An Investigation of Affective Timbres: Considering Affect Locus, Experimental Context, and Individual Differences

Iza Ray Korsmit

Music Technology Area, Department of Music Research,
Schulich School of Music, McGill University, Montreal, Canada

April 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Table of Contents

Abstract

This thesis aims to further our understanding of music's capacity to communicate and induce affect. Here, the focus lies on the musical feature of timbre, as it varies with orchestral instrument family and pitch height. As a holistic approach to understanding musical affect, both the experimental context as well as individual characteristics of the music listener are considered. Two online experiments collected ratings of perceived and induced affect (referred to as affect locus) in response to single notes (Exp. 1) and chromatic scales (Exp. 2), as well as questionnaire data on the individual differences of the participants (pre-existing mood, empathy, Big-Five personality, musical sophistication, and musical preferences). These two experiments were designed to answer the three main research questions. What is the most appropriate method for quantifying musical affect in this experimental context? What are the effects of instrument family, pitch register, and affect locus on the affective response to short sounds? And which timbre features describing the acoustic properties of a sound predict affect?

To answer the first research question, self-reported ratings on three-dimensional affect (valence, tension arousal, and energy arousal) were compared to discrete affect ratings (anger, fear, sadness, happiness, and tenderness) based on scale consistency, dimension reduction, and regression analyses. Correlation analysis investigated the relevance of individual differences. The results show that two dimensions or discrete categories are sufficient for the quantification of affect in response to short sounds. Furthermore, energy arousal captures affective variation that is not captured by any of the discrete affect categories. All sources of individual differences, in particular pre-existing mood, are moderately correlated with the affect measurements, particularly valence.

To answer the second research question, polynomial mixed-effects analysis investigated the effects of pitch register, instrument family, and affect locus on each of the affect ratings. The role of individual differences was explored as they moderate the effects of the preceding models. Pitch register shows a U-shaped effect on most affect scales, although it is linearly related to energy arousal and sadness. The percussion family is considered the least tense, scary, sad, or angry, and the most positively valenced, happy, and tender. Any affects that one may consider unpleasant (negative valence, tension, or sadness) are more strongly perceived than induced, particularly in response to chromatic scales. Finally, instrument family is most susceptible to influence from individual differences and pitch register is least susceptible.

To answer the third research question, several acoustic descriptors from the Timbre Toolbox are used to predict the affect ratings of the current experiments, as well as to re-analyze two previously published experiments by Eerola et al. (2012, *Mus. Percept.*) and McAdams et al. (2017, *Front. Psychol.*). The analyses reveal that the emergence of the fundamental frequency and attack components play a prominent role in perceived and induced affect across experiments. Whether a stimulus set contains variation in pitch register or attack quality also determines whether pitch-related or temporal timbre descriptors predict the affect ratings.

These findings further contribute to our understanding of affect quantification and the role of pitch register, instrument family, and timbre descriptors in the perception and induction of musical affect. Divergence from previous research argues for careful consideration of the stimulus selection, as this may influence one's findings. Consistencies with previous research, however, may encourage future studies to conduct their experiments online, to reach a more diverse and representative participant population, while also considering the characteristics of the individuals in that population.

Résumé

Cette thèse étudie la capacité de la musique à communiquer et à induire de l'affect, surtout le timbre musical, qui varie en fonction de la famille d'instruments d'orchestre et du registre de hauteur du son. Afin de comprendre l'affect musical, le contexte de l'expérience et les caractéristiques individuelles de l'auditeur sont pris en compte. Deux expériences en ligne ont recueilli des évaluations de l'affect perçu et induit (appelé locus d'affect) en réponse à des notes uniques (Exp. 1) et à des échelles chromatiques (Exp. 2), ainsi que des données sur les différences individuelles des participants (humeur préexistante, empathie, personnalité Big-Five, sophistication musicale, et préférences musicales). Nous posons trois questions. Quelle est la méthode la plus appropriée pour quantifier l'affect musical dans ce contexte expérimental ? Quels sont les effets de la famille d'instruments, du registre de hauteur et du locus de l'affect sur la réponse affective à des sons courts ? Et quelles propriétés acoustiques d'un son prédisent l'affect ?

Pour la question 1, les évaluations autodéclarées de l'affect tridimensionnel (valence émotionnelle, tension et énergie) ont été comparé aux évaluations de l'affect discret (colère, peur, tristesse, bonheur et tendresse) sur la base de la cohérence de l'échelle, de la réduction des dimensions et des analyses de régression. Une analyse de corrélation a permis d'étudier la pertinence des différences individuelles. Les résultats montrent que deux dimensions ou catégories discrètes suffisent à quantifier l'affect en réponse à des sons brefs. En outre, l'énergie permet de saisir des variations affectives qui ne sont prises en compte par aucune des catégories discrètes. Toutes les sources de différences individuelles, en particulier l'humeur préexistante, sont corrélées modérément avec les mesures de l'affect, surtout la valence.

Pour la question 2, une analyse polynomiale à effets mixtes a étudié les effets du registre de hauteur, de la famille d'instruments et du locus de l'affect sur chacune des évaluations de l'affect. Le rôle des différences individuelles a été exploré dans la mesure où elles modèrent les effets des modèles précédents. Le registre de hauteur montre un effet en forme de U sur la plupart des échelles d'affect, bien qu'il soit linéaire pour l'énergie et la tristesse. La famille des percussions est considérée comme la moins tendue, effrayante, triste ou coléreuse, et la plus positivement valencée, heureuse et tendre. Valence négative, tension et tristesse sont plus fortement perçus qu'induits, en particulier en réponse aux gammes chromatiques. Enfin, la famille d'instruments est

la plus susceptible d'être influencée par les différences individuelles et le registre de hauteur est le moins susceptible de l'être.

Pour la question 3, plusieurs descripteurs acoustiques de la Timbre Toolbox ont été utilisés pour prédire les évaluations d'affect, ainsi que pour réanalyser deux autres expériences par Eerola et al. (2012, *Mus. Percept.*) et McAdams et al. (2017, *Front. Psychol.*). L'émergence de la fréquence fondamentale et les composantes de l'attaque jouent un rôle prépondérant dans l'affect perçu et induit d'une expérience à l'autre. Le fait qu'un ensemble de stimuli contienne des variations du registre ou de la qualité de l'attaque détermine également si les descripteurs de timbre temporels ou ceux liés à la hauteur prédisent les évaluations.

Ces résultats témoignent du rôle du registre, de la famille d'instruments et des descripteurs de timbre dans la perception et l'induction de l'affect musical. Les divergences par rapport aux recherches antérieures plaident en faveur d'un examen attentif du choix des stimuli, qui peut influer sur les résultats. Les concordances avec les recherches antérieures sont encourageantes pour les expériences en ligne, afin d'atteindre des participants plus diversifiés et plus représentatifs, tout en tenant compte de leurs caractéristiques individuelles.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor, Stephen McAdams. We had only spoken in person once before I joined the Music Perception and Cognition Lab (MPCL), so I started this PhD journey with quite a few unknowns. I consider myself very lucky that Stephen turned out to be such a great supervisor. He gave me a lot of freedom but also gave me a little push when I needed one. He was always available for feedback and questions – I am not sure how he keeps on top of responding to all his students with so little delay. When any difficulties arose, he would stick up for me and make sure everything turned out okay. Thank you, Stephen, for everything!

Next, I would like to thank all the other MPCL members, past and current. I missed you all, this past year! It is very stimulating to be part of such a truly interdisciplinary research group, and I have learned so many new things from everyone. Special thanks to Eddy (Savvas) Kazazis for working on a CIRMMT research project on misophonia together, and persevering with me throughout the completion of this project. Also special thanks to Bennett K. Smith, who stood by my side throughout the online experimentation and Prolific adventure, making sure everything ran smoothly, but who more importantly was a constant source of entertaining and thoughtful conversation. Lena Heng, also thanks to you – I am lucky to have you as my buddy to simultaneously go through all the stages of this PhD with, where we can count on each other for questions or help, or to play boardgames and go to the brasserie!

Marcel Montrey, although not technically an MPCL member, has been extremely helpful to several lab members, including myself. Thank you, Marcel, for your integral role in the analysis of my experiments, and your continuous support and feedback during the completion of this dissertation. Thank you to Philippe Macnab-Séguin, who helped with the design of the second experiment by creating the stimuli for Experiment 2 in OrchSim.

Now if we go back in time, about ten years ago, I took a course by Henkjan Honing on music cognition, which ignited the spark that led me on this academic journey. His passion for music cognition was so inspiring, and I am grateful for being able to work with him for several years and still being able to count on his support now. Similarly, I am so grateful for having been able to work with Ashley Burgoyne throughout my time at the University of Amsterdam, and to still have his support to this day. Both Henkjan and Ashley have been of massive help in getting accepted for this PhD position at McGill University.

I am pleased to acknowledge the financial support from the Maurice Pollack Foundation, the Prins Bernhard Cultuurfonds, and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Schulich School of Music also supported me with funding for several years and awarded me with the Differential Fee Waiver. The research was further supported by funding from the Social Sciences and Humanities Reseach Council (SSHRC) and the National Sciences and Engineering Council (NSERC) awarded to Stephen McAdams, and the Fonds de Recherche du Québec (FRQ) awarded to both Stephen McAdams and Robert Hasegawa.

Finally, thank you to all my friends and family back home. Thank you to my parents, Ellen de Bruijne and George Korsmit, and bonus parents Saskia Janssen and Klaas Kloosterboer, for believing in me and showing me how to work hard for something that you are passionate about; my siblings, David, Naomi, and Stijn Kloosterboer, and Clarissa Kölgen, for all being such great examples of cool and kind human beings; and my dearest friends Elin Beckmans, Lydia Kooistra, and Kelsey Onderdijk (in alphabetical order!) for all the good times. But most importantly, thank you to the love of my life and best friend of all time, Valentijn Nieman. Thank you for joining me on this crazy Canadian adventure, for all the pep talks during this last writing period, and for regularly making me roll on the floor with laughter. I could not have done this without you, and I cannot wait to see what is in store for us next.

Contributions of Authors

All chapters for this manuscript were written by me. Chapter 3–5 will be prepared for submission with co-authorship from Marcel Montrey, Alix Yok Tin Wong-Min, and Stephen McAdams:

I. Korsmit, I. R., Montrey, M., Wong-Min, A. Y. K., & McAdams, S. (in preparation). Representing the perceived and induced experience of musical affect: Dimensional versus discrete.

II. Korsmit, I. R., Montrey, M., Wong-Min, A. Y. K., & McAdams, S. (in preparation). The effects of instrument family and pitch register on perceived and induced affect.

III. Korsmit, I. R., Montrey, M., Wong-Min, A. Y. K., & McAdams, S. (in preparation). The acoustic origins of affective timbre.

The background research, conception and design of the experiments, and the analysis and interpretation of the data were all completed by me. My supervisor Stephen McAdams provided guidance and feedback throughout all the stages of the dissertation research. Marcel Montrey helped develop the statistical approach for Chapters 3–5, wrote custom R scripts for the analysis, and gave feedback for the interpretation of the data. Alix Yok Tin Wong-Min was an undergraduate student who collaborated on the development, execution, and analysis of the two experiments for her thesis project.

List of Figures

List of Tables

Chapter 1
General Introduction

The goal of this thesis is to further our understanding of music's capacity to communicate and induce affect. More specifically, the studies investigate how the musical feature of timbre, as it varies with instrument family and pitch register, influences perceived and induced musical affect, as it is quantified with different methods and moderated by individual characteristics of the music listener. The research in this thesis is divided into three main research questions:

1. What is the most appropriate method for the quantification of perceived and induced affect in response to affectively ambiguous and relatively short musical sounds, and how is this related to individual differences?
2. What are the effects of instrument family, pitch register, and affect locus (i.e., perceived/induced affect) on the affective response to those musical sounds, and how are these effects moderated by individual differences?
3. Which acoustic properties that describe the timbre of a sound predict the perceived and induced affective response to the same musical sounds?

Two experiments were conducted, which were both used to answer all three research questions. Although Chapters 3–5 deal with each of the three research questions individually and are each meant to serve as articles to be submitted to peer-reviewed journals, in this thesis the methods are described separately in Chapter 2, to prevent unnecessary repetition in the succeeding chapters. Chapters 3–5 also each include their own introductions that describe the research background and motivations for the research questions. Thus, to further avoid unnecessary repetition, this general introduction does not describe all the relevant background to the research questions but will discuss four of the main topics in this thesis: affect, timbre, individual differences, and online experimentation, followed by an overview of the succeeding chapters.

1.1 Affect Representation & Quantification

Affect is an umbrella term that is used throughout this thesis to encompass emotions, moods, and other evaluative or valenced states. Emotions are generally considered to be more intense and shorter-term affective responses than moods, which are less intense and longer-term. Emotions also comprise a form of appraisal, which urges one to evaluate the affective state and assess what external or internal object (such as music) is the cause of that affective state, whereas lower-intensity mood changes do not warrant appraisal (Juslin & Västfjäll, 2008). Without arguing whether music can communicate or induce ‘real-life’, or ‘basic’, emotions at all, the overarching term of affect is used here to accommodate the fact that music could also lead to more minor perturbations in mood instead of emotions, especially in an experimental context with relatively short sounds. Regardless of whether an emotion, mood, or other valenced state is engendered, the assumption is made that the musical quality causing this change in affective response is the same.

Assessment of the external affect locus (i.e., perceived affect) concerns recognition of what feeling state an object is expressing, whereas assessment of the internal affect locus (i.e., induced affect) requires one to introspect and evaluate one’s own feeling state. Although some researchers question whether music is really able to induce affects in its listener (e.g., Kivy, 1990), participants are able to distinguish the two affect loci and report on them individually when given appropriate instructions (Zentner et al., 2008). Furthermore, self-reported induced musical affect is accompanied by physiological changes related to affective processing (e.g., Krumhansl, 1997; Nyklíček et al., 1997; Rickard, 2004). The two affect loci may overlap, but they may also show a negative relationship (e.g., a sad song that makes you happy) or relate to each other in different ways (Gabrielsson, 2001). Consequently, the musical features that are related to perceived affect need not be the same as the musical features that are related to induced affect.

In particular pertaining to induced affect, the best way to represent, and consequently quantify, affect is a matter of debate. In essence, one can never know with absolute certainty what another person is feeling or whether that other person is feeling anything at all (the ‘other minds problem’; see e.g., Harnad, 1991). Self-report measures of affect may be compromised by demand characteristics (i.e., participants change their response to fit potential experimenter hypotheses), self-representation bias (i.e., participants change their response to be more desirable or socially acceptable), limited awareness (i.e., participants are not fully aware of what they are perceiving or feeling; although this is not an issue if researchers are interested in the participants’ conscious experience), and verbalization (i.e., perceptions or feelings may be ineffable; Zentner & Eerola, 2010). Self-report can be seen as the verbalization of subjective feelings (though imperfect), which contains information about the conscious experience of perceived and induced affect (Barrett et al., 2007). Psychophysiological measures of induced affect, such as electrodermal or cardiovascular activation, may be deemed more objective and able to uncover pre-attentive or subconscious affective responses. However, these components can also be influenced by processes other than affect (Mauss & Robinson, 2009). Some researchers use combined measures of self-report and psychophysiology, although this raises the question of synchronization. The different components of affective processing are not always synchronized and correlation between measurements of the different components is moderate to small (Barrett, 2006a; Cacioppo et al., 2000; Mauss & Robinson, 2009; Zentner & Eerola, 2010).

In this thesis, the aim is to capture the subjective experience of perceived and induced affect. The question of which self-report measure is appropriate for quantifying musical affect revolves around the discussion of dimensional versus discrete affect. These two models are closely investigated in Chapter 3.

1.2 Timbre Definition & Description

Timbre is often described by what it is not, as a ‘wastebasket category’ (Bregman, 1990). Timbre is the quality of the sound that is not pitch and not loudness, or rather the quality that allows one to judge that two sounds with the same pitch and loudness are dissimilar. Siedenburg & McAdams (2017) describe four characteristics of what timbre is, instead of what it is not. First, timbre is a perceptual attribute. It is not merely the characteristic of a sound or its signal, but a subjective experience in the listener. Second, timbre is a sound quality and simultaneously a contributor to source identification. That is, two sounds can be perceived to have different timbres without associating them with different sound sources, but at the same time timbre contributes to the recognition and differentiation of sound sources. Third, timbre functions on different levels of detail, such that a musical instrument’s timbre may be, in terminology drawn from biology, described as a *genus*, whereas that same instrument can produce different *species* of timbres as it is played at different pitches, dynamics, and articulations. Fourth, timbre is a property of fused auditory events, such that multiple sound sources may lead to the perception of a single timbre if they are perceptually fused, or multiple timbres if they are not fused.

What further complicates the definition of timbre, is that it is a *multidimensional* perceptual phenomenon. That is, it cannot be described by a single parameter in the same way that a pitch can be described by frequency. One way to investigate the multidimensionality of timbre is with multidimensional scaling (MDS; Shepard, 1962). MDS is conducted on ratings of how dissimilar pairs of sounds are. Such dissimilarity ratings allow researchers to avoid any pre-conceived notions of what timbre is and do not require ambiguous verbal descriptions of timbre. MDS configures those dissimilarity ratings (or distances) in a timbre space where each dimension represents a perceptual dimension of timbre. Each sound has coordinates for each dimension in the MDS space, which allows one to uncover which perceptual attribute each dimension represents. The closer two sounds are on a particular dimension, the more similar their timbre perception. Although different MDS approaches and stimulus selections have led to variations in resulting MDS spaces, most studies find that two or three dimensions best describe timbre perception, and that attack time and spectral centroid most consistently correlate with the coordinates on the MDS dimensions (Siedenburg et al., 2019). These findings do not say that perception can *only* be observed along two or three dimensions, as one may imagine more ways in which timbre can vary. Rather, these dimensions of attack time and spectral centroid reveal the two most salient timbre dimensions that listeners tend to evaluate when considering the overall difference between two musical sounds.

Another way to describe timbre is by analyzing the verbal descriptions of timbral impressions used by both music professionals and naïve listeners. Descriptions such as dark, dull, bright, and rough provide information about how perceived timbres are conceptualized. The semantic differential (SD) method is utilized for timbre semantic research in which a set of sounds is judged on a variety of scales such as “dull-bright” or “not bright-bright” (Carron et al., 2017; Osgood et al., 1957; Solomon, 1958). Brightness, fullness, and roughness are the three dimensions that are most consistently found to capture timbre semantics (Kendall & Carterette, 1993; Siedenburg et al., 2019; Zacharakis et al., 2014). Brightness corresponds to the spectral energy distribution or the amount of energy in the upper frequencies of a sound, whereas attack components and spectral variation have also been associated with timbre semantics (Zacharakis et al., 2015). These three dimensions appear to be stable across languages and cultures, although this topic requires more thorough investigation (Alluri & Toiviainen, 2012; Zacharakis et al., 2014).

The perceptual (MDS) and semantic (SD) dimensions can be correlated with acoustic properties of the rated sounds to give them psychoacoustical meaning. A large set of audio descriptors can be extracted from the audio signal using toolboxes such as the MIR Toolbox (Lartillot & Toiviainen, 2007) or the Timbre Toolbox (Peeters et al., 2011; revised by Kazazis et al., 2021). Although not every descriptor may be perceptually relevant, the correlation of a descriptor with a perceptual or semantic dimension suggests that they are relevant. Temporal descriptors, such as attack or decay time, characterize the temporal domain of an entire sound. Spectral descriptors, such as spectral centroid or spread, characterize the spectral distribution of the timeframe of a sound. Harmonic descriptors are similar to spectral descriptors but consider a sinusoidal model describing a sound's partials instead of a spectral distribution. Spectrotemporal descriptors, such as spectral flux or modulation frequency, characterize the interaction of spectral and temporal descriptors as spectral descriptors may vary over time (Caetano et al., 2019). Spectral centroid, attack and decay time, and the deviation from a smooth spectral envelope have been most consistently associated with perceptual dimensions of instrumental sounds, but these findings may depend on the stimulus selection and further research is needed (McAdams, 2019a).

In this thesis, the sounds played by different instruments at different pitch heights give rise to varying timbre perceptions, and the affective responses that are caused by those timbres (alongside any learned associations of instrument or pitch with affect, for example) are measured. Chapter 4 looks at timbre with a lower level of detail, considering different *genera* of timbre that are grouped together by the comparison of instrument *families*, whereas Chapter 5 looks at timbre *species* with a higher level of detail, describing the acoustic origins of each sound.

1.3 Individual Differences

Perceived and induced affective responses to music have both been associated with various sources of individual differences. Several studies have found that empathy plays a role in affective processing. Empathy can be considered as a personality trait and distinguishes the act of empathizing with a composer, musician, or fictitious musical character from contagion which is a more unconscious mimicking of the affect expressed by the music (Miu & Vuoskoski, 2017). Variability in empathy has been most frequently associated with the enjoyment of sad music (see e.g., Eerola et al., 2016; Garrido & Schubert, 2011; Kawakami & Katahira, 2015), suggesting that increased dispositional empathy leads to greater enjoyment of sad music. Empathy also plays a role in the induced affect of *sublimity* (Balteș & Miu, 2014) and increased correspondence between perceived and induced affect (Egermann & McAdams, 2013).

Other personality traits have also been associated with affective processing. The Big-Five factor of agreeableness (characterized by high altruism, trust in other people, and concern with the well-being of others) is associated with more intense affective responses to unfamiliar music. Both agreeableness and neuroticism (characterized by frequent experiences of negative affect and difficulty coping with stress) are associated with increased feelings of sadness. Openness-to-experience (characterized by increased aesthetic sensitivity, curiosity, and creativity) and introversion (characterized by decreased tendency to be outgoing and assertive or seeking out excitement) are associated with increased enjoyment of music that induces sadness (Ladinig & Schellenberg, 2012). Extraversion (i.e., the opposite of introversion) is also associated with the overall presence of music-induced affects (Juslin et al., 2008), and openness-to-experience is related to a higher prevalence of music-induced chills (Nusbaum & Silvia, 2011). The mood that one is in before listening to music determines the perception and induction of affects congruent with said mood state (Balteș & Miu, 2014; Vuoskoski & Eerola, 2011b). Finally, musical sophistication has been less frequently and less directly investigated in response to musical affect, but some studies do suggest that there is a relationship. For example, preference for the genre of a piece of music increases the likelihood that awe is experienced in response to that music (Pilgrim et al., 2017) and musical sophistication is related to accuracy of decoding an intended affect (Akkermans et al., 2019).

In timbre perception research, research on individual differences is scarcer. There are different versions of MDS, some of which allow for individual and class differences by giving them different weights on the dimensions in a resulting timbre space (e.g., INDSCAL or CLASCAL). Two CLASCAL studies show that participants can be grouped into different latent classes (Caclin et al., 2005; McAdams et al., 1995). The first way in which these classes differ is in how they use the dissimilarity ratings scale (e.g., very limited or using the extremes of dissimilarity). Secondly, the different weights attributed to the different dimensions reflect differences in salience that are attributed to those dimensions. For example, whereas some listeners attribute more salience to variations in the temporal domain in the differentiation of timbres, others attribute more salience to variations in the spectral domain. Further class differences can be found in the kind of spectral features that are attributed more salience. These class differences could not be explained by musical training or musicianship, but no studies have further investigated which possible sources of individual differences may underly these class differences.

Another indication that individual differences underly timbre perception are found in neuroscientific studies. An fMRI study that compared the neural response of violinists and flutist to music played on the violin or flute found that expertise in an instrument affects the neural response to that instrument. Violinists showed more activity in the left superior temporal gyrus (related to auditory processing) in response to violin music, and the inverse was true for flutists (Margulis et al., 2009). Another study compared the auditory brainstem response of pianists and nonpianists to sounds produced by piano, bassoon, and tuba (Strait et al., 2012). These results showed that the amplitude envelope of piano sounds is more strongly correlated with the temporal envelope of the brainstem-response of pianists than is the case for nonpianists. The pianists' and nonpianists' brainstem responses, however, showed more similar correlations to the amplitude envelope of the bassoon and tuba sounds. Thus, both studies do indicate a role for musical expertise in timbre perception, even though these may not be related to dimensions of the timbre space.

Given the relevance of individual differences in affective processing and timbre perception, this thesis considers several sources of individual differences in the comparison of affect models (Chapter 3) and the effect of pitch register and instrument family on perceived and induced affect (Chapter 4), considering pre-existing mood, empathy and other personality traits, but also musical sophistication (Müllensiefen et al., 2013) and musical preferences (Rentfrow et al., 2011). The latter two are included in this exploration of individual differences, as musical sophistication (or expertise) and preferences for different genres of music are hypothesized to influence familiarity with certain timbres and the perceived and induced affective response.

1.4 Online Experimentation

The current experiments have been executed in an online testing environment. Unlike perhaps many experiments from the past few years, the COVID-19 pandemic was not the motivation for running the experiments online. The first practical motivation for online experimentation was that the inclusion of individual differences as moderating factors called for a large and diverse participant sample. In online experimentation, collecting data from hundreds of participants is done much quicker than in laboratory testing. More importantly, laboratory experiments often recruit university students for their experiments, who may not be representative of the global population. The majority of research in social and behavioural sciences is based on participant samples that are considered *WEIRD*: Western, Educated, Industrialized, Rich, and Democratic

(Henrich et al., 2010). Although the concept of WEIRD-ness does not deny the presence of universality, it does raise the question of generalizability when context and culture lead to variability. In 2014 and 2017, 95% of the published articles in *Psychological Science* used WEIRD participant samples (Rad et al., 2018). Although a truly balanced and diverse population may also not be reached by online experimentation, the participant samples are decidedly less WEIRD (Behrend et al., 2011; Chandler et al, 2019; Eerola et al., 2021; Gosling et al., 2010; Sheehan, 2019). The frequency of use of crowdsourcing with online platforms, such as *Amazon Mechanical Turk* (MTurk), *Prolific*, and *Crowdflower*, has steadily increased in published music and auditory studies since 2005 (Eerola et al., 2021). In addition to the increased diversity (or decreased WEIRD-ness) of the participant sample and the speed with which data from a large sample size can be collected, other benefits of online experimentation are the lower costs of online research (Armitage & Eerola, 2020) and the ability to conduct research when a laboratory is not at one's disposal.

One major concern with regards to online auditory experiments, however, is the sound delivery. Laboratory experiments allow for control over the environment, and importantly with regards to auditory research allow for consistent audio setup and attenuation of environmental noise or distractions. This control is largely lost online; participants use their own computer and audio setup, are required to adjust their own volume, are likely not in a sound-proofed environment, and may be more easily distracted by their environment. A few possibilities to regain that control are to set certain restrictions to the online participants' audio setup (e.g., only use headphones, not speakers) and run tests to confirm that participants are wearing headphones. One such test uses sounds with different dynamics due to phase differences, which are easy to discern with headphones but not, or more difficult, with speakers (Woods et al., 2017). A test developed by Milne et al. (2021) uses the Huggins pitch test to confirm that participants are wearing headphones. Another challenge in noisy environments or with less-than-ideal audio setups, is ensuring that participants can hear the auditory qualities relevant to the experiment, especially when considering auditory thresholds or just noticeable differences. In this case, experimenters may use test or calibration sounds that include variation on the auditory feature(s) of interest, and test whether participants are able to differentiate the sounds.

On the other hand, the home environment may also be considered more ecologically valid in its similarity to daily music listening, and the increased participant sample that can be reached with online experiments can negate any erroneous variability. Furthermore, several studies have shown that data obtained online through services such as Amazon Mechanical Turk lead to similar results as laboratory studies and have high internal and test-retest reliability (Berinsky et al., 2012; Klein et al., 2014; Paolacci et al., 2010). Even more so, an online study on the affective response to auditory stimuli finds results similar to laboratory responses to the same stimuli and good inter-rater and test-retest reliability (Seow & Hauser, 2022).

Finally, there are several checks that can be done during the experiment to ensure data quality. In a questionnaire, an attention check can be done by stating “In the following question, please answer *strongly agree*” when presenting participants with a Likert scale from strongly disagree to strongly agree. You can also check participants’ auditory attention by asking a question relevant to the auditory stimulus (e.g., Sauter et al., 2020). To test for consistency, items can be presented twice to the participants, although depending on the task some variation in response may be expected. When inclusion criteria are stated for the experiment and participants have declared to fulfill those criteria, it is possible to ask towards the end of the experiment whether they have been truthful about their eligibility while they are also assured that their answer will not affect their payment. A study on MTurk participants found that 2 to 28% of participants misrepresented their eligibility (MacInnis et al., 2020). It should be noted that lapses in attention or misrepresentation is also possible with participants of laboratory experiments, and the vigilance that is required by online experimenters may also be adapted by those running experiments in the lab.

1.5 Overview

Chapter 2 describes the methods of the two experiments that were both executed online Participants were asked to rate the perceived and induced affect in response to single notes (Experiment 1) and chromatic scales (Experiment 2), which varied in orchestral instrument family and pitch register. Additionally, participants filled in several questionnaires to provide measures of individual differences. When Chapters 3–5 are each submitted for publication, a (condensed) version of Chapter 2 will be incorporated as method sections.

The first question that arose when designing the experiments to test the effect of timbre on musical affect was; how do you measure perceived and induced affect? How do we best represent what a person is feeling and, consequently, how do we measure those induced feelings? Chapter 3 deals with this question by comparing the dimensional model of affect (valence, tension arousal, and energy arousal; Russell, 1980; Schimmack & Grob, 2000) with the discrete model (anger, fear, sadness, happiness, and tenderness; Ekman, 1999; Juslin & Timmers, 2010; Panksepp, 2007), as well as investigating which sources of individual differences are related to the self-reported affect. The context in which the affective response takes place may influence the manner in which affect is best quantified, and such short and affectively ambiguous sounds as used in the current experiments have not frequently been used in studies investigating different affect quantification methods. Consequently, before delving in the role of timbre on musical affect, a thorough examination of quantification methods is relevant not only to the current studies, but to all future studies that want to measure perceived and induced affect in a similar experimental context.

Chapter 4 covers the second research question which is designed as a direct follow-up to McAdams et al. (2017) investigating the effects of instrument family and pitch register on the perceived affective response to single notes. Timbre, in this chapter, is considered as it changes between orchestral instrument families of strings, woodwinds, brass, and percussion, and covaries with pitch register. Here, McAdams et al.'s findings are extended by the addition of induced affect measurements (testing the difference between affect loci), the inclusion of discrete alongside dimensional affect measurements, the response to longer musical stimuli than just single notes (Exp. 2), and the consideration of the role of individual differences. The current experiments also allow the comparison of McAdams et al.'s original findings as tested in a laboratory environment to an online testing environment.

Chapter 5 investigates timbre on a more detailed level, looking at timbre descriptors that summarize characteristics of the spectral, temporal, and spectrotemporal domains of the single notes and chromatic scales. The timbre descriptors are extracted from the sounds using the latest version of the Timbre Toolbox (Kazazis et al., 2021) and are used to predict the affective response. This procedure is not only executed on the current sounds and affect ratings, but also on the sounds and affect ratings from two previously conducted experiments by Eerola et al. (2012) and McAdams et al. (2017). These experiments are similar to the current experiments in that they measure the affective response to sounds produced by different musical instruments. The two studies showed some diverging results, which may in part be explained by differences in methodological approaches. Thus, the re-analysis of Eerola et al. and McAdams et al., and the analysis of the current experiments with the same methodological approaches throughout allows for close comparison of the results that describe the acoustic characteristics that predict the affective response to musical sounds.

Finally, Chapter 6 will summarize the findings of Chapters 3–5, describe how they relate to each other and previous research, discuss the major contributions to the research field, and suggest directions for future research.

Chapter 2
Methods

2.1 Introduction

We conducted two online experiments to answer the four research questions as described in the general introduction (Chapter 1). To avoid unnecessary repetition in the following chapters, the methods are described here in an individual chapter. As the overview of previous research in Chapter 1 has shown, the research area of musical affect and more specifically affective timbre shows discrepant findings, which may in part be explained by methodological differences. Therefore, the aim of the experimental design was to maintain a close resemblance to the previous experiment of McAdams et al. (2017) and to only make systematic changes in areas that are of interest.

In Experiment 1, we replicated the methodology from McAdams et al. (2017). Consequently, our stimulus sets and experiment interfaces are highly similar. We also employed additional methodological elements, such as induced and discrete affect as dependent variables, or individual differences as covariates. In Experiment 2, we aimed to make only one major change from Experiment 1; namely in the stimulus set. Instead of single-notes, Experiment 2 contained longer excerpts of chromatic scales. On the one hand, these excerpts more closely resemble music and allow more exposure time for the participants. On the other hand, we aimed to keep the excerpts as neutral as possible and not include more dominant variations in melody or tempo that might overshadow the effect of timbre.

2.2 Experiment 1

Inclusion & Exclusion

Participants were recruited through the online platform Prolific. Through custom Prolific inclusion criteria, we selected participants who indicated they were fluent in English and had no hearing problems. Participants were also asked to withdraw their participation if they were not able to use headphones or earphones for the experiment, and if they were not able to do the experiment on a laptop or desktop computer. To further protect the quality of data, we set up exclusion criteria that would suggest a participant's submission was not trustworthy. If a participant's reading time of the instructions (~450 words, depending on the participant group) was less than 10 seconds, their data were excluded from analysis. We also looked at the combination of trial reaction time and uniformity of ratings. If a participant's average time to complete a listening trial was relatively fast (i.e., one standard deviation under the group average time) and their rating for each scale appeared to be almost identical (i.e., rating variation across scales was one standard deviation below the group average variation), their data were also excluded from the analysis.

Participants

In total, 296 participants completed the experiment. The data for 33 of those participants (11%) were not used in further analysis, because of the exclusion criteria. The final 263 participants were randomly assigned to one of four groups of roughly equal size differing in the type of rating: perceived dimensional (n = 67), perceived discrete (n = 65), induced dimensional (n = 65), and induced discrete (n = 66). When asked for their gender identity, 161 listed male (61%), 96 female (37%), 4 were nonconforming or questioning their gender identity (2%), and 2 preferred to not answer (<1%). Their ages ranged from 18 to 68 years old (M = 29.0, SD = 10.2). Most participants grew up in Europe (71%), followed by North America (17%), Africa (5%), South America (3%), Asia (3%), and Oceania (<1%). The completed educational level of most participants was a bachelor's degree (33%), followed by high-school (31%), master's (19%), secondary school (7%), T-levels (UK-based; 7%), PhD (2%), and other/NA (1%).

Stimuli

Our stimuli were highly similar to those used by McAdams et al. (2017), as obtained from the Vienna Symphonic Library (VSL; *Vienna Symphonic Library GmbH*, 2022). The stimuli consisted of recorded samples playing a single note at pitch class D# at the *forte* dynamic. The stimuli varied in instrument family (woodwind, brass, strings, pitched percussion) and pitch register (1 to 7, where C4 is middle C with a fundamental frequency of 261.6 Hz). In contrast to McAdams et al., our stimuli lasted 3 s instead of 500 ms to provide the participants with more time for affect induction. To ensure that the experiment would not take too long, we used a sub-selection from the stimulus set from McAdams et al. Because their analyses showed that attack strength and playing technique did not significantly predict the affect ratings, we determined our sub-selection by eliminating the attack and technique variation. Furthermore, in contrast to McAdams et al., our instrument selection for the string family was less varied, as we were not able to obtain the solo viola and bass samples in the VSL. Our final selection contained 59 stimuli to which we applied a raised cosine-ramp to fade the sounds out in the final 100 ms. See Table 2.1 for an overview of the entire list of stimuli.

Table 2.1 *Experiment 1 Stimulus Selection*

		Pitch Register						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Woodwind	Contra Bassoon	■	■					
	Bassoon			■	■			
	Bass Clarinet		■	■				
	Clarinet Bb				■	■	■	
	Oboe				■	■		
	English Horn				■	■		
	Alto Flute				■			
	Flute					■	■	
Brass	Contrabass Tuba	■	■					
	Tuba			■	■			
	Contrabass Trombone	■	■					
	Bass Trombone			■				
	Tenor Trombone				■			
	Horn			■	■	■		
	Trumpet				■			
	Piccolo Trumpet					■	■	
String	Harp	■		■		■	■	■
	Cello		■	■	■	■	■	
	Violin				■	■	■	■
Percussion	Timpani		■	■				
	Gong			■		■		
	Celesta			■		■		■
	Glockenspiel						■	■
	Xylophone					■	■	
	Vibraphone				■		■	
	Crotales						■	■

Note. Colour-filled box signifies that the stimulus selection contained a sound from the given instrument (row) at the given pitch register (column).

Measures

Rating Scales

We used 9-point analogical-categorical scales to obtain the ratings (Weber, 1991). The participants in the dimensional group rated valence (negative – positive), tension arousal (tense – relaxed), energy arousal (tired – awake), and preference (dislike – like) on four separate scales. The participants in the discrete group rated anger, fear, sadness, happiness, and tenderness on five separate scales ranging from "no [anger/fear/sadness/happiness/tenderness]" to "a lot of [anger/fear/sadness/happiness/tenderness]." The participants in the perceived condition were asked "What emotional quality do you perceive in this sound?," and the participants in the induced condition were asked "What emotional quality do you feel in response to this sound?" To rate preference, participants were asked "How much do you like this sound?"

Questionnaires

Pre-existing Mood. We used the PANAS-X (Watson et al., 1988; Watson & Clark, 1994) to measure the participants' pre-existing mood. This questionnaire consists of 60 items that describe different feelings and emotions. Each item is rated on a 5-point Likert scale from "very slightly or not at all" to "extremely," reflecting the extent to which they feel a particular affect at the present moment. From this, we obtained scores of *general negative affect* and *general positive affect* to represent participants' pre-existing mood.

Big-5 Personality. We used the BFI-44 (John & Srivastava, 1999) to measure the participants' personality traits of extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The questionnaire consists of 44 items that list characteristics that may apply to participants. Responses are given on a 5-point Likert scale ranging from "disagree strongly" to "agree strongly."

Trait Empathy. The participants' dispositional empathy was measured with the Interpersonal Reactivity Index (IRI), which measures a *general empathy* score as well as four subscales of *perspective taking*, *fantasy*, *empathic concern*, and *personal distress* (Davis, 1983). The questionnaire consists of 28 items that reflect statements on the participants' thoughts and feelings in different situations. Participants respond on a 5-point Likert scale ranging from "does not describe me well" to "describes me very well."

Musical Sophistication. To assess the participants' musical expertise, we used the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014). The questionnaire consists of 39 self-report items. Most items are rated on a 7-point Likert scale ranging from "completely disagree" to "completely agree." The final 8 items are scored on a 7-point scale that reflects different measures such as years of musical training. This provided us with a score of *general musical sophistication*, but also subscales of *active engagement*, *perceptual abilities*, *musical training*, *emotion*, and *singing abilities*.

Music Preferences. We measured participants' musical preferences with the Short Test of Musical Preferences – Revised (STOMP-R; Rentfrow et al., 2011; Rentfrow & Gosling, 2003). It consists of a list of 23 musical genres, for each of which the participants indicate their preference on a 7-point Likert scale ranging from "dislike strongly" to "like strongly." This test provided us with five scores that represent the latent structure underlying musical preferences: *mellow*, *unpretentious*, *sophisticated*, *intense*, and *contemporary*.

