränen wirst du immer ruhn

Meinen Kummer kann ich nicht verbergen
Wie die Sonne hinter schweren Wolken

A.2.1. Gibberish for Sad

Lo gugu sore, yulo gu teeso
deila Gula yu guyo tesulo
mayu Meegulado, dei yuladeo
go ma degu, yu yugu latela

Yugu meela, Ua de i gula
guyu mela, gu ta me tesogu

A.3. Angry

Wenn ich dich mit ihm seh werd ich rasend
kann die Abscheu in mir kaum ertragen
Dachtest du echt, ich könnte damit Leben?
Dachtest du echt, ich würd dir einfach vergeben?

Dieses Feuer verbrennt mich von innen
Kann den Flammen in mir kaum entrinnen

A.3.1. Gibberish for Angry

Tee lalei o lei a yu lei mate
he fa Makireli o ga le a
leiyudei a yulei, gula tsoyu
leiyudei a yulei, magu yoli

fa Ole i legutee lalei
he tee Lamatee ga la gu tee

A.4. Tender

Sanft und sicher hältst du meine Hände
Spür einen Anfang und noch lang kein Ende
Bist mir so vertraut, sind Teil eines Ganzen
wenn wir langsam miteinander tanzen

Weiche Küsse, dich zärtlich berühren
Wünschte ich könnt dich für immer spüren

A.4.1. Gibberish for Tender

He li o ta madei li o Heyu
keli yugate Hemaray guya
li o la gu Fa yu ga Lei Mara
li o ta tee guyu mee fa gula

yore Lio, yu la e malagu
soyu magu to fagu loma

B. Distribution of ratings

B.1. Ratings for the four basic emotions in all combined stimuli

Note: Ratings for melancholy and irony can be found in the main text in the section on mixed emotions

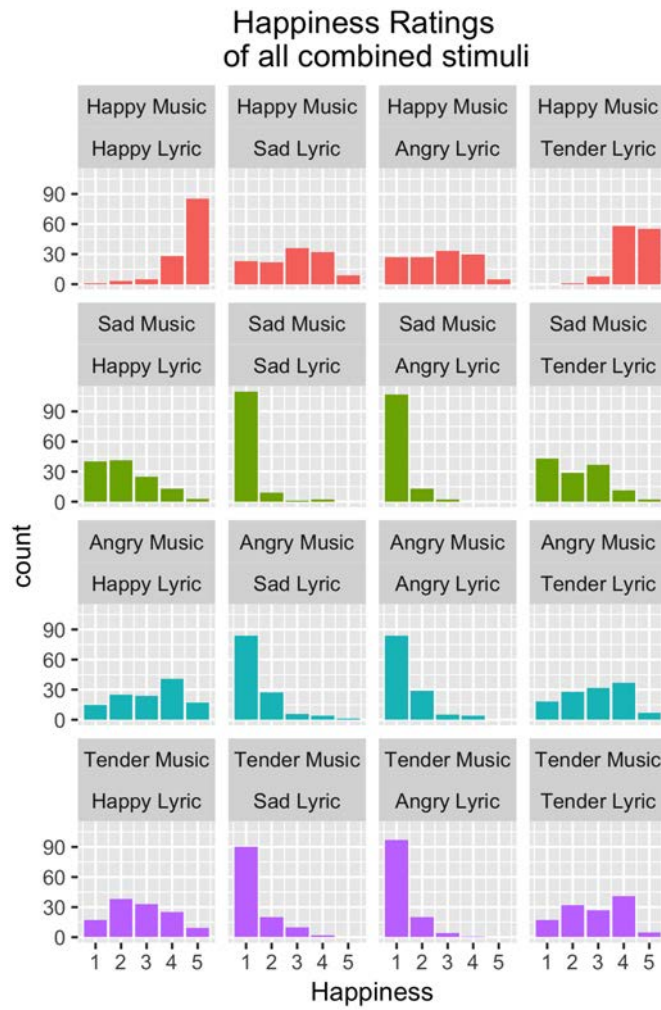


Figure B.1.: Ratings for happiness in all combined stimuli

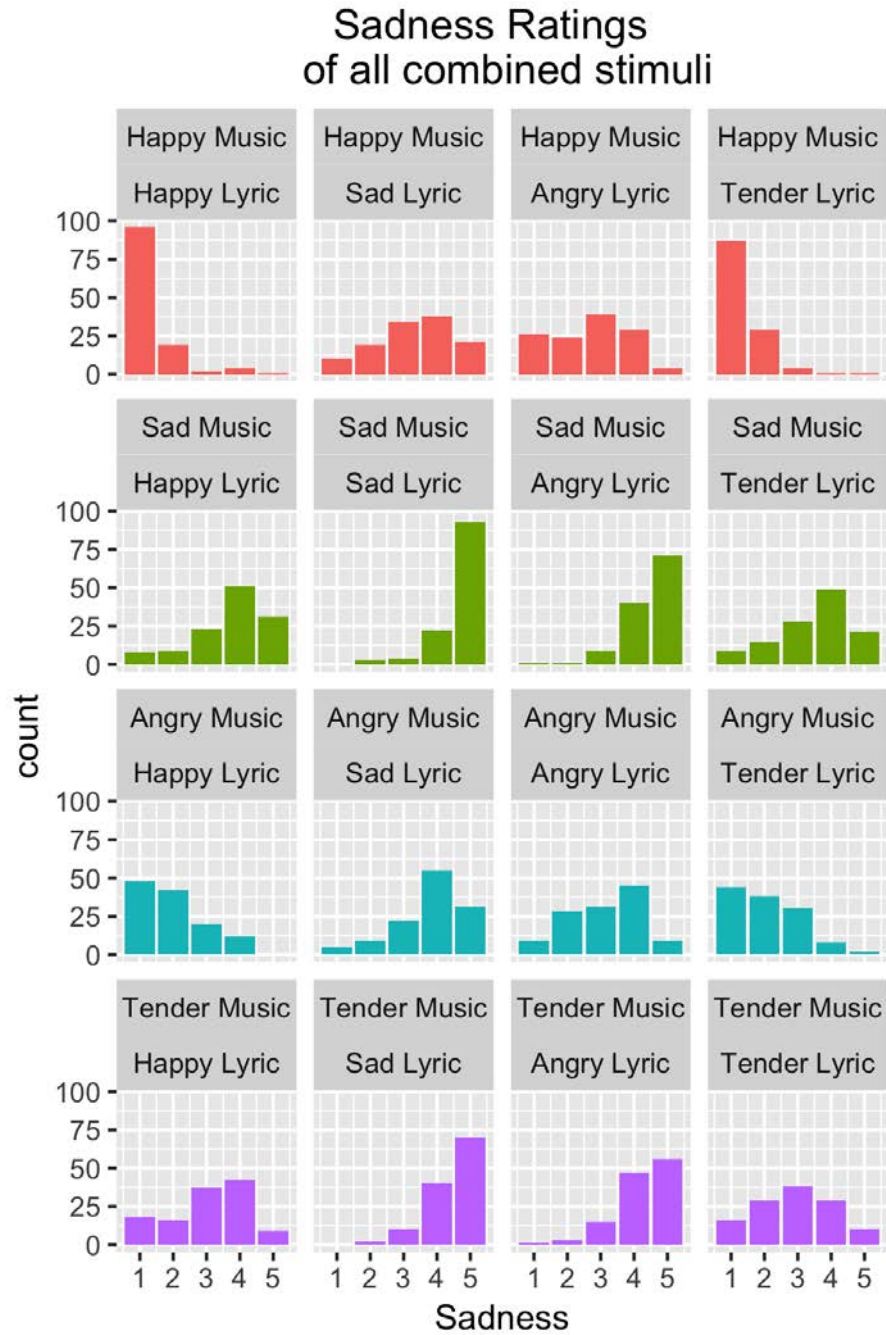


Figure B.2.: Ratings for sadness in all combined stimuli

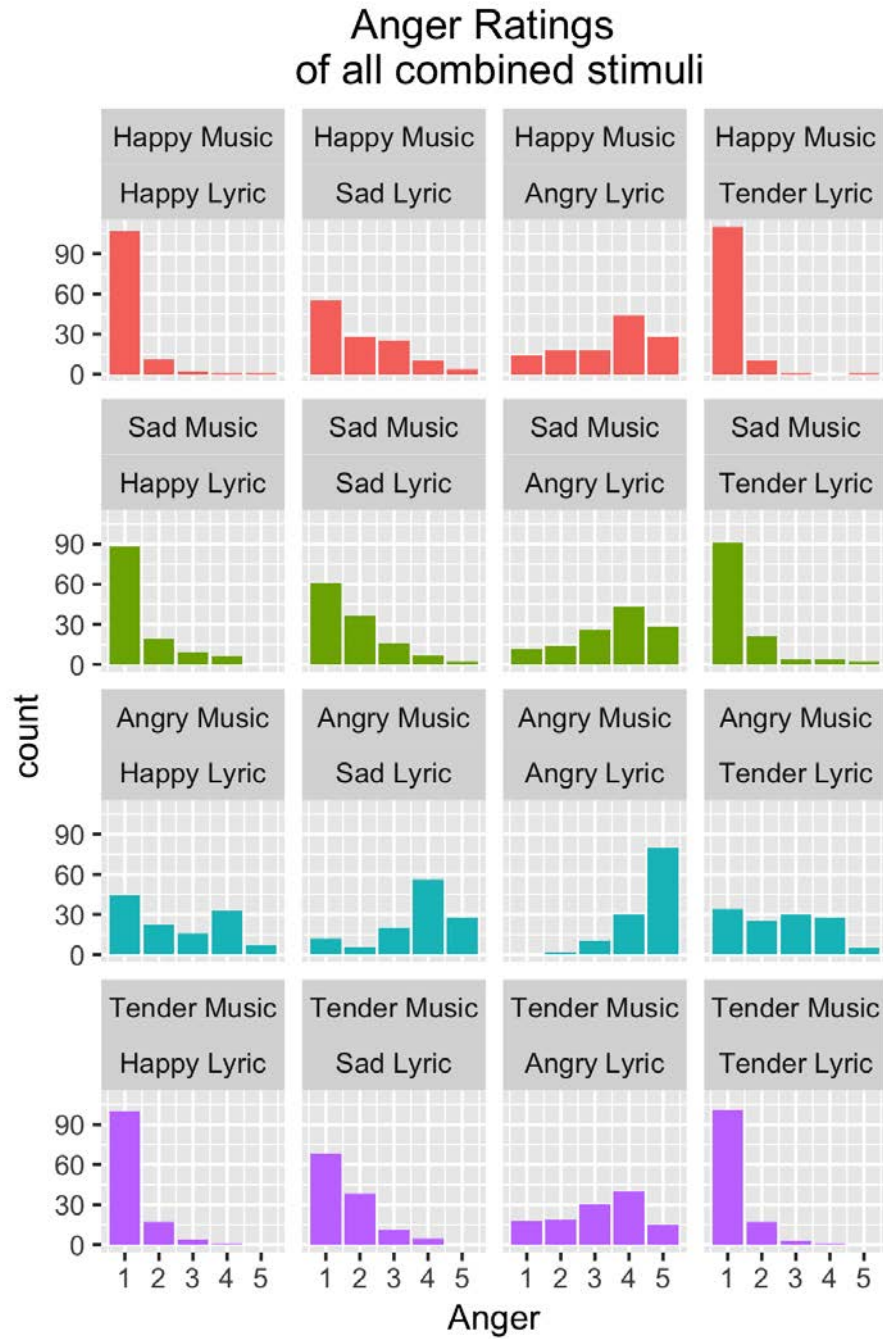


Figure B.3.: Ratings for anger in all combined stimuli

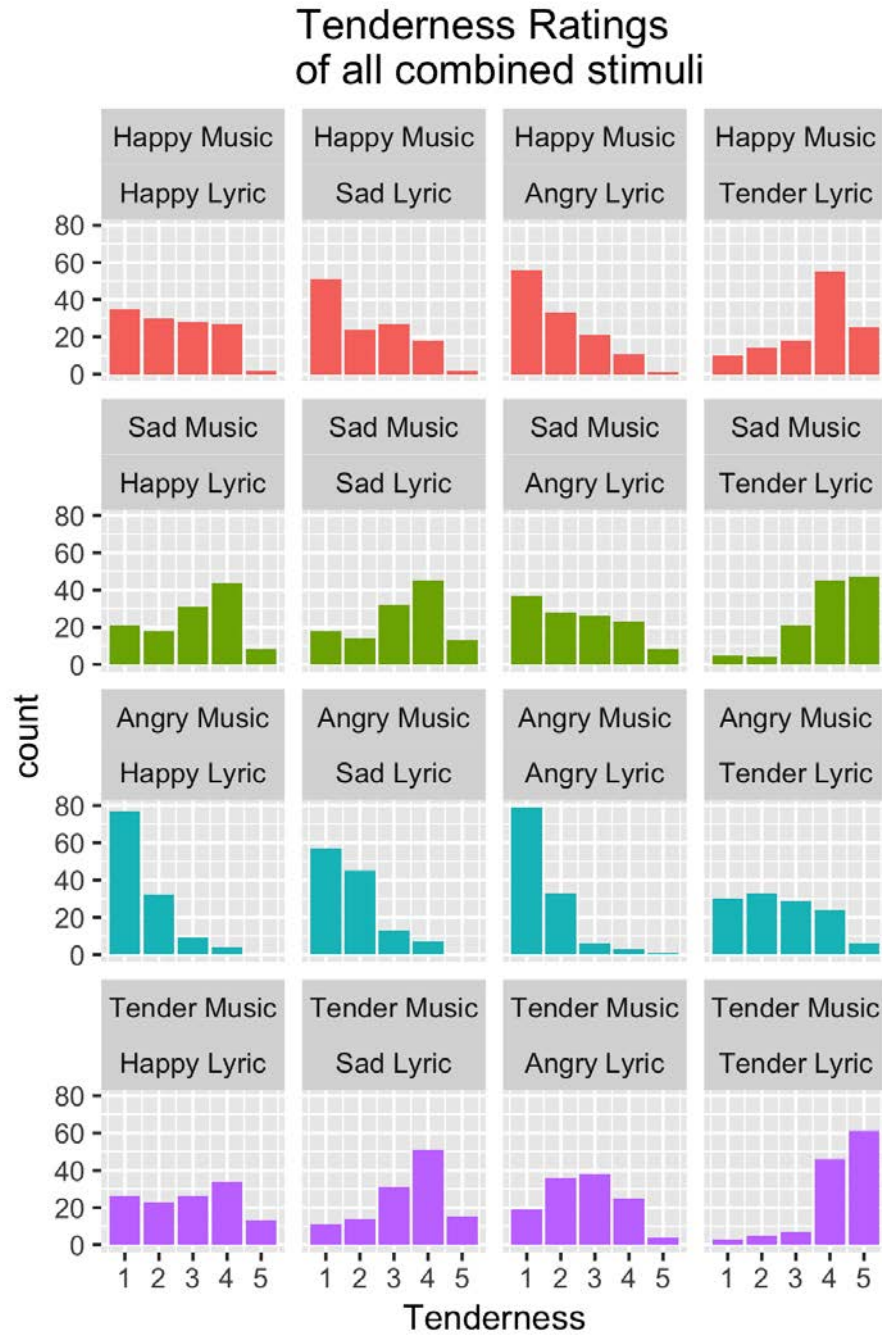


Figure B.4.: Ratings for tenderness in all combined stimuli