Procedure

Participants were invited through Prolific and then re-directed to an external link that was hosted on a secure webserver of the Music Perception and Cognition Laboratory at McGill University. The experimental interface was built with JavaScript. See Figures 2.1 and 2.2 for examples of the interface of perceived discrete and induced dimensional conditions. At the start of the experiment, participants gave their informed consent in an online form. Then, we tested participants' headphones by presenting them with two sounds, one each in their left and right ears, asking them to verify that they heard sound in both ears. The next step was to set the headphone volume to a comfortable level. We presented participants with four sounds similar to the stimulus set and asked them to set the headphones to a level that ensured that all sounds were audible, but not uncomfortably loud. Immediately after reading the instructions, the participants completed the PANAS-X to measure pre-existing mood. Then, they were presented with a review question that asked whether they were supposed to rate perceived or induced affect. Their response to the review question was not used as an exclusion criterion but acted as a reminder for the participant. After the review question, the listening task commenced with four practice trials. The main listening task (59 trials in random order) was completed in two blocks. The order of the rating scales as presented on the screen was randomized for each participant. In the dimensional condition, the

Figure 2.1 *Example of Induced Dimensional Condition User Interface*

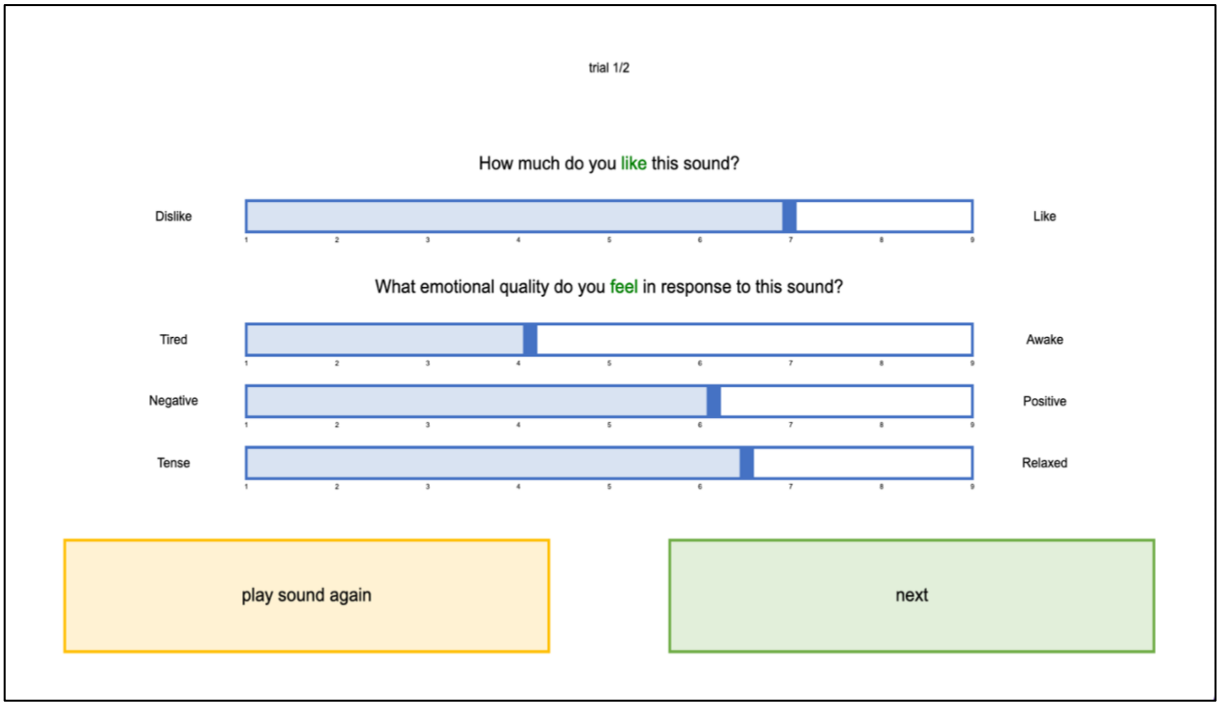
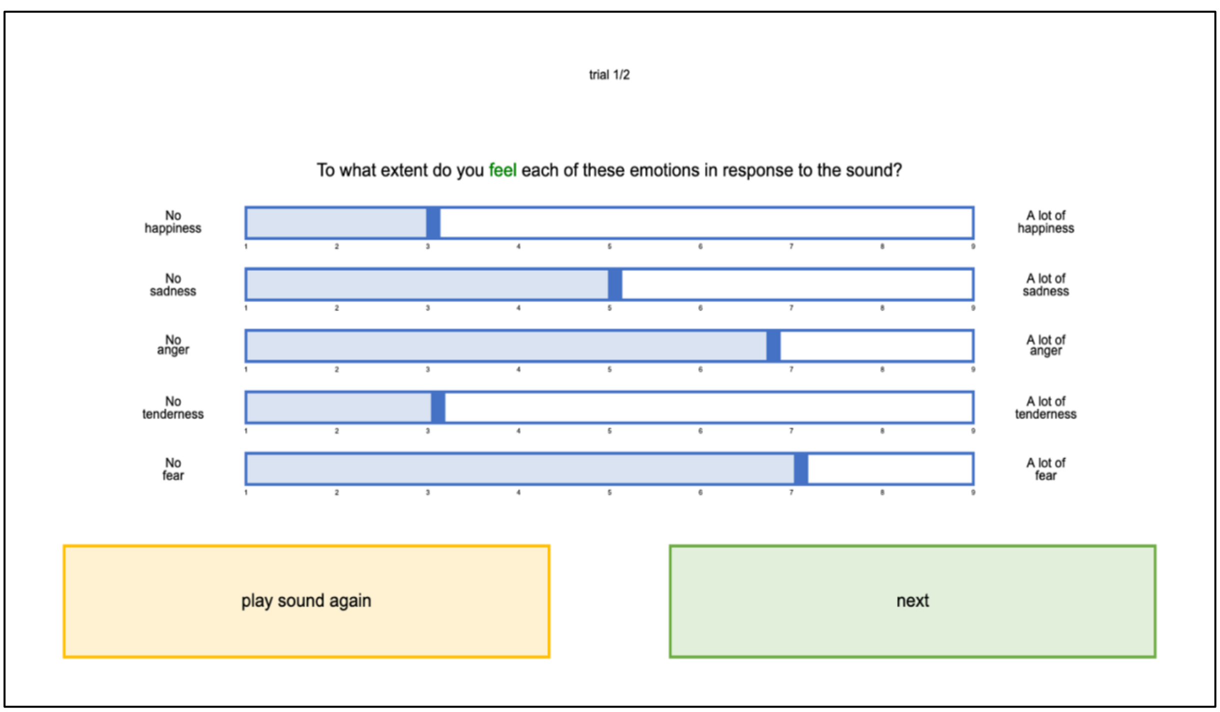


Figure 2.2 *Example of Induced Discrete Condition User Interface*



order of the three affect scales was randomized, and the preference scale was randomly either at the top or bottom of those affect scales. For the discrete conditions, the order of the five affect scales was also randomized. For each participant, this order did not change throughout the experiment. The participants could replay the sound two times in each trial if they so desired.

Finally, after completing the listening experiment, the participants completed the rest of the questionnaires as well as some demographic questions about their age, gender, and in which country they spent their formative years. During the main experiment, the participants were able to go back to the instructions page and a FAQ page in case they ran into issues. When the experiment was completed, participants were shown their scores on the Gold-MSI and STOMP-R questionnaires and provided with debriefing information. On average, it took participants 40 minutes to complete the experiment. To fully complete their participation, they were redirected to Prolific and received a financial compensation. This protocol was certified for ethical compliance by the McGill University Research Ethics Board II.

2.3 Experiment 2

Modifications from Experiment 1

The main difference between Experiments 1 and 2 lies in the stimulus design. Whereas Experiment 1 investigated single notes that last 3 seconds, Experiment 2 investigated chromatic scales that last around 8 seconds. Furthermore, whereas the two conditions of affect locus (perceived and induced) were rated by separate groups of participants in Experiment 1 (between-subjects), in Experiment 2 these were rated by the same group of participants, in separate counter-balanced blocks (within-subjects design). This was done with the aim that participants were better able to distinguish between perceived and induced affect, as they were required to actively switch the locus of affect. Due to the change in design (longer stimuli and within-subjects affect locus), we reduced the total number of stimuli in Experiment 2 to ensure that the experiment would not take too long. We also aimed for half the number of participants compared to Experiment 1, as this should result in the same statistical power with the change from between- to within-subjects. A more complete summary of the differences between Experiments 1 and 2, as well as the differences with McAdams et al. (2017), can be found in Table 2.3.

Inclusion & Exclusion

We used the same inclusion criteria (language, hearing, and setup) as in Experiment 1. In contradistinction to Experiment 1, the instructions were no longer one long page of text, but chopped up into shorter paragraphs of easily digestible excerpts of 2-3 sentences on separate pages that a participant would click through. The *next* button to continue to the subsequent instruction excerpt was only activated after a certain amount of time had passed to allow for normal reading speed, depending on the length of the text, before the program would allow the participant to proceed to the next page. Therefore, we no longer assessed the instruction reading time as an exclusion criterion. We did still assess the combination of trial reaction time and uniformity of ratings as exclusion criteria. We assessed these separately for the perceived and induced blocks (participants on average were slower on the perceived trials), but if a participant fulfilled the criteria for timing and uniformity for only one block, their data were still excluded from analysis. We experienced some technical issues, which led some participants to get stuck during the experiment, or several participants being assigned to identical versions of the randomization of stimulus order. We did not include the data from these participants in our analysis.

Participants

In total, 181 participants completed the experiment. The data for 29 of those participants (16%) were not used in further analysis. The final 152 participants were randomly assigned to rate either dimensional (n = 76) or discrete (n = 76) affect. When asked for their gender identity, 84 listed male (55%), 67 female (44%), and 1 preferred to not answer (<1%). Their ages ranged from 18 to 68 years old (M = 31.6, SD = 10.4). Most participants grew up in Europe (48%), followed by North America (36%), Africa (14%), Asia (2%), and South America (<1%). The majority of participants completed a bachelor's degree (36%), followed by master's (25%), high school (18%), T-levels (UK-based; 9%), secondary school (8%), PhD (2%), and other/NA (3%).

Stimuli

We created 32 stimuli with OrchSim[1] (McAdams & Goodchild, 2017a; *OrchPlayMusic*, 2022) in which a chromatic scale (ascending and descending) spanning a perfect fifth from C to G and back

[1] Philippe Macnab-Séguin created the stimuli in OrchSim, which uses several musical sound databases.

Table 2.2 *Experiment 2 Stimulus Selection*

		Pitch Register					
		1	**2**	**3**	**4**	**5**	**6**
Woodwind	Contra Bassoon	■					
	Bassoon			■			
	Bass Clarinet		■				
	Clarinet Bb					■	
	Oboe					■	
	English Horn				■		
	Alto Flute				■		
	Flute						■
Brass	Contrabass Tuba	■					
	Tuba		■				
	Contrabass Trombone			■			
	Bass Trombone	■					
	Tenor Trombone			■			
	Horn				■		
	Trumpet					■	
	Piccolo Trumpet						■
String	Harp	■					■
	Cello		■			■	
	Violin				■		■
	Bass	■					
	Viola			■			
Percussion	Timpani		■				
	Marimba		■				
	Celesta			■			
	Glockenspiel						■
	Xylophone				■		
	Vibraphone				■	■	
	Crotales						■

Note. Colour-filled box signifies that the stimulus selection contained a sound from the given instrument (row) at the given pitch register (column).

to C was created. The scales were played in tenuto style (each note was re-attacked), at a *mezzo-forte* dynamic level, with added reverb of a medium-sized room (i.e., not completely dry nor a concert hall). The sound sequences lasted ~8 seconds each. Because these stimuli lasted longer than those in Experiment 1, we again reduced the total number of stimuli to ensure that the experiment would not take too long to complete. See Table 2.2 for an overview of the stimuli we used. Note that here we were able to include solo viola and bass stimuli in the string family. In the percussion family, we replaced the gong with the marimba, as it is more suited to play melodic phrases. The registers also no longer spanned seven registers, but six, to ensure that the chromatic scale did not surpass the instruments' tessituras.

Measures

All rating scales and questionnaires used in Experiment 2 are identical to those used in Experiment 1.

Procedure

Again, the experimental procedure was very similar to the procedure in Experiment 1, and only the notable changes will be discussed here. As described in the *Inclusion & Exclusion* section, the instructions were no longer one long page of text as in Experiment 1. The review question would again ask participants if they were to rate induced or perceived affect, but now the correct answer was that they were to rate both in separate blocks. Before each block, they were reminded which locus of affect they were rating. For the dimensional affect group, preference was only rated in the perceived affect block. On average, it took participants 45 minutes to complete the experiment. We removed the option to refer (back) to the instructions and FAQ during the experiment, as none of the participants in Experiment 1 utilized that function.

2.4 Overview

Throughout the following chapters, this chapter can be referred to for details on the design of Experiments 1 and 2, how they were similar and different from each other. In Table 2.3 is a final overview of the two experiments, as well as the McAdams et al. (2017) experiment, which will show the main similarities and differences in the sample size, stimulus selection, affect measurements, independent variables, inclusion of covariates, and experimental procedure. Table 2.4 shows descriptive statistics of each of the questionnaires' (sub-)scores. This shows, for example, that the PANAS-X Negative Affect subscale showed high positive skewness and high positive kurtosis, indicating that most scores were low (no negative affect) and most outliers were on the high end (a lot of negative affect). Other moderate positive skewness scores are found for Gold-MSI Musical Training, and negative skewness scores are found for IRI Perspective Taking, IRI Empathic Concern, Gold-MSI Emotion, and STOMP-R Contemporary. These summary statistics and indications of nonnormality may be taken into account in the interpretation of results.

Table 2.3 *Differences Between McAdams et al. (2017), Experiment 1, and Experiment 2*

	McAdams et al. (2017)	**Experiment 1**	**Experiment 2**
Participants	*n* = 40	*n* = 263	*n* = 152
Stimuli	*n* = 137	*n* = 59	*n* = 32
	Single note	Single note	Chromatic scale
	VSL	VSL	OrchSim
	500 ms	3 s	~8 s
	D#	D#	C-G
Dependent variables	3 dimensional affects	3 dimensional affects	3 dimensional affects
	–	5 discrete affects	5 discrete affects
	Preference	Preference	Preference
	Familiarity	–	–
Independent variables	Register (1-7)	Register (1-7)	Register (1-6)
	Instrument family	Instrument family	Instrument family
	Musicianship	–	–
	–	Affect locus	Affect locus
Covariates	–	Pre-existing mood	Pre-existing mood
	–	Personality	Personality
	–	Empathy	Empathy
	–	Musical sophistication	Musical sophistication
	–	Music preferences	Music preferences
Procedure	In lab	Online	Online
	–	Affect locus between subjects	Affect locus within subjects

Table 2.4 *General statistics from questionnaire results from Experiments 1 (E1) and 2 (E2)*

	Min		Max		Mean		SD		Skewness		Kurtosis	
	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2
Perspective Taking	10	12	30	35	21.98	26.32	3.91	4.88	-0.50	-0.47	-0.04	0.03
Empathic Concern	9	7	35	30	26.52	23.54	4.77	4.79	-0.52	-0.64	0.31	0.12
Personal Distress	7	7	34	35	20.28	20.00	5.30	5.30	0.03	-0.23	-0.08	-0.21
IRI Total	56	62	135	130	97.06	94.92	12.76	12.83	-0.38	0.04	0.10	-0.23
Agreeable-ness	18	11	50	45	36.37	34.33	6.23	5.78	-0.08	-0.48	-0.42	0.99
Extra-version	9	8	43	39	25.34	23.57	7.86	6.95	0.02	0.00	-0.76	-0.54
Conscient-iousness	10	16	45	45	30.96	33.90	7.09	6.90	0.06	-0.18	-0.43	-0.66
Neuroti-cism	8	8	40	39	24.57	22.85	6.97	7.03	-0.06	-0.07	-0.60	-0.46
Openness	17	19	49	49	35.53	37.18	6.50	6.14	-0.17	-0.15	-0.38	-0.03
Musical Engagement	9	12	62	54	35.52	36.68	9.82	8.76	-0.13	-0.13	-0.41	-0.50
Perceptual Abilities	16	25	63	63	44.83	46.64	8.65	7.82	-0.38	0.04	0.25	-0.25
Musical Training	7	7	48	45	19.56	18.99	10.04	9.13	0.61	0.61	-0.46	-0.26
Singing Abilities	7	12	49	49	26.51	27.91	7.86	8.11	-0.14	0.29	-0.26	-0.31
Emotion	17	22	42	42	32.14	32.89	5.13	4.60	-0.52	0.07	0.00	-0.72
Gold-MSI Total	21	26	126	119	68.19	69.69	18.63	17.52	0.15	0.33	-0.26	-0.24
Mellow	6	0	100	94	54.85	56.51	16.90	16.97	-0.25	-0.47	0.03	0.27
Unpreten-tious	11	6	100	94	53.32	54.38	17.58	17.93	0.03	0.15	-0.33	-0.22
Sophisti-cated	0	2	88	88	53.51	50.91	16.84	17.40	-0.46	-0.30	-0.08	-0.14
Intense	4	4	100	100	60.85	61.09	20.65	20.60	-0.25	-0.49	-0.59	-0.19
Contem-porary	0	0	96	100	60.94	62.17	17.63	19.62	-0.67	-0.34	0.21	0.12
Positive Affect	10	8	50	35	27.88	20.36	7.93	6.25	0.16	0.18	-0.56	-0.64
Negative Affect	10	10	46	36	14.10	13.13	5.97	4.79	2.52	2.21	7.73	5.69

Chapter 3
Representing the Perceived and Induced Experience of Musical Affect: Dimensional versus Discrete

3.1 Introduction

The ability of music to convey and induce moods and emotions is one of the main reasons why people listen to music (Juslin & Laukka, 2004). One source of discussion and discrepancy in findings in musical emotion research is the theoretical representation and practical operationalization of emotion. These discussions mirror the discourse in general emotion research and mostly revolve around the distinction between discrete (or categorical) basic emotions versus multidimensional affect. The differences in theoretical and practical approaches complicate the comparison of findings and inhibit the discourse on the musical features and mechanisms that underlie the affective response to music. Affect, here, is used as an umbrella term that covers both (subtle, long-term) moods and (big, short-term) emotions, as well as other valenced states (Juslin & Västfjäll, 2008).

Theories of basic emotions as discrete categories in their strictest form suggest that there are a number of basic emotions that are both universal and innate, such as anger, fear, sadness, happiness, and disgust. Each of these emotions are supported by independent neural systems and are also expressed in a categorical manner (i.e., physiologically, behaviourally, and subjectively; Ekman, 1999; Panksepp, 2007). Research shows that certain basic emotions are readily perceived and induced when listening to music (Juslin & Laukka, 2003, 2004; Juslin & Timmers, 2010). However, the evidence to support the basic emotion hypothesis that independent neural systems underlie discrete emotions has been unreliable and inconsistent (Barrett, 2006b; Cacioppo et al., 2000). Furthermore, there are certain basic emotions that are less readily perceived or induced in music, such as disgust. This category is often replaced by more musically applicable emotions such as peacefulness or tenderness (Gabrielsson & Lindström, 1995; Juslin & Timmers, 2010; Vieillard et al., 2008). The Geneva Emotional Music Scale (GEMS) suggests an independent set of categorical emotions that are specifically relevant to music; wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension, and sadness (Zentner et al., 2008).

During the last couple of decades, studies on musical emotions have increasingly employed a version of the dimensional model of affect (Eerola & Vuoskoski, 2013). The most common dimensional affect representation is that of valence (displeasure/pleasure) and arousal (deactivation/activation) in the circumplex model (Russell, 1980). Like discrete emotions, affects that span the four quadrants of two-dimensional affect (low or high on valence or arousal) are readily recognized in and induced by music (Eerola et al., 2012; Ilie & Thompson, 2006; McAdams et al., 2017; Rickard, 2004; Vieillard et al., 2008). In music research, the two dimensions of valence and arousal are generally represented to be orthogonal, or independent, although some studies suggest other relations between the two dimensions, such as a v-shape where increased positive or negative valence is related to an increase in arousal (e.g., Kuppens et al., 2017).

One criticism of the two-dimensional affect model is that it is not able to distinguish between certain emotions. For example, although anger and fear are two distinct emotions, they would be positioned on the same valence and arousal coordinates. To counter the argument, additional affect dimensions have been suggested in general affect research, such as approach-avoidance (Carver, 2004), potency-control (feelings of power or weakness) and predictability (feelings of surprise or familiarity; Fontaine et al., 2007), or tension arousal (Schimmack & Grob, 2000). Although there is some evidence that potency is relevant to musical emotion (Rodà et al., 2014), neither potency nor approach-avoidance have received much attention in music research. Predictability, or violation of expectation, has received more attention, although it is considered as a mechanism in emotion induction rather than a third dimension in the affect space (Juslin, 2013; Meyer, 1956; Steinbeis et al., 2006). Some studies on musical emotion do employ tension arousal as a third dimension in the affect space. However, although some studies find that tension arousal is a useful addition (Ilie & Thompson, 2006; McAdams et al., 2017), others find that it does not improve the performance of two-dimensional models in perceived and induced affect (Eerola et al., 2012; Eerola & Vuoskoski, 2011; Vuoskoski & Eerola, 2011a). Finally, models that separate valence into independent positive and negative affect can account for mixed emotions (Watson et al., 1988), which are especially common in music (Hunter et al., 2010).

Hybrid models combine the discrete emotion and dimensional affect models by suggesting that fundamentally, core affect is dimensional in nature, but the conscious interpretation or appraisal of these affects is best described in categorical emotion terms (Barrett, 2006b; Russell, 2003). In their 2013 review, Eerola and Vuoskoski report that around 10% of the studies on musical emotion used hybrid methods to quantify affect. This percentage of studies may have increased further in the last decade (see e.g., Lahdelma & Eerola, 2016) and has also been formulated in theoretical accounts of musical emotion (Cespedes-Guevara & Eerola, 2018). Thus, hybrid models suggest that depending on whether one is investigating core affect or emotional episodes, a dimensional or discrete representation is more suitable. However, there are also individual differences in how people label affective states, some fitting better into a dimensional and others into a discrete representation (Barrett, 1998). Furthermore, describing one's own induced feelings compared to perceived feelings may shift a person from arousal focus (more variation in deactivation-activation) to valence focus (more variation in displeasure-pleasure; Barrett, 2004).

Stimulus selection is important when comparing the suitability of different affect models, because they determine the kind of affective variation that can be observed in or felt in response to the stimuli. Eerola and Vuoskoski (2011) note that nearly all the studies that combine discrete and dimensional models of musical affect choose stimuli that are exemplars of discrete emotions (e.g., Gosselin et al., 2006, 2007; Khalfa et al., 2008; Vieillard et al., 2008). Eerola and Vuoskoski (2011) created a stimulus set that consists of examples that were moderately and highly representative of discrete emotions, and examples that were moderately and highly representative of the extremes of valence, energy arousal, and tension arousal, as confirmed by listener judgments. They found that for perceived affect, a two-dimensional model of affect (valence and arousal) provided the highest correspondence, and that the discrete emotion model performed especially more poorly on characterizing the affective content of emotionally ambiguous examples. Using the stimuli from the same corpus, they also compared discrete, dimensional, and the GEMS model for induced musical affect (Vuoskoski & Eerola, 2011a). Here, too, they found that the two-dimensional model of affect outperformed the other two models in terms of rating consistency and discrimination of music excerpts.

Although Vuoskoski and Eerola put together a corpus that spans both discrete and dimensional continua and confirmed their validity with listener judgments, there is a possibility that the selections lacked variation on a dimension or category that had not been pre-determined by the authors. Here, we employ a selection of sounds, single notes and chromatic scales, that have not been selected based on their affective intent, but rather on their variation in timbre, pitch, and stimulus length (number of notes). Although such stimuli are less ecologically valid, they allow us to investigate the representations of affect in response to musical sound without bias towards a hypothesized representation of musical affect. Here, we employ a stimulus set that is exemplary of stimuli that are used in affect research to investigate local musical features, like timbre and pitch.

The aim of this study is to investigate the applicability of the discrete and dimensional affect models on musical sounds that are not selected or designed with a specific affective intent. We will test this on two different stimulus sets (single notes from Experiment 1 and chromatic scales from Experiment 2, see Chap. 2) and on the two affect loci (perceived and induced) in an online environment. We expect that an increased stimulus length from single notes to chromatic scales would increase the likelihood of emotion categorization and thus might make a discrete affect representation more suitable than a dimensional one. Furthermore, perception of emotions may also warrant more discrete categorization than feelings of smaller changes in mood, which may be represented as dimensional core affect. Finally, we will investigate the role of individual differences – personality, mood, and musical preferences – in the affective response to music.

3.2 Analysis

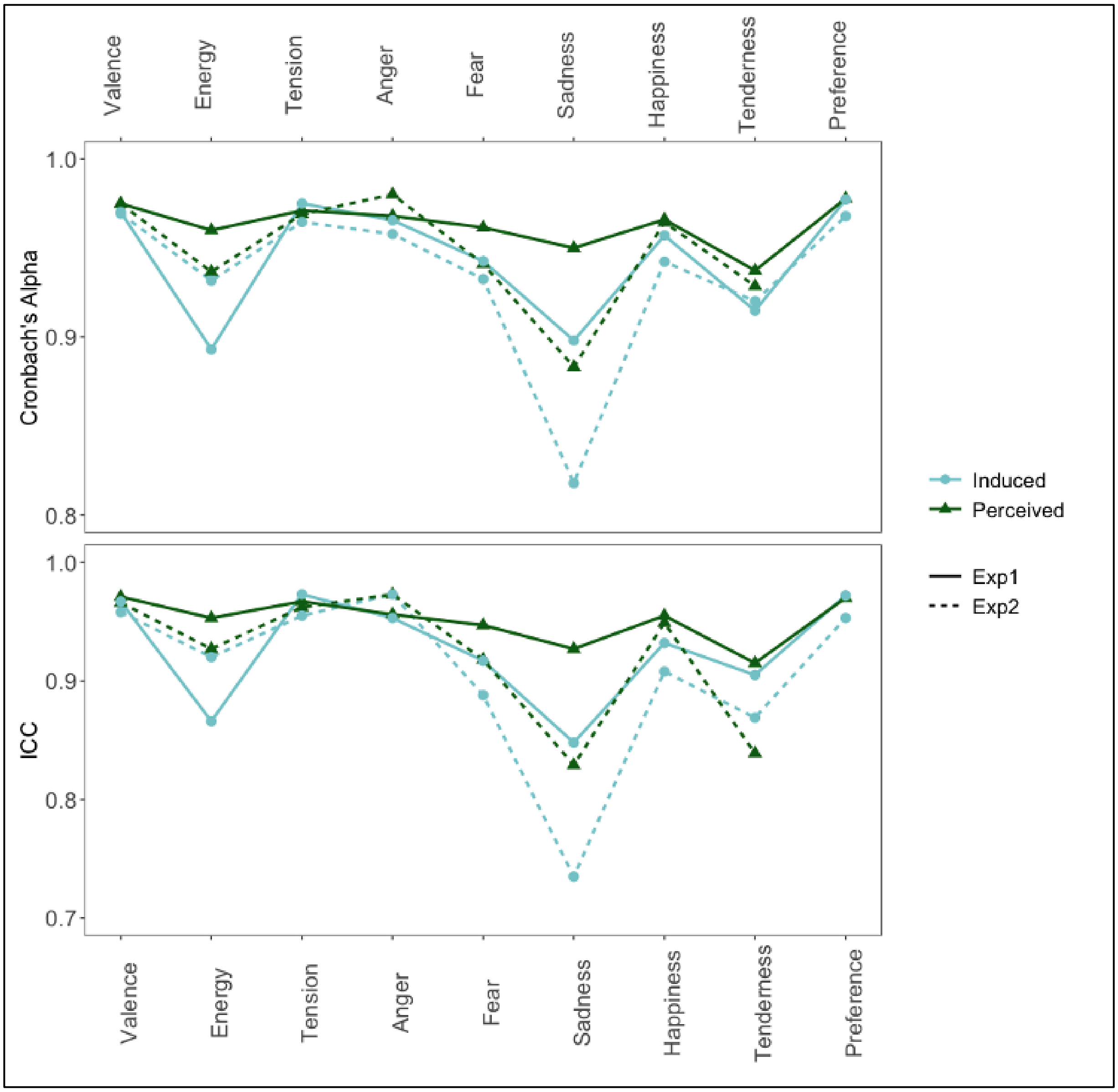
All analyses were done in R version 4.2.1 (www.r-project.org). Cronbach's alpha was calculated with the *CronbachAlpha* function from the *performance* package (Lüdecke et al., 2022), and intraclass correlations (ICC) were calculated with the *icc* function from the *irr* package (Gamer et al., 2019). The testing of the difference between correlations was done with the *cocor* function of the *cocor* package (Diedenhofen & Musch, 2015). All linear mixed modelling was done with the *lmer* and *anova* functions from the *lmerTest* package (Kuznetsova et al., 2020). Post-hoc analyses were done with the *emtrends* and *emmeans* functions from the *emmeans* package (Lenth et al., 2022). Correlation ellipse plots were created with the *corrplot* function from the *corrplot* package (Wei et al., 2021). The principal component analysis was conducted with the z*prcomp* function and the biplot was created with the *biplot* function, both from the *stats* package (inbuilt). Canonical correlation analysis was performed with the *cc* function, and its loadings were extracted with the *comput* function, both from the *CCA* package (González & Déjean, 2021). The lasso regressions were computed with the *islasso* function from the *islasso* package (Sottile et al., 2022). Finally, data were visualized with the *ggplot* function from the *ggplot2* package (Wickham et al., 2022).

3.3 Results

Internal Consistency & Inter-Rater Agreement

Figure 3.1 shows the internal consistency (Cronbach's alpha) and the inter-rater agreement (intraclass correlation; ICC) for each of the affect scales. Whereas Cronbach's alpha measures consistency regardless of individual participants' mean ratings (some participants tend to give higher scores than others), ICC measures the correlation between participants' ratings and how

Figure 3.1 *Cronbach's Alpha and Intraclass Correlations (ICC) Separated by Affect Locus and Experiment*



similar their mean ratings are. Both Cronbach's alpha and ICC show highly similar results with good consistency and agreement ratings (range alpha = [.82, .98]; range ICC = [.74, .97]). Thus, overall participants' ratings on the stimuli were strongly correlated and highly similar. We can see, however, that averaged over affect loci and experiments sadness scores the lowest on both Cronbach's alpha (M = .89; SD = .05) and ICC (M = .83, SD = .08), with an especially large dip in internal consistency for induced sadness in Experiment 2 (alpha = .82, ICC = .74). We see a similar but smaller dip in Figure 3.1 for energy arousal and tenderness. Consequently, sadness, energy arousal, and tenderness may be the most susceptible to individual differences, or their concepts may have been less clear to the participants. Preference shows the highest consistency and agreement overall ($M_{alpha\ \&\ ICC}$ = .97, $SD_{alpha\ \&\ ICC}$ = .01), followed by valence and tension arousal.

When we compare the average alpha and ICC of the different affect loci, experiments, and affect models (Table 3.1), overall there are only minor differences in consistency and agreement. Perceived affect, Experiment 1, and the dimensional affect model score slightly higher than induced affect, Experiment 2, and the discrete affect model, respectively, which indicates that participants are slightly more consistent and agree more on those affect ratings.

Table 3.1 *Means and Standard Deviation for the Comparisons of Affect Locus, Experiment, and Model on Cronbach's Alpha and ICC*

Variable	Groups	Cronbach's alpha (SD)	ICC (SD)
Affect Locus	Induced	.94 (.04)	.92 (.06)
	Perceived	.96 (.02)	.94 (.04)
Experiment	Exp. 1	.95 (.03)	.94 (.04)
	Exp. 2	.94 (.04)	.91 (.06)
Model	Dimensional	.96 (.02)	.95 (.03)
	Discrete	.94 (.04)	.91 (.06)

Note. The comparison of dimensional and discrete models excludes the preference ratings.

Model Redundancy

Tables 3.2 and 3.3 show the Pearson correlations between the dimensional and discrete affect scales, respectively, for each experiment and each affect locus. Note that Table 3.2 does not list correlations with the preference scale in the second experiment's perceived condition, as the participants gave no preference ratings in that condition. We tested the difference between each pair of correlations comparing the perceived and induced affect loci, and the first and second experiments. The comparison of affect loci entailed testing the difference between two nonoverlapping (i.e., different variables) dependent (i.e., same stimuli) groups. The comparison of experiments entailed testing the difference between independent groups (i.e., different variables and different stimuli). For both kinds of comparisons, multiple different tests are considered, as discussed by Diedenhofen and Musch (2015). Here, all potential tests agreed on the statistical (non)significance of the difference between correlations.

In Experiment 1, all correlations between the dimensional affect scales were significantly stronger in the induced affect locus than in the perceived affect locus (all $p < .001$), whereas this was not the case for any of the dimensional affect scales in Experiment 2. For the discrete affect scales, statistically significant differences between affect loci were also more rare. In Experiment 1, the correlations between anger and fear, and anger and tenderness, were stronger for perceived than induced affect (all $p < .01$). In both experiments, the correlations between fear and happiness were also stronger in the perceived affect locus (all $p < .01$). When comparing the correlations between experiments, there were also only rarely significant differences between the correlations. The induced tension and energy correlation was significantly stronger in the first experiment ($p < .05$), whereas the correlations between induced anger and fear, and induced anger and tenderness, were stronger in the second experiment (all $p < .05$). Because these correlations in both experiments and affect loci are so close to each other, and the statistical differences we do find concern differences in strength and not direction, we can expect that further analysis on the model redundancy will lead to similar results as well. Thus, for the further comparisons of the affect models, we will combine the data of the two affect loci and experiments, although we will take into consideration the difference between affect loci in the first experiment.

Figure 3.2 visualizes the correlation of all the affect scales, with the ratings averaged over participant, affect locus, and experiment. Nearly all scales were significantly correlated to each other. Within the dimensional affect scales, there is a strong negative correlation between valence

Table 3.2 *Pearson Correlations Between Dimensional Affect Scales and Preference per Affect Locus and Experiment*

Scale Pairs			Experiment 1		Experiment 2	
			Perceived	Induced	Perceived	Induced
Valence	–	Tension	–.89***	–.96***	–.95***	–.97***
Valence	–	Energy	.22	–.23	.28	.18
Valence	–	Preference	.88***	.97***	–	.94***
Tension	–	Energy	.22	.45**	–.02	.04
Tension	–	Preference	–.94***	–.98***	–	–.94***
Energy	–	Preference	–.14	–.38**	–	–.00

Note. * $p < .05$; ** $p < .001$; *** $p < .0001$

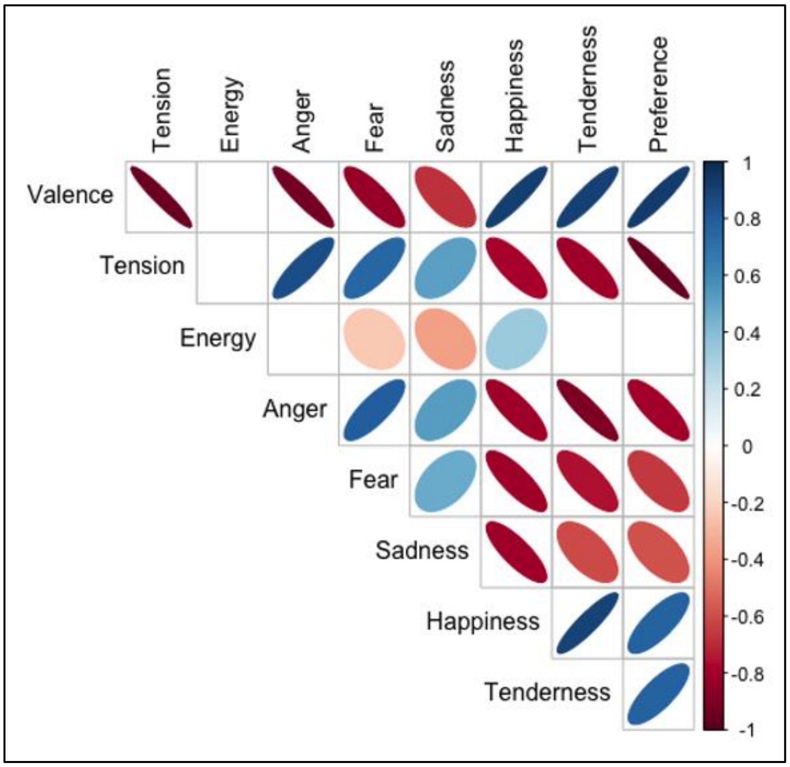
Table 3.3 *Pearson Correlations Between Discrete Affect Scales per Affect Locus and Experiment*

Scale Pairs			Experiment 1		Experiment 2	
			Perceived	Induced	Perceived	Induced
Anger	–	Fear	.90***	.70***	.87***	.90***
Anger	–	Sadness	.47**	.56***	.48*	.45*
Anger	–	Happiness	–.85***	–.80***	–.82***	–.81***
Anger	–	Tenderness	–.90***	–.76***	–.91***	–.90***
Fear	–	Sadness	.51***	.59***	.49*	.38*
Fear	–	Happiness	–.84***	–.74***	–.88***	–.76***
Fear	–	Tenderness	–.84***	–.79***	–.89***	–.81***
Sadness	–	Happiness	–.80***	–.82***	–.77***	–.76***
Sadness	–	Tenderness	–.56***	–.63***	–.62**	–.54*
Happiness	–	Tenderness	.90***	.92***	.93***	.92***

Note. * $p < .05$; ** $p < .001$; *** $p < .0001$

and tension arousal, $r(89) = –.94$, $p < .0001$, but no significant correlation with energy arousal. Within the discrete affect scales, anger, fear, happiness, and tenderness were relatively strongly correlated with each other, $|r|(89) = [.75,.90]$, $p < .0001$. Although sadness was also strongly correlated to happiness, $r(89) = -.81$, $p < .0001$, it was less strongly correlated to the other discrete affect scales, $|r|(89) = [.48, .60]$, $p < .0001$. This is our first indication that in both models, some of the three dimensions and five categories may be redundant.

Figure 3.2 *Correlation Plot of All Affect Scales*



Note. Here, ratings from the different experiments and affect loci are combined. Narrowness of ellipses and saturation of colours signify the strength of the correlation, whereas the orientation and colour (red/blue) signify the direction of the correlation. No ellipses are shown when correlation was non-significant.

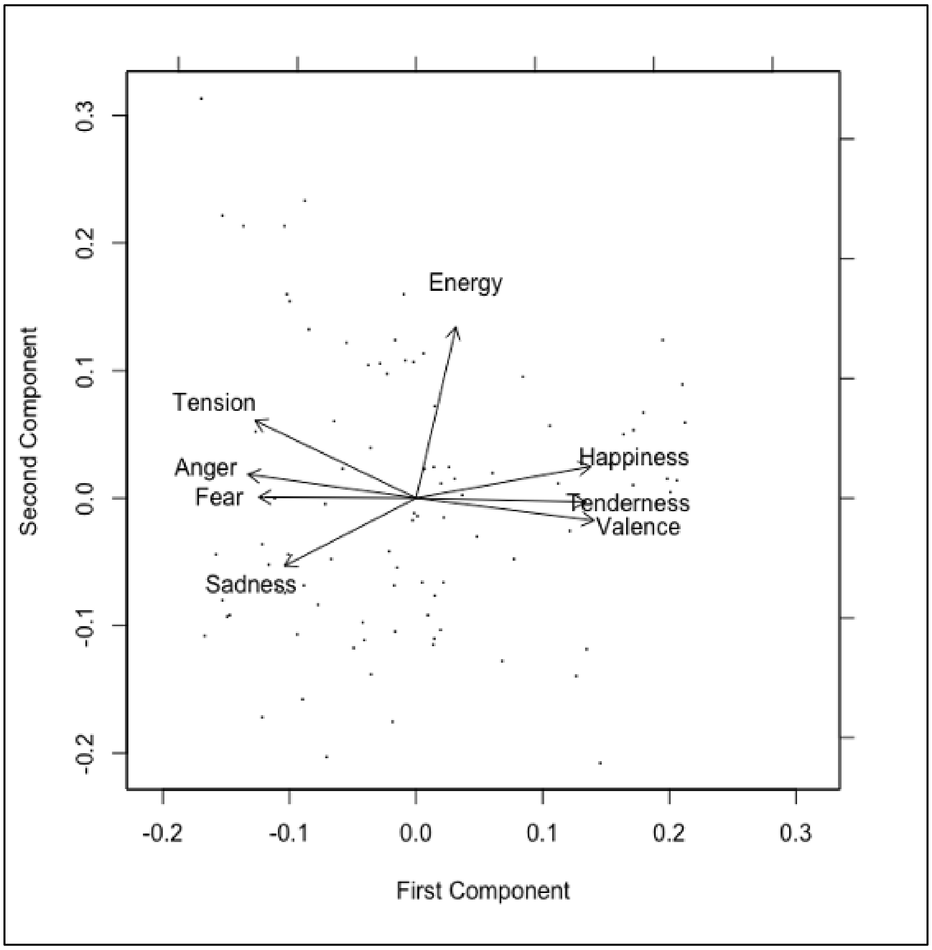
There are also strong correlations between the dimensional and discrete affect models. Valence and tension were strongly correlated with anger, fear, happiness, and tenderness, $|r|(89) = [.76,.92]$, $p < .0001$, but less strongly with sadness, $|r|(89) = [.51, .68]$, $p < .0001$. Energy did show correlations with fear, sadness, and happiness, but these were relatively weak, $|r|(89) = [.24, .36]$, $p < .01$). Preference was also correlated with most affect scales, except for energy arousal. It was most strongly correlated with tension arousal, $r(89) = -.95$, $p < .0001$, and less strongly correlated with sadness, $r(89) = -.59$, $p < .0001$. To summarize, nearly all affect scales were strongly correlated with each other, except for energy arousal and sadness, which showed non-significant or less strong correlations.

We ran principal component analysis (PCA) to investigate whether the three dimensional and five discrete affect scales could be reduced to a lower number of principal components (PCs). We ran this for the dimensional and discrete affect scales individually, as well as combined (hybrid). Table 3.4 shows the results of the three PCA analyses. In all three cases, two PCs explained most of the variance in the rating data. For the dimensional model, the loadings of the three dimensional scales indicate that the first component is represented by valence and tension in opposite directions, whereas the second component is represented by energy arousal. We also considered the PCA for the perceived and induced dimensional model in the first experiment separately (not described in Table 3.4), because of our earlier finding that the induced condition in Experiment 1 showed stronger collinearity. The results were highly similar to the results described in Table 3.4, with only a stronger loading of energy arousal on PC1 in the induced condition (–.38) than in the perceived condition (.00). For the discrete affect model, we find that most scales

Table 3.4 *PCA Results for Dimensional, Discrete, and Hybrid Model Showing Variance Explained for Each PC and The Loadings of Each Variable on Those PCs*

	Dimensional		**Hybrid**	
	PC 1 (64.9%)	**PC 2 (34.8%)**	**PC 1 (71.8%)**	**PC 2 (15.7%)**
Valence	.70	–.20	.41	–.11
Tension	–.71	–.10	–.37	.38
Energy	–.07	–.97	.09	.84
	Discrete			
	PC 1 (79.9%)	**PC 2 (12.3%)**		
Anger	–.46	.34	–.39	.13
Fear	–.43	.39	–.36	.00
Sadness	–.38	–.82	–.30	–.33
Happiness	.49	.18	.40	.15
Tenderness	.47	–.16	.39	–.02

Figure 3.3 *Loadings of Scales and Stimuli on the Two Rotated Principal Components for the Hybrid PCA*



load relatively equally onto the first component. This component may also be summarized as a form of valence, as it is differentiated by anger, fear, and (less strongly) sadness in one direction and happiness and tenderness in the other direction. For the second PC, sadness shows the strongest loading. Thus, the three dimensions may be reduced to a valence and energy model, and the five discrete affects may be reduced to a valence and sadness model. Finally, when we take all eight affect scales and run a PCA, we find two dimensions where, based on the strongest loadings, the first may be called a valence dimension, and the second an energy dimension. These loadings of the hybrid PCA are also visualized in Figure 3.3., where we see tension, anger, and fear, versus valence, happiness, and tenderness differentiating on the first PC, and energy arousal strongly loading on the second PC. Based on the loadings of sadness, we may conclude it is a combination of the two PCs, or of valence and energy arousal.

Discrete and Dimensional Mapping

Canonical correlation analysis (CCA) tests the relationship between two sets of variables, here the dimensional and discrete affect scales, by finding linear combinations of the two sets of variables that are maximally correlated. It is another form of dimension reduction that aims to find maximum overlap. Table 3.5 shows the results of the CCA. All three canonical variates were found to be significant (Wilk's lambda), but the first canonical variate explained nearly all of the variance in the data, whereas the following two canonical variates explained only 1% or less. We find a similar pattern if we look at the redundancy. Redundancy represents the proportion of variance in one set of variables, that is explained by the canonical variate of the other set of variables. Thus, the first canonical variate of the discrete variables, explained 58% of the variance in the dimensional variables. Vice versa, the first canonical variate of the dimensional variables, explained 76% of the variation in the dimensional variables. The second and third variates explained 3% or less of the variance, even though the two models did show significant canonical correlations of .46 and .32. Consequently, we focus on interpreting the first canonical variable, as it captures almost the entirety of the relationship between the dimensional and discrete affect scales. Most variables show a high loading on the first canonical variate, which may be interpreted as a valence dimension. Sadness, but especially energy, show a lower loading on the first canonical variate, which indicate that they behave more independently from the other sets of variables. Thus, the two models correspond most strongly on the valence dimension, where the redundancies of the first canonical variable suggest that the dimensional scales explain more of the variance in the discrete affect scales than vice versa.

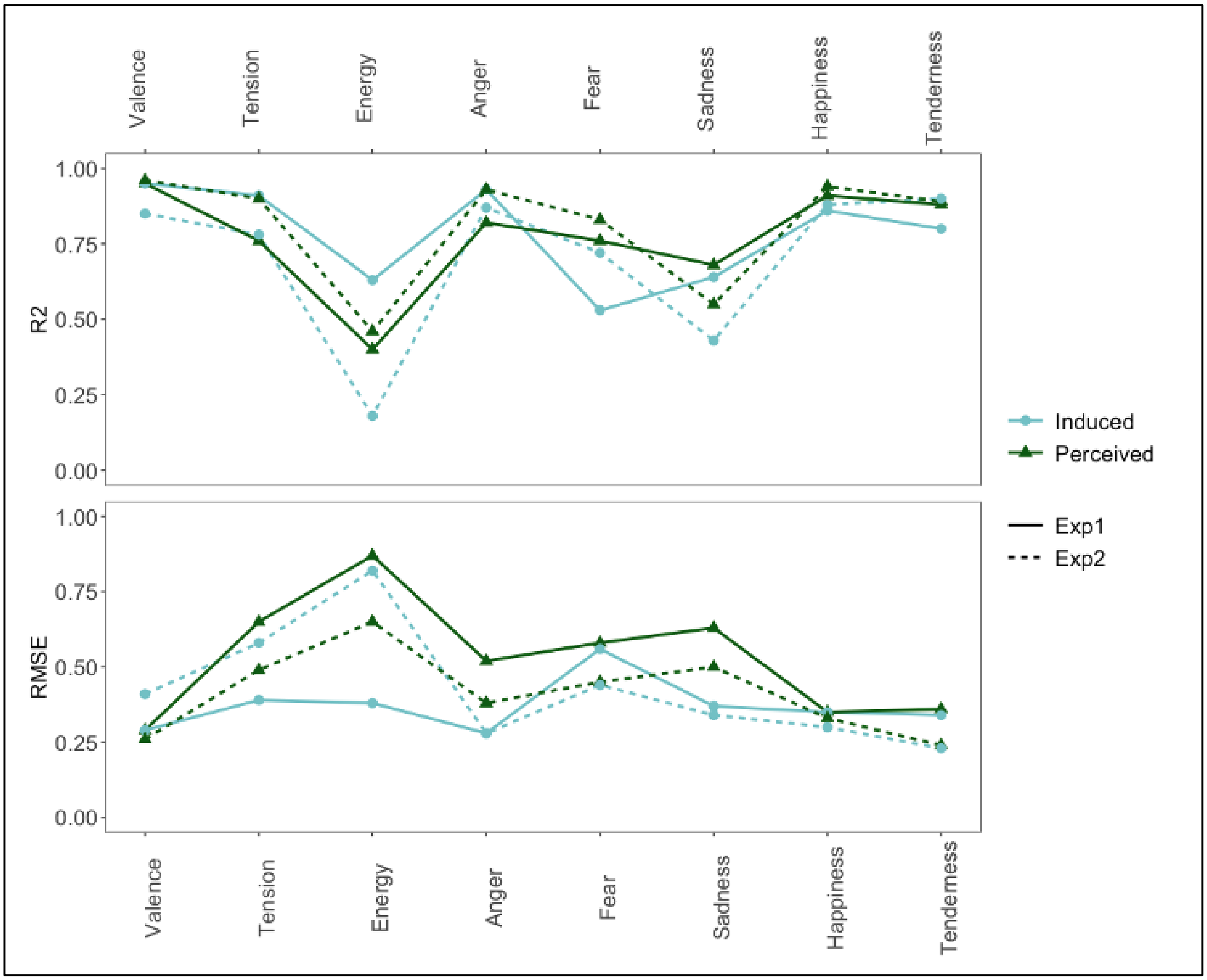
We conducted lasso regressions to further analyze how the dimensional and discrete affect scales correspond to each other. Lasso (**l**east **a**bsolute **s**hrinkage and **s**election **o**perator) regression conducts variable selection by applying penalties to the magnitudes of coefficients (shrinkage) and eliminating predictors whose coefficients are reduced to zero (Tibshirani, 2011). We used the discrete affect scales to predict each of the scales in the dimensional affect model to select which discrete scales were most important in predicting each of the dimensional affect scales. For example, all five discrete scales were predictors (independent variables) in the lasso regression

Table 3.5 *Significance, Correlation, and Variation Explained of Canonical Variates, as Well as Loadings and Redundancy of the Two Affect Models*

	CV 1	**CV 2**	**CV 3**
Wilk's lambda	$F(15, 229.5) = 39.1$, $p < .0001$	$F(8, 168) = 3.9$, $p < .001$	$F(3, 85) = 3.3$, $p < .05$
Can. Corr.	.98	.46	.32
Var Explained	.98	.01	.005
Valence	.98	–.09	–.15
Tension	–.87	.35	.35
Energy	.28	.39	.88
Redundancy	.58	.02	.03
Anger	–.93	.19	.30
Fear	–.86	.33	–.22
Sadness	–.74	–.63	–.17
Happiness	.97	.12	.19
Tenderness	.93	–.01	–.17
Redundancy	.76	.02	.005

predicting valence (dependent variable). We did the inverse by using the dimensional affect scales to predict each of the discrete affect scales. The affect scales that were selected by the lasso regression were then used in a linear regression to enable us to interpret the relative contribution of each scale (standardized coefficients). The 5-fold cross-validated model performance was obtained from linear regression models before lasso selection (i.e., including all three dimensional or all 5 discrete predictors) to allow an even comparison. Figure 3.4 shows the R^2 and *RMSE* values for each of the regression models. Especially for energy arousal, there is a dip in R^2 and a spike in *RMSE* indicating that the model predicting energy arousal performs relatively poorly, i.e., the discrete affect scales cannot fully capture or predict the dimensional scale of energy arousal. A similar, but smaller, trough and peak can be seen for the regression performance of the dimensional affect scales in predicting the discrete scale of sadness.

Figure 3.4 *The R^2 and RMSE Values for Each Regression Model Predicting Each of the Affect Scales, Separated by Experiment and Affect Locus*



Comparing the mean R^2 and *RMSE* as summarized in Table 3.6, we see that the differences in performance between group comparisons is relatively small, and sometimes behaves in opposite directions for R^2 and *RMSE* (comparing perceived and induced affect for Experiments 1 and 2), but both R^2 and *RMSE* show better performance for the models in which the dimensional scales are used to predict the discrete scales than vice versa. Regression models that contain more predictors generally show better model performance (higher R^2, lower RMSE) and consequently the discrete scales (five predictors) predicting the dimensional scales have an advantage over the dimensional scales (three predictors) predicting the discrete scales. It is thus noteworthy that in spite of that, the dimensional model is actually more successful at predicting the discrete affect ratings than vice versa.

Table 3.6 *Means and Standard Deviations for the Groups Summarizing Affect Locus, Experiment, and Model on R^2 and RMSE*

Variable	**Groups**	**R^2 (SD)**	**RMSE (SD)**
Affect Locus	Induced	.74 (.21)	0.40 (0.15)
	Perceived	.79 (.18)	0.47 (0.17)
Experiment	Exp. 1	.78 (.16)	0.45 (0.17)
	Exp. 2	.74 (.23)	0.42 (0.16)
Model	Dis → Dim	.73 (.26)	0.51 (0.21)
	Dim → Dis	.80 (.14)	0.39 (0.11)

Note. Dis → Dim summarizes the models where discrete affect was the independent variable and dimensional affect the independent variable, and Dim → Dis summarizes the inverse case.

Figure 3.5 visualizes the standardized regression coefficients of each model where the discrete affect scales predict the dimensional affect scales, separated by affect locus and experiment. Valence showed consistent results in the first and second experiments and also showed the best performance of the three dimensional affect scales (mean R^2 = .95, mean *RMSE* = .29). Induced valence was predicted by induced anger and happiness, whereas perceived valence was predicted by a combination of all perceived discrete affects, except for happiness. Induced tension was predicted by anger (Experiment 1) or happiness (Experiment 2), and perceived tension by a combination of fear, tenderness, as well as happiness in the first experiment. Finally, induced energy arousal was predicted by anger (Exps. 1 & 2 in opposite directions) and sadness (Exp. 2), and perceived energy arousal was predicted by a combination of all discrete affects, except sadness. The performance of the energy arousal model, however, was relatively poor (see Figure 3.4).

Figure 3.6 visualizes the inverse models, i.e., the standardized regression coefficients of each model where the dimensional affect scales predict each of the discrete affect scales, separated by affect locus and experiment. Fear was consistently predicted by valence in both affect loci and

Figure 3.5 *Standardized Regression Coefficients of Discrete Affect Predicting Each Dimensional Affect Scale, Separated by Affect Locus and Experiment*

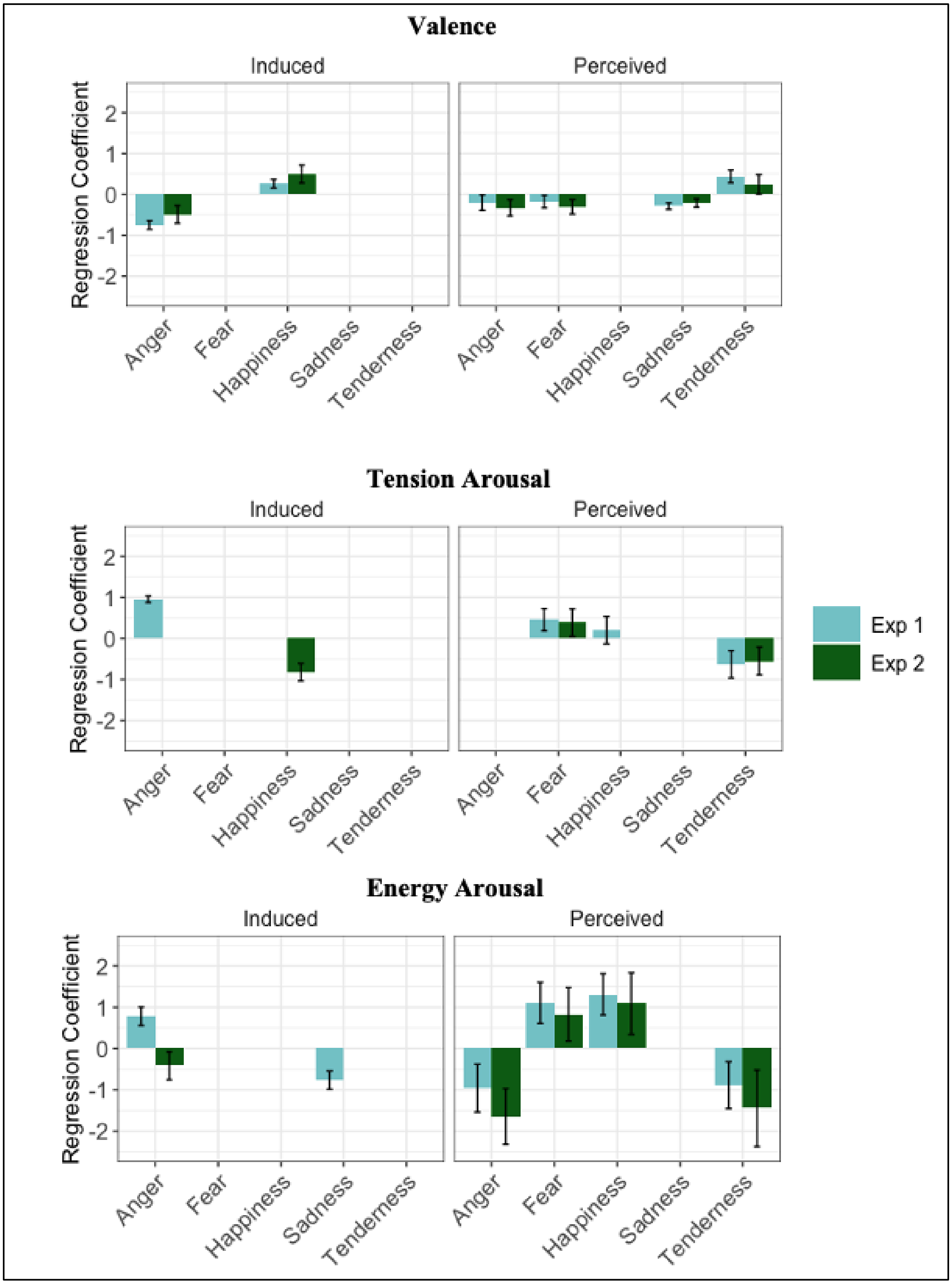
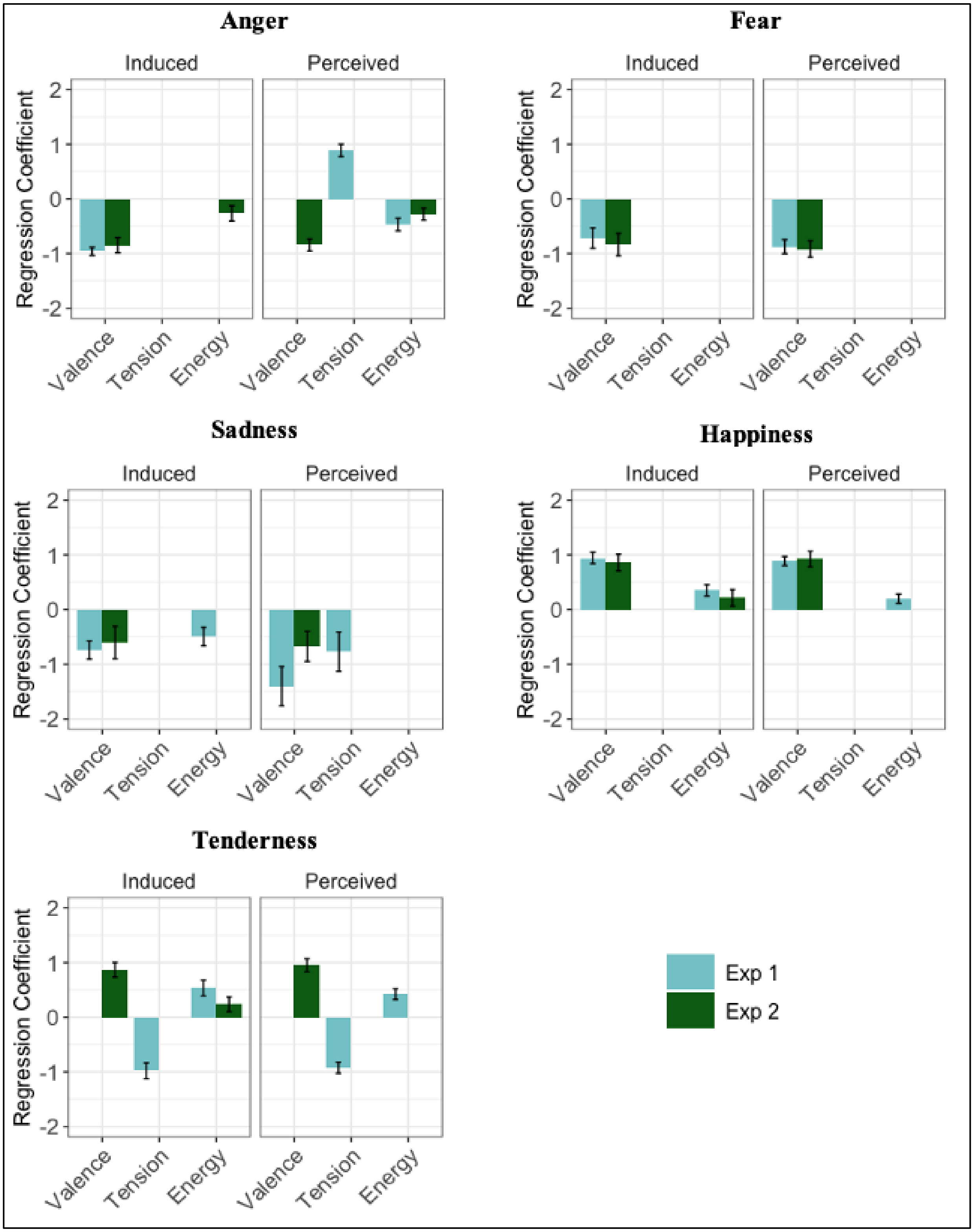


Figure 3.6 *Standardized Regression Coefficients of Dimensional Affect Predicting Each Discrete Affect Scale, Separated by Affect Locus and Experiment*



experiments (although performance is poorer for the induced model in Experiment 1, see Figure3.4). Valence was a significant predictor in most other models, except for tenderness only in the second experiment. Besides fear, the other discrete affects were predicted by a combination of two of the three dimensional affects. Which two dimensional affects were significant predictors, differed among the discrete affects, the affect loci, and the experiments, but any consistent patterns are not easily observed. The model for happiness showed the best performance of all the discrete scales (mean R^2 = .90, mean *RMSE* = .33). It was predicted by valence and energy arousal in most of the models, except for the perceived affect of the second experiment where it was only predicted by valence. Sadness showed the poorest performance of the discrete affect models (mean R^2 = .60, mean *RMSE* = .45): the dimensional scales of valence (Experiment 2), valence and energy (Experiment 1, induced), or valence and tension (Experiment 1, perceived) can neither fully capture nor predict the variation in sadness ratings.

To further explore the scales of energy arousal and sadness, which showed relatively poor performance in the regression models and appeared to behave more independently in the PCA, we visually mapped the discrete affects onto the dimensional affect scales and vice versa. For Figure 3.7, we took the stimuli that fell in the lowest (first) and highest (fourth) quantile of each of the discrete affects and calculated the mean valence and energy rating for those stimuli. Here, we can see that negative and positive valence are clearly distinguished by the discrete affects, as low happiness and tenderness, and high anger, fear, and sadness, are mapped onto the lower, more negative, end of the valence scale. High sadness, however, was not as negatively valenced as the other affects. Conversely, high happiness and tenderness, as well as low anger, fear, and sadness, are mapped onto the higher, more positive, end of the valence scale. For energy, however, such distinctions are much less extreme. The high and low extremes of the discrete affects, all hover around the midpoint of the energy arousal scale. Thus, any variation in energy arousal that was present in the affect ratings, was not captured by the discrete affect scales.

Similarly, for Figure 3.8, we took the stimuli that fell in the lowest (first) and highest (fourth) quantile of each of the dimensional affects and calculated the mean anger and sadness ratings for those stimuli. Here we see that the extremes of tension and valence map onto the extremes of the anger dimension. Energy hovers more around the midpoint of anger. Now for sadness, we do see more distinction than with energy arousal in Figure 3.7. Lower tension and more positive valence present a lower degree of sadness, whereas higher tension and more negative valence have a higher

Figure 3.7 *Mapping of the Stimuli in the First (Low) and Fourth (High) Quantile of the Discrete Affects onto the Valence and Energy Dimensions*

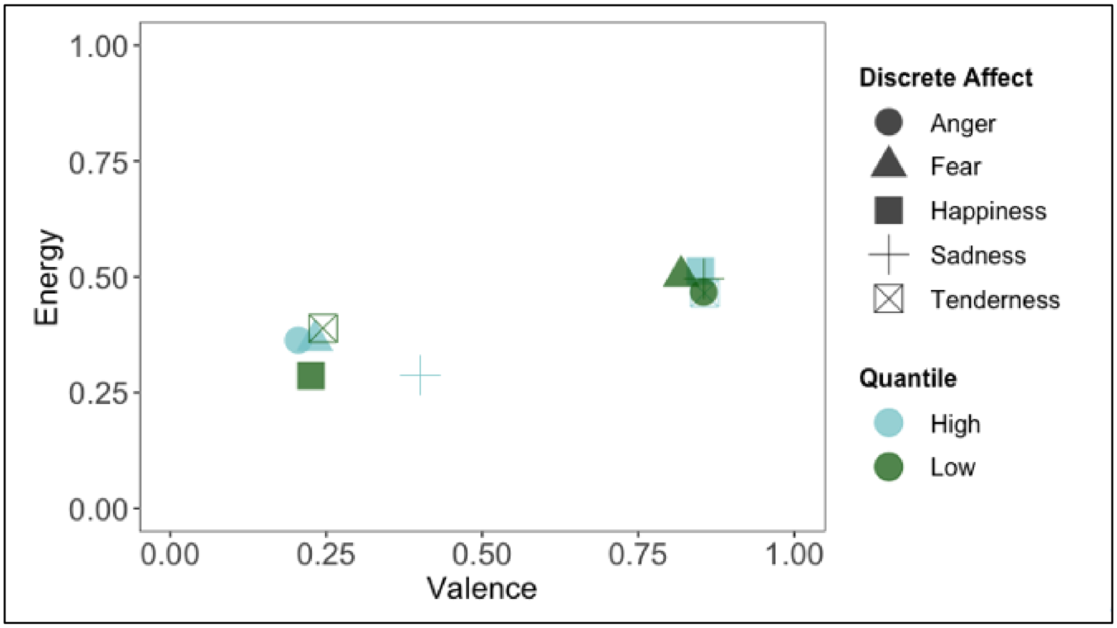
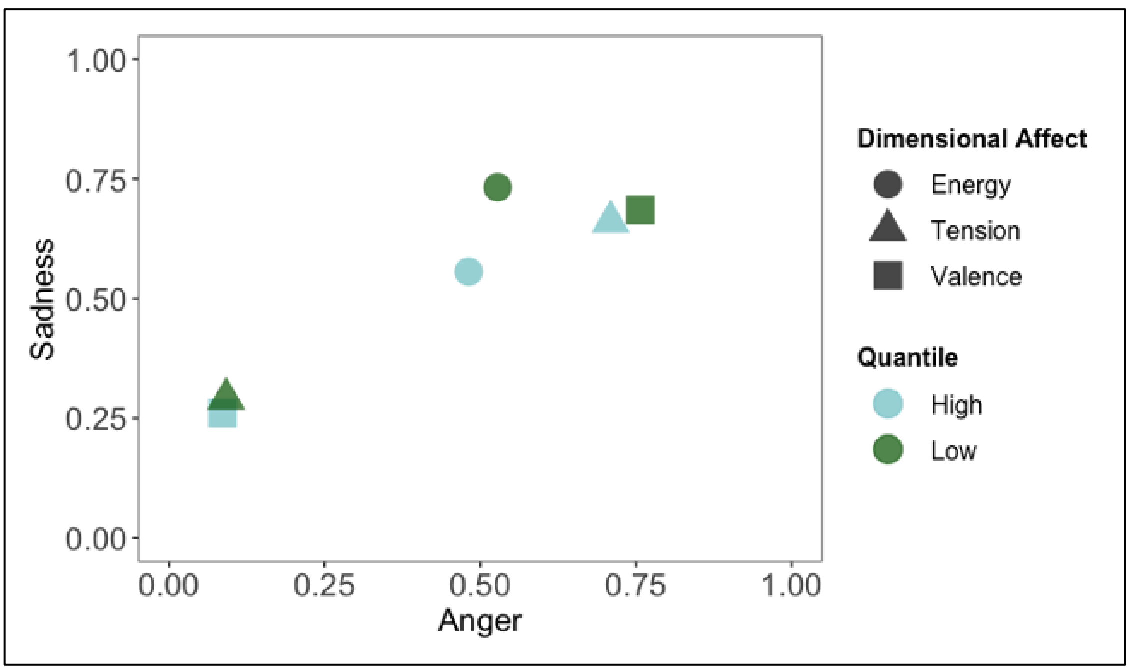


Figure 3.8 *Mapping of the Stimuli in the First (Low) and Fourth (High) Quantile of the Dimensional Affects onto the Anger And Sadness Categories*

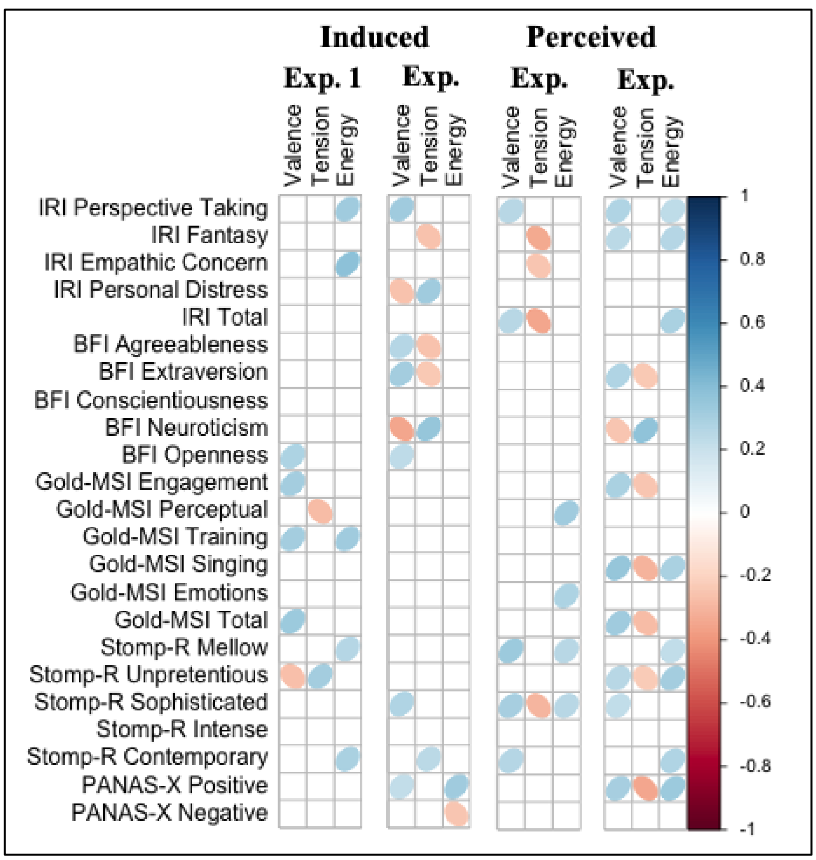


degree of sadness. Both high and low energy map onto the higher end of sadness, although low energy is more increased in sadness.

Individual Differences

We calculated Pearson correlations for each questionnaire score and sub-score with each of the affect scales, separated by experiment and affect locus. Figures 3.9, 3.10, and 3.11 visualize the correlations of each questionnaire score with the dimensional and discrete affect scales, as well as the preference scale, respectively. Table 3.7 summarizes the absolute strength and frequency of occurrence of the correlations of the questionnaires (sub-)scales and the affect ratings, separated by affect locus, experiment, and affect model. The first thing we note is that none of the correlations were particularly strong, ranging from $|r| = .23$ to $|r| = .44$. When we consider the frequency of significant correlations, we find that of all the questionnaire sub-scores, pre-existing

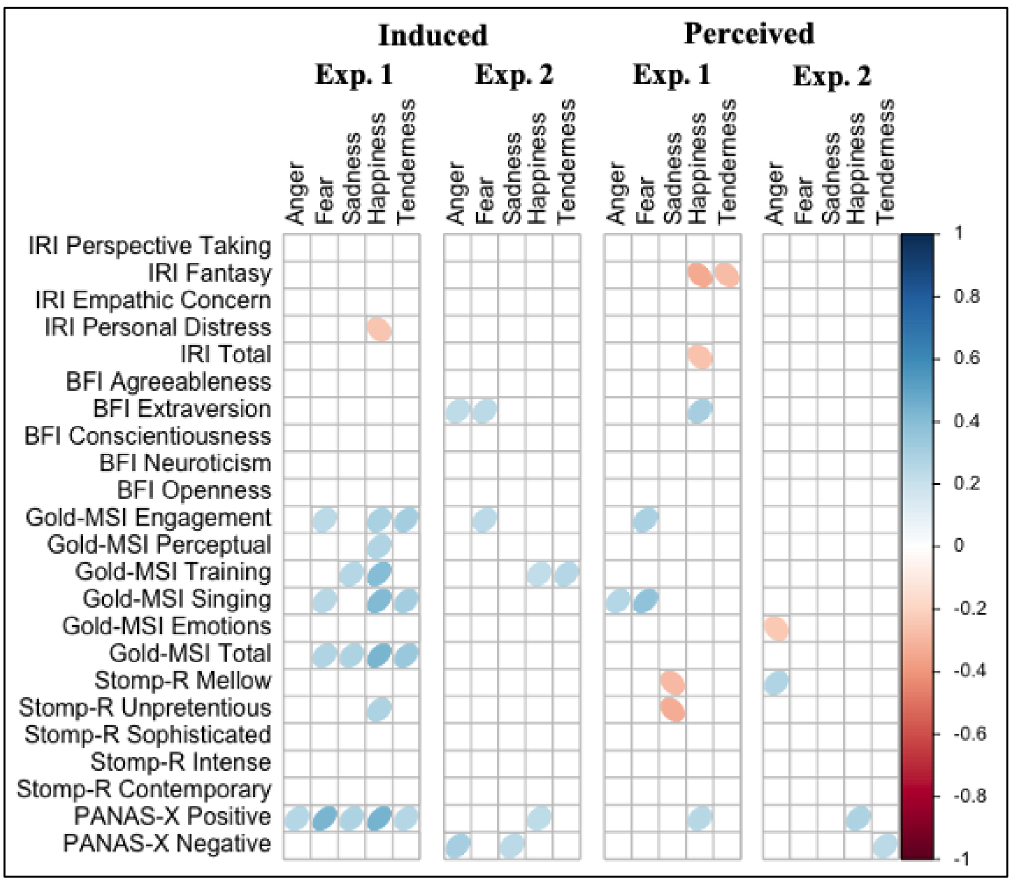
Figure 3.9 *Correlation Plot of Questionnaire Scores with Dimensional Affect Scales, Separated by Experiment and Affect Locus*



Note. See Figure 3.2.

positive mood (PANAS-X positive affect) was most frequently correlated with the affect scales. This was especially noticeable in the induced ratings of discrete affect in Experiment 1: positive mood was positively correlated to anger, fear, sadness, happiness, and tenderness. Note that the direction of correlation does not change, but rather positive mood was related to increased affect ratings overall. When we average the frequencies of the subscales of each questionnaire together, we again find that pre-existing mood was the most prevalent (PANAS-X; mean(*f*) = 9), followed by musical sophistication (Gold-MSI; mean(*f*) = 6.7), empathy (IRI; mean(*f*) = 6), musical preferences (STOMP-R; mean(*f*) = 5.2), and Big-Five personality (BFI; mean(*f*) = 3.4). When we consider each of the affect scales, we find that averaged over affect locus and experiment, valence most frequently correlated with the questionnaires (mean(*f*) = 7). There isn't one questionnaire (sub-score) that specifically stands out with respect to valence. Rather, depending on affect locus

Figure 3.10 *Correlation Plot of Questionnaire Scores with Discrete Affect Scales, Separated by Experiment and Affect Locus*



Note. See Figure 3.2.

and experiment, empathy, personality, musical sophistication, preferences, and pre-existing mood all correlated with valence in one way or another.

Further considering frequency and strength of correlation, we find that there was only a small difference in frequency between experiments, with a slightly higher frequency and correlation strength for Experiment 1 than Experiment 2 (see Table 3.7). Affect ratings in the induced condition also correlated slightly more frequently than ratings in the perceived condition, but the correlation strengths were the same. Finally, when we consider the two affect models (without preference), we find that the dimensional affect scales correlated more frequently with the questionnaires than the discrete affect scales, but between correlation strengths there is only a difference of .01.

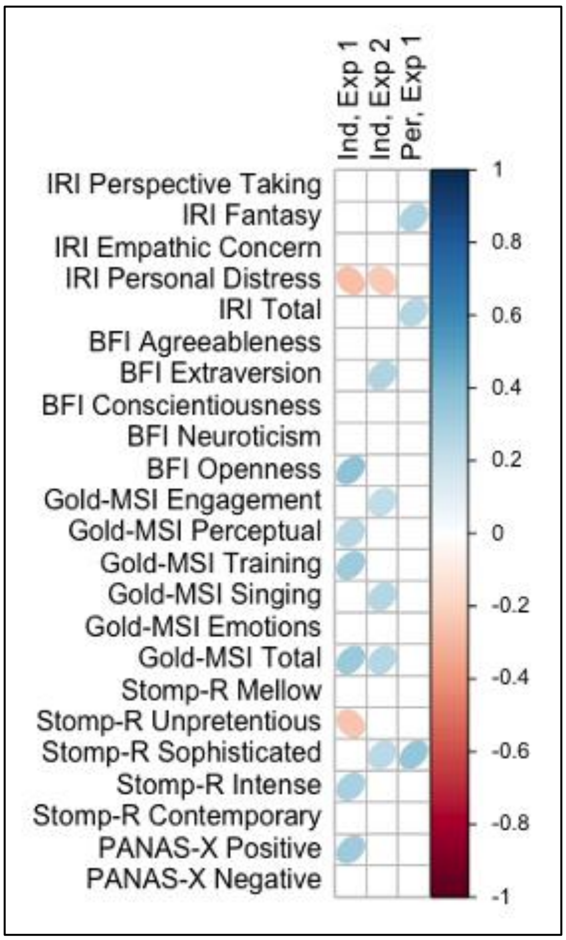
Table 3.7 *Average Absolute Strength and Frequency of Correlations Between Questionnaires and Affect Ratings Separated by Affect Locus, Experiment, and Affect Model*

Variable	**Groups**	***f***	**|*r*| *(SD)***
Affect Locus	Induced	3.9	.29 (.05)
	Perceived	3.2	.29 (.04)
Experiment	Exp. 1	3.7	.30 (.05)
	Exp. 2	3.5	.27 (.04)
Model	Dimensional	5.5	.29 (.04)
	Discrete	2.1	.30 (.06)

Note. Model groups exclude the preference ratings.

Taking the frequency and strength of correlations together, it appears that the experiment with single notes is especially susceptible to the influence of individual differences. Comparing the two affect models, the dimensional model showed more frequent correlations than the discrete model, and thus also appeared to be more susceptible to the influence of individual differences. We find that nearly all questionnaires were correlated with some affect scales, most frequently the pre-existing positive mood. A few more patterns are detectable in Figures 3.9 and Figure 3.10. For example, like PANAS-X Positive, several Gold-MSI scores also positively correlated with the discrete affect scales in the induced affect of Experiment 1, suggesting that increased musical sophistication also leads to increased intensity of induced affect. Note that we inverted the scores for tension; whereas a participants' higher rating reflected increased relaxation in the experimental procedure, for analysis we inverted the scores so that a higher rating reflected increased tension. Thus, the negative correlations with tension could be inverted to reflect relaxation, consequently rendering nearly all correlations as positive. That is, a higher score on most of the questionnaire (sub-)scores is associated with higher ratings of positive valence, relaxation, energy, anger, fear, sadness, happiness, and tenderness, perhaps indicating a form of affective reactivity regardless of affective content.

Figure 3.11 *Correlation Plot of Questionnaire Scores with Preference, Separated by Experiment and Affect Locus*



Note. See Figure 3.2.

Finally, we investigated the preference scale independently, as it is separate from our main focus on the dimensional and discrete affect models. First, we found that after valence, the preference ratings most frequently correlated with the questionnaire (sub-)scores, especially when participants were rating their induced affect. Figure 3.11 shows all the correlations with the preference scale. Increased scores of IRI Personal Distress were related to a decreased preference for the musical stimuli when rating induced affect, although IRI Fantasy and IRI Total when rating perceived affect were related to increased preference for the musical stimuli. That is, participants with the tendency to feel anxious and uneasy during tense interpersonal settings (i.e., Personal Distress) were less likely to enjoy the musical stimuli in both experiments, whereas participants with the tendency to transpose themselves imaginatively into the feelings and actions of fictitious others and who are generally more empathetic, were more likely to enjoy the musical stimuli in the first experiment. Personality traits of extraversion and openness were also related to increased preference in the induced condition. Several Gold-MSI scores were positively correlated with preference for the stimuli in the induced condition, as well as PANAS-X positive. Considering musical preferences (STOMP-R), preference for sophisticated music (jazz, classical, opera) and intense music (rock, punk, metal) were related to increased preference for our stimuli, but preference for unpretentious music (pop, country, religious) was related to a decreased preference for our stimuli.