B.2. Ratings for congruent combinations and control stimuli

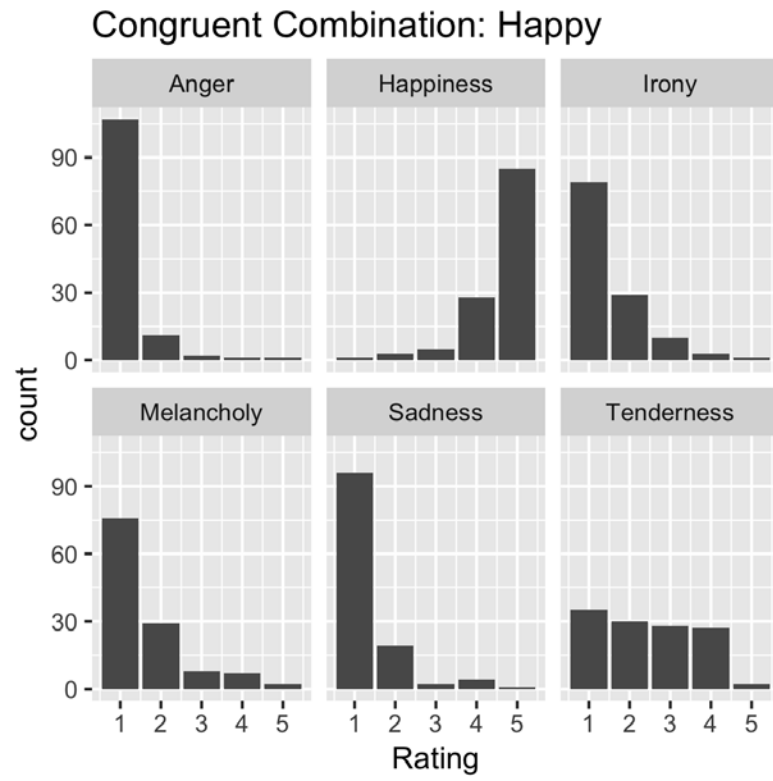


Figure B.5.: Ratings for Congruent Combination: Happy

B. Distribution of ratings

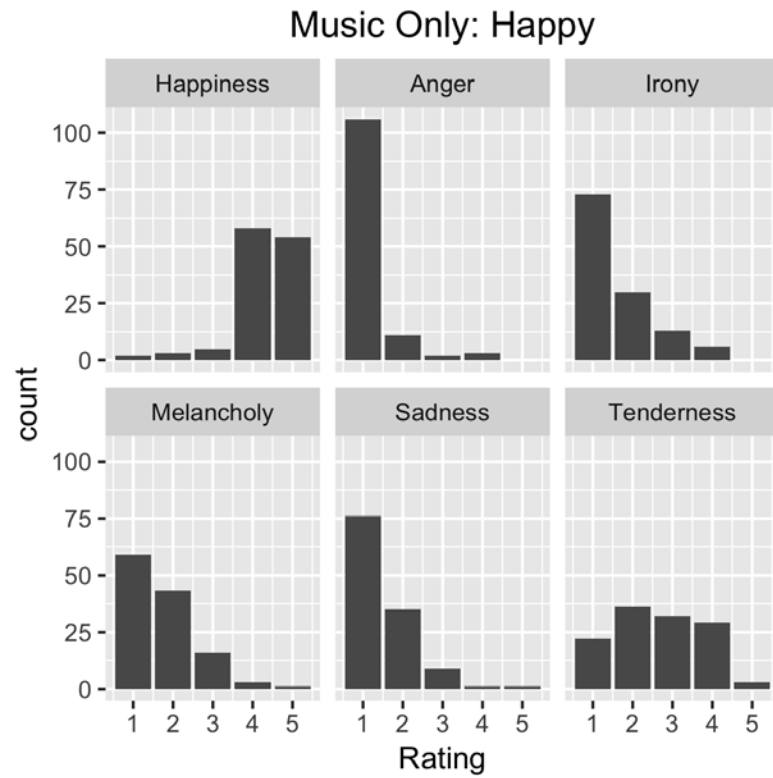


Figure B.6.: Ratings for Music Only: Happy

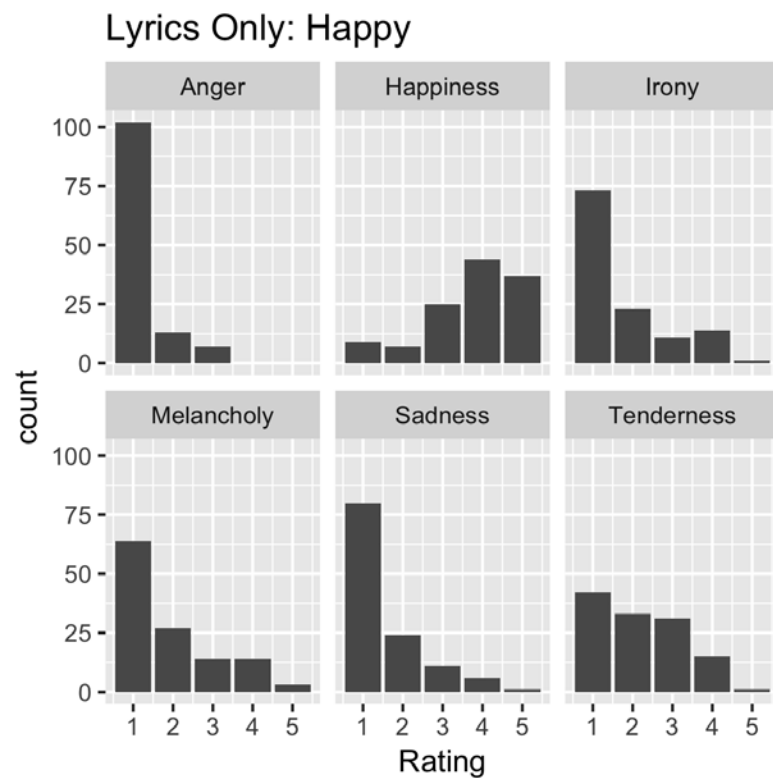


Figure B.7.: Ratings for Lyrics Only: Happy