3.4 Discussion

Based on the internal consistency and inter-rater reliability, we found that sadness scored the lowest and potentially was the most susceptible to individual differences compared to the other dimensional and discrete affect scales. Energy arousal and tenderness also scored low on both measures, but to a lesser extent than sadness. Vuoskoski and Eerola (2011a) found a similar pattern in their study on music-induced emotions, with lower inter-rater reliability for energy arousal, sadness, and tenderness. They also found that anger was less consistent than other scales, which in our induced condition scored a little bit lower as well. Overall, there was higher consistency and agreement on perceived than on induced affect, on single notes than on chromatic scales, and on dimensional than on discrete affect scales, although the differences are minor. Vuoskoski and Eerola (2011a) and Zentner et al. (2008) similarly found higher consistency and agreement for the dimensional than the discrete affect model.

Correlation analyses between affect scales indicated that dimension reduction in both the dimensional and discrete affect models was warranted. For the dimensional model, valence and tension were highly correlated and dimension reduction showed that two dimensions representing valence/tension and energy arousal explained most of the variation in the ratings. The inclusion of tension arousal as a third affect dimension has been a point of contention. Whereas some studies also find that valence and tension arousal are highly collinear (Eerola et al., 2012; Eerola & Vuoskoski, 2011; Krumhansl, 1997; Vuoskoski & Eerola, 2011a), interestingly, the study with stimuli and experimental design closest to our first experiment, did not find high collinearity between these two affect dimensions (McAdams et al., 2017; $r(135) = .46$). The key difference between the current experiments and the one by McAdams and colleagues are the participant pool and testing environment, i.e., university-based versus world-wide and in-lab versus online. Our findings on the correlation between energy and tension arousal are, however, more similar to McAdams et al.'s findings; we found no (significant or strong) correlation, whereas Eerola et al. (2012) did. McAdams et al. explain this incongruence by the lack of pitch variation in Eerola et al.'s 2012 experiment, which was present in McAdams et al.'s and the current experiments, where energy arousal appears to be mostly related to spectral variability due to changes in pitch. Furthermore, whereas Eerola et al. (2012) did find a strong correlation between valence and preference, as did the current study, McAdams et al. (2017) did not. Preference may be considered a fundamental manifestation of affect (Zajonc, 1980) and has been previously used as a substitute measure for valence, alongside arousal (Brown et al., 2004). The similarities and differences we find in results further argue for the systematic comparison of affect models in different experimental contexts, as there may not be a one-size-fits-all model appropriate for representing perceived and induced affect in music.

For the discrete affect model, anger, fear, happiness, and tenderness all correlated strongly with each other, and sadness did so to a lesser extent. Similarly, a PCA showed that two components representing anger/fear/happiness/tension and sadness explained most of the variation in the ratings. Although Eerola and Vuoskoski (2011) also found that most of the discrete ratings of perceived affect correlated with each other, the correlations were not as strong as in the present study. Nevertheless, they did find that most of the discrete affect categories could be mapped onto a single valence dimension, and sadness onto a second dimension. Similarly, in their study on induced affect, Vuoskoski and Eerola (2011a) found that most of the variance in the discrete affect model could be represented by two components of valence (or tension) and energy. Furthermore, when putting all affect scales together in the current experiment, again two components explained most of the variation in ratings, with the highest loading for valence on the first component, and energy arousal on the second component. Based on the correlations and PCA, we conclude that, be they dimensional or discrete, two components captured most of the variance in induced and perceived affect of single notes and chromatic scales.

Canonical correlations and predictive modelling showed that the correspondence of the two affect models is relatively high, in particular on the valence dimension. The regression modelling performance overall was higher for perceived than induced affect, especially in response to single notes. Thus, especially with short stimuli when assessing perceived affect, the dimensional and discrete affect models appear to be mostly capturing the same affective responses and may thus be considered interchangeable. Although the differences were small, the average model performance measures suggest that dimensional scales are better at predicting discrete scales than vice versa. Furthermore, the differences in redundancy of the canonical covariates indicate that a dimensional model of affect is capturing more variance in the discrete affect ratings than vice versa. This suggests that the dimensional affect model is more appropriate for capturing the affective response to music, which is concurrent with previous dimensional and discrete comparisons in perceived and induced affect (Eerola & Vuoskoski, 2011; Vuoskoski & Eerola, 2011a). Prediction error was highest when the discrete scales predicted energy arousal, followed by the dimensional scales predicting sadness. Further investigation of those scales revealed that energy in particular varied in a manner that was not captured by any of the discrete affect scales. Although the same was true for sadness, valence and tension still appeared to capture some of the variation that was present in the sadness ratings, whereas both low and high energy were associated with increased sadness.

Interestingly, recent studies on sadness suggest that there are different kinds of musical sadness, which may be distinguished in their energetic level; melancholy (low energy) versus grief (high energy; Warrenburg, 2020). In the current study, sadness also appeared to be loading on both principal components, as perhaps a combination of valence and energy. Additionally, Eerola and Vuoskoski (2011) found that sadness did not correlate with valence, and although these two scales did correlate in the present study, the sadness scale showed the weakest correlations of all the discrete scales overall. Sadness also correlated negatively with preference, but this was less strong than the correlation of preference with other unpleasant affects such as tension, anger, and fear.

We were not able to directly test the role of individual differences in the preference for sadness, a topic of particular interest in musical affect research (Eerola et al., 2018), as sadness and preference were rated by different groups of participants. Our results and previous studies suggest that musical sadness is not necessarily associated with negative valence and behaves more independently from the other discrete affect measures (Bigand et al., 2005; Kreutz et al., 2008).

Finally, our exploration of the role of individual differences showed that all the measures correlated with affect ratings in one way or another, with a moderate correlation strength at most. Of all the affect scales, valence was significantly correlated the most frequently with the measures, indicating that it was the most susceptible to influence by individual differences. Comparing the dimensional and discrete models, on average dimensional affect scales correlated more frequently, also indicating they are more susceptible to influence by individual differences. Similarly, the first experiment showed more frequent and stronger correlations than the second experiment, suggesting that a short exposure time allowed for more individual variability. This contrasts our findings on internal-consistency and inter-rater reliability as discussed above, which suggested that the discrete model, the second experiment, and specifically the scales of sadness, energy, and tenderness, would be more susceptible to individual differences. The differences in consistency and agreement perhaps do not reflect the individual differences we measured but may be caused by other sources of individual differences. Alternatively, these differences may rather suggest that the participants lacked conceptual clarity to consistently rate their responses. That is, they may not have fully understood or agreed on what it means to perceive or feel, for example, tenderness in (response to) musical stimuli. After valence, preference was most influenced by individual differences from all included sources (pre-existing mood, dispositional empathy, personality, musical sophistication, and musical preferences).

Pre-existing positive mood most frequently influenced the affect ratings. Interestingly, this influence was unidirectional; a positive mood led to higher ratings on most scales, suggesting a higher emotional reactivity, or intensity, overall. Previous research has also shown that extraversion, related to experiencing more positive affects, and pre-existing mood were related to affective processing or intensity (Vuoskoski & Eerola, 2011a, b). Following positive mood, several musical sophistication scores were also frequently correlated with the affect ratings, mostly unidirectionally. Whereas some studies found little or no difference in affective response between musicians and nonmusicians (Bigand et al., 2005; Frego, 1999; Filipic et al., 2010), other studies did (Egermann & McAdams, 2013; McAdams et al., 2017). Part of this discrepancy may be a result of the ambiguous definition of musicianship. The broader and more inclusive use of musical sophistication in the current study does corroborate the importance of musical expertise in the affective response to music, although musical sophistication may not be directly related to degree of formal musical training. After pre-existing mood and musical sophistication, we found that dispositional empathy, musical preferences, and Big-Five personality were most frequently correlated with the affect scales. The manner in which individual differences correlated varied between affect scales, experiments, and affect loci, which makes it difficult at this point to hypothesize about the underlying mechanisms connecting individual differences and affective responses. This does again indicate, however, that the experimental context can influence one's findings. Furthermore, these results show that many sources of individual differences may influence the affective response to music.

3.5 Conclusion

The aim of this study was to compare performance of the discrete and dimensional affect models, measuring perceived and induced affect, in response to relatively short, affectively ambiguous, musical sounds, as tested in an online environment. We found that perceived affect, single notes, and the dimensional affect model all performed slightly more consistently than induced affect, chromatic scales, and the discrete affect model. Both affect models' scales could be reduced to two components. This indicated that regardless of one's theoretical preference for the discrete or dimensional affect model, the inclusion of tension arousal or all five discrete categories is superfluous in a similar experimental context to the one used in this study. Furthermore, whereas the dimensional and discrete affect models appeared to be relatively interchangeable, energy arousal varied in a manner that was not captured or predicted by the discrete affect model. Sadness was also shown to play a distinct role from the other discrete affects but was more effectively captured by the dimensional affect scales than energy arousal was captured by the discrete affect scales. The models showed highest correspondence on the valence dimension. Future research may further explore the role of energy compared to discrete affects (Warrenburg, 2020) and how it relates to changes in pitch height (McAdams et al., 2017). Finally, our exploration of individual differences showed that pre-existing mood, dispositional empathy, Big-Five personality, musical sophistication, and musical preferences, all played a role in the affective response to music. Future studies may thus either consider these factors in their investigation of the mechanisms underlying musical affect or ensure their participant population is appropriately represented on those factors to ensure generalizability.

Chapter 4
Effects of Instrument Family and Pitch Register on Perceived and Induced Affect

4.1 Introduction

A number of different musical factors may be involved when musical affect is perceived or induced. Many studies on perceived or expressed affect, have focussed on factors such as mode, rhythm, tempo, or melody, which describe the global musical structure of music (for a review, see Gabrielsson & Lindström, 2010). Concerning induced affect, studies that investigate the effect of musical features are scarcer. They either concern structural music features similar to those in perceived affect studies (Gomez & Danuser, 2007; Kawakami et al., 2013) or even more global features such as musical passages that were structurally analyzed post-hoc (Guhn et al., 2007; Sloboda, 1991) and pre-determined orchestral gestures (Goodchild et al., 2019).

More local, nonstructural, features can also influence perceived affect. Listeners have been able to consistently judge the affective content of extremely short sound samples, that do not provide any or much structural musical information. At 250 ms, listeners can tell whether an excerpt is emotionally moving or neutral (Filipic et al., 2010). At 500 ms, listeners are able to distinguish between happy and sad excerpts (Peretz et al., 1998) and show high internal consistency when judging perceived valence, tension arousal, and energy arousal (McAdams et al., 2017). At 1 s, listeners who were asked to judge the *induced* emotional dissimilarity showed similar patterns as they did in response to 30 s excerpts (Bigand et al., 2005).

In the case of induced affect judgments of 1-s excerpts, one can raise the question of whether listeners were in fact able to distinguish between perceived and induced affect, or whether they were just reporting their perceived affect, or a combination of both affect loci. A pilot experiment by Bigand et al. (2005) showed that listeners' ratings of perceived affect were highly correlated with their ratings of induced affect. However, other self-report studies have shown that although perceived and induced affect are mostly positively related, they can show some minor differences (for a review, see Schubert, 2013). The validity of self-reported induced affect measures is often a point of discussion, although it remains one of the most direct methods to investigate a participants' subjective experience. See Chapter 1.1 for a more detailed discussion of affect quantification.

Timbre is one of the nonstructural acoustic features that may be recognized in very short musical excerpts (Robinson & Patterson, 1995), and may thus influence listeners' affective judgments. Timbre is often described as the sound quality (or "colour") that allows one to distinguish between two sound sources (e.g., musical instruments) when they are playing at the same pitch and loudness. However, an instrument does not consist of one single timbre, but rather a "constrained universe of timbres that covary with other musical parameters" (McAdams & Siedenburg, 2019, p. 71). In fact, a single instrument's perceived timbre can change with variation in pitch, dynamics, and playing technique (Marozeau et al., 2003; McAdams & Goodchild, 2017b; Risset & Wessel, 1999).

In a review of the musical features that are most commonly associated with the expression of discrete affect, Juslin and Laukka (2004) note that a bright timbre is associated with happiness, a dull timbre with sadness, a sharp timbre with anger, and a soft timbre with both fear and tenderness. Huron et al. (2014) investigated the acoustic properties of sadness asking one group of participants to judge the acoustic properties of a set of Western instruments, and another group of participants to judge the ability of that same set of instruments to convey sadness. As a result, the authors found that several percussion instruments, as well as the banjo, were considered least able to convey sadness. String instruments were the most able to convey sadness, which was associated with timbral darkness. This experiment, however, tested the knowledge and learned associations of the participants, instead of the affective effect of the instruments' timbres in sound events. Hailstone et al. (2009) used four existing instruments as well as four newly synthesized instruments to study perceived happiness, sadness, fear, and anger. They found that both groups of instruments were associated with different affect perceptions, and thus showed that the affective effect of timbre goes beyond simple learned associations. In a forced-choice paradigm with only nonsustained instruments, Chau and Horner (2015) found that the harpsichord, marimba, vibraphone, and xylophone were perceived as more positive; guitar, harp, and plucked violin perceived as more negative; and piano was found to be more emotionally neutral.

For induced affect, studies on timbre are scarcer, they focus on other features in addition to timbre, and they measure the presence of chills or self-reported affective intensity. Sloboda (1991) found that musical passages that were especially chill-inducing featured textural changes.

Similarly, Guhn et al. (2007) found that textural changes, specifically the change from full orchestra to solo (or the reverse), induced chills in their participants. Goodchild et al. (2019) systematically investigated such textural changes, or orchestral gestures, and found that the sudden-addition gesture induced the most chills. Reductive and additive gestures resulted in increased feelings of intensity. To the best of our knowledge, there are no studies looking directly at the effect of specific instruments or instrument families on induced affect.

Pitch height or register, another musical property that can be assessed in very short sounds (although not as quickly as timbre; Robinson & Patterson, 1995), has also been directly related to perceived affect. Eerola et al. (2013) compared several musical properties and found that a higher pitch register was related to higher perceived happiness, whereas a lower pitch register was related to higher perceived fearfulness. They also reported a slight quadratic trend of register for perceptions of scariness, sadness, and peacefulness. Timbre was also assessed and contributed relatively little to perceived affect alongside other factors such as mode, tempo, and pitch register. The timbre factor consisted of the comparison between only three instruments, selected for their differences in perceived brightness – flute, French horn, and trumpet. Timbre contributed the most to the perception of peacefulness, with flute expressing the most peacefulness, trumpet the least, and the French horn falling in between. Ilie & Thompson (2006) looked at both music and speech excerpts; whereas lower pitch was associated with higher perceived pleasantness in music, the reverse was true for speech. In their review, Gabrielsson and Lindström (2010) also found some apparent contradictions with regards to pitch height. A high pitch level has been associated with a range of emotions of a different nature; graceful, serene, happy, and joy, but also anger, fear, and increased tension arousal.

McAdams et al. (2017) studied the effect of instrument family and pitch register on perceived affect to investigate timbre and the pitch-related changes that may occur in timbre. They compared the perceived affect of musician and nonmusician listeners on tones that varied in instrument family, pitch register, attack strength, and playing technique. Neither attack strength nor playing technique were statistically significant predictors of any of the affect ratings. However, the authors did find an inverted U-shaped (concave) register trend for valence, meaning that middle registers were perceived to be most positive, except for percussion with showed a linearly increasing register trend. Overall, strings were perceived as the most positive, and nonmusicians perceived more negative valence than musicians in the lower registers. In a similar, but opposite, fashion, there was a U-shaped (convex) register trend for tension arousal, with the middle register rated as least tense, except for percussion which did not show a register effect for musicians. Brass was overall perceived as the tensest instrument family. Energy mostly showed a linearly increasing register trend, where instrument families overall behaved similarly, except in the lower registers for nonmusicians, who rated percussion as the most awake and brass and woodwinds as the least awake.

The current study was designed to be a direct follow-up to McAdams et al. (2017). The findings will be extended in five ways. Firstly, we did not only look at perceived affect, but also self-reported induced affect, to investigate the ways in which the two affect loci might overlap or diverge. We expect that register and instrument family effects may be less pronounced in induced affect, and some more unpleasant affects may be more strongly perceived than induced (e.g., Kawakami et al., 2013).

Secondly, in our first experiment the sound samples were nearly identical to the samples in McAdams et al. (2017), but in our second experiment they consisted of longer, multi-tone sound samples that are a (slightly) more ecologically valid representation of music. We expect that the longer stimuli will cause some effects to be more pronounced, as there is more time for differences to become apparent. On the other hand, the affect locus effect may decrease, as there is more time for any affects to be induced and approach the levels of perceived affect.

Thirdly, although McAdams et al. used dimensional affect scales (valence, tension arousal, and energy arousal), we also included discrete affect scales (anger, fear, sadness, happiness, and tenderness). In the debate surrounding the best way to quantify perceived and induced affect, we want to consider that the experimental context, such as stimulus characteristics or affect locus, matters. Chapter 3 deals with this research topic in depth.

The fourth way in which we extend the experiment of McAdams et al. (2017) is by considering several sources of individual differences as moderating factors in the effect of instrument family, pitch height, and the affect locus. We considered factors that have previously been associated with musical affect: musical expertise, empathy, personality traits, and pre-existing mood (Balteș & Miu, 2014; Egermann & McAdams, 2013; Garrido & Schubert, 2011; Juslin et al., 2008; Ladinig & Schellenberg, 2012; McAdams et al., 2017; Vuoskoski & Eerola, 2011b). We also consider musical preference, which has not been directly considered in relation to musical affect to the best of our knowledge. We argue that a general preference for certain musical genres will influence the affective response to stimuli that are most prevalent in classic orchestral music. We took an exploratory approach to uncover potential moderators and provide future starting points for studies that explore the role of the music listener in understanding musical affect.

Fifth and finally, these experiments were conducted in an online setting, which allows people from a wider variety of cultural and socio-economic backgrounds to participate and provides a more ecologically valid listening environment. This should make our findings more representative of the general population, although there is an obvious lack of experimental control compared to the lab experiments. See Chapter 1.4 for a more in-depth review of online versus lab experiments.

We did not look at attack strength or playing techniques, as they did not show an effect in McAdams et al. (2017). Musicianship was also not included as a categorical predictor, although we do analyze another form of musical expertise, musical sophistication (Müllensiefen et al., 2013), as one of the individual factors.

To summarize, in this project we investigate the effect of pitch register, instrument family, and affect locus on the affective response to music. Hereby we consider both single-note stimuli as used in previous experiments, and longer chromatic scales to approach a more musical stimulus set and allow for longer exposure without introducing variations in musical mode or tempo. The affective response is measured in both dimensional and discrete affect self-report representations, given that both models feature prominently in musical affect research and may not be mutually exclusive. Finally, we will explore how individual differences moderate the effects of pitch register, instrument family, and affect locus.

4.2 Analysis

We performed polynomial mixed-effects analyses with pitch register, instrument family, and affect locus as independent variables to predict each of the affect ratings. We included the polynomial effects because previous research also found quadratic relationships between pitch register and affect (e.g., McAdams et al., 2017). The models are fit using orthogonal polynomials, which reduces the correlation, as much as possible, between the quadratic and the linear trends. Mixed effects analyses include variables as both fixed and random effects. We included random intercepts for participants and stimuli, as well as random register slopes for participants. Specifying these random effects allows us to consider the random variance that may be associated with, for example, individual differences, instead of combining these sources of random variation together in one error term. Furthermore, we aimed to fit maximal random effects, with random slopes for as many independent variables as possible, because the sole inclusion of random intercepts has been shown to lead to high Type I error rates (i.e., false positives; Barr et al., 2013; Schielzeth & Forstmeier, 2009). However, we did not include random effects for instrument family, as this resulted in singular models, which indicated that the models may be overfitted with maximal random effects (Bates et al., 2018). We compared models with and without three-way interactions (register × instrument family × affect locus) because models that include three-way interactions (even if they are statistically non-significant) affect the two-way interactions, which can result in Type I and Type II errors (false positives and false negatives). We also compared models that included only linear register effects, or both quadratic and linear register effects. Based on the Akaike information criterion (AIC), Bayesian information criterion (BIC), and log likelihood, we selected which of the four models showed the best fit: models with or without quadratic register effects, and with or without three-way interactions. With the selected model, we continued our Bonferroni-corrected post-hoc comparisons.

A measure of variance explained as R^2 is not straightforward in mixed-effects modelling, because there is more than one error (variance) term, which shows more complex behaviours. Several pseudo-R^2 methods have been developed for mixed-effects modelling that approach the conventional R^2. Here, we report the Nakagawa pseudo-R^2 values consisting of the marginal R^2 (proportion of variance explained by the fixed effects) and the conditional R^2 (proportion of variance explained by both fixed and random effects; Nakagawa et al., 2017). Note, however, that these pseudo-R^2 values are not directly comparable to R^2. For example, unlike in regular regression, adding variables may actually reduce pseudo-R^2, depending on how it is calculated. Instead, the pseudo-R^2 measures can be used to compare the different affect models in this chapter to each other.

To analyze the effect of individual differences as moderators for the effects found in the polynomial mixed-effects models, we wrote a custom script that iteratively tested each potential moderator (individual difference) on each main or interaction effect of pitch register, instrument family, and affect locus, on each affect scale. Given that there is no commonly agreed upon procedure to correct for Type I error (false positive) rates in model selection (as opposed to post-hoc testing), we took a highly conservative approach for flagging relevant moderators. Potential moderators were only flagged when their inclusion was statistically significant at $p < .01$ and improved the model performance by a decrease of AICc (i.e., AIC with a correction to prevent overfitting) larger than 10. We included all 23 questionnaire (sub-)scores, except for the Gold-MSI instrument question, because it was an open-ended question not suited for this analysis, and the Gold-MSI genre question because genre preferences were already represented by STOMP-R. Demographic data of age, gender, education, country of residence, and country of formative years were also not included. The final selection of potential moderators consists of the IRI total score and subscores, the BFI subscores, the Gold-MSI general and subscores, the STOMP-R subscores, and the PANAS-X positive and negative affect scores.

All analyses were done in R version 4.2.1 (www.r-project.org). Modeling was done with the *lmer* and *anova* functions from the *lmerTest* package (Kuznetsova et al., 2020). Post-hoc analyses were done with the *emtrends* and *emmeans* functions from the *emmeans* package (Lenth et al., 2022). The Nakagawa pseudo-R^2 measures were calculated with the *r2* function from the *performance* package (Lüdecke et al., 2022). AICc was calculated with the *MuMIn* package (Bartoń, 2022). Finally, data were visualized with the *ggplot* function from the *ggplot2* package (Wickham et al., 2022).

4.3 Results: Pitch Register, Instrument Family, Affect Locus

As became apparent in Chapter 3, several affect scales were highly intercorrelated (see p. 29 for the correlation table). Among dimensional scales, valence and tension arousal were highly correlated in both experiments, $r_{Exp1}(57) = –.93$, $r_{Exp2}(30) = –.96$. Several of the discrete scales were highly correlated, with the average absolute inter-correlations between anger, fear, happiness, and tenderness at $|r| = .88$ ($SD = .04$). Sadness only showed a similarly high correlation to happiness, $r_{Exp1}(57) = –.83$, $r_{Exp2}(30) = –.79$, and more moderate correlations to the other scales, $|r|_{average} = .55$, SD = .07. Table 4.1 shows the explained variance for the models of each scale in each experiment (Nakagawa pseudo R^2). The marginal R^2 (variance explained by the fixed effects) shows us that overall the difference between scales and experiments is minimal (3–9%). Comparison of the marginal and conditional R^2 (variance explained by both fixed and random effects) reveals that a large proportion of the variation in scores can be explained by random effects, which reflect factors such as participant idiosyncrasy. This motivates closer investigation of individual differences.

Table 4.1 *Nakagawa Pseudo-R^2 for Each Affect Model*

	Marginal R^2		**Conditional R^2**	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**
Valence	.20	.23	.51	.52
Tension	.17	.21	.50	.45
Energy	.15	.11	.41	.39
Anger	.15	.24	.53	.54
Fear	.16	.13	.52	.44
Sadness	.16	.08	.50	.42
Happiness	.15	.14	.52	.49
Tenderness	.08	.10	.50	.50
Preference	.18	.19	.58	.60

Table 4.2 shows which main and interaction effects are statistically significant for each affect model (see Table A1 in Appendix A for the full statistics). It specifies the significance level as well as the consistency of significance between experiments. This also shows us that although the highly correlated scales show similar significance patterns, there are some minor differences in which effects are significant or not. Compare, for example, the results of happiness and tenderness, or anger and fear. In blue asterisks we marked the significant findings that were consistently found in both experiments, demonstrating that for all affect scales, there was a consistent significant main effect of register, either linear, quadratic, or both. For valence, anger, and fear, we see a strongly significant linear register trend in addition to the quadratic register trend in the second experiment (chromatic scales) which is not significant in the first experiment (single notes). The main effect of instrument family was consistently significant for all affect scales except energy arousal (Exps. 1 & 2) and anger (Exp. 2). Affect locus was only a consistent main predictor for fear, sadness, and tenderness. It was also more often a significant main predictor in Experiment 2 than in Experiment 1. However, even when it was not a significant main predictor, it featured in an interaction effect. Next, we will analyze the post-hoc results for each of the affect scales in more detail. Table A2–A5 in Appendix A list the full results of the pairwise comparisons.

Table 4.2 *Significance Results from Polynomial Mixed-Effects Models for Each Affect Scale and Experiment*

	Valence		**Tension**		**Energy**	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**	**Exp. 1**	**Exp. 2**
Register–linear (R_{LN})	–	••••	*	****	****	****
Register–quadratic (R_{QD})	**	****	***	****	–	–
Family (F)	****	***	****	****	–	–
Affect Locus (AL)	–	••••	–	••••	–	•
R_{LN} x F	–	•••	–	••••	•••	–
R_{QD} x F	–		–	•	–	–
R_{LN} x AL	****	****	••	–	•••	–
R_{QD} x AL	–	–	–	–	–	–
F x AL	–	–	••	–	–	–
R_{LN} x F x AL	••	–	•	–	••	–
R_{QD} x F x AL	••	–	•	–	–	–
	Anger		**Fear**		**Sadness**	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**	**Exp. 1**	**Exp. 2**
R_{LN}	–	••••	–	••••	****	**
R_{QD}	****	****	****	****	–	–
Family (F)	•••	–	**	*	****	****
Affect Locus (AL)	–	••••	****	****	****	****
R_{LN} x F	–	–	***	****	–	–
R_{QD} x F	–	–	–	–	–	–
R_{LN} x AL	***	****	–	–	•••	–
R_{QD} x AL	****	**	–	–	••	–
F x AL	••••	–	****	*	****	**
R_{LN} x F x AL	–	–	•••	–	–	–
R_{QD} x F x AL	–	–	–	–	•	–

Table 4.2 (continued)

	Happiness		Tenderness		Preference	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**	**Exp. 1**	**Exp. 2**
R_{LN}	***	****	*	****	*	*
R_{QD}	*	**	**	****	*	****
Family (F)	****	***	**	***	****	****
Affect Locus (AL)	–	–	*	****	–	n/a
R_{LN} x F	–	••••	–	–	–	–
R_{QD} x F	–	–	–	–	–	–
R_{LN} x AL	**	***	–	–	–	n/a
R_{QD} x AL	–	–	–	–	–	n/a
F x AL	••	–	–	–	•••	n/a
R_{LN} x F x AL	–	–	–	–	–	n/a
R_{QD} x F x AL	–	–	–	–	–	n/a

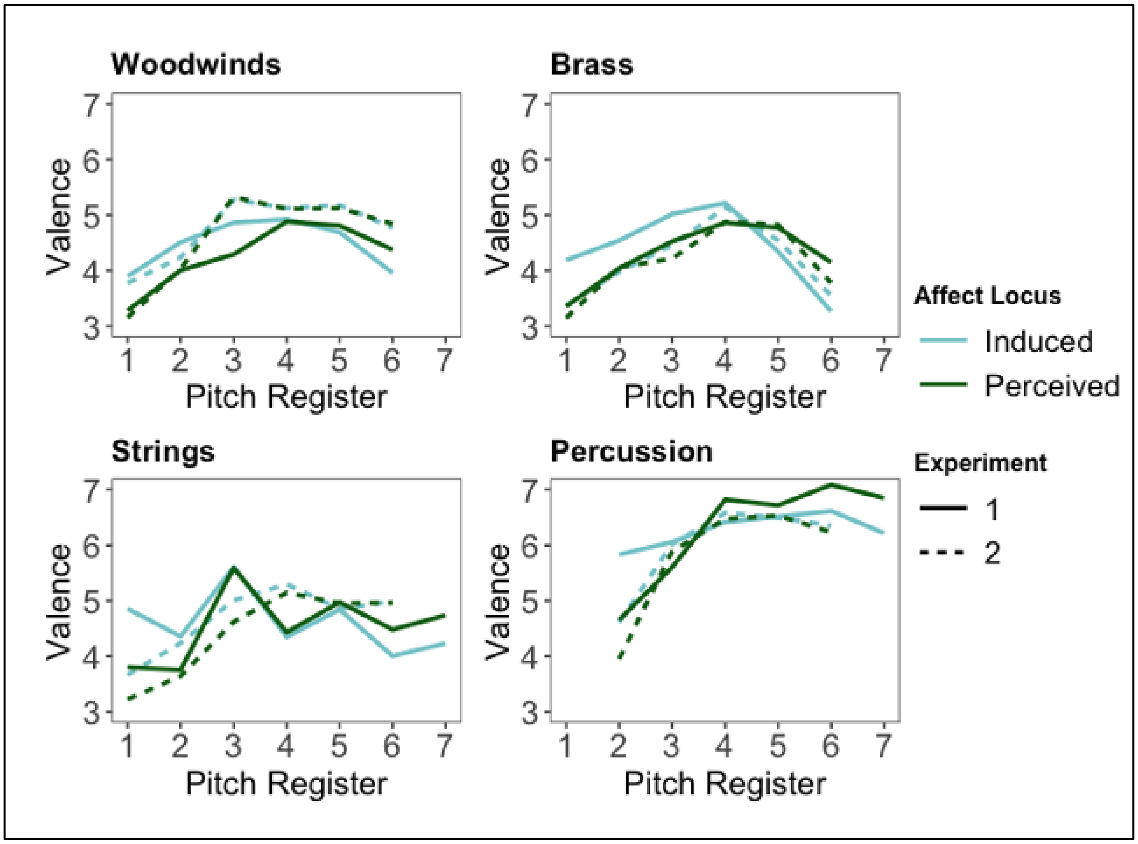
Note. • $p < .05$; •• $p < .01$; ••• $p < .001$; •••• $p < .0001$. Black dots (•) indicate the effect is statistically significant in one experiment and blue asterisks (*) indicate significance in both experiments. In Experiment 2, preference was only rated in the induced affect condition, so AL and its interactions cannot be considered as predictors (n/a).

Valence

In both experiments, valence shows a significant concave register effect as is shown by the inverted U-shapes in Figure 4.1. In Experiment 2, this register effect is also linearly increasing, where the higher registers are rated as more positive than the lower registers. In both experiments, there is an effect of instrument family with percussion as the most positive instrument family (M_{Exp1} = 6.5, SE_{Exp1} = 0.32; M_{Exp2} = 6.3, SE_{Exp2} = 0.24). In Experiment 2, woodwinds are also significantly more positive than brass ($M_{woodwinds}$ = 5.2, $SE_{woodwinds}$ = 0.22; M_{brass} = 4.7, SE_{brass} = 0.23). Again in the second experiment, percussion show a stronger positive linear register effect than the other instrument families ($b_{percussion}$ = 0.2, $SE_{percussion}$ = 0.11). A main effect of affect locus was also only found in Experiment 2, with induced valence rated as slightly more positive than perceived valence overall ($M_{induced}$ = 5.3, $SE_{induced}$ = 0.15; $M_{perceived}$ = 5.2, $SE_{perceived}$ = 0.15). In both experiments, perceived valence also showed a stronger linear register effect than induced valence. Finally, a three-way interaction in Experiment 1 reveals that whereas for brass induced valence has a stronger concave register trend than perceived valence, the inverse is true for percussion and strings with perceived valence showing a stronger concave register trend than induced valence.

In summary, the middle registers are rated as the most positive. Percussion instruments are also rated as the most positive, especially in the higher pitch registers. Although induced valence is slightly more positive than perceived valence overall, Figure 4.1 shows that this is especially the case in the lower pitch registers, and the affect loci are inversely related, or more similarly valenced, in the higher pitch registers. We can also see that perceived valence of percussion instruments seems to vary relatively little with register for the single notes in Experiment 1.

Figure 4.1 *Register Trends for Perceived and Induced Valence Separated by Instrument Family*

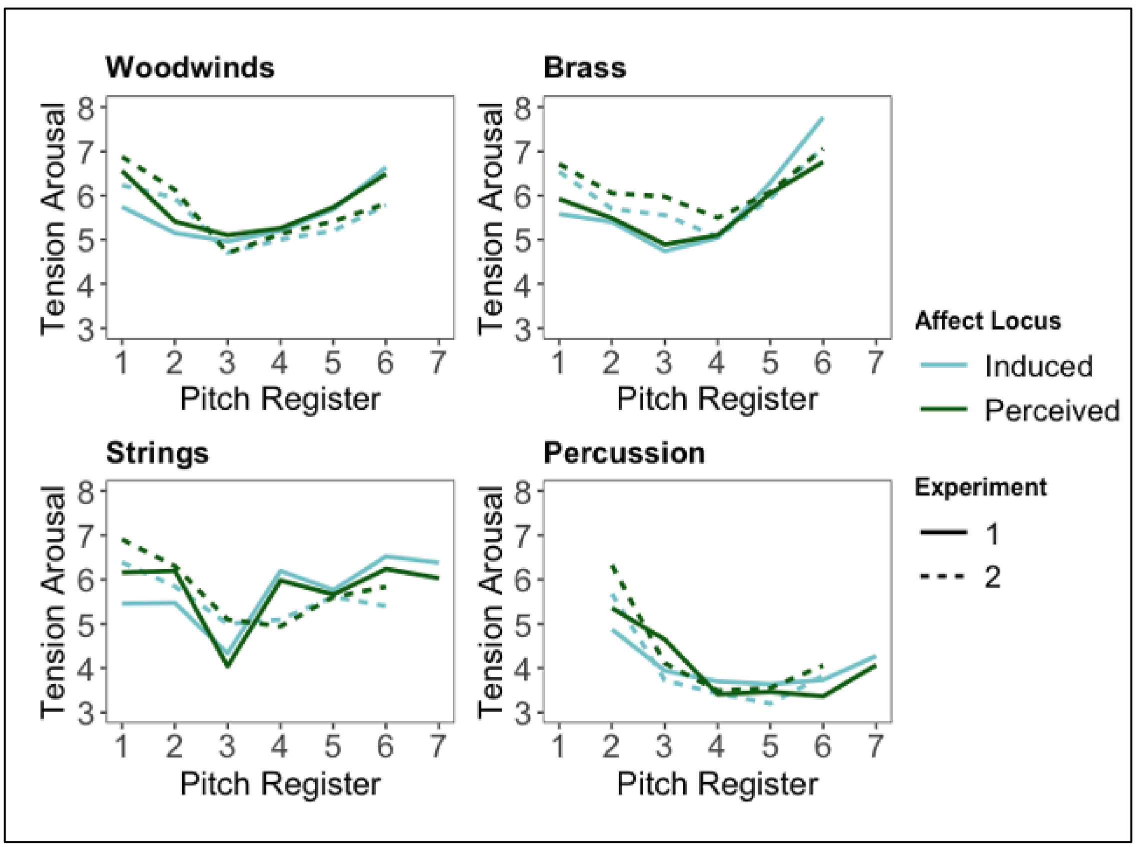


Tension Arousal

Inversely to valence, tension arousal shows a convex register effect in both experiments, as is shown by the U-shapes in Figure 4.2. There is also a linear register trend, but interestingly this is slightly increasing for Experiment 1 ($b = 0.2$, $SE = 0.11$), and slightly decreasing for Experiment

2 (b = –0.3, SE = 0.05). In both experiments, percussion (M = [3.4, 3.60], SE = [0.2, 0.39]) are rated as less tense than the other three instrument families (M = [5.0, 5.6], SE = [0.2, 0.4]). In Experiment 2, percussion are also more strongly linearly decreasing than the other instrument families, and significantly more convex than strings. Further in Experiment 2, perceived tension is higher than induced tension overall ($M_{induced}$ = 4.7, $SE_{induced}$ = 0.15; $M_{perceived}$ = 5.0, $SE_{perceived}$ = 0.15), and affect locus does not significantly interact with any of the other effects. In Experiment 1, affect locus does not appear as a main effect, but rather interacts with the other effects; firstly, induced tension is more strongly linearly increasing than perceived tension ($b_{perceived}$ = 0.3, $SE_{perceived}$ = 0.05; $b_{perceived}$ = 0.1, $SE_{perceived}$ = 0.11); secondly, percussion show the biggest difference in affect locus (contrast of *induced – perceived*; contrast$_{I-P}$ = –0.3, SE_{I-P} = 0.15, p = .04); and thirdly, for brass, induced affect has a stronger convex register trend than perceived affect, whereas

Figure 4.2 *Register Trends for Perceived and Induced Tension Arousal Separated by Instrument Family*



for the other instrument families, perceived affect has a stronger convex register trend than induced affect.

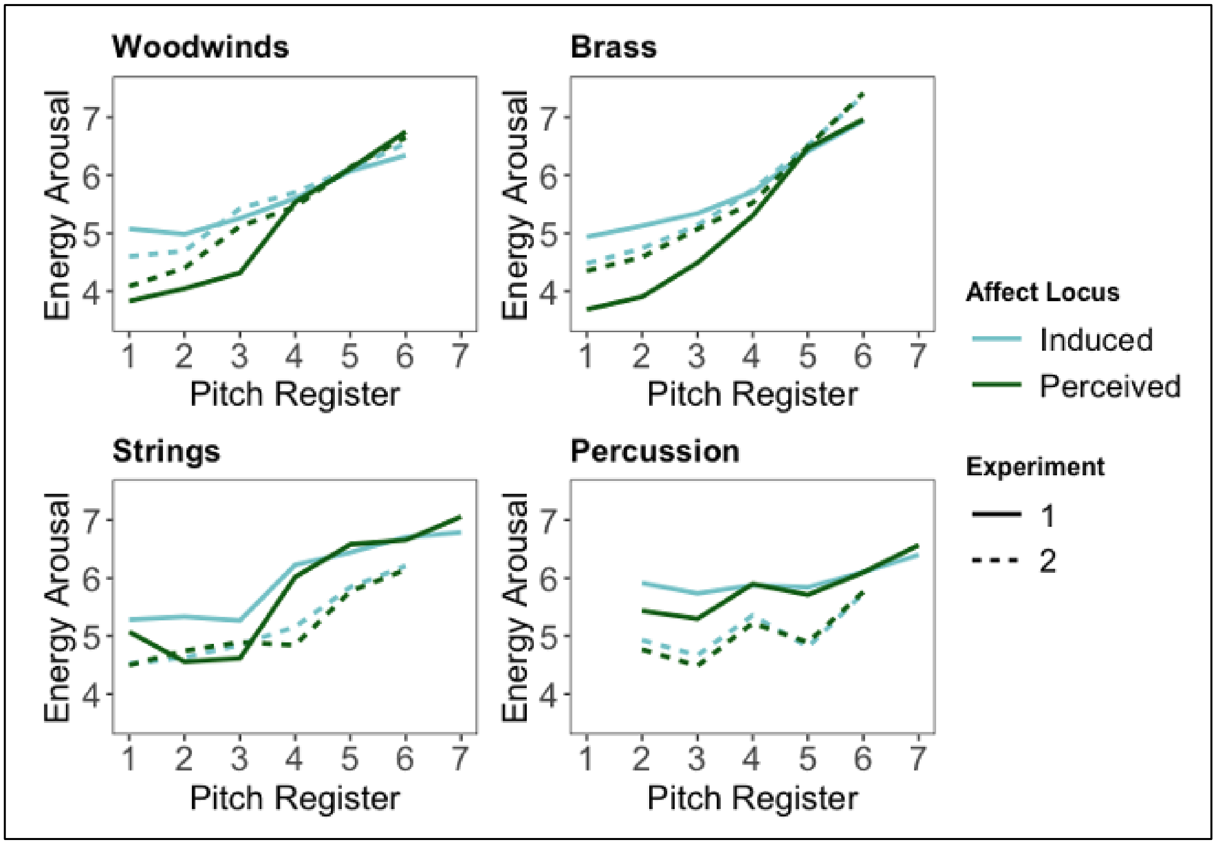
In summary, the middle registers are rated as the least tense. Percussion are rated as the least tense instrument family overall, especially in the higher pitch registers. Although with the longer excerpts, perceived tension is rated higher overall than induced tension, in both experiments Figure 4.2 shows that especially in the lower pitch registers, perceived tension is higher than induced tension, but in the higher pitch registers, this relation is inversed, or the affect loci are more similarly tense.