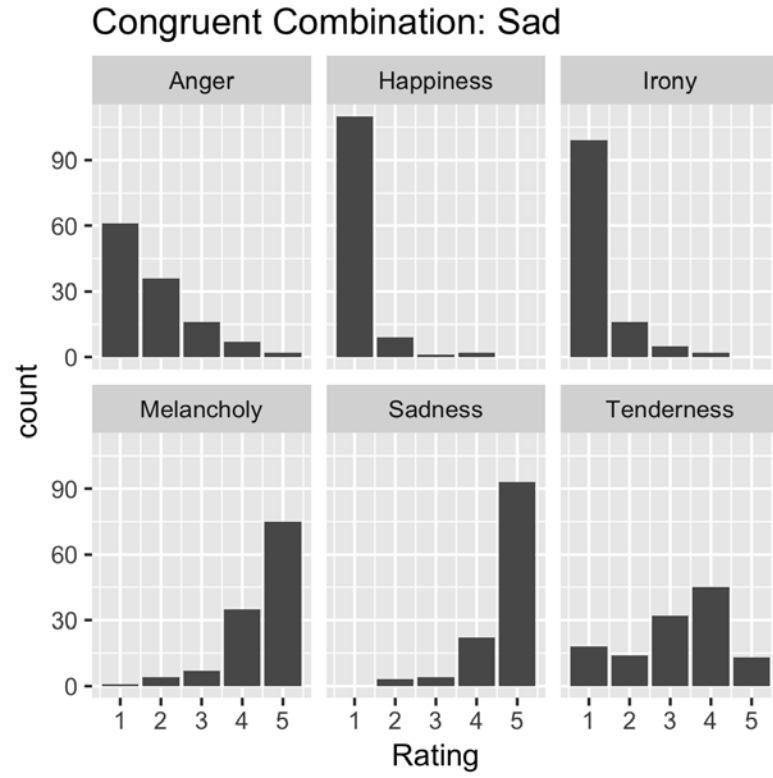


Figure B.8.: Ratings for Congruent Combination: Sad

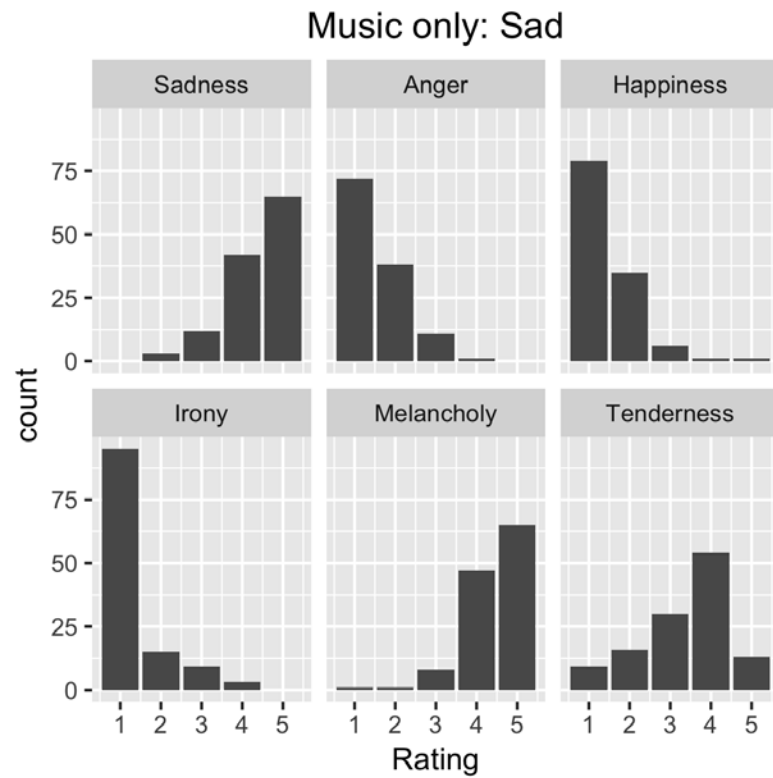


Figure B.9.: Ratings for Music Only: Sad

B. Distribution of ratings

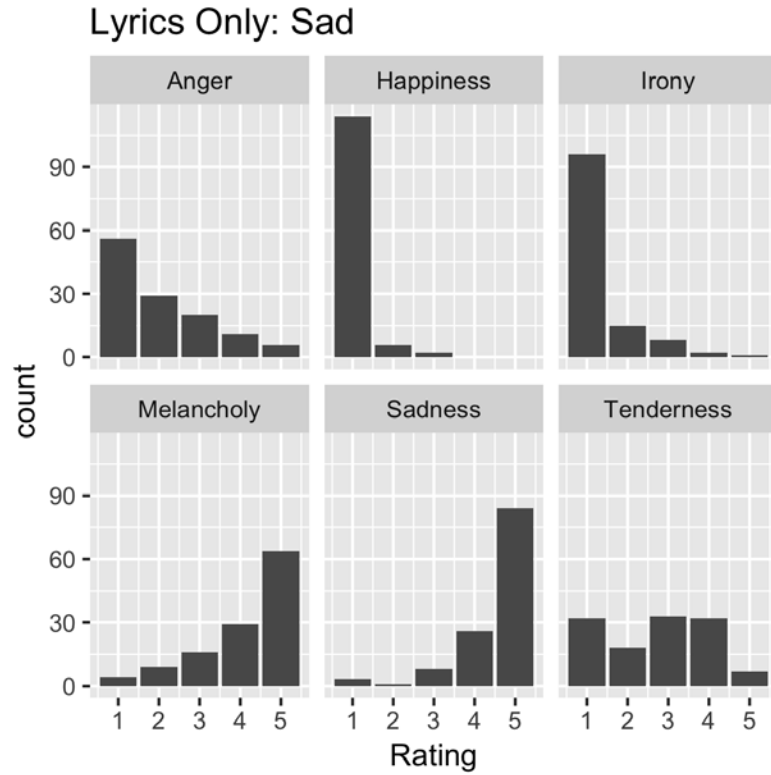


Figure B.10.: Ratings for Lyrics Only: Sad

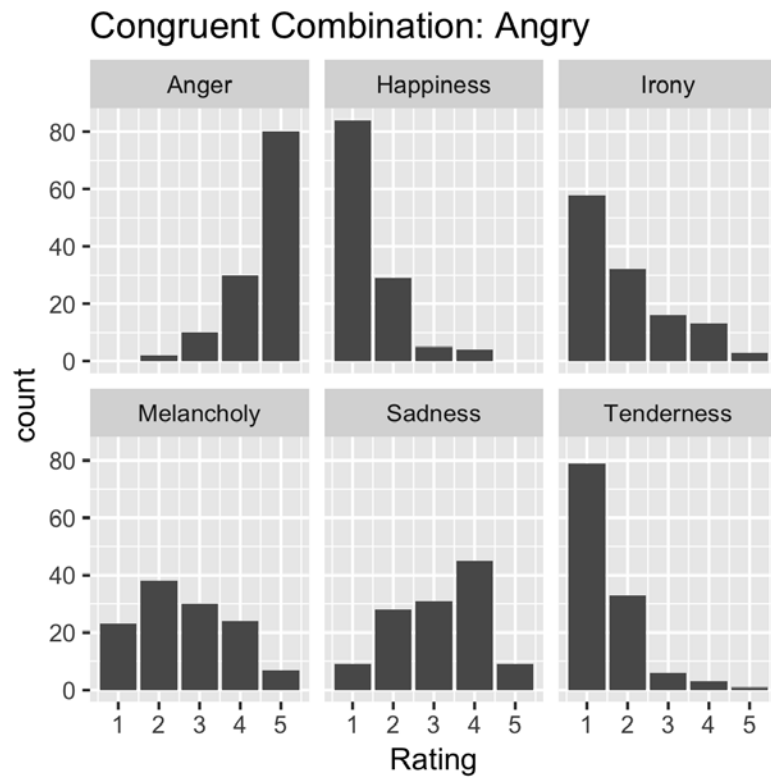


Figure B.11.: Ratings for Congruent Combination: Angry

B. Distribution of ratings