Energy Arousal

Energy arousal does not show a quadratic register effect in either experiment but is found to be linearly increasing, i.e., it is rated as more awake as pitch register increases. There is no main effect of instrument family, but in Experiment 1 percussion are the least increasing with register (b = 0.1, SE = 0.08) compared to the other instrument families (b = [0.4, 0.7], SE = [0.07, 0.11]) as visualized in Figure 4.3. Our analysis revealed a significant main effect of affect locus in Experiment 2, although the contrast between induced and perceived ratings (I – P) only shows a near significant difference ($contrast_{I-P}$ = 0.1, SE = 0.07, p = .07). In Experiment 1, perceived energy is found to be more strongly linearly increasing with register than induced energy ($b_{perceived}$ = 0.6, SE = 0.06; $b_{induced}$ = 0.3, SE = 0.06). Experiment 1 also shows a three-way interaction: the difference in linear register effect between affect loci, where perceived energy is more increasing than induced energy, is biggest for woodwinds and smallest for strings. The brass family also shows a large difference between perceived and induced register trend like woodwinds, but it is not significantly contrasted from the other instrument families.

In summary, energy arousal appears to mostly vary linearly with pitch register. With longer stimuli, induced energy is rated as slightly (near-significantly, p = .07) more awake than perceived energy. With shorter stimuli, percussion show the least variation with register, and induced energy is especially more awake (or less tired) than perceived energy in the lower pitch registers, as can be seen in Figure 4.3.

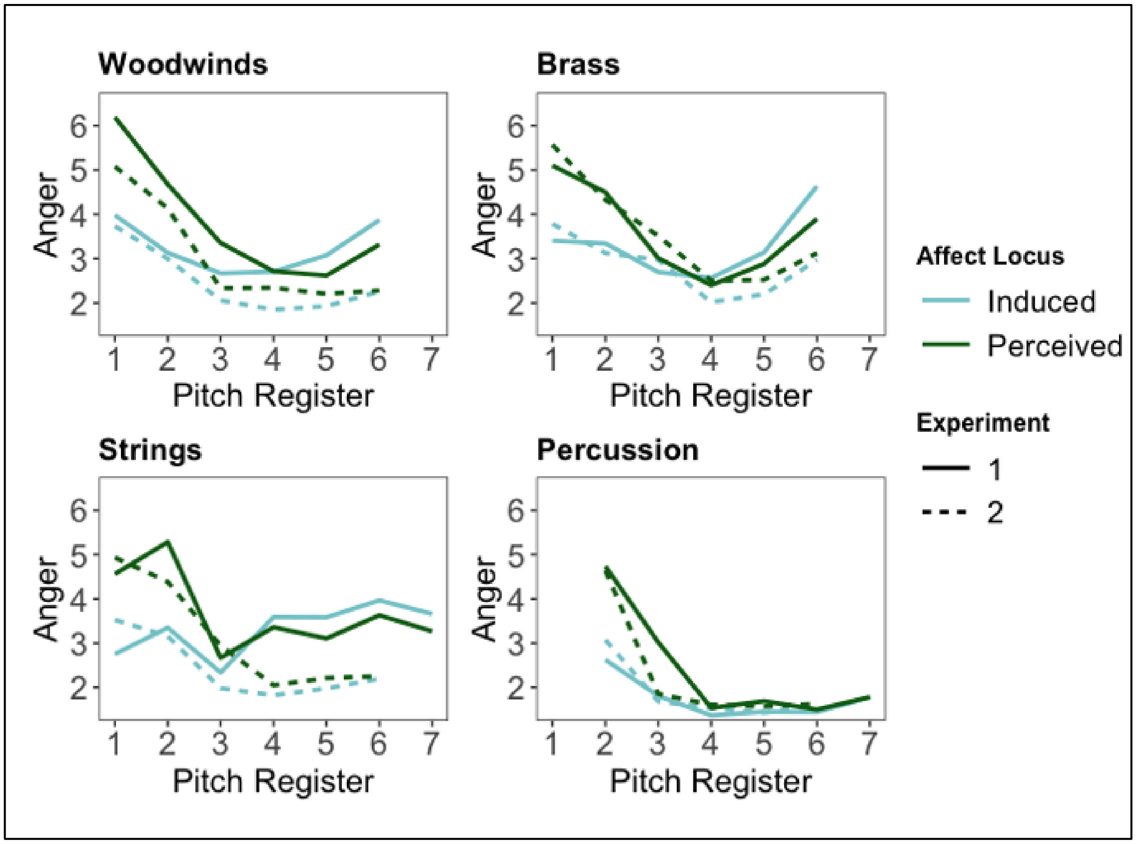
Figure 4.3 *Register Trends for Perceived and Induced Energy Arousal Separated by Instrument Family*



Anger

Anger shows a convex (U-shaped) register effect in both experiments and is also decreasing with pitch register in the second experiment. In Experiment 1, percussion are rated as the least angry and strings as the most angry instrument family ($M_{percussion}$ = 1.7, $SE_{percussion}$ = 0.34; $M_{strings}$ = 3.3, $SE_{strings}$ = 0.32). In Experiment 2, there are no main or interaction effects related to instrument family. We do see that perceived affect is rated as more angry than induced affect overall ($M_{perceived}$ = 2.5, $SE_{perceived}$ = 0.17; $M_{induced}$ = 2.0, $SE_{induced}$ = 0.17), and affect locus does not interact with other effects. In the first experiment, there is no main effect of affect locus. However, perceived anger has a slightly stronger convex register effect than induced anger, indicated by a more positive coefficient ($b_{perceived}$ = 0.2, $SE_{perceived}$ = 0.05; $p_{perceived}$ < .0001; $b_{induced}$ = 0.1, $SE_{induced}$ = 0.05, $p_{induced}$ = .003). Finally, percussion show the biggest and only significant difference between perceived and induced anger ($contrast_{I-P}$ = –0.9, SE = 0.20, p < .0001), compared to the other instrument families ($contrast_{I-P}$ = [– 0.4, –0.2], SE = 0.20 for all three).

Figure 4.4 *Register Trends for Perceived and Induced Anger Separated by Instrument Family*



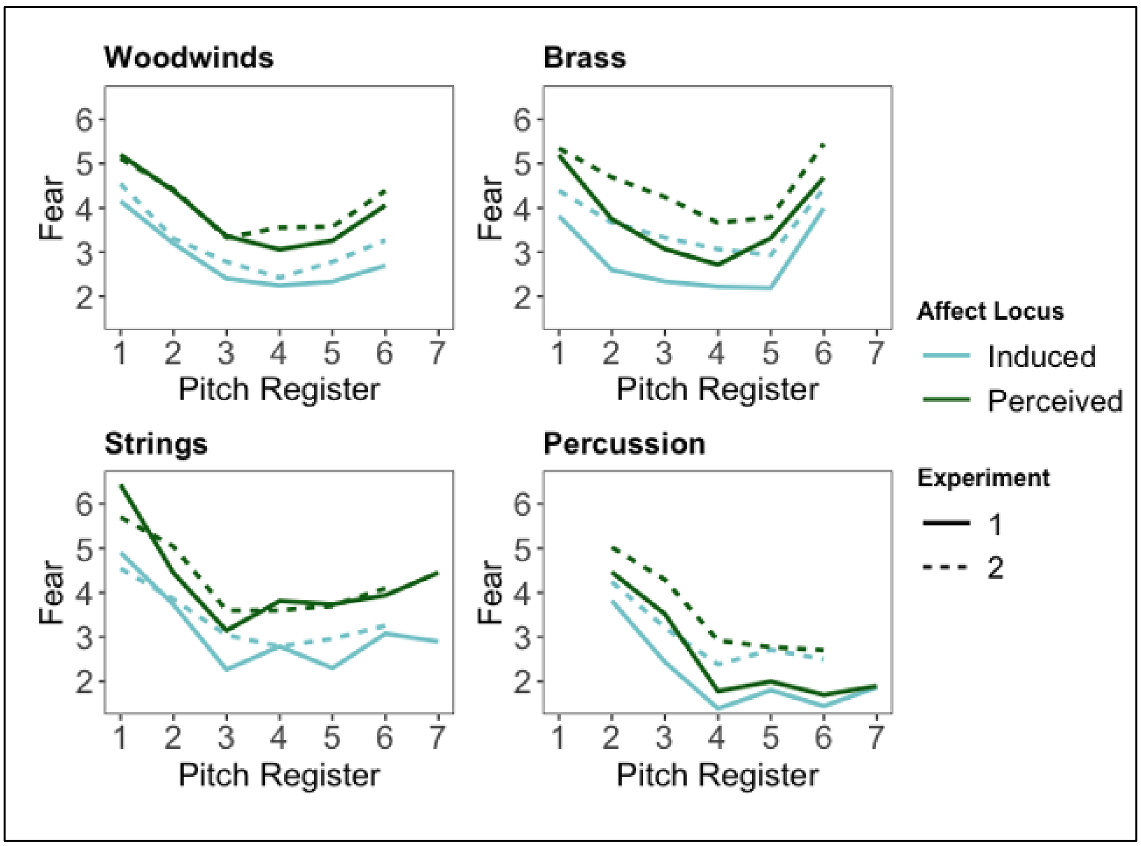
Overall, in Figure 4.4 we can see that for both experiments, the middle registers are rated as least angry. With shorter stimuli, percussion are considered the least angry instrument family, but with longer stimuli these family differences are no longer significant. With longer stimuli, perceived affect is rated as more angry than induced affect throughout the pitch registers and instrument families, whereas with shorter stimuli, this is mostly the case in the lower pitch registers and especially so for the percussion instruments.

Fear

Similarly to anger, fear shows a convex (U-shaped) register effect in both experiments, which also decreases slightly with register in Experiment 2. In both experiments, there is a main effect of Instrument family, although the contrasts are only significant for the first experiment; percussion (M_{Exp1} = 2.0, SE_{Exp1} = 0.26) is rated significantly lower on fear than strings (M_{Exp1} = 2.9, SE_{Exp1} =

0.25). Instrument family also interacts with register in that percussion are the most strongly linearly decreasing (b = [–0.7, –0.5], SE = [0.10, 0.11]) compared to the other instrument families (b = [–0.3, 0.3], SE = [0.07, 0.17]). Perceived fear is rated higher than induced fear ($M_{perceived}$ = [2.9, 3.6], $SE_{perceived}$ = [0.17, 0.19]; $M_{induced}$ = [2.15, 2.8], $SE_{induced}$ = [0.17, 0.19]). Again in both experiments, but with significant contrasts only in Experiment 1, instrument family and affect locus interact. Percussion show a significantly smaller difference between perceived and induced fear ($contrast_{I-P}$ = –0.6, SE = 0.23) than strings and woodwinds ($contrast_{I-P}$ = [–1.2, –1.1], SE = [0.22]). Finally, Experiment 1 shows a three-way interaction, where percussion show a significant difference in linear register trend between affect loci (perceived fear is more strongly decreasing than induced fear), whereas the other families show no significant difference.

Figure 4.5 *Register Trends for Perceived and Induced Fear Separated by Instrument Family*

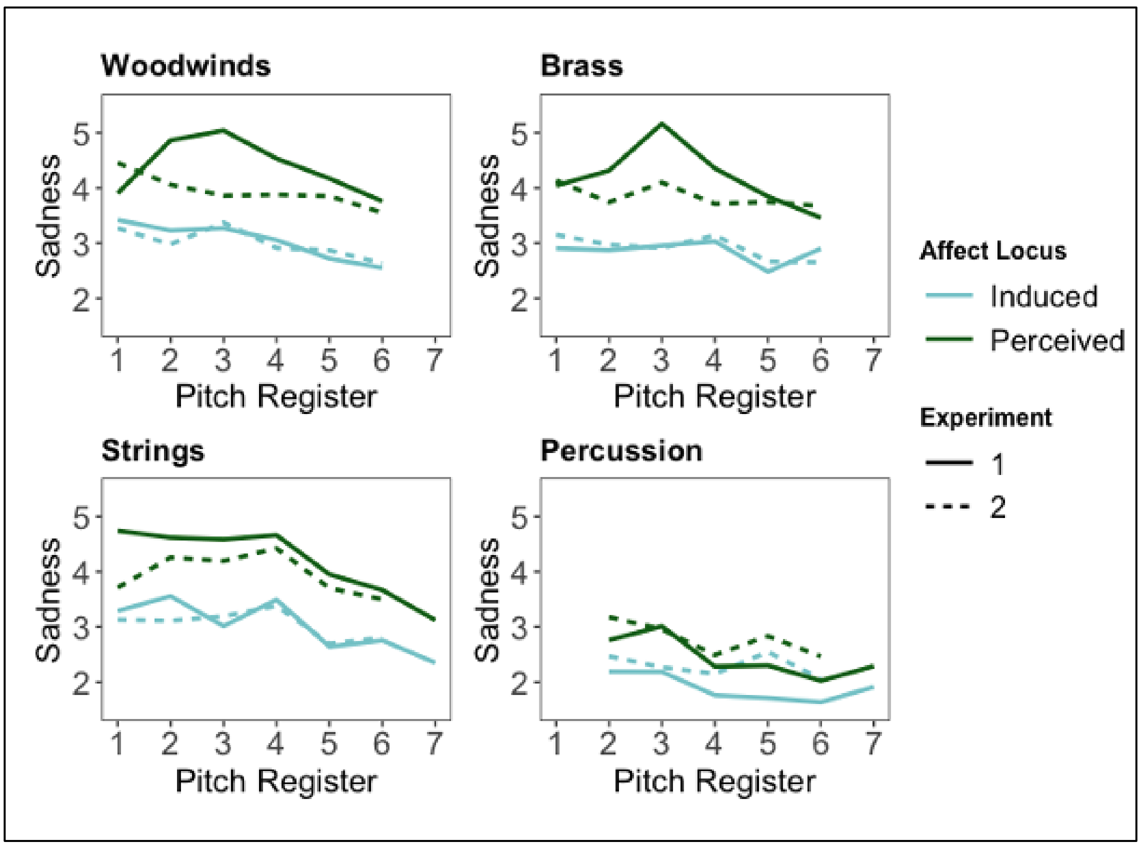


In summary, Figure 4.5 shows that the middle pitch registers are rated as the least fearful in both experiments. Percussion are also considered the least fearful. Furthermore, although perceived affect is rated more fearful than induced affect over experiments and instrument families, this difference is the smallest for percussion, especially in the higher pitch registers.

Sadness

For sadness, we only find a linearly decreasing register trend. Percussion are rated as the least sad instrument family overall (M = [2.1, 2.5], SE = [0.20]), whereas the other instrument families are not significantly different from each other (M = [3.4, 3.8], SE = [0.17, 0.19]). Perceived sadness (M = [3.7, 4.0], SE = [0.16, 0.18]) is rated higher than induced sadness (M = [2.7, 2.9], SE = [0.16, 0.18]). In Experiment 1, perceived sadness is also more strongly linearly decreasing with register than induced sadness ($contrast_{I-P}$ = 0.16, SE = 0.06), and perceived sadness is convex in shape

Figure 4.6 *Register Trends for Perceived and Induced Sadness Separated by Instrument Family*



($b_{perceived}$ = –0.1, SE = 0.03, p = .002), whereas induced sadness shows no quadratic effects (p = .85). A three-way interaction in Experiment 1 further reveals that this difference in convex register trend is biggest for woodwinds and brass and not significant for strings and percussion. This is visualized in Figure 4.6 by the perceived affect peak in the middle registers for woodwinds and brass in the Experiment 1. In both experiments, the difference between perceived and induced sadness is smallest for percussion ($contrast_{I-P}$ = [–0.6, –0.5], *SE* = [0.11, 0.26]), compared to the other instrument families, which show a similar difference between affect loci ($contrast_{I-P}$ = [–1.6, –0.9], *SE* = [0.11, 0.26]).

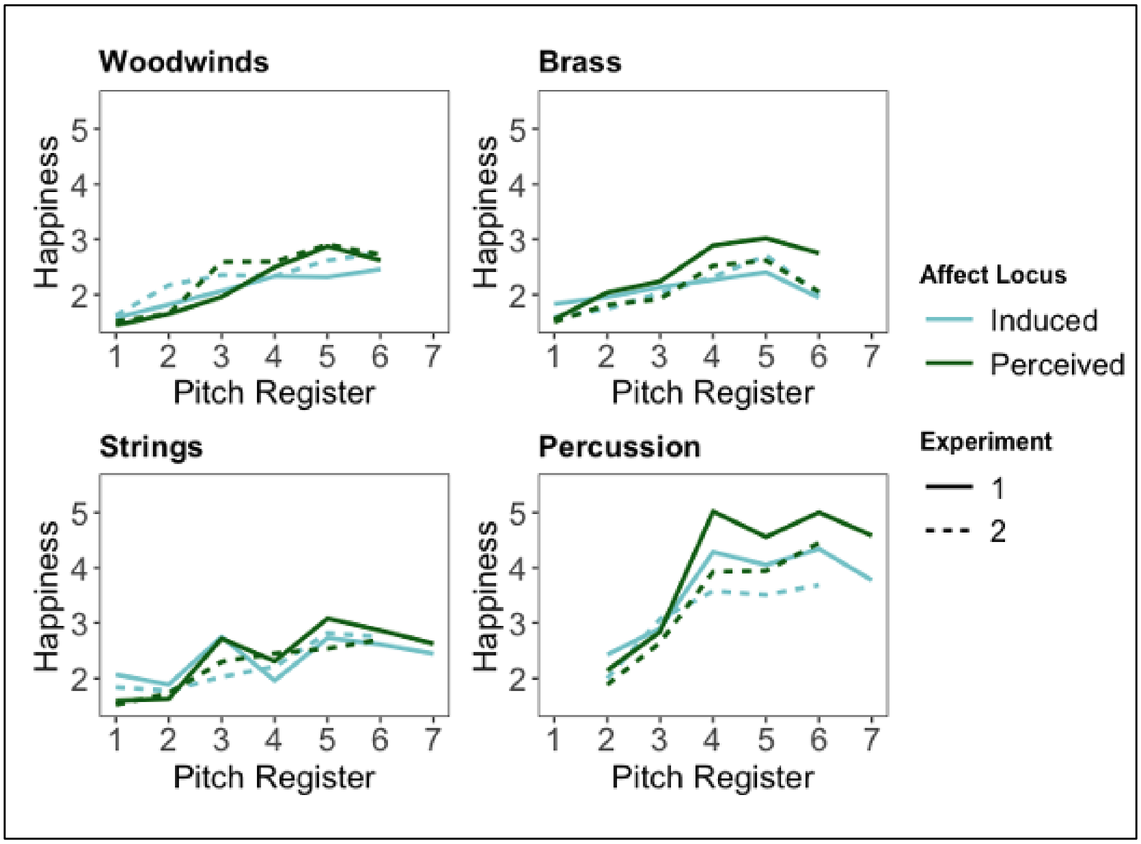
In summary, as pitch register increases, sadness ratings decrease. For single-note stimuli, however, there is a peak in sadness in the middle registers for woodwinds and brass. There is also a clear difference between perceived and induced sadness across pitch registers and instrument families. The difference is the smallest for percussion, which is also rated as the least sad instrument family overall.

Happiness

In both experiments, happiness shows a linearly increasing and slightly concave (inverted U-shape) register effect. Percussion are rated as the most happy instrument (*M* = [3.4, 4.1], *SE* = [0.20, 0.27]), whereas the other instrument families are not significantly different from each other in overall happiness (*M* = [2.2, 2.5], *SE* = [0.19, 0.26]). In Experiment 2, percussion are also more strongly linearly increasing with pitch register (*b* = 0.7, *SE* = 0.08) than the other instrument families (*b* = [0.1, 0.2], *SE* = [0.06, 0.07]). There are no main effects of affect locus, but it does interact with pitch register in both experiments, and instrument family in Experiment 1. Perceived happiness (*b* = [0.3, 0.4], *SE* = [0.05, 0.07]) is more strongly increasing with pitch register than induced happiness (*b* = [0.2, 0.3], *SE* = [0.02, 0.07]), and brass show a significant difference where perceived affect is more happy than induced affect across registers ($contrast_{I-P}$ = –0.5, *SE* = 0.23, *p* = .02), whereas the other instrument families do not (*p* = [.07, .45]; near significance for percussion, *p* = .07).

In summary, happiness increases mostly linearly with pitch register, with a slight curve. Percussion are rated as the most happy and happiness ratings increase with register the most strongly. Affect locus appears to have less of an effect on happiness, although in both experiments perceived happiness increases more strongly than induced happiness. We can see this visualized in Figure 4.7: especially in the higher pitch registers, perceived happiness is higher than induced happiness, whereas in the lower registers the affect loci are more similar, or perceived happiness is in fact lower than induced happiness. Perceived affect is also significantly more happy than induced affect for the brass instrument family across pitch registers, with the single-note stimuli only.

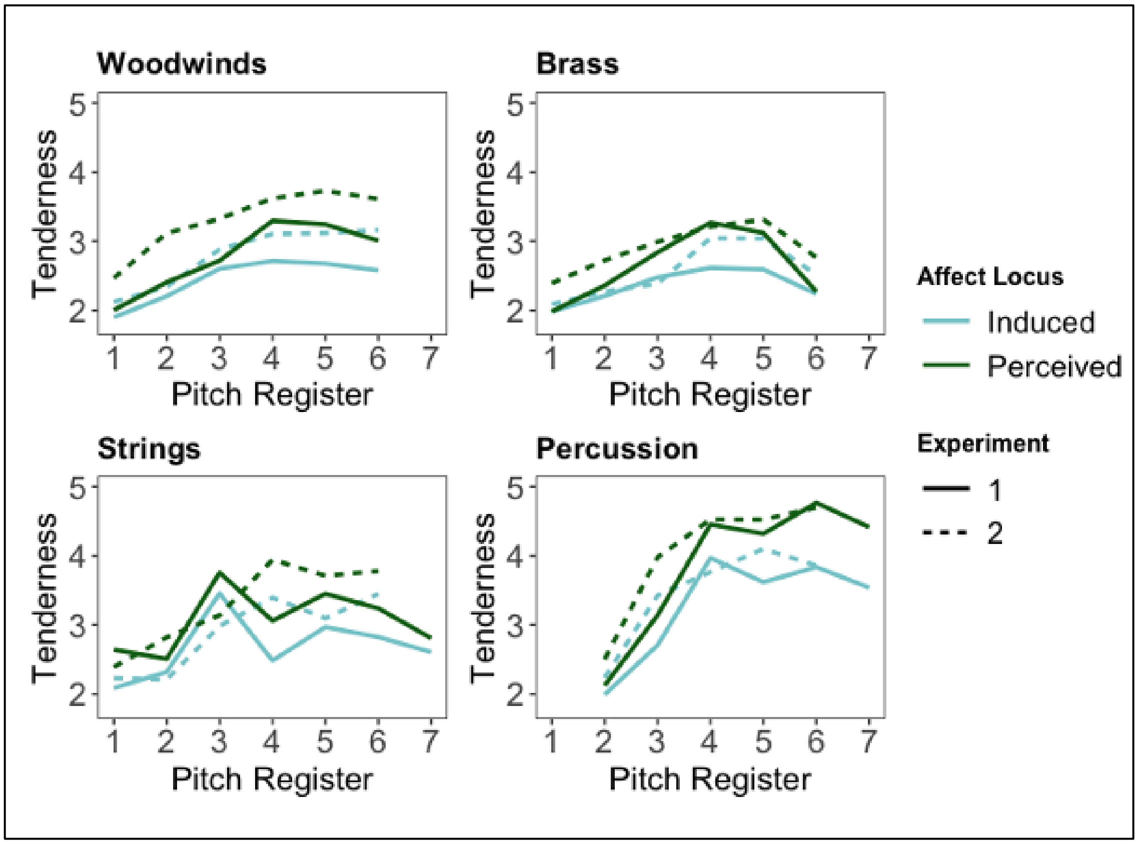
Figure 4.7 *Register Trends for Perceived and Induced Happiness Separated by Instrument Family*



Tenderness

For tenderness, we find the same effects of pitch register, instrument family, and affect locus in both experiments. Tenderness has a linearly increasing and slightly concave (inverted U-shape) register trend. Percussion are rated as the most tender (M = [3.9, 4.0], SE = [0.19, 0.26]), whereas the other instrument families are not significantly different from each other in overall tenderness (M = [2.9, 3.2], SE = [0.18, 0.25]). Furthermore, perceived affect is rated as more tender than induced affect ($M_{perceived}$ = [3.5, 3.6], $SE_{perceived}$ = [0.16, 0.20]; $M_{induced}$ = [3.0, 3.1], $SE_{induced}$ = [0.16, 0.20]). In Experiment 2, percussion are the most strongly linearly increasing with pitch register ($b_{percussion}$ = 0.6, *SE* = 0.07; b_{rest} = [0.1, 0.3], *SE* = [0.05, 0.07]), and shows the most strong concave shape ($b_{percussion}$ = –0.2, *SE* = 0.04; b_{rest} = [–0.1], *SE* = [0.03]).

Figure 4.8 *Register Trends for Perceived and Induced Tenderness Separated by Instrument Family*

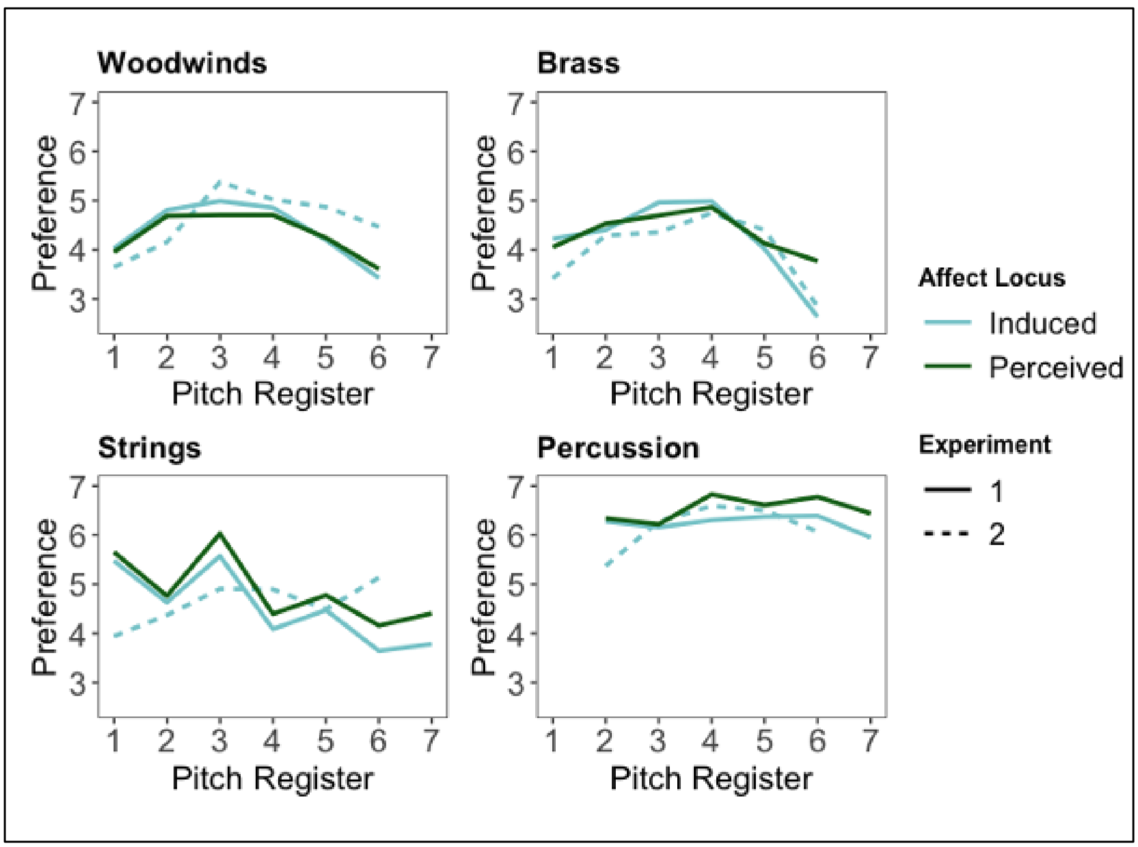


In summary, we can see in Figure 4.8 that the middle pitch registers are rated as the most tender, and the upper registers are also rated as more tender than the lower registers (i.e., the inverted U-shape is slightly tilted counter-clockwise). We can see that percussion are the most tender family for longer melodic stimuli, especially so in the higher pitch registers. Finally, perceived affect is more tender than induced affect throughout families and experiments.

Preference

Finally, for preference we find a linearly decreasing register effect in Experiment 1 (b = –0.3, SE = 0.11), and increasing in Experiment 2 (b = 0.1, SE = 0.06). In both experiments, there is also a concave register effect. Percussion again stand out as the most liked (M = [6.5], SE = [0.25, 0.39]), and the other instrument families are not significantly different from each other (M = [4.6, 5.1], SE = [0.24, 0.37]. Affect locus cannot be considered in Experiment 2, as participants only rated their preference in the induced affect condition. In Experiment 1, we see that strings show the biggest difference in perceived and induced affect ($contrast_{I-P}$ = – 0.3, SE = 0.19, p = .11), and brass the smallest ($contrast_{I-P}$ = 0.02, SE = 0.19, p = .90). Although these contrasts comparing perceived and induced affect are significantly different *between* instrument families, the p-values for each individual contrast show that the differences between perceived and induced affect are not significant *within* each individual instrument family (see Table A4 for full statistics).

Figure 4.9 *Register Trends for Perceived and Induced Preference Separated by Instrument Family*



Thus, in summary, the preference for stimuli was the highest in the middle pitch registers. Interestingly, single notes and chromatic scales showed opposite linear register effects, albeit small. We can see in Figure 4.9, especially for single notes (Exp. 1), that besides the overall preference for middle registers, the lower registers are also more preferred than the upper registers. Percussion are the most liked overall. Although the difference in preference when participants were rating either perceived or induced affect is rather small, and non-significant for all instrument families, we can see it is the biggest across pitch registers for strings (higher preference when rating perceived affect), and smallest for brass instruments.

Summary: Pitch Register

Pitch register is a consistent predictor in all affect models in both experiments (see Table 4.2). As Table 4.3 shows, most scales show both a quadratic and a linear register trend. Describing the quadratic register effect, the middle registers are rated to be the most positive, happy, and tender, and least tense, angry, and fearful. Concurrently, the middle registers are also most liked. Interpreting linear trends in the context of quadratic trends is somewhat challenging. If the quadratic trend is weak, interpretation of the linear trend is similar to that of a standard linear regression. However, if the quadratic trend overpowers the linear trend, the linear trend is much less interpretable. For happiness and tenderness there is a consistent linear register trend alongside the quadratic trend such that the higher pitch registers are happier and more tender than the lower pitch registers. A similar combination of quadratic and linear trends is present in valence, anger, and fear, but only in the second experiment with the longer melodic excerpts. Tension and preference also show a linear register trend alongside the quadratic register trend in both experiments, but in opposite directions. Energy arousal and sadness consistently show a nonquadratic register effect; they are linearly increasing and decreasing with register, respectively. The single-note and chromatic scales especially show noticeable differences in the linear register trend. As mentioned, valence, anger, and fear do not change linearly with register in Experiment 1, but in Experiment 2, the linear effects are relatively large and strongly significant. The quadratic register effects appear more consistent between experiments. Tension arousal and fear show the relatively strongest quadratic effects.

Table 4.3 *Register Trends and Standard Errors for Linear and Quadratic Effects of Each Affect Scale*

	Linear trend (*SE*)		Quadratic trend (*SE*)	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**
Valence	——	0.34 (0.06)	–0.13 (0.04)	–0.19 (0.03)
Tension	0.24 (0.11)	–0.25 (0.06)	0.18 (0.05)	0.24 (0.04)
Energy	0.44 (0.05)	0.40 (0.07)	——	——
Anger	——	–0.48 (0.07)	0.18 (0.04)	0.20 (0.04)
Fear	——	–0.28 (0.05)	0.22 (0.03)	0.20 (0.03)
Sadness	–0.24 (0.05)	–0.11 (0.04)	——	——
Happiness	0.25 (0.07)	0.31 (0.04)	–0.08 (0.03)	–0.07 (0.02)
Tenderness	0.14 (0.07)	0.30 (0.04)	–0.11 (0.03)	–0.10 (0.02)
Preference	–0.26 (0.10)	0.14 (0.06)	–0.11 (0.05)	–0.16 (0.03)

Note. All listed register trends are statistically significant. For detailed significance levels, see Table 4.2. A negative quadratic trend indicates a concave shape (inverted-U), and a positive quadratic trend indicates a convex shape (U).

Summary: Instrument Family

Table 4.4. summarizes the paired comparisons of the main effect of instrument family as well as its interaction with pitch register. Here we can see that instrument family is a main predictor for many affect scales consistently in both experiments for valence, tension, sadness, happiness, tenderness, and preference. Instrument family is not significant for energy arousal in both experiments and is only significant in Experiment 1 (single notes) for anger. In instances in which instrument family is a significant main predictor, the effect is consistently caused by percussion. The percussion family is rated to be less tense, angry, fearful, and sad, as well as more positive, happy, tender, and it is more liked than the other families. In Experiment 1, strings are also rated to be the angriest. Fear shows a consistent interaction effect, with percussion showing the strongest linearly decreasing register trend. The other interaction effects appear mostly in the second experiment (chromatic scales), except for energy arousal which has a significant family × register interaction only in Experiment 1. Here too, the percussion family shows the *strongest* linear

register effect (either increasing or decreasing depending on the scale). However, for energy arousal, percussion show the *weakest* linear register effect, and barely vary in energy across pitch registers. For tension and tenderness, percussion also show the strongest quadratic trend of register.

Summary: Affect Locus

Affect locus is a significant main predictor for fear, sadness, and tenderness in both experiments. It is only significant for valence, tension, energy, and anger in the second experiment with the

Table 4.4 *Summary of Paired Comparisons for Instrument Family Main Effect and its Interaction with Register for Each Rating Scale*

	Instrument Family		Instrument Family × Register	
	Exp. 1	**Exp. 2**	**Exp. 1**	**Exp. 2**
Valence	P > B ≈ W ≈ S	P > B ≈ W ≈ S	—	LN: P > B ≈ W ≈ S
Tension	P < B ≈ W ≈ S	P < B ≈ W ≈ S	—	LN: P > B ≈ W ≈ S QD: P > S
Energy	—	—	LN: P < B ≈ W ≈ S	—
Anger	P < W ≈ B < S	—	—	—
Fear	P < S		LN: P > B ≈ W ≈ S	LN: P > B ≈ W ≈ S
Sadness	P < B ≈ W ≈ S	P < B ≈ W ≈ S	—	—
Happiness	P > B ≈ W ≈ S	P > B ≈ W ≈ S	—	LN: P > B ≈ W ≈ S
Tenderness	P > B ≈ W ≈ S	P > B ≈ W ≈ S	—	LN: P > B ≈ W ≈ S QD: P > B ≈ W ≈ S
Preference	P > B ≈ W ≈ S	P > B ≈ W ≈ S	—	—

Note. P: Percussion, B: Brass, W: Woodwinds, S: Strings, LN: Linear register effect, QD: Quadratic register effect

chromatic scales and is not significant in either experiment for happiness and preference. There is a general trend where affects that are generally considered to be unpleasant or undesirable, are more strongly perceived than induced. Perceived affect in general was rated to be more negatively valenced, tense, tired, angry, fearful, and sad than induced affect. Perceived ratings of tenderness, however, are higher than induced ratings. Table 4.5 shows the pairwise comparisons of the main effects of affect locus. These results show that in Experiment 2, with chromatic scales, there is always a main effect of affect locus, whereas in Experiment 1, with single-note stimuli, this is only the case for fear, sadness, and tenderness. The $I - P$ contrasts show that the difference between induced and perceived affect is biggest for fear and sadness, and smallest for valence and energy arousal.

Table 4.5 *Affect Locus Pairwise Comparisons*

	Experiment 1				Experiment 2			
Scale	**I – P**	***SE***	***z***	***p***	**I – P**	***SE***	***z***	***p***
Valence	——	——	——	——	0.14	0.06	2.40	.022
Tension	——	——	——	——	–0.21	0.07	–2.90	.004
Energy	——	——	——	——	0.13	0.07	1.81	.07
Anger	——	——	——	——	–0.47	0.07	–7.05	<.0001
Fear	–0.77	0.22	–3.49	.0005	–0.81	0.08	–9.76	<.0001
Sadness	–1.28	0.24	–5.43	<.0001	–0.85	0.07	–11.38	<.0001
Tenderness	–0.53	0.25	–2.15	.032	–0.51	0.07	–7.25	<.0001

Note. The $I - P$ column shows the results of the average induced rating minus the average perceived rating.

Whenever we did not find a significant main effect for affect locus, affect locus interacted with register or instrument family. Table 4.6 shows a summary of the pairwise comparisons for both interaction effects (see Tables A3 and A4 in Appendix A for full statistics). In the *Affect Locus × Register* columns, we see that affect locus mostly interacted with the linear register trend. There is a lot of variation in whether induced affect is more/less increasing/decreasing with register than perceived affect. A more consistent trend becomes clear in the previously discussed Figures 4.1–4.9. The affect locus trend where the less desirable or more unpleasant affects are more strongly perceived than induced is especially apparent in the lower pitch registers where perceived affect is more negatively valenced, tense, tired, angry, and less happy. In the higher pitch registers, this relationship is either inversed or perceived and induced affect converge. Sadness behaves differently in that perceived affect is always sadder than induced affect. However, in the first experiment, this difference is greatest in the middle pitch registers.

In the *Affect Locus × Family* columns, we see that consistently in both experiments, the difference between perceived and induced fear and sadness (Per > Ind) is smallest for the percussion instrument family. These instruments are already considered to be the least fearful and sad, so this may be interpreted as a floor effect; because percussion are rated low on perceived fear and sadness, the rating scale does not allow for a much lower induced experience of fear and sadness. We see a similar effect for anger in Experiment 1. For tension in Experiment 1, we see that overall, for brass, induced tension is in fact slightly (though non-significantly) higher than perceived tension, and for woodwinds and percussion this relation is inversed (but only significant for percussion). For happiness in Experiment 1, we see that where perceived happiness is higher than induced happiness, the difference is largest and only significant for the brass instruments.

Table 4.6 *Summary of Affect Locus × Register and Affect Locus × Family pairwise comparisons*

	Affect Locus × Register		**Affect Locus × Family**	
Scale	**Exp. 1**	**Exp. 2**	**Exp. 1**	**Exp. 2**
Valence	LN: Ind– < Per+	LN: Ind+ < Per+	——	——
Tension	LN: Ind+ > Per+	——	B+ > P– ≈ W–	——
Energy	LN: Ind+ < Per+	——	——	——
Anger	LN: Ind+ > Per– QD: Ind+ < Per+	LN: Ind– > Per– QD: Ind+ < Per+	P – < B– ≈ S– ≈ W–	——
Fear	——	——	P– > S– ≈ W–	P– > S– ≈ W–
Sadness	LN: Ind– > Per– QD: Ind– > Per–	——	P– > B– ≈ S– ≈ W–	P– > B– ≈ S– ≈ W–
Happiness	LN: Ind+ < Per+	LN: Ind+ < Per+	B– < S– ≈ W–	——
Preference	——	——	S– < P– ≈ W– ≈ B+	——

Note. In the Affect Locus × Register columns, linear (LN) and quadratic (QD) register effects are specified. The +/– symbols refer to increasing/decreasing (LN) or convex/concave (QD) register trends. The < and > symbols specify whether these trends are bigger/smaller for induced (Ind) than perceived (Per) affect. In the Affect Locus × Family columns, the + and – symbols represent the result of the *induced – perceived* subtraction. The result is negative (–) when Per > Ind, and positive (+) when Per < Ind. The < and > symbols specify the relative size of the affect locus differences between instruments. Additionally, the ≈ symbol indicates no statistically significant difference of affect locus effect between instrument families.