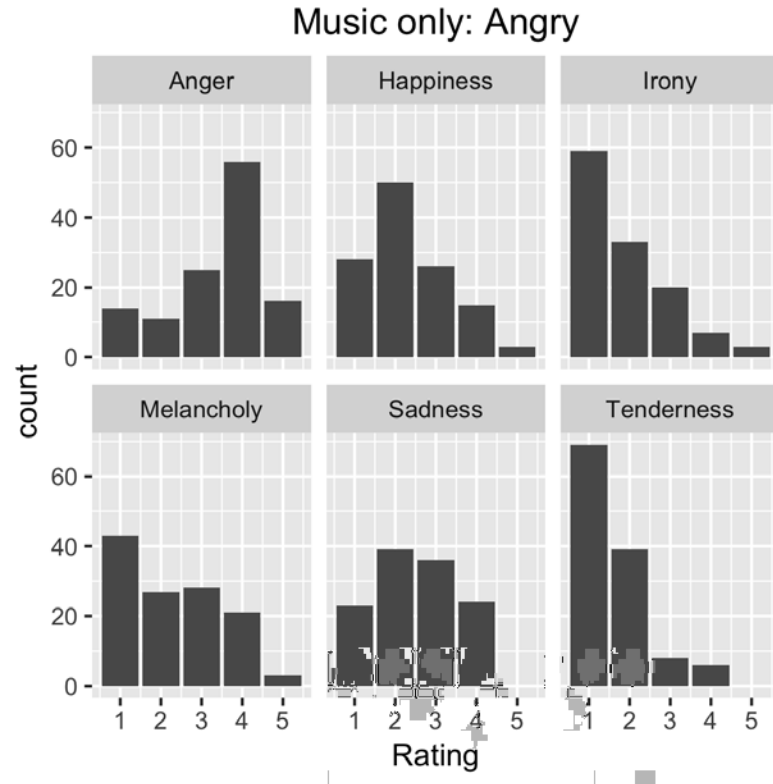


Figure B.12.: Ratings for Music Only: Angry

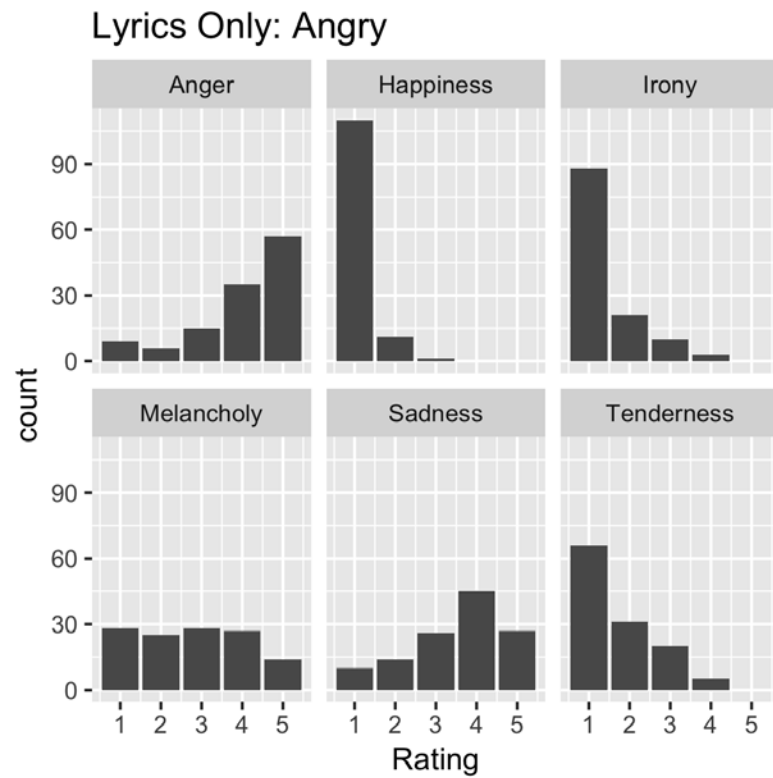


Figure B.13.: Ratings for Lyrics Only: Angry

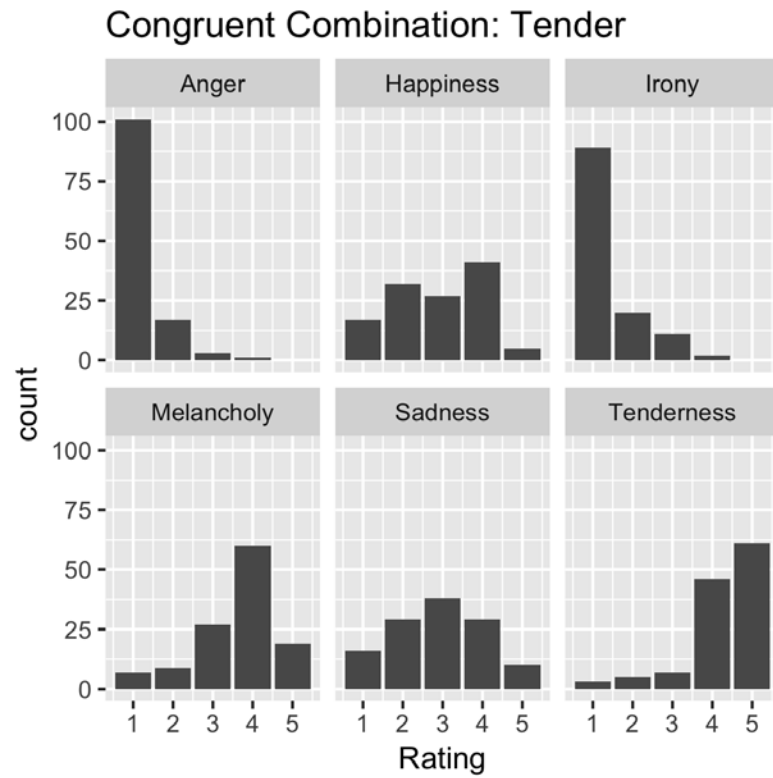


Figure B.14.: Ratings for Congruent Combination: Tender

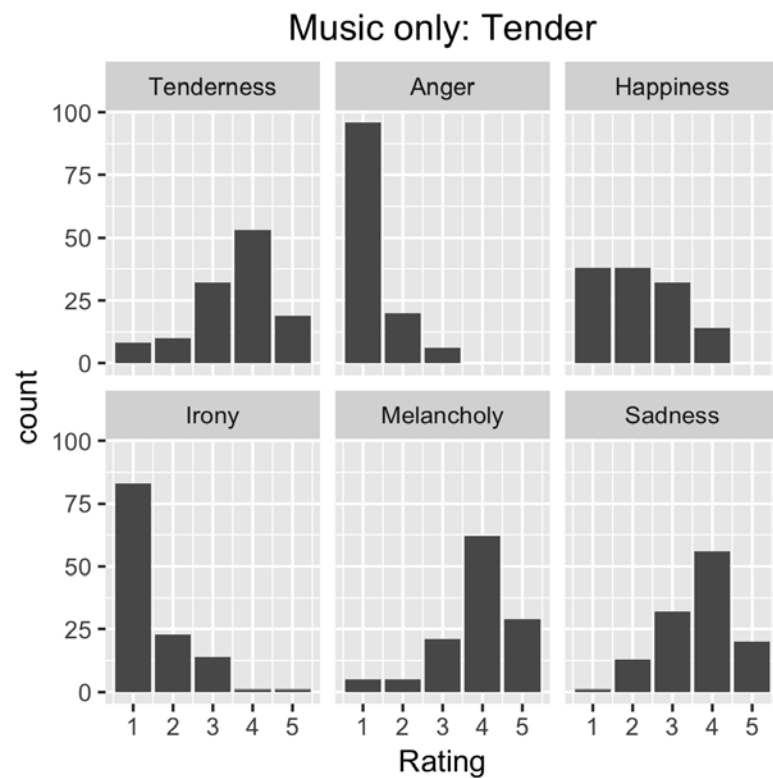


Figure B.15.: Ratings for Music Only: Tender