Finally, in Experiment 1, we see that although there is not a significant effect of affect locus for any instrument family, this difference (Per > Ind) is largest for the string instrument family. That is, when participants rated perceived affect, strings were more liked than when participants rated induced affect.

Summary: Three-way interactions

The two-way interactions can be further contextualized by exploring the three-way interactions. These interactions were only significant in the first experiment. Table 4.7 gives a summary of the pairwise comparisons of the three-way interactions. Table A5 in Appendix A shows more detailed statistics. The instrument families that play a role in the three-way interactions are varied, but in

Table 4.7 *Summary of Three-Way Interactions in Experiment 1*

Scale	Family	Register × Affect Locus
Valence	B	QD: Ind– < Per–
	S & W	QD: Ind– > Per–
Tension	B	QD: Ind+ > Per+
	S & P	QD: Ind+ < Per+
Energy	W	LN: Ind+ < Per+, biggest
	S	LN: Ind+ < Per+, smallest
Fear	P	LN: Ind– < Per –
	B, W, S	LN: Ind– ≈ Per –
Sadness	W & B	QD: Ind– > Per–
	P & S	QD: Ind– ≈ Per–

Note. Symbolics are as explained in Table 4.6. Additionally, the ≈ symbol indicates no statistical significant difference in register trend between affect loci.

general all the interactions are related to the trend in which perceived and induced affect diverge in the lower pitch register and either converge more or invert their relationship in the higher pitch registers. For example, the crossing of the perceived and induced valence lines (Figure 4.1) happens at a higher pitch register for the brass instruments than for the string and percussion instruments, represented by differences in quadratic effects. For tension, brass show a smaller difference between perceived and induced affect in the lower pitch register compared to strings and percussion (Figure 4.2). For energy, the difference between perceived and induced affect in the lower pitch registers is greatest for woodwinds and smallest for strings. Fear only shows a stronger linearly decreasing register effect for perceived than induced affect for percussion, which is visualized in Figure 4.5 by a convergence of the perceived and induced lines. Finally, sadness is higher in perceived than in induced ratings in the middle pitch registers, but this is only the case for woodwinds and brass, not for strings and percussion.

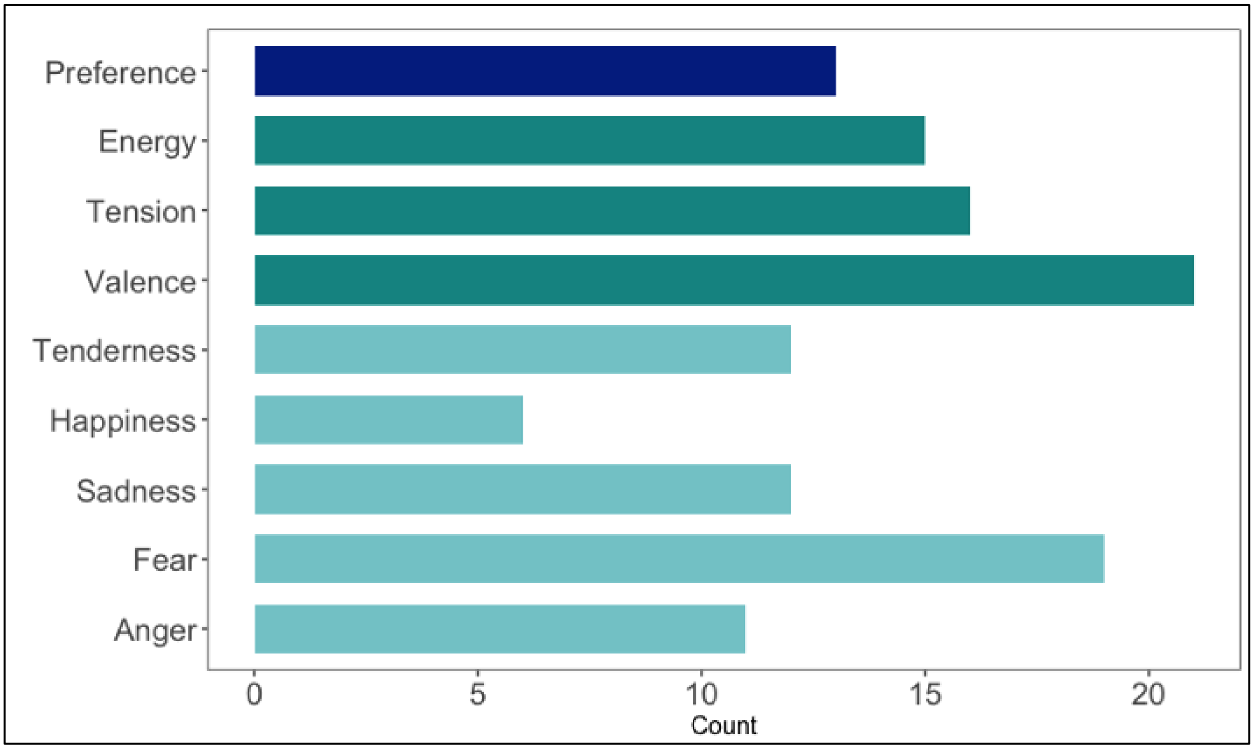
4.4 Results: Individual Differences

Table A6 in Appendix A shows the statistically significant results for each potential individual-differences moderator, on each effect, for each dependent variable (affect scale). Many fixed effects had several significant moderators. Furthermore, due to the large number of comparisons, the chance of finding false positives is increased. Consequently, to assess the general trends, we will examine the frequencies of occurrence across the 23 components of the individual difference measures for all significant fixed effects (see previous section) for each dependent variable to find the main trends.

Figure 4.10 shows the frequency at which each of the affect scales, as well as preference, was influenced by a moderation effect of individual differences on the fixed effects. Valence is the most often influenced by a moderation effect, followed by the discrete scale of fear, and then the other dimensional affect scales of tension and energy arousal. The average frequency of occurrence for the dimensional affect scales (M = 17.3, SD = 3.2) is higher than for the discrete affect scales (M = 12.0, SD = 4.6). Happiness shows the lowest frequency of moderation effects of all the scales.

The three panels on the left of Figure 4.11 show the frequency with which each of the potential moderators was found to be significantly ($p < .01$) moderating an effect and improving the models' performance (Δ *AICc* > 10). The upper panel labelled *All* includes both the counts for

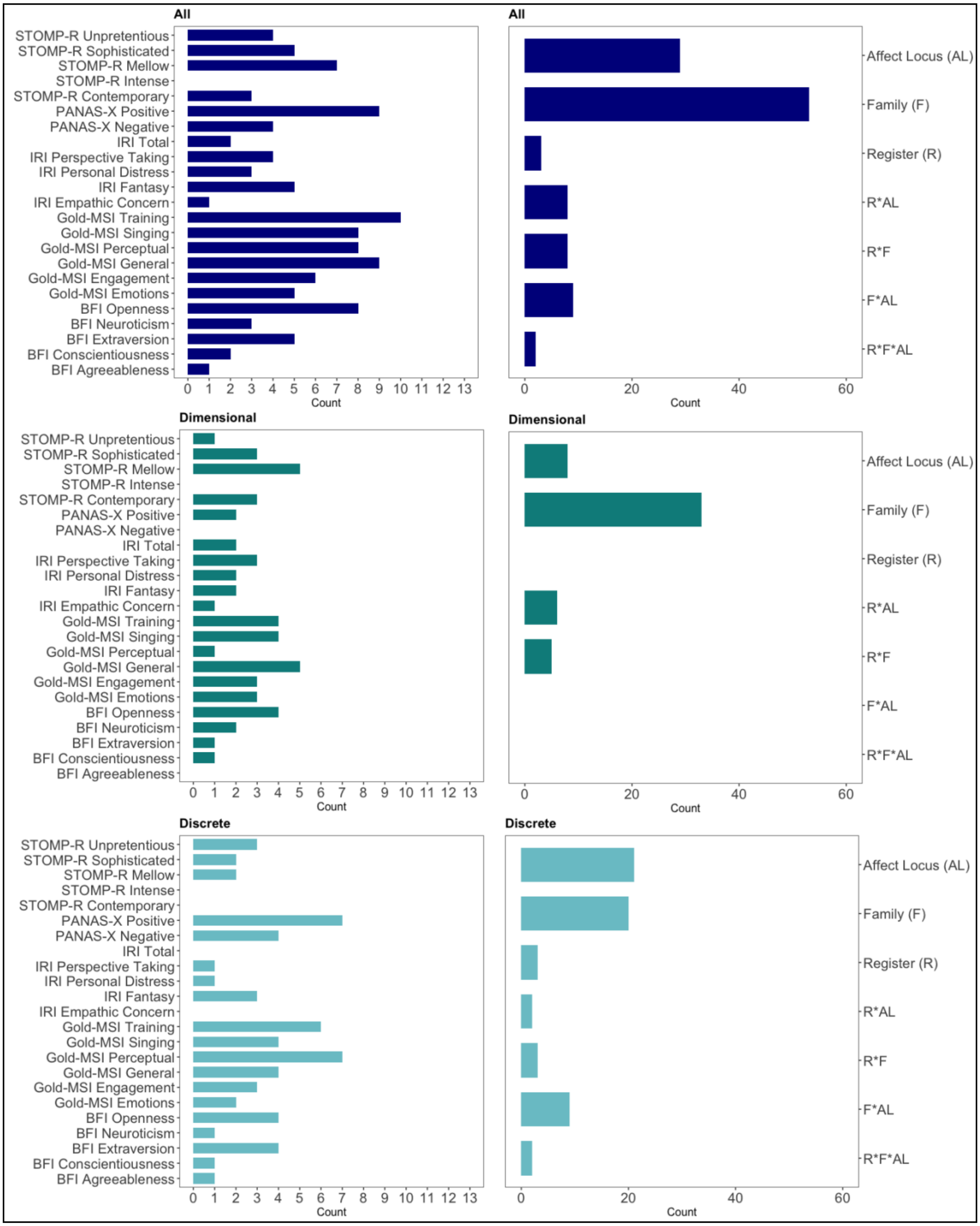
Figure 4.10 *Number of Individual-Differences Components that Moderated a Fixed Effect for Each Affect Scale*



the dimensional and discrete affect models, as well as the preference model. The results for the dimensional and discrete models are separated in the panels below that, excluding preference.

Overall, the average frequency of significant moderation by musical sophistication (Gold-MSI) is the highest (M = 7.7, SD = 1.9), followed by pre-existing mood (PANAS-X; M = 6.5, SD = 3.5), then musical preferences (STOMP-R; M = 3.8, SD = 2.6), Big Five personality (BFI-44; M = 3.8, SD = 2.8), and dispositional empathy (IRI; M = 3.0, SD = 1.6). The training sub-score of the Gold-MSI features most prominently, and, interestingly, the emotions sub-score to a lesser extent. Pre-existing positive mood (PANAS-X Positive) also occurs as one of the most frequent moderators. For genre preferences, mellow (electronica/dance, new age, and world music) shows the highest count. For Big-Five personality (BFI), openness (imaginative, curious, and open-minded traits) is most frequently found as a significant moderator. Looking at the dimensional affect models specifically, STOMP-R mellow and Gold-MSI general are the most frequent moderators. For the discrete affect models, pre-existing positive mood and the Gold-MSI perceptual (self-reported music-listening skills) sub-score play a more prominent role.

Figure 4.11 *Frequency Count of Moderation Effects*



Note. All individual-differences components (left) and all moderated fixed effects; top, all affect scales (right)

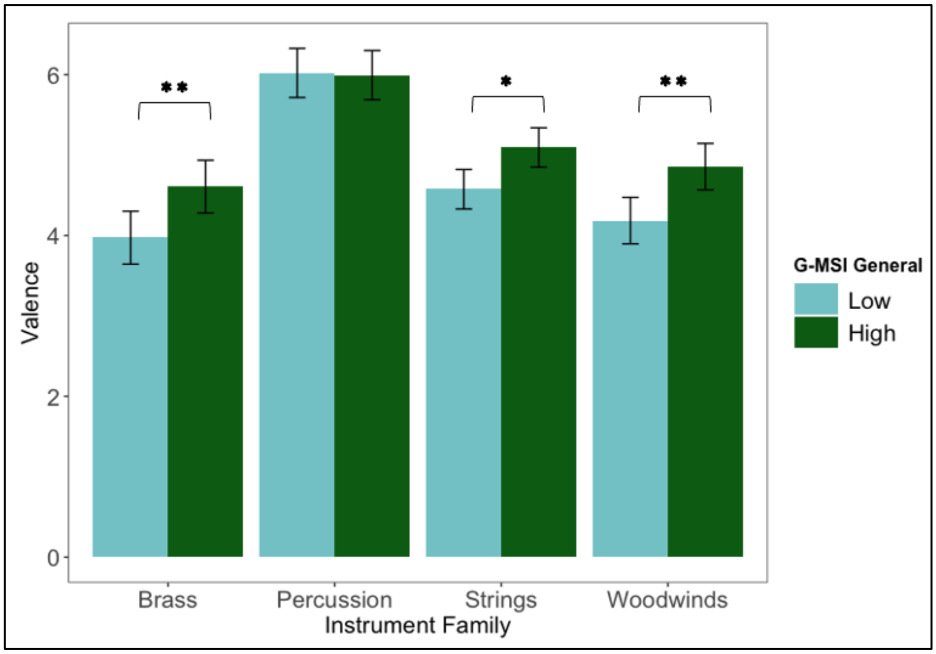
The three panels on the right of Figure 4.11 show the frequency with which each of the main or interaction effects of affect locus, instrument family, and pitch register, were moderated by one of the individual differences. Again, the upper right panel showing *All*, combines the results for the dimensional, discrete, and preference models. The following two panels show the results separated by dimensional and discrete without preference. Both overall and in the dimensional affect models, instrument family is the effect most often moderated by individual differences. For the discrete affect models, both affect locus and instrument family are most often influenced by individual differences. Pitch register, as well as the two- and three-way interactions occur less frequently in the moderation effects.

Although we took a conservative approach to analyzing the individual differences, the large number of tests to analyze the moderation effects requires us to look at the general trends instead of considering each moderation effect in detail. However, for illustration, we will explore a few moderation effects here, featuring the most frequently occurring moderators, fixed effects, and dependent affect scales. For each of the significant moderators, we selected the participants who scored the lowest (first quantile) and highest (fourth quantile) on the relevant moderator (individual differences questionnaire). Using their data, we investigated the moderation effects.

Figure 4.12 illustrates the moderation effect of the Gold-MSI General score on instrument family for the dimensional affect scale of valence in Experiment 1. Whereas participants scoring high on Gold-MSI General always rated stimuli from brass, strings, and woodwind instruments as more positively valenced than people scoring low on Gold-MSI General ($p < .01$), the two participant groups both rated percussion instruments as equally the most positive instrument family ($p = .90$). We may interpret this as a ceiling effect; percussion are considered to be such a positively valenced instrument family that any differences between participant groups are eliminated. The other instrument families do show that participants with a low Gold-MSI General score generally perceive and feel less positive (or more negative) valence in response to single notes.

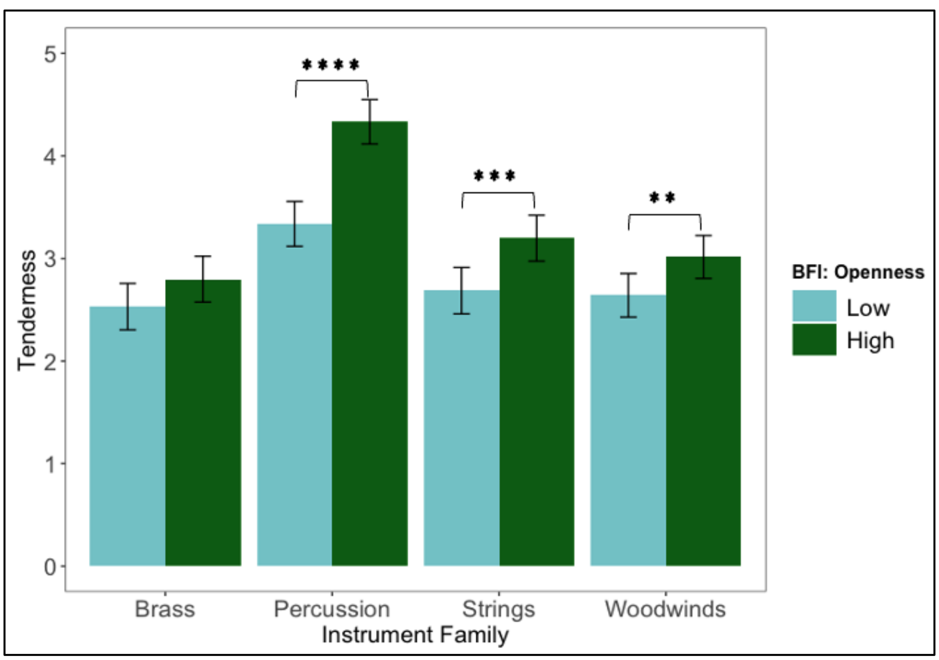
Figure 4.13 shows the moderation effect of the Big-Five personality factor of openness to experience on instrument family for the discrete affect scale of tenderness, also in Experiment 1. Here, we see that for percussion, strings, and woodwinds, participants who scored high on openness rated stimuli as more tender than those who scored low on openness. This difference is the largest for percussion instruments and the smallest (non-significant) for brass instruments. Here, there is no suggestion of a ceiling effect. Rather, in all cases participants with a high Gold-

Figure 4.12 *Gold-MSI General Score Moderating the Instrument Family Effect on Valence in Experiment 1*



Note. Significance between groups indicated by * ($p < .05$) and ** ($p < .01$)

Figure 4.13 *BFI-44 Openness Score Moderating the Instrument Family Effect on Tenderness in Experiment 1*

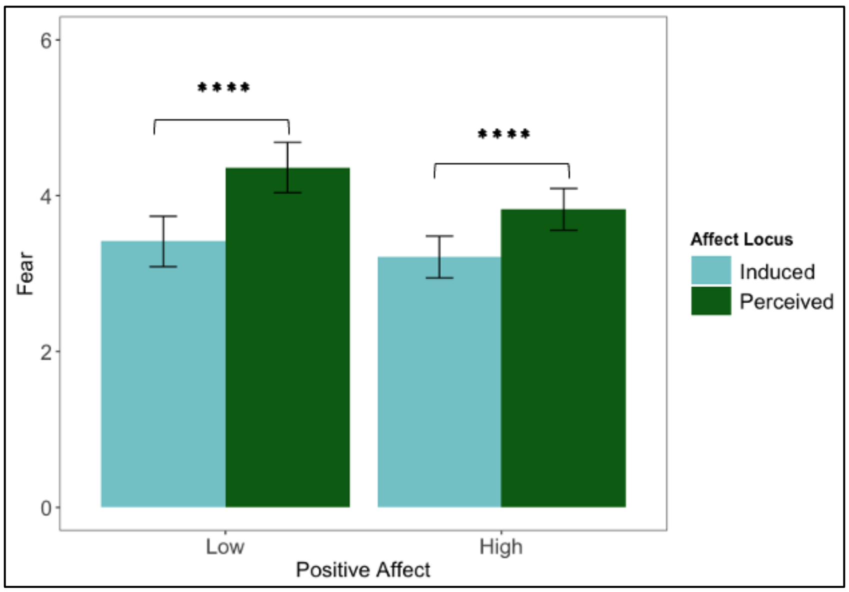


Note. Significance between groups indicated by ** ($p < .01$), *** ($p < .001$), and **** ($p < .0001$).

MSI General score perceive and feel more tenderness in response to single-notes, and this is especially the case for percussion instruments, but not for brass instruments.

Figures 4.14 and 4.15 show the moderating effects of individual differences on affect locus in Experiment 2. In Figure 4.14, we see how pre-existing positive mood moderates the effect of affect locus on the discrete affect scale of fear. Although for both participant groups perceived fear is rated higher than induced fear ($p < .0001$), this difference is bigger for participants that have a less positive pre-existing mood ($p < .05$). This difference appears to be mostly caused by differences in perception of fear; those in a more positive pre-existing mood *perceived* less fear than those in a less positive pre-existing mood, but their *feelings* of fear were more similar.

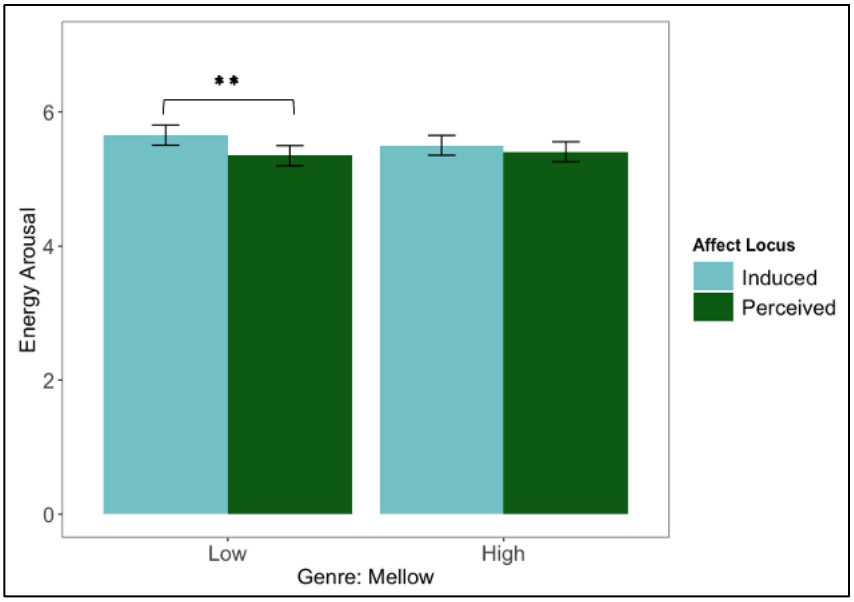
Figure 4.14 *PANAS-X Positive Affect Score Moderating the Affect Locus Effect on Fear*



Note. Significance between groups indicated by **** ($p < .0001$).

In Figure 4.15, we see how a preference for mellow music (electronica/dance, new age, and world music) moderates the effect of affect locus on the dimensional affect scale of energy arousal. Participants with a lower preference for mellow music, significantly *feel* more awake in response to chromatic scales than they *perceive* awake energy (or they perceive more tiredness than they feel tiredness). There is no difference in affect locus for the participants with a higher preference for mellow music. Intuitively, “mellowness” is related to “energy,” and it appears that for those with a preference for mellow music, the induced energy is more similar to the perceived energy.

Figure 4.15 *STOMP-R Mellow Score Moderating the Affect Locus Effect on Energy Arousal*



Note. Significance between groups indicated by ** (p < .01).

4.5 Discussion

Pitch register influenced perceived and induced affect mostly in its quadratic component: the middle registers were considered to be the most positive, happy, and tender, and least tense, angry, and fearful. Energy arousal and sadness, however, increased and decreased linearly with pitch, respectively. The register trends of the dimensional affect ratings are highly similar to those found in McAdams et al. (2017), who also found quadratic register effects for valence and tension arousal, but a linear register effect for energy arousal. Considering the discrete affect ratings, Eerola et al. (2013) found perceived emotions of scary, happy, sad, and peaceful to be mostly linearly related to register, whereas a quadratic trend of register contributed relatively little or not at all. Here, we cannot compare the relative contribution of linear versus quadratic trend of registers, as the inclusion of the quadratic trend in our regression analyses effectively modified the linear register trends. However, whenever we included the quadratic register trend, we did so because it improved model performance significantly. We find that sadness decreases only linearly with register, which is in line with the very small quadratic contribution (0.05%) in Eerola et al.'s models. However, they found no quadratic register contribution to the happiness predictions as we have. Eerola et al.'s stimuli differed from ours in duration (3 or 8 s here vs. their 25 s), instrument variation (flute, horn, and trumpet, i.e., no percussion or strings), as well as variation in other musical features that we did not include. The increased variation in musical features in their study may have obscured the importance of the quadratic trend of registers we clearly find here.

The percussion family is distinctively different from the other instrument families on all affect scales, either as a main effect or interacting with pitch register and/or affect locus. Percussion expressed and induced the least tension, scariness, sadness, or anger, the most positive valence, happiness, tenderness, and was the most preferred. Although McAdams et al. (2017) found similar instrument family effects to the dimensional affect results presented here, percussion plays a slightly less prominent role in their data. They found that strings were more positively valenced than percussion. Percussion was more convex decreasing with tension arousal for nonmusicians but showed a more neutral (flat register) trend for musicians. The lower pitch registers of percussion instruments were perceived to be higher in energy than the other instrument families, especially for nonmusicians, similar to our findings. The participant population in the current study is likely to be similar in musical sophistication to the nonmusician participant group in McAdams et al.'s study, because the average general Gold-MSI score and musical training sub-scores we found were towards the lower end of the percentile norms (see Table 4.8). Percussion may also have a bigger effect in our studies because our stimuli were longer in duration (3/8 s vs. 500 ms in McAdams et al.). Furthermore, we used a subset of the stimulus set in McAdams et al., excluding variation in playing technique and attack strength. Chau et al. (2015) also found that pitched percussion (marimba, vibraphone, and xylophone) were perceived as more positive. Participants

Table 4.8 *Raw Gold-MSI Scores and Corresponding Percentiles of the Participant Sample*

Gold-MSI		**Mean raw scores (SD)**	**Percentile**
General	Exp. 1	68.2 (18.6)	26
	Exp. 2	69.7 (17.5)	29
Musical Training	Exp. 1	19.6 (10.0)	31–32
	Exp. 2	19.0 (9.1)	29–30

Note. Percentile estimations are based on the table in Appendix A3 of the Gold-MSI documentation (Müllensiefen et al., 2013).

in Huron et al. (2014) also judged that pitched percussion were uncommonly used and relatively less able to convey sadness. Thus, in addition to the banjo, our results suggest that one cannot play a sad (or tense, fearful, angry, negative, unhappy, non-tender, unlikeable) song on pitched percussion either.

The paradoxical enjoyment of sad music has been a topic of research interest in the recent decades (see Eerola et al., 2018, for a review). Although here we cannot allude to the mechanisms through which sad music is enjoyed, we do see that any affects that one may consider unpleasant or undesirable are more strongly perceived than they are induced. This is mostly true in the lower pitch registers, and, specifically for sadness, we see that throughout all pitch registers perceived sadness is rated higher than induced sadness. Thus, we see that unpleasant affects that are perceived are not as strongly induced, in line with findings by Zentner et al. (2008), but the divergence is not so extreme that the induced affect veers towards the pleasant end of the scale. In a review, Schubert (2013) concludes that induced affect is always rated lower than perceived affect, but here we find that this depends on the overall “pleasantness” of the relevant affect, as well as the pitch register. Consider the bipolarity of valence; negative affect is more strongly perceived than induced, but positive affect is more strongly induced than perceived. Tenderness is the only affect that behaves differently in this respect. Tenderness is more strongly perceived than induced. Perhaps tenderness, as an emotion category, does not translate to a listener’s feelings as well as it does to perceived affective intent.

We did not directly test the differences between Experiment 1 and 2, that is, between short sound examples and longer musical excerpts, because that would introduce a fourth dependent variable and compromise the interpretability of our results. We can, however, observe some (dis)similarities in the modelling results. The register and family effects are consistently found in both experiments, although the linear or quadratic nature of the register effect may show some differences. The affect locus effect is more consistently found in Experiment 2 with longer musical excerpts. The more consistent affect locus effect we found with the longer stimuli suggests a bigger divergence between perceived and induced affect than with short stimuli. This is contrary to what we expected, namely that the increased stimulus duration would allow the induced affect to further align with the perceived affect. One explanation for this may be a confound due to a change in experimental design. In Experiment 1 the affect locus was varied between groups, whereas, for practical reasons, we varied this factor within groups in Experiment 2. Because each participant rated both affect loci in Experiment 2, although separated by blocks, they may have been more aware of the differences in perceived and induced affect locus than the participants in Experiment 1 who only attended to one of the affect loci. Schubert (2013) refers to this as "contrasting," although he considers perceived and induced affect ratings that are discussed directly after each other, instead of separated by blocks. Further research is needed to consider the effects of stimulus duration and the experimental design independently.

Finally, we find general trends that reveal the importance of considering individual differences in musical affect research. Whereas the affective response to pitch register seems to be relatively independent of the individual differences we considered here, the affective response to instrument family appears to be more variable between participants. This is somewhat contrary to McAdams et al.'s (2017) results, which show that the pitch register effects interact with musicianship. Again, this difference may be caused by the relatively low score on the Gold-MSI questionnaire for our participant population. We do find, however, that Gold-MSI plays the most important role in moderating the effects of instrument family and affect locus, which further confirms the effect that musical expertise has on affective processing of musical stimuli. Thus, future studies investigating the mechanisms of affective responses to music may further investigate the role of musical expertise and ensure that their participant population shows a broad distribution on this questionnaire if they want to generalize their findings. Furthermore, given our finding on the moderating role of musical preferences, the often-narrow selection of musical stimuli from classical music by experimenters may prove problematic. A wider variety in musical genres may appeal to a wider variety of participants and thus allow us to investigate general trends in affective processing more accurately. Or, if a narrower stimulus selection is warranted, future studies are encouraged to investigate musical preferences as a moderator. Interestingly, empathy played the least frequent role as moderator, although other studies have consistently found it to play an important role in affective processing (Balteș & Miu, 2014; Egermann & McAdams, 2013; Garrido & Schubert, 2011). Here we too find that effects are moderated by dispositional empathy, but other factors of musical expertise, musical preferences, personality, and pre-existing mood feature more prominently. Finally, future studies may consider the role of individual differences especially when they are investigating the dimensional response of valence or the discrete response of fear.

4.6 Conclusion

This study aimed to follow up on and extend the findings from McAdams et al. (2017). We investigated the effects of pitch register and instrument family in both perceived and induced affect. Affect was not only considered in a dimensional representation but also in a discrete representation. In addition to single-note sounds, we also looked at the effects of longer, more musical, sound examples. Finally, the experiments were carried out online, to reach a wider and more varied participant population. We find results that are consistent with previous studies, on the effects of pitch register and instrument family. We also contribute new findings on the difference between perceived and induced affect by directly comparing the two affect loci. The results are also mostly consistent between shorter and longer stimuli, except for a stronger role of affect locus. Although the previous chapter discussed the two affect representations in more detail, the added representation here provides information for those specifically interested in, for example, the enjoyment of sad music. Our exploration of the moderating effects of individual differences reveals which individual factors may be particularly influential, which effects may be particularly moderated, and which affective responses may be particularly influenced. Future studies ought to consider these when they investigate general tendencies of musical affect or the specific mechanisms through which affect is perceived and induced.

Chapter 5

The Acoustic Properties of Affective Timbres

5.1 Introduction

Music has the ability to communicate and evoke musical emotions. Structural features of musical mode and tempo have been most often studied and appear to have the greatest influence on perceived musical affect. Local features, such as the timbre of a sound, have received less attention and appear to have a smaller influence on perceived musical affect compared to mode and tempo (Eerola et al., 2013; Gabrielsson & Lindström, 2010; Grimaud & Eerola, 2022). Nevertheless, in orchestration practices, timbre is a fundamental tool in communicating musical affect (McAdams, 2019b; Schutz et al., 2008). The finding that reliable affective judgments can be made on sounds as short as 250 ms, during which mode or tempo cannot be established, indicates that local features like timbre can influence musical affect (Bigand et al., 2005; Filipic et al., 2010; McAdams et al., 2017; Peretz et al., 1998). Here, we will dive deeper into the role of timbre, and specifically its acoustic origins, in perceived and induced musical affect. This will consist of the re-analysis of two previously published studies by Eerola et al. (2012) and McAdams et al. (2017) who investigated the timbral origins of perceived musical affect, which will be described in further detail below, and the comparison of those results to the current experiments that extend the previous studies in multiple ways.

Several studies have investigated the role of different instruments and their influence on perceived or induced affect. String instruments like the violoncello or violin were judged to be most capable of expressing sadness, whereas most percussion instruments like the snare drum or tambourine were judged the least capable (Huron et al., 2014). A perception study found that happiness was less likely perceived in violin sounds, sadness in synthesizer sounds, and anger in trumpet sounds (Hailstone et al., 2009). Regarding nonsustaining instrument sounds, more positive affects were associated with pitched percussion instruments, and more negative affects with string instruments (Chau et al., 2015). Perceived affect of common orchestral instruments that varied in pitch register, attack, and playing technique showed that strings were perceived to be the most positively valenced and least tense, whereas woodwinds were the most negatively valenced and brass the tensest (McAdams et al., 2017). Furthermore, in the previous chapter we showed that in an online testing environment, pitched percussion were rated as least tense, scary, sad, or angry, and most positive, happy, and tender on both perceived and induced affect. Strings were considered the most angry and fearful. Both the previous chapter and McAdams et al. (2017) showed that the effects of instrument family also interacted with pitch register. Finally, in an experiment investigating features that are relevant in both music perception and production, string instruments were associated with communicating anger and fear (Grimaud & Eerola, 2022).

Timbre, alongside other features such as pitch and loudness, can be used to discriminate and identify different sound sources (McAdams, 1993). However, it is incorrect to say than an instrument has ‘a timbre’. A single instrument may be perceived to have very different timbres when they are playing at different pitches or dynamics (Marozeau et al., 2003; Risset & Wessel, 1999). An instrument rather has a “constrained universe of timbres” (McAdams & Goodchild, 2017b, p. 129). Consequently, investigating how different instruments influence affective processing does not give a complete picture of the role of timbre. The multidimensional perception of timbre can be described by looking at the characteristics of the temporal envelope, spectral envelope, and spectrotemporal envelope of a sound, i.e., timbre descriptors. These computational timbre descriptors can be related to the perceptual and semantic dimensions of timbre, such as the spectral centroid, which has been associated with perceptions of brightness (Schubert & Wolfe, 2006).

Two studies have investigated how such timbre descriptors can be related to affective judgments. First, Eerola et al. (2012) investigated the role of timbre descriptors in the prediction of perceived affect. In their first experiment, participants were asked to rate the perceived valence, tension arousal, and energy arousal in response to 1-s single-note stimuli at a D#4 pitch. The stimuli consisted of a wide variety of instruments, both common (e.g., piano, guitar, clarinet) and less common (e.g., shawm and crumhorn), from different eras and music genres. They found that energy and tension arousal were strongly correlated, $r(107) = .84$, and continued their analyses only on the valence and energy arousal responses (a second experiment further corroborated the validity of a two-dimensional valence/arousal approach). For each of the 110 instrument sounds, the means of 26 timbre descriptors were calculated with the MIRToolbox in MATLAB (Lartillot & Toiviainen, 2007). Through principal components analysis (PCA), the 26 descriptors were reduced to seven components. Robust regression analysis showed that the ratio of high-frequency to low-frequency energy, temporal centroid, and spectral skewness predicted perceived valence.

Positively valenced sounds showed a high envelope centroid in the temporal domain (i.e., more sustain) and contained more energy in the lower frequencies. Energy arousal was predicted by attack slope, temporal centroid, and ratio of high-frequency to low-frequency energy. Energetic sounds instead showed faster attacks and were less sustained, with more energy in the higher frequencies. In a third experiment, these results were mostly corroborated with a more restricted set of sounds (orchestral instruments) with spectral and loudness manipulations. These findings were in line with previous research on speech and emotion, which showed that valence was related to less relative energy in the higher frequencies, and activation was related to more relative energy in the higher frequencies and steeper attack slope (Laukka et al., 2005).

The second study, by McAdams et al. (2017), extended the experiments from Eerola et al. (2012) by analyzing the perceived affective response to stimuli at varying pitch registers, instead of notes at a single pitch height. This allowed the researchers to study the effect of timbre and pitch-related differences in timbre on perceived affect. The selected 137 stimuli consisted of orchestral instrument sounds of 500-ms duration at the D# pitch that ranged between registers 1–7 based on each instruments' tessitura. Participants similarly rated the perceived affect on valence, tension arousal, and energy arousal scales. Here, the two arousal dimensions did not correlate strongly, $r(135) = -.29$, and, consequently, all three affect measures were used for further analysis. The timbre descriptors were obtained with the Timbre Toolbox (Peeters et al., 2011), which gave median and interquartile-range measures of each descriptor for each stimulus. First, based on hierarchical cluster analysis, the initial set of 23 timbre descriptors was reduced to 17 because of high collinearity. Then, a partial least squares regression (PLSR) analysis, which couples multiple linear regression with PCA, was run for each of the affect scales. Furthermore, neural network (NN) modelling was conducted as a nonlinear analysis of the relationship between the timbre descriptors and the affect ratings. The nonlinear approach appeared to have a moderate advantage over the linear approach in terms of model fitness and predictive power. Looking at the linear and nonlinear results combined, more positively valenced sounds showed more high-frequency energy (lower spectral slopes), a greater emergence of strong partials, a sharper attack, and earlier decay. This is nearly opposite of the findings by Eerola et al. (2012), who found that positively valenced sounds had more low-frequency energy and a sustained temporal development. Highly tense sounds contained more high-frequency energy, showed more spectral variation, and more gentle attacks. Highly energized (awake) sounds were mostly predicted by pitch-related spectral descriptors, which is somewhat in line with Eerola et al., although they also found an attack component. The discrepant findings in energy arousal may be explained by the differences in pitch variation; however, it is unclear what caused the stark discrepancy in the findings on valence. They may be caused by differences in stimulus selection, differences in computational approach (MIRToolbox vs. Timbre Toolbox), or differences in analytical approach (PCA & robust regression vs. PLSR and NN).

The first aim of this project is to re-analyze the results from both studies with identical computational and analytical approaches to enable a closer comparison of the findings. The second aim is to further extend the above experiments by analyzing the data from two new experiments that 1) measure both perceived and induced affect, 2) on both dimensional and discrete affect scales, 3) in response to both single notes (Experiment 1) and chromatic scales (Experiment 2), 4) in an online testing environment. As was shown in the previous chapter, pitch register and instrument family influence perceived and induced affect in a slightly different manner, and consequently, the timbre descriptors that predict the two affect loci may also be different, although perhaps only in a subtle manner. We will compare the results of the perceived ratings, in response to single-note stimuli (Experiment 1) on the dimensional affect scales, with those of Eerola et al. (2012) and McAdams et al. (2017). We expect our findings to be most similar to those from McAdams et al., as our stimuli were a subset of their stimulus set. The consistency in computational and analytical approach is expected to reveal findings that are consistent across all three stimulus sets. Then, we will extend those findings by analyzing the results on induced affect, the chromatic-scale results, and the discrete affect scales. Finally, we expect some overlap between the findings on dimensional versus discrete scales, given the results of Chapter 3. However, we expect that sadness may show timbral patterns that have not been revealed by the dimensional affect scales.

5.2 Analysis

Table 5.1 summarizes the methodological approaches of Eerola et al. (2012; henceforth referred to as Eerola12)[2], McAdams et al. (2017; henceforth referred to as McAdams17), and the current Experiment 1. These three experiments overlap in many aspects, but also show some key

[2] The stimuli and rating data for Eerola12 are available at https://github.com/tuomaseerola/timbre2012

Table 5.1 *Comparison of Methodological Approaches of Eerola et al. (2012), McAdams et al. (2017), and Experiment 1*

	Eerola12	McAdams17	Experiment 1
Stimuli	MUMS	VSL	VSL
	Single note	Single note	Single note
	D#4	D#1–7	D#1–7
	1 sec	1 sec	3 sec
	Wide variety of instruments	Orchestral instruments	Orchestral instruments
	Different playing techniques	Different playing techniques & attacks	No different playing techniques or attacks
Affect Ratings	Perceived	Perceived	Perceived & Induced
	Valence: unpleasant/pleasant	Valence: negative/positive	Valence: negative/positive
	Tension: tense/relaxed	Tension: tension/relaxation	Tension: tense/relaxed
	Energy: tired/awake	Energy: tired/awake	Energy: tired/awake
Environment	In lab	In lab	Online
	English & Finnish	English	English
Timbre Descriptors	MIRToolbox 1.2.4	Timbre Toolbox (2016)	Timbre Toolbox (2021)
	Spectral & Temporal	Spectral & Temporal	Harmonic, Spectral, & Temporal
	Mean	Median & IQR	Median & IQR
Analysis	Principal component analysis, correlation & robust regression	Partial least squares regression & neural net modelling	Lasso regression & random forest modelling

differences. Eerola12's stimulus set is obtained from a different source than both McAdams17's and the current stimulus set. The Eerola12 set also does not vary in pitch register but does contain a greater variety of instruments. The current stimulus set contains no variations in playing technique or attack, as this was not shown to have an effect in McAdams17. Whereas Eerola12's affect scales diverge on the valence labels, McAdams17 diverges on the tension labels. The current experiment was run online, whereas the other two were executed in a controlled laboratory environment. All three experiments took different approaches to describe the timbre of the stimuli and to use those predictors to predict affect ratings.

The differences in experimental design may cause discrepancies in results, but this cannot be changed. However, the differences in analytical approach may also cause discrepancies in results. Here we re-analyze the data in a uniform manner. In the current experiment, we decided to describe the acoustic origins of the stimuli using timbre descriptors from the Timbre Toolbox. We analyzed the relationship between those descriptors and the affect ratings in two ways; (linear) lasso regression and (nonlinear) random forest regression. Concurrently, we also re-analyzed the results from Eerola12 and McAdams17 with an identical approach.

The Timbre Toolbox was originally created by Peeters et al. (2011) but has since undergone modifications. Here we used the latest version (TimbreToolbox-R2021a; Kazazis et al., 2021), which is a completely reprogrammed version of the Timbre Toolbox. The computations of audio descriptors diverge from the ones described by Peeters et al. (2011) and the toolbox includes new descriptors as well. The toolbox first preprocesses the audio input, further analyzes the temporal and spectral parameters required for audio representation, extracts each audio descriptor, and summarizes the time-varying values to summary statistics. Here, we analyzed descriptors from the power spectrum of the Short-Term Fast-Fourier Transform (STFT; spectral domain), the harmonic and inharmonic partials (harmonic domain), and the temporal energy envelope of the audio signal (temporal domain). For the descriptors of the harmonic and spectral domain of the single notes we used the median and interquartile range (IQR) summary statistics. For the chromatic scales of the second experiment, we also included the range of the harmonic and spectral descriptors, as they provide additional information relevant to longer stimuli. Table 5.2 shows all timbre descriptors we used in our analysis, and their relevant descriptions.

Table 5.2 *List of Timbre Descriptors Used From the Timbre Toolbox*

Domain	Descriptor	Description
Harm	Pitch	Pitch estimation based on algorithm of Camacho & Harris (2008)
	Spectral Deviation	Deviation of partial's amplitudes for a global spectral envelope (average of three adjacent harmonic amplitudes)
	TriStimulus 1	Amplitude of 1st harmonic normalized by the sum of amplitudes of all harmonics
	TriStimulus 2	Amplitude of 2nd, 3rd, and 4th harmonics normalized by the sum of amplitudes of all harmonics
	TriStimulus 3	Amplitude of the 5th and higher harmonics normalized by the sum of amplitudes of all harmonics
	Odd:Even Ratio	Ratio of odd partials to even partials
	Inharmonicity	Weighted sum of deviation of each partial from its ideal harmonic rank
	Harmonic Energy	Amount of energy in the partials (one partial per harmonic rank)
	Noise Energy	Partial's energy subtracted from total energy
	Noisiness	Ratio of noise energy to total energy
	Harmonic:Noise Energy	Ratio of harmonic energy to noise energy
	Partials:Noise Energy	Ratio of partials energy to noise energy
Spect	Centroid	Spectral center of gravity
	Spread	Spread of the spectrum around its mean value
	Skewness	Asymmetry of the spectrum around its mean value
	Kurtosis	Flatness of the spectrum around its mean value
	Flatness	Measure of noisiness of the spectrum
	Crest	Measure of "peakiness" (or tonalness) of the spectrum
	Slope	Slope of the spectral envelope over frequency
	Decrease	Steepness of the decrease of the spectrum
	Roll Off	Frequency below which 95% of the signal's energy is contained
	Variation	Variation of the spectrum over time, based on correlation between spectra in successive timeframes
	Flux	Variation of the spectrum over time, based on Euclidian distance between spectra in successive time frames
Temp	Attack Time	Time it takes the waveform to reach its maximum level
	Log Attack Time	$\log_{10}$ of the attack time
	Attack Slope	Rate of increase of the signal energy during the attack time
	Decrease Slope	Rate of decrease of the signal energy during the sustained part of the sound
	Temporal Centroid	Center of gravity of the energy envelope
	Effective Duration	Crude approximation to the perceived duration of a sound, i.e., the time the energy envelope is above –10 dBFS
	Frequency of Energy Modulation	Frequency of modulation of energy over time, for the sustained part of a sound
	Amplitude of Energy Modulation	Amplitude of modulation of energy over time, for the sustained part of a sound

Since the accuracy of the harmonic descriptors depends on the accuracy of the estimated pitch, we excluded the harmonic descriptors for some stimuli whose pitch estimation was inaccurate. From the first experiment, we excluded harmonic descriptors for two timpani stimuli playing the second and third register, which were highly inharmonic sounds. From McAdams17's stimuli, we excluded harmonic descriptors for highly inharmonic percussion sounds as well (five timpani D#2 and D#3, two bowed gong D#3 and D#4, a glockenspiel D#6, a crotale D#6 stimulus). Eerola12's stimulus set only required exclusion of harmonic descriptors from one stimulus, which was an organ sound with several audible organ stops.

To analyze how the timbre descriptors linearly predicted the affect ratings, we decided to use lasso regression. We did not use PCA followed by regression like Eerola12 because conducting a PCA and then choosing the highest loading descriptor on each PCA as regression predictor may result in different predictors for each stimulus set. We also did not use PLSR like McAdams17, as the interpretation of each component as predictors, with all descriptors loading on each component, is somewhat arbitrary and similarly complicates comparison between datasets. Lasso (**l**east **a**bsolute **s**hrinkage and **s**election **o**perator) regression minimizes sum of squares (goodness of fit) under the constraints of l_1 regularization by applying penalties to the magnitudes of coefficients (shrinkage) and eliminating predictors whose coefficients are reduced to zero (selection; Tibshirani, 2011). It trades of a slightly higher sum of squares in exchange for a sparser model. This is well suited for a predictor set that contains high multicollinearity (which is the case for the timbre descriptors) and automates variable selection without a priori assumptions. After variable selection through lasso regression, we performed multiple linear regression (LR) with only the lasso-selected predictors to obtain standardized coefficients and measure fivefold cross-validated R^2 trained on the full dataset. Henceforth, the term *lasso regression* will refer to the standard regressions performed on the lasso-selected variables, rather than to the lasso regressions themselves.

McAdams17 found that the nonlinear method of supervised feedforward artificial neural networks with back propagation provided better model fit than PLSR for affect prediction. Although the downside of nonlinear methods is that they are less easily interpretable than linear methods (black box), we did decide to include a nonlinear analysis alongside the linear analysis so we could compare the performance and results. Here, we used random forest regression (RF) because, compared to neural network modelling, it is less prone to overfitting, less computationally intensive, and requires less tweaking of hyperparameters. A random forest is a collection of decision trees each based on a random selection of different observations and predictors that are then averaged to provide a single averaged prediction estimate (Biau & Scornet, 2016; Breiman, 2001). Relative variable importance (RVI) allowed us to interpret the RF results, as it ranks the variables in order of how much they improved the model when decision splits were made on a given variable, with the most important variable receiving a score of 100 (Liaw & Wiener, 2002). Note, however, that this does not tell us *how* these variables related to the relevant prediction, like a standardized coefficient in linear regression would. Finally, for the RF models we also conducted fivefold cross-validation to compare R^2 of the model trained on the full data.

As additional model performance measures for the LR and RF models, the *testing* R^2 and *testing RMSE* of each five-fold iteration (where the four folds predicted the fifth fold) were averaged as measures of predictive relevance. All *RMSE* values were then also converted to normalized *RMSE* (*NRMSE*), because the affect ratings of the different experiments were made on scales with different ranges. This allowed for closer comparison. *NRMSE* was obtained by dividing the *RMSE* values by the mean rating of the relevant scale.

The Timbre Toolbox was used in Matlab version 2020b (The MathWorks Inc., Natick, MA). All further analysis were conducted in R version 4.2.1 (www.r-project.org). We used the *glmnet* package to perform lasso regression analysis, with added modifications that repeated the lasso 100 times with tenfold cross-validation (Friedman et al., 2022). We used the *standardize_parameters* function of the *effectsize* package to calculate standardized regression coefficients (Ben-Shachar et al., 2022). We used the *lm* function for the linear regression, the *train* function for the RF regression, and the *varImp* function for the RVI values from the *caret* package (Kuhn et al., 2022).

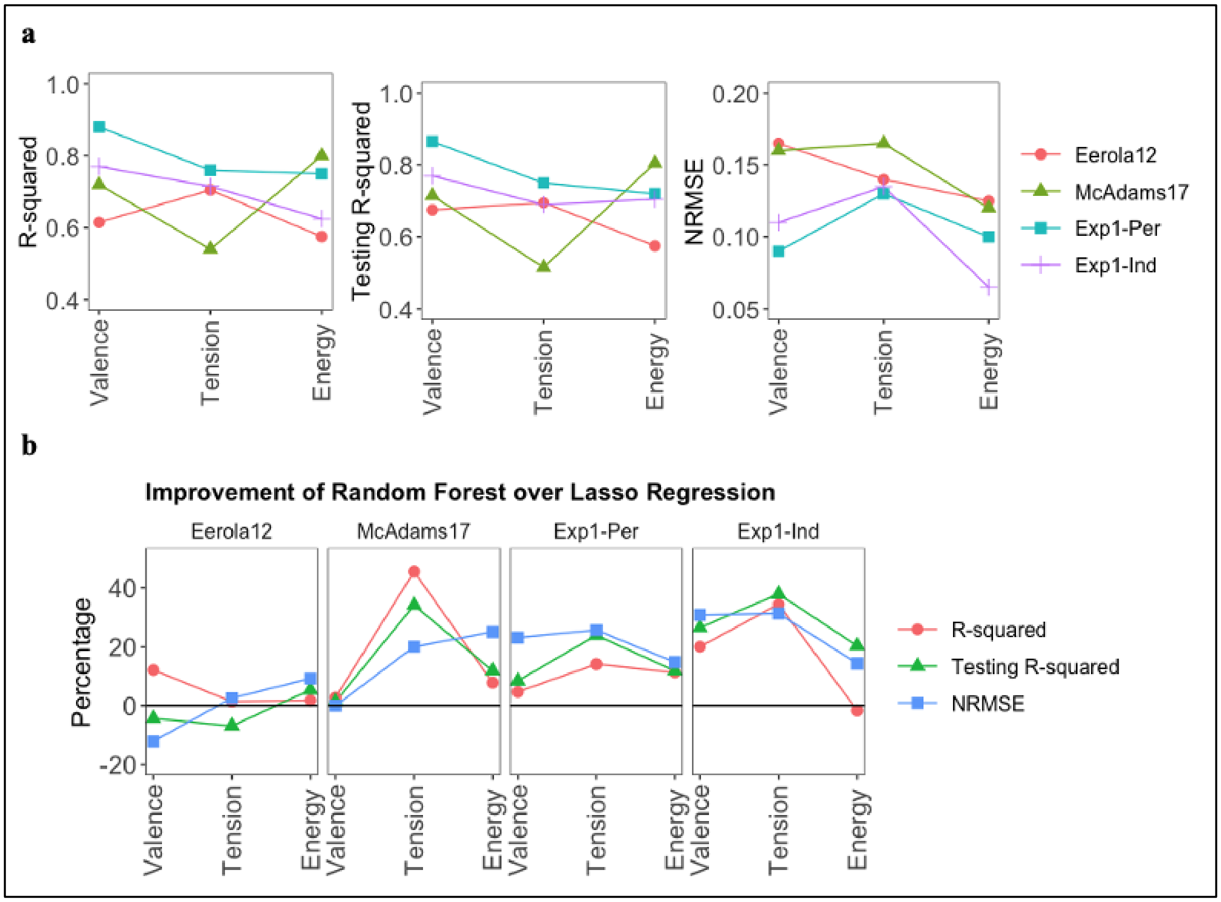
5.3 Results

Eerola12, McAdams17, and Experiment 1 Perceived & Induced

Model Performance

First, we will discuss the results from Eerola12, McAdams17, and the perceived and induced dimensional affect ratings from Experiment 1 (single notes), as these were the most similar to each other in experimental design. Figure 5.1 shows the model performance of the lasso regression (linear) and random forest regression (nonlinear) in predicting the affect ratings of the four datasets, as well as the percentage improvement of RF over lasso (see full results in Tables B1 and

Figure 5.1 *R^2, Testing R^2, and NRMSE for the Prediction of Eerola12, McAdams17, and Experiment 1 (Perceived & Induced) (a), and Percentage Improvement of RF over Lasso Regression (b)*



Note. a) Measures are averaged over Lasso and RF regressions; b) a positive percentage means that RF performance is improved compared to Lasso, and a negative percentage indicates deterioration in model performance.

B2 in Appendix B). When we compare the performance of the four datasets, taking together the results from both the lasso and random forest regressions (since these models did not show extremely divergent results), we see that the perceived affect model for Experiment 1 shows the highest explained variance on the three rating scales (average $R^2_{Valence}$ = .88, average $R^2_{Tension}$ =.76, average R^2_{Energy} = .75), followed by the induced affect ratings (average $R^2_{Valence}$ = .77, average $R^2_{Tension}$ =.72, average R^2_{Energy} = .63), then by McAdams17 (average $R^2_{Valence}$ = .72, average $R^2_{Tension}$ = .54, average R^2_{Energy} = .80), and finally Eerola12 (average $R^2_{Valence}$ = .62, average $R^2_{Tension}$ = .71, average R^2_{Energy} = .58). Inversely, when we compare the three dimensional affect scales, we find that valence is best predicted by the timbre descriptors (average R^2 = .75), followed by energy arousal (average R^2 = .69), and tension arousal (average R^2 = .68). We see similar trends for the *testing* R^2 and *NRMSE*.

Given that lasso is effectively sacrificing goodness of fit in exchange for parsimony whereas RF keeps all predictors in its analysis, it is to be expected that RF will result in higher R^2 values than lasso because the greater the number of predictors included, the higher the predictive accuracy will be. Indeed, there is an improvement in explained variance of nonlinear over linear modelling across all four datasets and all three affect scales, although relatively small (13% improvement). Averaged across all three scales and measures, the McAdams17 data showed the most improvement from nonlinear analysis (16% improvement on average), particularly for the tension ratings (33% improvement). This is in line with their original findings comparing the NN and PLSR analyses, although the improvement they find of NN over PLSR is slightly larger. The current perceived and induced data from Experiment 1 show a decent improvement from nonlinear analysis as well (perceived, 15% improvement; induced, 24% improvement), also mostly for the tension ratings (perceived, 21% improvement; induced, 35% improvement). Eerola12's data benefit the least from a nonlinear approach (1% improvement), with the most overall improvement on the energy data (7% improvement). Furthermore, although RF is more prone to overfitting, the improvement in *testing* R^2 results for Eerola12 and Experiment 1 show that this is not the case as the average R^2 is higher even with different combinations (folds) of the data. Thus, overall, we see slight improvements in modelling performance when taking a nonlinear approach. The lack of pitch variation in Eerola12, and our finding in Chapter 4 that pitch shows nonlinear (i.e., quadratic) relationships with affect ratings, may explain why Eerola12 shows the least benefit from the nonlinear RF regression, as it is lacking the nonlinear pitch-related variation in timbre.

Thus, model performance was good overall, with decent R^2 (range = [.44, .90], mean = .71) and *testing* R^2 (range [.44, .90], mean = .70). The current data from the perceived affect of Experiment 1 and the valence data are best predicted by the timbre descriptors. We find that the nonlinear RF regression shows a slight improvement over the linear lasso regression, particularly for McAdams17 and the tension arousal data. We will continue with the interpretation of both the LR and RF results.

Valence

Table 5.3 shows the results for the lasso and RF regression predicting the valence data, showing the standardized coefficients of the timbre descriptors that were selected by the lasso regression and the rankings of the timbre descriptors with an RVI > 20 from the RF regression.

Table 5.3 *Lasso-Selected Linear and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Valence for Eerola12, Mcadams17, and the Perceived and Induced Ratings of Experiment 1*

	Eerola12		**McAdams17**		**Exp1-Per**		**Exp1-Ind**	
	LR	**RF**	**LR**	**RF**	**LR**	**RF**	**LR**	**RF**
Inharmonicity Median	–	–	–0.26	2	–	–	–	–
Tristimulus 1 Median	0.29	4	–	–	–	–	–	–
Tristimulus 3 Median	–	6	–0.27	1	–0.26	4	–	4
Spectral Centroid Median	–	3	–	–	–	–	–	5
Spectral Decrease Median	–	5	–	–	–	–	–	6
Spectral Flatness Median	–	8	–	–	–	–	–	–
Spectral Roll Off Median	–	1	–	–	–	–	–	–
Spectral Spread Median	–0.41	2	–	–	–	–	–	–
Spectral Centroid IQR	–	–	0.13	–	–	–	–	–
Spectral Flux IQR	–	–	–	–	–0.26	–	–0.32	–
Spectral Variation IQR	–	–	–0.15	–	–	–	–	–
Attack Time	–	–	–	–	–	3	–	2
Frequency of Energy Modulation	–	7	–	–	–	–	–	–
Log Attack Time	–	–	–	–	–	2	–0.67	3
Temporal Centroid	–	–	–0.24	3	–0.37	1	–	1
R^2	.58	.65	.71	.73	.86	.90	.70	.84
Test-R^2	.69	.66	.71	.72	.83	.90	.68	.86
NRMSE	.18	.15	.16	.16	.10	.08	.13	.09

Note. The LR column shows the standardized coefficients of the timbre descriptors that were selected by the lasso regression, all at $p < .05$. The RF column shows the rankings of the timbre descriptors with an RVI > 20.

We see that the Eerola12 valence ratings are linearly predicted by median spectral spread and median tristimulus 1, which are both also mirrored in the nonlinear results. Stimuli that have a narrower spectral spread and more relative energy in the first harmonic are rated as more positively valenced. Based on the nonlinear RF, several other spectral and spectrotemporal descriptors also play a role. In their original paper, Eerola et al. found that valence was predicted by the ratio of high-frequency to low-frequency energy, spectral skewness, and temporal centroid. Here, the temporal component was not selected by the lasso or RF regression. The ratio of high-to-low frequency energy and spectral skewness can, however, be associated with our findings on tristimulus 1, as they all reflect an increase of energy in the lower frequencies (or first harmonic), compared to higher frequencies (or upper harmonics), as stimuli are perceived as more positive.

The McAdams17 results show that valence is linearly and nonlinearly predicted by median inharmonicity, median tristimulus 3, and temporal centroid. The IQR of spectral centroid and variation also play a role, albeit smaller and only linearly. Thus, stimuli that are perceived to be more positively valenced are more harmonic, contain less relative energy in the upper harmonics, and contain more energy in the beginning of the temporal envelope. In their original paper, McAdams et al. also found that valence was predicted by high-frequency energy, due to the predictor of median spectral decrease In their original publication they concluded that this means that positively valenced sounds contain more higher-frequency energy, which seems to contradict our findings on median tristimulus 3. However, spectral decrease and tristimulus 3 are not significantly correlated. Median tristimulus 3 is more strongly correlated to median inharmonicity, $r(135) = .75$, $p < .001$, and median spectral crest, $r(135) = –.76$, $p < .001$, thus rather suggesting that the current findings show that positive sounds have a more clear emergence of the fundamental frequency. Median spectral crest also positively predicted valence in the original McAdams17 analysis. The relative higher-frequency energy based on median spectral decrease was not found here, which could be due to different settings in the Timbre Toolbox, the exclusion of harmonic descriptors in the original analysis, or other differences in analytical approach. The original McAdams17 finding that positively valenced sounds have sharper attacks and earlier decays is however, consistent with the current findings; attack slope is strongly correlated with temporal centroid, $r(135) = .83$, $p < .001$.

The results for the perceived ratings of the current first experiment show that valence is also linearly and nonlinearly predicted by median tristimulus 3 and temporal centroid. We also find a linear predictor of the IQR of spectral flux and nonlinear predictors of attack time and log attack time (LAT). Thus, stimuli that are perceived to be more positively valenced show a more clearly emerging fundamental harmonic, more relative energy in the earlier part of the temporal envelope, and a narrower range of spectral variability over time. The results for induced valence are mostly similar, although there is a stronger contribution of the attack component, and median tristimulus 3 is only a nonlinear predictor alongside other spectral descriptors. The results on relative energy in the upper harmonics and the temporal component are consistent with McAdams17. Furthermore, spectral variation and spectral flux are both measures of the variability of the spectrum over time (though not correlated). So some aspects of spectrotemporal properties are found in both McAdams17 and Experiment 1. We thus show that these findings are mostly consistent with a different participant population, experimental design, analytical approach, and affect locus. The results of tristimulus 1 for Eerola12 and tristimulus 3 in the other two datasets are also directly related, and thus indicate that the finding that positively valenced sounds contain more energy in the first harmonic compared to the upper harmonics is also consistent with different stimulus sets, participant populations, experimental design, and analytical approaches.

Tension Arousal

Table 5.4 shows the modelling results for the tension arousal ratings. Several harmonic and temporal timbre descriptors are significant linear predictors of Eerola12's tension ratings: median tristimulus 1 and 3, median harmonic-to-noise and partials-to-noise energy ratio, attack time, decrease slope, and frequency of energy modulation. RF regression suggests nonlinear contributions of median spectral rolloff and spread, but this approach did not show much improvement over the more easily interpretable linear lasso regression approach. Stimuli that were perceived as more tense showed more noisy energy, a less clearly emerging fundamental harmonic, slower attacks and decays, and a slower modulation frequency. Tension arousal was strongly correlated with valence, $r(108) = -.88$, $p < .001$, and energy arousal, $r(108, p < .001) = .84$. Consequently, one might expect the lasso and RF findings for tension arousal to be similar to the findings on valence and energy arousal. Although there is some overlap on the median tristimulus 1, valence and tension diverge in their temporal component, although a temporal component was found in Eerola12's original results. Noise energy also plays a role in tension, but not in valence.

Table 5.4 *Lasso-Selected Linear and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Tension for Eerola12, Mcadams17, and the Perceived and Induced Ratings of Experiment 1*

	Eerola12		**McAdams17**		**Exp1-Per**		**Exp1-Ind**	
	LR	**RF**	**LR**	**RF**	**LR**	**RF**	**LR**	**RF**
Harmonic:Noise Energy Median	–0.36	–	–	10	–	–	–	–
Inharmonicity Median	–	–	–	–	–	6	–	7
Noisiness Median	–	–	–	9	–	–	–	–
Partials:Noise Energy Median	–0.24	–	–	7	–	–	–	–
Pitch Median	–	–	0.22	–	–	–	–	8
Tristimulus 1 Median	–0.24	–	–	–	–	–	–	–
Tristimulus 3 Median	0.43	–	–	–	–	3	–	6
Spectral Centroid Median	–	–	–	6	–	5	–	5
Spectral Decrease Median	–	–	–0.32	5	–	–	–	4
Spectral Flatness Median	–	–	–	1	–	–	–	–
Spectral Flux Median	–	–	0.20	3	–	–	–	9
Spectral Roll Off Median	–	2	–	2	–	–	–	10
Spectral Spread Median	–	1	–	–	–	–	–	–
Spectral Variation Median	–	–	–	8	–	–	–	–
Spectral Flux IQR	–	–	–	–	0.32	–	0.37	1
Attack Time	–0.18	–	0.23	–	–	1	–	–
Decrease Slope	–0.13	–	–	–	–	–	–	–
Frequency of Energy Modulation	–0.15	–	–	–	–	–	–	–
LAT	–	–	–	–	0.66	2	0.57	2
Temporal Centroid	–	–	–	–	–	4	–	3
R^2	.70	.71	.44	.64	.71	.81	.61	.82
Test-R^2	.72	.67	.44	.59	.67	.83	.58	.80
NRMSE	.14	.14	.18	.15	.15	.11	.16	.11

Note. See Table 5.3.

The tension ratings were not analyzed in Eerola12's original publication, so these cannot be compared.

RF modelling was better at predicting McAdams17's tension ratings than was lasso regression. Linearly, stimuli that were perceived as tense showed a higher pitch, less steep spectral decrease (i.e., a more flat frequency distribution), higher median spectral flux, and slower attack time. The median spectral decrease and flux predictors were also selected by the RF regression, but neither attack time nor any other temporal components were. Several spectral and harmonic descriptors nonlinearly predicted tension arousal in addition to spectral decrease and flux, in particular spectral flatness and rolloff, as well as spectral centroid and measures of noisiness derived from the harmonic input representation. The findings mostly overlap with the findings in McAdams et al.'s original paper with regards to effects of spectral variation and attack slope. Moreover, they also found linear and nonlinear effects of median spectral centroid. Although here we only found it as a nonlinear predictor, median spectral centroid is correlated with median spectral roll off, $r(135) = .91$, $p < .001$, median pitch, $r(135) = .66$, $p < .001$, and median spectral decrease, $r(135) = -.50$, $p < .001$. Thus, with both statistical approaches and in both versions of the Timbre Toolbox, we see that increased tension is predicted by higher pitches with more energy in the higher frequencies, more spectral variation, and slower attacks.

The perceived and induced tension ratings of the current Experiment 1 were also better predicted by RF than lasso regression. The linear and nonlinear methods show similar findings for the timbre descriptor LAT, which is also in line with McAdams17's findings: stimuli that had higher perceived and induced tension ratings had a slower (log) attack time. Interestingly, this is opposite to what we find for Eerola12, where a faster attack is associated with increased perceived tension. Thus, there may be some differences in stimulus selection that led to these diverging results. Linearly, the current perceived and induced tension ratings are also predicted by IQR spectral flux, which is correlated with median spectral flux, $r(57) = .73$, $p < .001$, and thus coincides with McAdams17's finding in which tense sounds show more spectral variation. The RF regression finds further contributions of attack time and temporal centroid, as well as median spectral centroid, tristimulus 3, and inharmonicity. The findings on median spectral centroid may overlap with McAdams17's and Eerola12's findings on higher-frequency energy and a less clearly emerging fundamental harmonic. Note that here, similar to Eerola12, tension arousal was strongly correlated with valence, $r_{per}(57) = -.89$, $p < .0001$; $r_{ind}(57) = -.96$, $p < .0001$, and thus we may expect some overlap in timbre descriptor results. Indeed we see an overlap in the temporal component (temporal centroid and LAT) and IQR spectral flux; as the attack is slower, and IQR spectral flux increases, there is less positive and more tense affect perception and induction. The linear analysis with median tristimulus 3 in the valence results and the nonlinear analysis with the same descriptor in the tension results suggests that an increase in upper-harmonic energy is associated with more negative valence and increased tension.

Comparing the four datasets, we see that Eerola12 and McAdams17 found that increased tension is associated with more energy in the upper harmonics or higher frequencies, which can be associated with a less clearly emerging fundamental harmonic and a more flat frequency distribution. The findings on spectral centroid and tristimulus are reflected in the nonlinear findings of the current experiment. The findings on attack clearly diverge between Eerola12, on the one hand, and McAdams17 and the current experiment, on the other hand. The main differences in experiment design between Eerola12 versus McAdams17 and the current experiment are the variation in pitch register and the instrument selection. Upon listening, the more relaxed sounds from Eerola12 are produced by woodwind instruments with slower attacks, whereas the more relaxed sounds from McAdams17 and the current experiment are produced by percussion instruments with faster attacks. All three experiments, however, contained woodwind and percussion instruments.

Energy Arousal

The energy arousal ratings of Eerola12 are both linearly and nonlinearly predicted by median tristimulus 3, attack time, median spectral slope, and IQR spectral spread. Only in the linear regression is median harmonic-to-noise energy an additional significant predictor, and in the nonlinear regression several harmonic, temporal, spectral, and spectrotemporal descriptors predicted energy arousal. Thus, stimuli that are perceived as more awake have more relative energy in the upper harmonics, a sharper attack, steeper spectral slope, and more variability of the spectral spread over time. As mentioned, Eerola12's energy ratings correlated strongly with the tension ratings, and we do see some overlap in timbre descriptors. Both an increase in tension and an increase in energy are associated with more relative energy in the upper harmonics and a sharper attack. In their original paper, Eerola et al. found that energy arousal was predicted by attack slope, temporal centroid, and ratio of high- to low-frequency energy. These results overlap in that sounds that were perceived as more awake had sharper attacks and more higher frequency or upper harmonic energy.

Table 5.5 *Lasso-Selected Linear and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Energy for Eerola12, Mcadams17, and the Perceived and Induced Ratings of Experiment 1*

	Eerola12		**McAdams17**		**Exp1-Per**		**Exp1-Ind**	
	LR	**RF**	**LR**	**RF**	**LR**	**RF**	**LR**	**RF**
Inharmonicity Median	–	–	–0.52	2	–0.57	2	–0.49	2
Harmonic:Noise Energy Median	–0.14	–	–	–	–	–	–	–
Harmonic Spectral Deviation Median	–	3	–	–	–	–	–	–
Pitch Median	–	–	0.35	1	0.35	1	0.39	1
Tristimulus 3 Median	0.37	6	–	–	–	–	–	–
Spectral Decrease Median	–	–	–	3	–	–	–	3
Spectral Flatness Median	–	10	–	–	–	–	–	–
Spectral Roll Off Median	–	2	–	–	–	–	–	–
Spectral Slope Median	0.23	11	–	–	–	–	–	–
Spectral Spread Median	–	1	–	–	–	–	–	–
Spectral Flatness IQR	–	7	–	–	–	–	–	–
Spectral Spread IQR	0.15	5	–	–	–	–	–	–
Spectral Centroid IQR	–	4	–	–	–	–	–	–
Attack Time	–0.28	9	–	–	–	–	–	–
LAT	–	8	–	–	–	–	–	–
Temporal Centroid	–	12	–0.21	–	–	–	–	–
R^2	.57	.58	.77	.83	.71	.79	.63	.62
Test-R^2	.56	.59	.76	.85	.68	.76	.64	.77
NRMSE	.13	.12	.14	.10	.11	.09	.07	.06

Note. See Table 5.3.

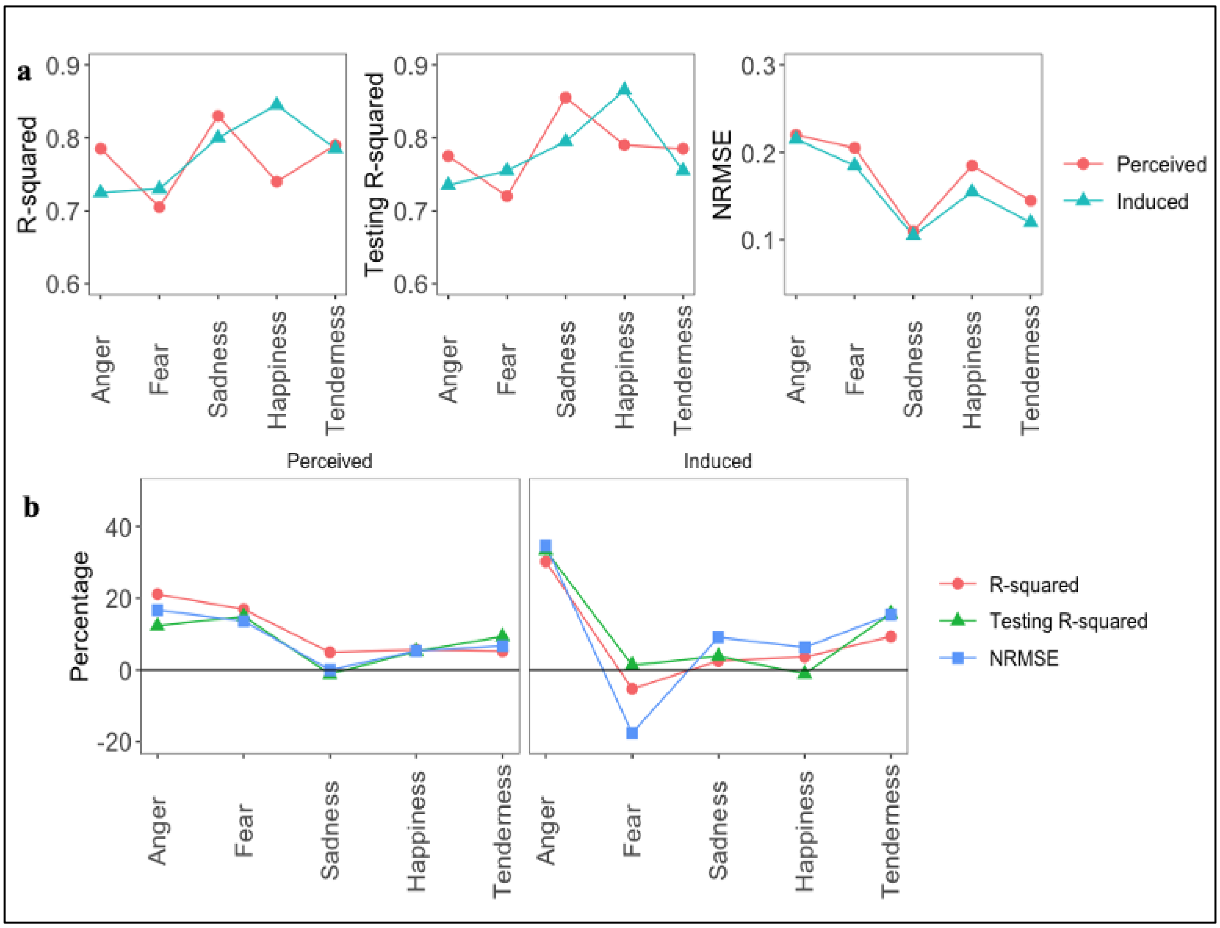
The timbre descriptors predicting the perceived energy arousal ratings of McAdams17 and the current experiment's perceived and induced energy ratings are highly similar; they are both linearly and nonlinearly predicted by median pitch and inharmonicity. Stimuli that resulted in increased perceived and induced energy arousal were higher in median pitch and lower in median inharmonicity. Interestingly, although median pitch and inharmonicity only show moderately strong correlations, $r(135)_{\text{McAdams17}} = -.54$, $p < .001$; $r(57)_{\text{Exp1}} = -.59$, $p < .001$, the two descriptors are computationally related. Due to the Nyquist limit and the mechanics of instruments which cannot produce very high-frequency modes of vibration, there are more audible harmonics in lower pitched sounds than in higher pitched sounds. Consequently, on average, the ability to detect and produce inharmonicity decreases as pitch increases. On a perceptual level, it has also been found that listeners are better at detecting inharmonicity at lower fundamental frequencies (Järveläinen et al., 2001). For McAdams17, we also see that perceptually awake stimuli contained more energy in the beginning of the temporal envelope, which is in line with the findings by Eerola12 on sharper attacks and the original findings by McAdams et al. in their NN model. In their original paper, McAdams et al. did not include pitch as a predictor in their analyses, but did find that several spectral descriptors that are associated with pitch height predicted energy arousal. Thus, whereas the relation of energy arousal with pitch height and inharmonicity are clearly present when there is pitch variation in the stimulus set (McAdams17 & Exp1), the relation with attack time is present when there is playing technique variation in the stimulus set (Eerola12 & McAdams17).

Summary

All four datasets show that positive valence is associated with more relative energy in the lower harmonics compared to the upper harmonics, i.e., a more clearly emerging fundamental frequency. The experiments with the most overlap in stimulus selection (McAdams17 and Exp1), also find that positive valence is predicted by more energy at the beginning of the temporal envelope, i.e., a more impulsive timbre. Tension, inversely to valence, is associated with more energy in the higher frequencies or upper harmonics in all four datasets, which can be associated with a flatter spectrum or less clearly emerging fundamental frequency. The perceived and induced dimensional affect ratings of Experiment 1 are generally predicted by the same descriptors, with only a more prominent role for attack time and a less prominent role for median tristimulus 3 in the prediction of induced valence. There is, however, a clear divergence on the attack component of tension. Whereas Eerola12 find that a fast attack predicts an increase in tension, McAdams17 and the current experiment show that a slower attack predicts an increase in tension. For energy arousal we can clearly see that the datasets that included pitch variation find that more awake sounds are higher in pitch, as was also suggested by the previous chapter and the original findings of McAdams et al. This increase in pitch is accompanied by a decrease in inharmonicity, i.e., tired sounds are highly inharmonic. The datasets that included more variation in attack or playing techniques (Eerola12 & McAdams17) also found that a faster attack predicted energy arousal, whereas the current experiment with a lack of attack or playing technique variation did not.

Experiment 1 – Discrete Perceived & Induced Affect

Figure 5.2 *R^2, Testing R^2, and NRMSE for the Prediction of the Perceived and Induced Discrete Affect Ratings of Experiment 1 (a), and Percentage Improvement of RF over Lasso Regression (b)*



Note. See Figure 5.1.

Model Performance

Figure 5.2 shows the model performance in predicting the perceived and induced discrete affect ratings, and the improvement of RF over lasso regression (see Table B3 in Appendix B for full results). Overall, the nonlinear method shows better performance than the linear method (9.4% improvement of R^2 and *test*-R^2, 9.0% improvement in *NRMSE*), particularly for the prediction of anger (>20% improvement) and less so for the other affect scales (<9% improvement). Regardless of statistical method, model performance is relatively good (R^2 = [.63, .86], *test*-R^2 = [.63, .87], *NRMSE* = [.10, .26]). There is not a single scale that stands out in terms of good or bad performance and there are no major differences in performance between the prediction of perceived and induced affect, except for happiness, which is better predicted in the induced affect locus (R^2_{lasso} = .83, R^2_{RF} = .86) compared to the perceived affect locus (R^2_{lasso} = .72, R^2_{RF} = .76).

Timbre Descriptors

Table 5.6 shows the selected timbre descriptors for predicting perceived and induced anger. A predictor that is both linearly and nonlinearly relevant in both affect loci is log attack time (LAT); stimuli that are rated as angrier have slower attacks. Whereas perceived anger is also linearly and nonlinearly predicted by median tristimulus 3, this is only a nonlinear predictor for induced anger. Perceived anger is furthermore nonlinearly predicted by pitch, inharmonicity, and the IQR of several harmonic descriptors. Given that anger is strongly correlated with the dimensions of valence, ($r_{per}(57) = –.89$; $r_{ind}(57) = –.96$, and tension, $r_{per}(57) = .78$; $r_{ind}(57) = .96$, we also expect some overlap in the timbre descriptor results. Indeed, the results overlap in terms of attack time and relative energy in the higher frequencies and linear contribution of the IQR of spectral flux. Thus, an increase in anger is characterized by slower attacks and more relative energy in the upper harmonics (or a less clearly emerging fundamental frequency), with an additional role in perceived anger for variability over time of several harmonic- and noise-related descriptors.

Table 5.7 shows the selected timbre descriptors for predicting fear. The improvement of RF over lasso, here, was smaller than for the models predicting anger. However, we do expect the results to be similar to the findings on anger, as anger and fear were strongly correlated, particularly in the perceived affect locus, $r_{per}(57) = .90$; $r_{ind}(57) = .70$. Consequently, and again particularly in the perceived affect locus, we also expect high overlap with valence, $r_{per}(57) = –.87$; $r_{ind}(57) = –.72$, and tension, $r_{per}(57) = .82$; $r_{ind}(57) = .64$. Indeed, we see that LAT and the IQR of spectral flux

Table 5.6 *Lasso and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Perceived and Induced Anger in Experiment 1*

Descriptor	**Perceived**		**Induced**	
	Lasso	**RF**	**Lasso**	**RF**
Inharmonicity Median	–	3	–	–
Pitch Median	–	2	–	–
Tristimulus 3 Median	0.56	1	–	5
Odd:Even IQR	–	7	–	–
Harmonic:Noise Energy IQR	–	8	–	–
Noisiness IQR	–	6	–	–
Partials:Noise Energy IQR	–	5	–	–
Tristimulus 3 IQR	–	4	–	–
Spectral Centroid Median	–	–	–	4
Spectral Decrease Median	–	–	–	7
Spectral Roll Off Median	–	–	–	6
Spectral Flux IQR	0.33	–	0.36	–
Attack Time	–	9	–	3
LAT	0.27	10	0.60	2
Temporal Centroid	–	–	–	1
R^2	.71	.86	.63	.82
Test-R^2	.73	.82	.63	.84
NRMSE	.24	.20	.26	.17

Note. See Table 5.3.

significantly predict perceived and induced fear. Here, spectral flux is also selected by the RF regression. Stimuli that are rated as more fearful have slower attacks and show more variability in variation of the spectrum over time. Perceived fear is also predicted by more relative energy in the upper harmonics (median tristimulus 3) and induced fear is predicted by more inharmonicity.