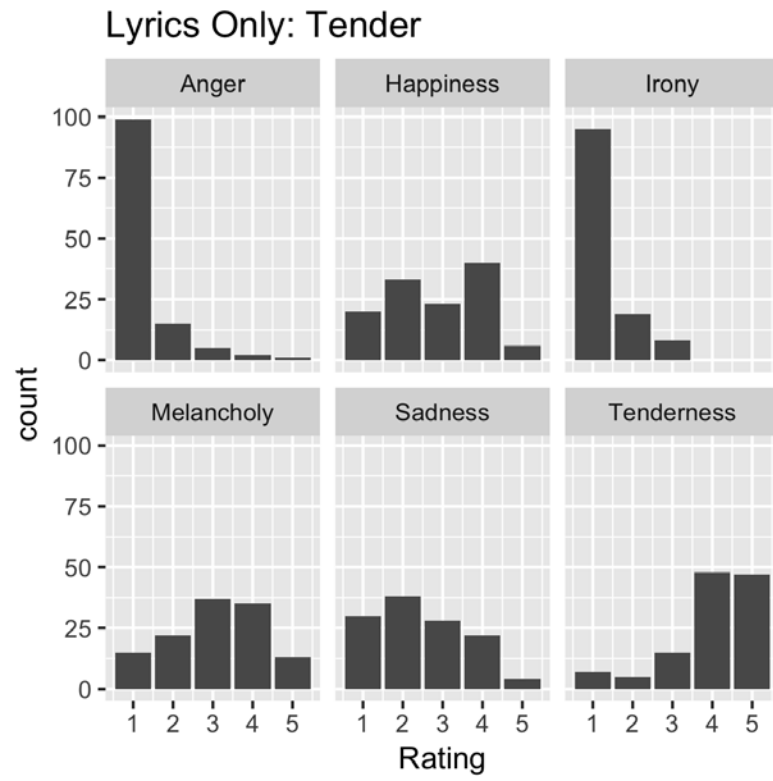


Figure B.16.: Ratings for Lyrics Only: Tender

C. STOMP: Median Split Differences

| Emotion | Parameter | Full data | L: Upbeat | DL: Upbeat | L: Reflective | DL: Reflective |
|------------|-----------|-----------|-----------|------------|---------------|----------------|
| Happiness | Intercept | -0.079 | -0.0298 | -0.1195 | -0.1096 | -0.0626 |
| | Song | 0.546 | 0.5188 | 0.5674 | 0.5418 | 0.5481 |
| | Lyric | 0.485 | 0.4778 | 0.4925 | 0.4859 | 0.4844 |
| Sadness | Intercept | 0.291 | 0.2987 | 0.2761 | 0.1714 | 0.3448 |
| | Song | 0.526 | 0.4976 | 0.5535 | 0.5545 | 0.5126 |
| | Lyric | 0.430 | 0.4536 | 0.4087 | 0.4276 | 0.4334 |
| Anger | Intercept | 0.112 | 0.1061 | 0.1075 | 0.0233 | 0.1633 |
| | Song | 0.501 | 0.5288 | 0.4787 | 0.5331 | 0.4817 |
| | Lyric | 0.587 | 0.5658 | 0.6056 | 0.5819 | 0.5891 |
| Tenderness | Intercept | 0.279 | 0.2871 | 0.2646 | 0.0967 | 0.3690 |
| | Song | 0.481 | 0.4950 | 0.4712 | 0.5195 | 0.4611 |
| | Lyric | 0.406 | 0.3944 | 0.4167 | 0.4235 | 0.3992 |

Table C.1.: Median split of the data set by liking or disliking a genre: Differences in the basic LMER models (L =Like; DL = Dislike)

D. Digital Resources