Table 5.7 *Lasso and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Perceived and Induced Fear in Experiment 1*

Descriptor	Perceived		Induced	
	Lasso	RF	Lasso	RF
Inharmonicity Median	–	–	0.34	4
Pitch Median	–	–	–	6
Tristimulus 3 Median	0.38	2	–	3
Harmonic:Noise Energy Median	–	9	–	–
Noisiness IQR	–	6	–	–
Partials:Noise Energy IQR	–	7	–	–
Tristimulus 3 IQR	–	1	–	5
Spectral Flux Median	–	3	–	2
Spectral Flatness IQR	–	8	–	–
Spectral Flux IQR	0.47	4	0.52	1
LAT	0.32	10	0.26	–
Temporal Centroid	–	5	–	–
R^2	.65	.76	.75	.71
Test-R^2	.67	.77	.75	.76
NRMSE	.22	.19	.17	.20

Note. See Table 5.3.

Table 5.8 shows the selected timbre descriptors for predicting perceived and induced sadness. As became evident from the previous chapter, sadness behaves more independently from the other discrete affects, although it was negatively correlated with happiness, $r_{per}(57) = –.80$; $r_{ind}(57) = –.82$, and valence, $r_{per}(57) = –.72$; $r_{ind}(57) = –.63$. Both perceived and induced sadness are linearly and nonlinearly predicted by temporal centroid, indicating slower attacks or more sustain as sadness increases. This is accompanied by the attack slope predictor for perceived sadness. Perceived sadness is also predicted by spectral crest and spectral decrease, indicating less tonalness and less higher-frequency energy for sad sounds. Induced sadness is predicted by an increase in inharmonicity, which is also correlated with median spectral decrease, $r(57) = .65$.

Thus, perceived and induced sadness seem to mostly overlap in that stimuli that are rated as sadder have slower attacks and less prominent tonalness but also less higher-frequency energy.

Table 5.8 *Lasso and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Perceived and Induced Sadness in Experiment 1*

Descriptor	**Perceived**		**Induced**	
	Lasso	**RF**	**Lasso**	**RF**
Harmonic Energy Median	–	–	–	3
Inharmonicity Median	–	–	0.25	–
Noise Energy Median	–	–	–	6
Tristimulus 3 IQR	–	4	–	7
Spectral Crest Median	–0.15	–	–	–
Spectral Decrease Median	0.19	–	–	–
Spectral Flux Median	–	–	–	2
Spectral Slope Median	–	6	–	1
Spectral Skewness IQR	–	5	–	–
Attack Slope	–0.30	3	–	–
Effective Duration	–	2	–	4
Temporal Centroid	0.49	1	0.63	5
R^2	.81	.85	.79	.81
Test-R^2	.86	.85	.78	.81
NRMSE	.11	.11	.11	.10

Note. See Table 5.3.

The timbre descriptors that were selected as significant predictors of perceived and induced happiness are shown in Table 5.9. Models predicting induced happiness performed better than those predicting perceived happiness. Furthermore, happiness was strongly correlated with valence, $r_{per}(57) = .93$; $r_{ind}(57) = .86$, anger, $r_{per}(57) = –.85$; $r_{ind}(57) = –.80$, fear, $r_{per}(57) = –.84$; $r_{ind}(57) = –.74$, sadness, $r_{per}(57) = –.80$; $r_{ind}(57) = –.82$, and tenderness, $r_{per}(57) = .90$; $r_{ind}(57) = .92$. We see that both perceived and induced happiness are predicted by effective duration and median tristimulus 3; stimuli that are rated as happier are perceptually shorter in duration and have a more clearly emerging fundamental frequency. Perceived happiness is also linearly predicted by a steeper attack slope, and induced happiness is linearly predicted by a narrower range of spectral flux and a wider range of spectral skewness. The attack slope component of perceived happiness is mirrored in the nonlinear findings for induced happiness (attack time, LAT, temporal centroid).

Table 5.9 *Lasso and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Perceived and Induced Happiness in Experiment 1*

Descriptor	**Perceived**		**Induced**	
	Lasso	**RF**	**Lasso**	**RF**
Tristimulus 3 Median	–0.26	–	–0.16	3
Tristimulus 3 IQR	–	1	–	2
Spectral Slope Median	–	2	–	–
Spectral Flux IQR	–	–	–0.13	–
Spectral Skewness IQR	–	–	0.19	–
Attack Slope	0.28	4	–	–
Attack Time	–	–	–	4
Effective Duration	–0.63	–	–0.52	6
LAT	–	–	–	5
Temporal Centroid	–	3	–	1
R^2	.72	.76	.83	.86
Test-R^2	.77	.81	.87	.86
NRMSE	.19	.18	.16	.15

Note. See Table 5.3.

Finally, the descriptors that were selected for the prediction of perceived and induced tenderness are shown in Table 5.10. Tenderness, too, shows strong correlations with valence, $r_{per}(57) = .93$; $r_{ind}(57) = .82$, tension, $r_{per}(57) = –.83$; $r_{ind}(57) = –.73$, anger, $r_{per}(57) = –.90$; $r_{ind}(57) = –.76$, fear, $r_{per}(57) = –.84$; $r_{ind}(57) = –.84$, and happiness, $r_{per}(57) = .90$; $r_{ind}(57) = .92$. Both perceived and induced tenderness are predicted by median tristimulus 3 and the IQR of spectral flux; stimuli that are rated as more tender have a more clearly emerging fundamental frequency and a narrower range of variation of the spectrum over time. Perceived tenderness is also predicted by an increase in spectral crest (tonalness) and a decrease in effective duration, whereas induced tenderness is predicted by an earlier temporal centroid.

Table 5.10 *Lasso and Random Forest Regression Results and Performance for Timbre Descriptors Predicting Perceived and Induced Tenderness in Experiment 1*

Descriptor	**Perceived**		**Induced**	
	Lasso	**RF**	**Lasso**	**RF**
Tristimulus 3 Median	–0.27	1	–0.25	1
Tristimulus 3 IQR	–	2	–	2
Spectral Crest Median	0.23	–	–	–
Spectral Flux IQR	–0.28	–	–0.26	–
Attack Time	–	4	–	–
Effective Duration	–0.32	–	–	–
LAT	–	3	–	–
Temporal Centroid	–	5	–0.43	–
R^2	.77	.81	.75	.82
Test-R^2	.75	.82	.70	.81
NRMSE	.15	.14	.13	.11

Note. See Table 5.3.

Summary

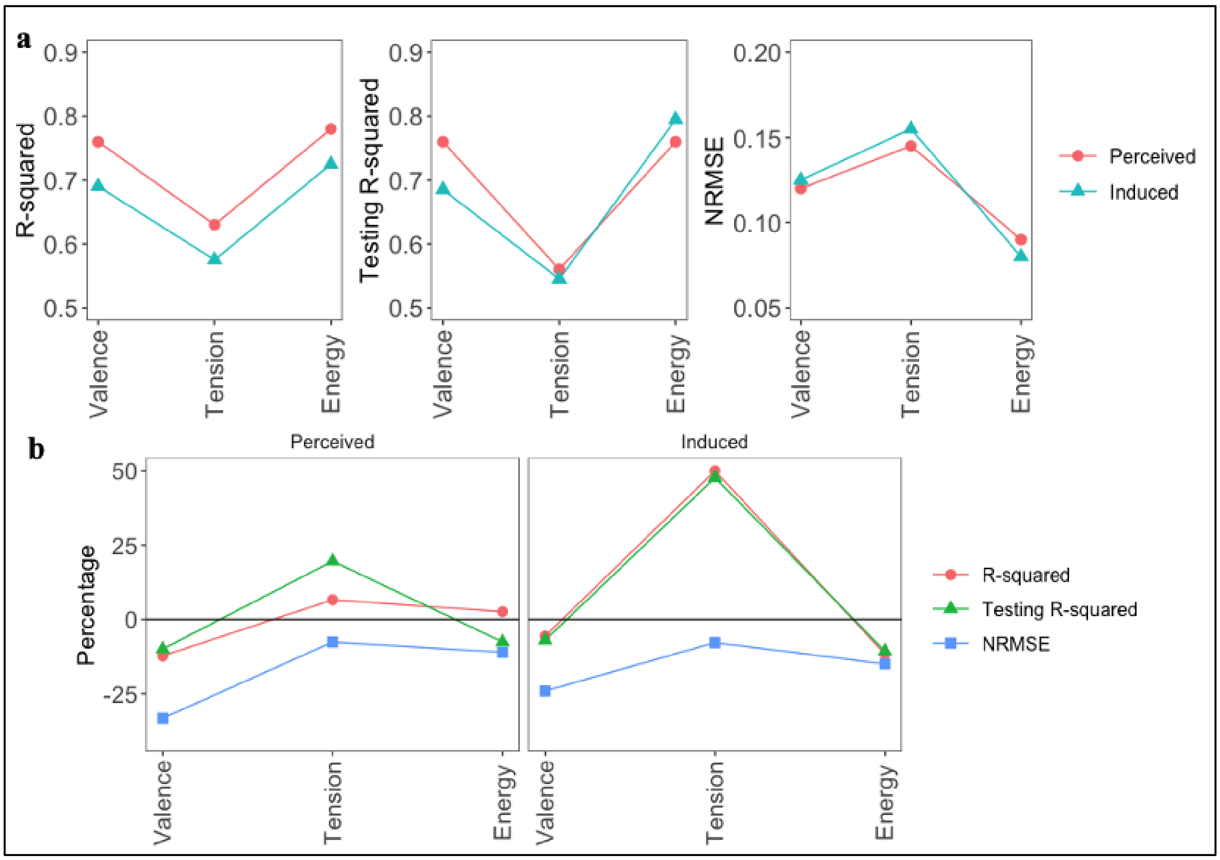
Performance for predicting the perceived and induced discrete affect ratings is also relatively good, with a slight improvement of nonlinear over linear methods, in particular for the prediction of anger. There are no major differences between affect loci, although induced happiness is better predicted by the timbre descriptors than is perceived happiness. Given the findings in Chapter 3, we expect overlap in results between anger, fear, happiness, and tenderness, as well as with valence and tension arousal. Indeed, we see that all are predicted by an attack component and the relative energy in the upper harmonics, as well as IQR spectral flux. In addition, happiness and tenderness ratings are predicted by effective duration; i.e., sounds judged as such are perceptually shorter. Sadness is also characterized by slower attacks, as well as less prominent tonalness and less higher-frequency energy.

Experiment 2 – Discrete & Dimensional, Perceived & Induced Affect

Model Performance

Figure 5.3 shows the performance of the regression models predicting perceived and induced valence, tension, and energy in Experiment 2, and the improvement of RF over lasso regression (see detailed results in Table B4 in Appendix B). Here, we do not see a clear improvement of RF over lasso regression. In fact, for valence and energy arousal, performance is worse with the RF regression. Only tension arousal, in particular induced tension arousal, is improved by the nonlinear regression method. Performance overall is again good (R^2 = [.46, .81], *testing* R^2 = [.44, .84], *NRMSE* = [.07, .16]), with relatively poor performance predicting tension arousal (mean R^2

Figure 5.3 R^2, *Testing* R^2*, and NRMSE for the Prediction of the Perceived and Induced Dimensional Affect Ratings of Experiment 2 (a), and Percentage Improvement of RF over Lasso Regression (b)*

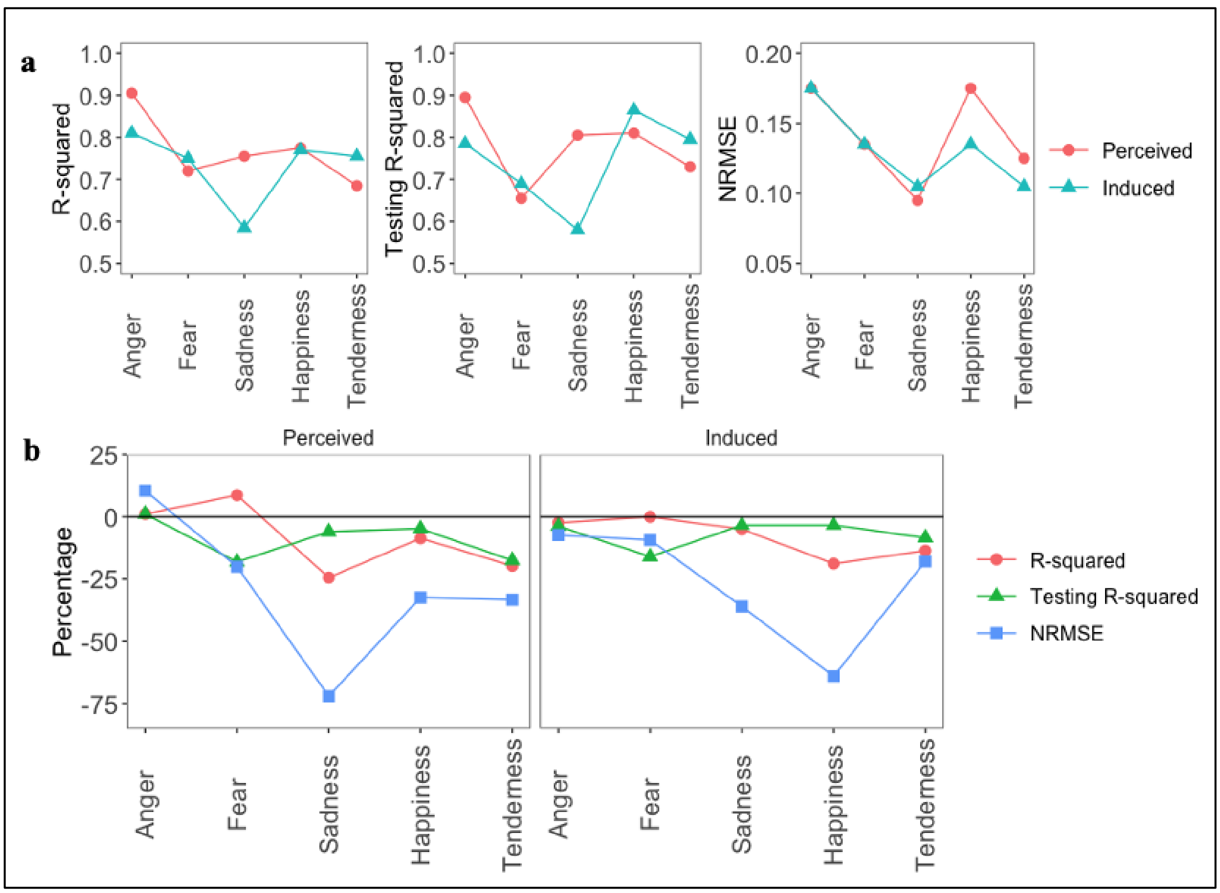


Note. See Figure 5.1.

= .60, mean *testing* R^2 = .55, mean *NRMSE* = .15). Perceived dimensional affect (mean R^2 = .73, mean *test*-R^2 = .70, mean *NRMSE* = .12) is also slightly better predicted by the timbre descriptors than induced dimensional affect (mean R^2 = .67, mean *testing* R^2 = .68, mean *NRMSE* = .12).

Figure 5.4 shows the performance for the regression models predicting the discrete affect ratings in Experiment 2 and the improvement of RF over lasso regression (see Table B5 in Appendix B for full results). Similar to the dimensional affect ratings, again we do not see a clear improvement of RF over lasso regression. Overall, linear lasso regression performs the best. The timbre descriptors also perform slightly better at predicting the perceived affect ratings (mean R^2 = .77, mean *test*-R^2 = .78, mean *NRMSE* = .14) than the induced affect ratings (mean R^2 = .74,

Figure 5.4 *R^2, Testing R^2, and NRMSE for the Prediction of the Perceived and Induced Discrete Affect Ratings of Experiment 2 (a), and Percentage Improvement of RF over Lasso Regression (b)*



Note. See Figure 5.1.

mean *test*-R^2 = .75, mean *NRMSE* = .13). Overall the model performance is good (R^2 = [.57, .91], *test*-R^2 = [.57, .90], *NRMSE* = [.07, .20]), with relatively worse performance for the prediction of induced sadness (mean R^2 = .59, mean *test*-R^2 = .58, mean *NRMSE* = .11).

For the discussion of the timbre descriptors that predict dimensional and discrete affect in Experiment 2, we will only discuss the results from the linear lasso regression, as these showed the best performance and provided more interpretable results. We will, however, discuss the RF results for tension arousal, which were improved by the nonlinear method. All other RF results are available in the Appendix B.

Timbre Descriptors

Table 5.11 shows the descriptors that were selected by the lasso regression for the prediction of valence, tension, and energy arousal and their standardized coefficients from the linear regression models. As in the previous experiment, valence and tension arousal were strongly correlated, $r_{perceived}(30) = -.95$, $p < .0001$, $r_{induced}(30) = -.97$, $p < .0001$, and consequently we expect them to be predicted by similar sets of timbre descriptors. We do see that in both affect loci, valence and tension are predicted by the range of spectral spread. This predictor is also strongly correlated with the median and range of spectral rolloff, $r_{median}(30) = .77$, $p < .0001$; $r_{range}(30) = .94$, $p < .0001$, and the median and range of spectral centroid, $r_{median}(30) = .73$, $p < .0001$; $r_{range}(30) = .86$, $p < .0001$. Thus, when a stimulus is rated as more positively valenced and less tense (more relaxed), it contains more energy in the higher frequencies of the spectrum, which may be related to perceptual brightness. This appears to be somewhat opposite to the findings in Experiment 1, which showed that positive valence is predicted by a more clearly emerging fundamental frequency and that tension is additionally predicted by more energy in the higher-frequencies of the spectrum. However, induced valence and perceived tension are also predicted by a respective increase and decrease of spectral crest (tonalness), which is strongly correlated positively with median tristimulus 1, $r(30) = .78$, $p < .001$, and negatively with tristimulus 3, $r(30) = -.80$, $p < .001$, thus also suggesting a more clearly emerging fundamental frequency for positively valenced and relaxed chromatic scales, alongside the brightness-related spectral descriptors.

Table 5.11 *Lasso Standardized Coefficients and Performance for Timbre Descriptors Predicting Perceived and Induced Dimensional Affect in Experiment 2*

Descriptor	**Valence**		**Tension**		**Energy**	
	Per	**Ind**	**Per**	**Ind**	**Per**	**Ind**
Inharmonicity Median	–.28	–.26	–	–	–.44	–.52
Pitch Median	–	–	–	–	.44	.37
Spectral Crest Median	–	.38	–.44	–	–	–
Spectral Skewness Range	–	–	–	–	–.31	–.30
Spectral Spread Range	.43	.47	–.46	–.60	–	–
Spectral Variation Median	–.27	–	–	–	–	–
R^2	.81	.71	.61	.46	.77	.77
Test-R^2	.80	.71	.51	.44	.79	.84
NRMSE	.10	.11	.14	.15	.09	.07

Note. All listed descriptors significantly predicted affect in the linear regression at $p < .05$.

Valence and tension diverge in the findings on median inharmonicity and spectral variation. However, spectral variation is selected as a nonlinear predictor of tension. It is also strongly correlated with median and IQR noisiness, $r_{median}(30) = .94$, $p < .0001$; $r_{IQR}(30) = .90$, $p < .0001$. Thus, valence and tension are also predicted by the variability of noisiness of a sound. Whereas we found descriptors from the temporal domain predicted valence and tension in Experiment 1, we do not find that here. However, the attack components may be less accurate descriptors of the chromatic scales, as they only consider the entire stimulus and not the temporal component of each individual note. Finally, like valence, energy arousal is predicted by a decrease in median inharmonicity, but also an increase in median pitch and a decrease in the range of spectral skewness. The findings on inharmonicity and pitch overlap with those of Experiment 1, showing that the role of pitch in energy arousal is consistent for both single-note and chromatic-scale stimuli.

Table 5.12 shows the standardized coefficients that were selected by the lasso regression to predict the discrete affect ratings of Experiment 2. Again, there were some strong intercorrelations and thus we may find some overlap in the predictive descriptors for anger, fear, happiness, and tenderness, with more divergence for sadness. As for valence and tension, we see that the range of spectral spread, which may be associated with perceptual brightness, significantly predicts the ratings of all discrete affects, except fear. Furthermore, as inharmonicity increases, stimuli are rated as more angry, and less happy and tender. Noisiness and spectral variation which were strongly correlated also play a role in predicting anger, fear, and tenderness, although less consistently. Fear and perceived sadness are predicted by a decrease in attack slope. Finally, a decrease in the range of spectral kurtosis (related to the spectral skewness findings for energy arousal) predicts perceived sadness, whereas a decrease in weight on the first harmonic predicts induced sadness.

Table 5.12 *Lasso Standardized Coefficients and Performance for Timbre Descriptors Predicting Perceived and Induced Discrete Affect in Experiment 2*

Descriptor	Anger		Fear		Sadness		Happiness		Tenderness	
	Per	Ind	Per	Ind	Per	Ind	Per	Ind	Per	Ind
Inharmonicity Median	.57	.56	–	–	–	–	–.34	–.32	–.42	–.53
Noisiness IQR	.22	.31	–	–	–	–	–	–	–	–
Tristimulus 1 Median	–	–	–	–	–	–.26	–	.21	–	–
Spectral Kurtosis Range	–	–	–	–	–.43	–	–	–	–	–
Spectral Spread Range	–.26	–.26	–	–	–.50	–.68	.65	.65	.49	.45
Spectral Variation Median	–	–	.30	–	–	–	–	–	–	–
Spectral Variation IQR	–	–	–	–	–	–	–	–	–.25	–
Attack Slope	–	–	–.46	–.38	–.35	–	–	–	–	–
R^2	.90	.82	.69	.75	.86	.60	.81	.85	.76	.81
Test-R^2	.89	.80	.72	.75	.83	.59	.83	.88	.80	.83
NRMSE	.18	17	.12	.13	.07	.09	.15	.10	.11.	10

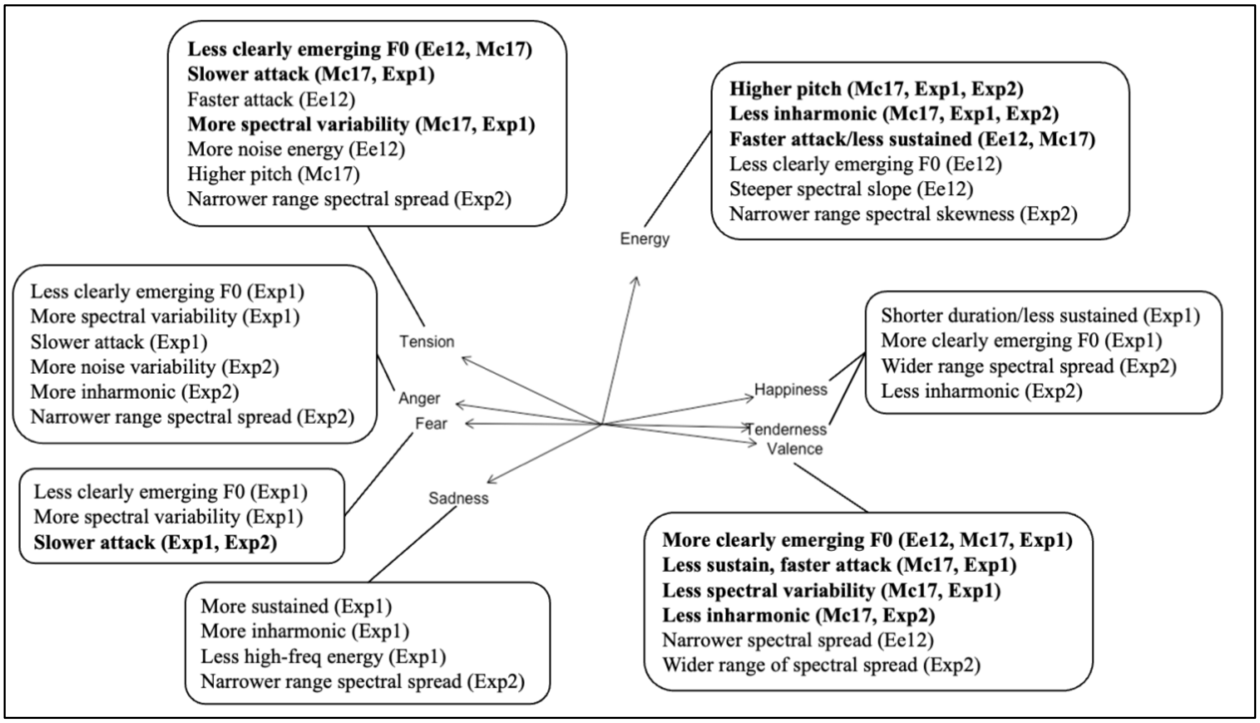
Note. All listed descriptors significantly predicted affect in the linear regression at $p < .05$.

5.4 Discussion

In the comparison of the results of different experiments, we find overlap across datasets, as well as discrepancies that may be explained by differences in stimulus set and methodological approaches. Figure 5.5 shows a diagram highlighting the relevant timbre descriptors that predict the affect ratings. It overlays the principal component analysis loading plot based on the ratings of

Experiment 1 and 2 (see Chapter 3), which further aids in seeing how certain timbre descriptors are (dis)similar for the different affect scales. The results for the perceived and induced affect ratings are combined here, as they did not show many large differences.

Figure 5.5 *Diagram Summarizing the Relevant Timbre Descriptors in Predicting the Affect Ratings of Eerola12 (Ee12), Mcadams17 (Mc17), Experiment 1 and 2 (Exp. 1 & Exp. 2)*



Note. Diagram is overlayed on the principal component analysis loading plot obtained from the ratings of Exps. 1 & 2 as described in Chapter 3, with descriptors that are relevant to two or more experiments highlighted in **bold**.

We see across all three datasets concerning the perceived dimensional affect of single notes, that sounds that are perceived as more positive and relaxed have a more clearly emerging fundamental frequency (F0). These findings are mostly in line with the original published results from Eerola12 and McAdams17. Similar timbre descriptors related to the emergence of the fundamental frequency predict the induced valence and tension in Experiment 1, as well as the discrete representation of affect (anger, fear, happiness, and tenderness). In Experiment 2, the range of spectral spread appears to be related to perceptual brightness, where the locus of energy is higher for sounds that are more positive, relaxed, happy, tender, and less angry and sad. Thus, the role of the emergent fundamental frequency is found consistently with different stimulus sets, experimental designs, participant populations, and affect representations for short sounds, although with longer stimuli, perceptual brightness appears to have a more important role. Future research may further investigate the relationship between local and global features of longer musical excerpts in relation to higher-frequency energy or the relative balance in energy of harmonics.

When pitch variation is present in the stimulus set, energy arousal is consistently predicted by an increase in pitch height and a decrease in inharmonicity. This is also the case for induced energy with chromatic scales. Other dimensional and discrete affects are not consistently predicted by pitch height (although we did find this to be the case in Chapter 4), nor are Eerola12's ratings, with a lack of pitch variation in their stimulus set. Median inharmonicity, however, is relevant to other dimensional and discrete affects as well, particularly in the second experiment. Furthermore, whereas McAdams17 did not include pitch as a predictor in their analyses, they did find pitch-related timbre descriptors predicted energy arousal and suggested this was due to the variation in pitch register. Thus, consistently both in lab and online, with different participant populations, analytical approaches, and stimulus lengths, sounds are considered more awake as they are higher in pitch and lower in inharmonicity. The high similarity between McAdams17 and the current Experiment 1 is in line with the previous findings that the use of affective auditory stimuli lead to comparable results in-lab and online (Seow & Hauser, 2022).

Some other discrepancies in findings between Eerola12 and both the current experiments and McAdams17 may additionally be explained by the differences in stimulus sets. For example, McAdams17 and the current Experiment 1 (perceived and induced) found that positive valence and relaxation were also predicted by a decrease in temporal centroid or attack time (sharper attack/less sustain). Although the current analytical method did not find that a temporal descriptor predicted the Eerola12 valence ratings, their original publication did, and furthermore, the current analysis did find that temporal descriptors predicted the Eerola12 tension ratings. However, the findings are opposite to those in the current experiment and McAdams17; positive valence and relaxation were predicted by an increase in temporal centroid and attack time (slower attack/more sustain). Future research may investigate whether the variations in pitch height or instrument selection caused variations in attack that may explain these diverging results on perceived and induced valence and tension arousal.

Interestingly, although this cannot explain the above discrepancy in attack descriptors predicting valence and tension, Eerola12 and McAdams17 included variation in attack for a given instrument in their stimulus sets, whereas the current experiments did not. What this can explain is the discrepancy in findings for the prediction of energy arousal. Eerola12's and McAdams17's energy ratings are predicted by a decrease in attack time or temporal centroid (sharper attack/less sustain), whereas the energy ratings from the current Experiments 1 and 2 are not predicted by this feature in either affect locus. This is also reflected in Eerola12's original findings and only in McAdams17's NN analysis. These results suggest that when sufficient attack variation is present, energy arousal is not only predicted by pitch and pitch-related spectral descriptors, but also a temporal component, which is consistent across attack-varied stimulus sets, participant populations, and analytical approaches.

Findings on the discrete affect ratings in Experiment 1 could be mostly translated to the findings on the dimensional ratings in Experiment 1. They were all predicted by the relative emergence of the fundamental frequency and a temporal component in the form of attack or sustain. Happiness and tenderness were additionally predicted by effective duration; happy and tender sounds are perceptually shorter in duration. Although playing techniques and stimulus lengths were kept constant across stimuli, the effect of *perceptual* duration may be related to differences between staccato and legato as previous research has associated staccato playing techniques with more perceived happiness (Carr et al., 2023). Sadness was also predicted by the relative energy in the upper frequencies, such that sad sounds are less harsh but also have a less clearly emerging fundamental frequency. The correspondence between the dimensional and discrete results was expected from our findings from Chapter 3, which showed that the dimensional and discrete affect ratings mapped on to each other quite well.

Finally, although the findings for energy arousal from Experiment 1 are extended to Experiment 2, there is some divergence when comparing the results from single notes to those from chromatic scales. The emergence of the fundamental frequency was less important in predicting the affective response to chromatic scales, although tonalness did play a role. We also find that median inharmonicity is much more prominent in the second experiment, except for tension, fear, and sadness, which can also be related to the tonalness. The attack components are also much less prominent in the second experiment. However, this was to be expected, since the temporal descriptors are only extracted from the entire stimulus by the Timbre Toolbox routines and not each individual note in the chromatic scale. Continuous measurements could further translate the role of local to more global timbral features in perceiving and inducing affect.

5.5 Conclusion

In the re-analysis of previous experiments and the extension to the current experiments, we have been able to pinpoint affective timbres that are consistent across stimulus sets, experimental designs, participant populations, affect loci, and analytical approaches. The relative emergence of the fundamental frequency plays a prominent role in perceiving and inducing musical affect. Changes in variation of sound characteristics like attack techniques or pitch height led to clear divergence in results, arguing for future studies to include maximal variation in their stimulus sets. The difference between perceived and induced affect were small. As was also shown by the previous chapter, although there are differences between the two affect loci, they are rather found in the magnitude of effects rather than the directionality. Findings on local characteristics of affective timbre (single notes) are to some extent translated to global characteristics of affective timbre (chromatic scales). Furthermore, whereas a nonlinear model shows clear model improvement over a linear model, this is only the case for the prediction of single notes and not chromatic scales. Finally, there were no large differences between the online and in-lab experiments, suggesting that auditory studies that focus on fine-grained changes in timbre can also be carried out online, reaching a more diverse and representative participant population.

Chapter 6
General Conclusion

The aim of this thesis was to further advance our knowledge and understanding of music's capacity to express and evoke affect. Here, the focus was on the musical feature of timbre and how it influences musical affect. The approach can be considered *holistic*, as it considers different methodologies, contexts, and individual differences to approximate a complete picture of the affective response to timbre. Two models of affect quantification, dimensional and discrete, are considered for the quantification of perceived and induced affect. Furthermore, timbre is approached at different levels of granularity: by considering timbre differences between orchestral instrument families, and by describing the acoustic origins of the sounds through timbre descriptors. Different contexts are considered such as the experimental context (in-lab vs. online) and the stimulus context (single notes and chromatic scales). Finally, the individual characteristics of the participant are explored as they may influence the affective response. To this end, two behavioural online experiments were conducted to answer the following three research questions:

1. What is the most appropriate method for the quantification of perceived and induced affect in response to affectively ambiguous and relatively short musical sounds, and how is this related to individual differences?
2. What are the effects of instrument family, pitch register, and affect locus on the affective response to those musical sounds, and how are these effects moderated by individual differences?
3. Which acoustic properties that describe the timbre of a musical sound predict the perceived and induced affective response to that sound?

In this final conclusion, first Chapters 3–5 will be summarized, followed by a discussion of how the results contribute to our existing knowledge, what the limitations of the research were and consequently how future research could tackle these. Finally, I will offer some concluding remarks.

6.1 Summary

The methods for both experiments were described in Chapter 2. The experiments were designed to closely follow McAdams et al. (2017). The two experiments presented the participants with sounds played by different orchestral instruments at different pitch registers, and in response to those sounds they rated the perceived and induced affect. In Experiment 1, the sounds consisted of single notes obtained from the Vienna Symphonic Library (VSL; *Vienna Symphonic Library GmbH*, 2022). In Experiment 2, the sounds consisted of chromatic scales spanning a perfect fifth, created in OrchSim using several sound databases (McAdams & Goodchild, 2017a; *OrchPlayMusic*, 2022). The perceived and/or induced affective response was rated on three dimensional affect scales (valence, tension arousal, and energy arousal) or five discrete affect scales (anger, fear, sadness, happiness, and tenderness). Preference for the stimuli was rated in addition to the dimensional affective ratings. Before the listening task, participants filled in the PANAS-X (Watson et al., 1988; Watson & Clark, 1994) to assess their pre-existing mood, and after the listening task they filled in questionnaires that measured their personality traits (BFI-44; John & Srivastava, 1999), dispositional empathy (IRI; Davis, 1983), musical sophistication (Gold-MSI; Müllensiefen et al., 2014), and musical preferences (STOMP-R; Rentfrow et al., 2011; Rentfrow & Gosling, 2003). Following the inclusion and exclusion criteria, we were able to obtain complete data from 263 participants in Experiment 1 and 181 participants in Experiment 2.

Chapter 3 answers the first research question by comparing the dimensional and discrete models of affect in response to single notes and chromatic scales, and by investigating the relation between the affect ratings and individual differences. The results showed that participants are more consistent and display more agreement in their ratings on perceived than induced affect, single notes than chromatic scales, and dimensional than discrete affect scales. These findings suggest that the personal experience of induced affect is more susceptible to individual differences, and similarly the longer exposure time of chromatic scales allows for more disagreement between participants. The lower score of the discrete affect model is mostly caused by the sadness scale, which showed the lowest rating consistency and agreement overall. Dimension reduction methods revealed that both the three dimensional and five categorical scales could be reduced to two components and still explain nearly all of the variability in the affective responses. Predictive modelling showed that the dimensional and discrete models map onto each other well, with the exception of energy arousal and sadness. The dimensional scales were slightly better at predicting the discrete affect ratings than the inverse case. Upon further investigation, energy arousal in particular varied in a way that was not captured by any of the discrete affect measures. Investigation of individual differences showed that each of the sources of individual differences included here correlated with the affect scales in multiple ways, although only moderately strong at best. Valence correlated the most frequently with individual differences, and pre-existing mood was the factor that correlated most frequently with the affect scales. Based on these results, we can conclude that in response to affectively ambiguous short sounds, two dimensions of valence and energy arousal best capture the affective response, although the role of individual differences needs to be considered when applying these measures.

Chapter 4 investigated the affective response to timbre as it varies between instrument family and along pitch register, and how these effects are moderated by individual differences, thus answering the second research question. Effects of pitch register were found consistently and mostly showed a quadratic relation to affect such that the middle pitch registers were considered to be the most positive, happy, and tender, and least tense, angry, and fearful. As became evident in Chapter 3, sadness and energy arousal behaved differently and showed a mostly linear relationship to pitch register. Comparing instrument families, percussion expressed and induced the least tension, fearfulness, sadness, and anger, and the most positive valence, happiness, and tenderness. Again, following the previous chapter, energy arousal stands out from the other affect scales; the perceived and induced energy response to percussion sounds varied the least with pitch register, whereas for the other instrument families, energy was more awake as pitch register increased. Furthermore, comparison of affect loci showed that any affect that may be considered unpleasant, particularly sadness, was less strongly induced than perceived, particularly in the lower pitch registers. Finally, whereas the effect of pitch register was least influenced by individual differences, the effect of instrument family was more frequently moderated by individual differences. Again, following the results from Chapter 3, effects on valence ratings were most frequently moderated by individual differences.

Finally, Chapter 5 approached timbre with a finer granularity by analyzing the effect of timbre descriptors on perceived and induced affect ratings. This chapter not only analyzed the data obtained from Experiments 1 and 2, but also reanalyzed the sounds and affect ratings of two previous studies that similarly investigated the role of timbre descriptors in predicting affect: Eerola et al. (2012) and McAdams et al. (2017). Here, pitch register (alongside inharmonicity) was only found to influence energy arousal, and only for the studies that included variations in pitch register (i.e., not for Eerola et al.). Attack components, however, influenced energy arousal for the studies that included within-instrument attack variations (i.e., not for Experiment 1 and 2). Furthermore, the relative energy in the first and upper harmonics, interpreted as the emergence of the fundamental frequency, predicted nearly all affect ratings, except for sadness. Increased sadness was characterized by sustained sounds, inharmonicity, and less high-frequency energy. Spectral variability also influenced tension arousal and valence in opposite directions for McAdams et al. and Experiment 1. Thus, spectral, temporal, and spectrotemporal properties are shown to influence the perceived and induced affective response to short sounds.

6.2 Contributions to Knowledge

The current dissertation contributes to the existing knowledge of affective timbres by considering different affect quantification methods, perceived and induced affect loci, single notes and chromatic scales, individual differences, and an online testing environment.

Chapter 3 contributes to the existing research on the representation and quantification of musical affect. The two often referred-to studies here from Eerola et al. (2012) and McAdams et al. (2017) showed diverging results on the dimensionality of affect. Based on correlation patterns, Eerola et al. concluded that two dimensions of valence and tension arousal sufficiently represented the affect ratings, whereas McAdams et al. found that three dimensions of valence, tension arousal, and energy arousal each captured different aspects of the affective response. Here, a deeper investigation beyond correlation analysis was taken to characterize the affective response, while also considering the discrete model of affect. Discrete and dimensional affect models have been compared in previous music research (see e.g., Eerola & Vuoskoski, 2011; Vuoskoski & Eerola, 2011a), but the selection of stimuli in musical affect studies generally consists of excerpts with a longer duration and rarely a duration shorter than 10 seconds as in the current experiments (Eerola & Vuoskoski, 2013). However, in the investigation of affective response to timbre, short sounds are necessary to isolate timbre from other musical features such as tempo and mode. Following the conceptual act theory of affect (Barrett, 2014), (core) affect may be perceived or felt multi-dimensionally, and then conceptualized into discrete categories of emotions. Furthermore, variations in the context and the individual can determine how affect is perceived, felt, and expressed (Barrett, 1998, 2004). Thus, Chapter 3 has taken a holistic approach by comparing dimensional and discrete affect models, in the experimental context of short and affectively ambiguous sounds that vary in instrument family and pitch height, while taking individual differences into account. As such, its primary contribution in this context is that the two dimensions of valence and energy arousal most effectively capture the variation in the affective response, although all the sources of individual differences that were included in the experiments influenced those affect ratings, in particular participants' pre-existing mood. This result corroborates the findings by McAdams et al. that the dimension of energy arousal is relevant in capturing the affective response to pitch-varying stimuli, but also those of Eerola et al. in that two dimensions are sufficient. This study further corroborates the studies that found that the dimensional model outperforms the discrete model of affect (Eerola & Vuoskoski, 2011; Vuoskoski & Eerola, 2011).

In all three Chapters 3–5 there is a direct comparison of the external and internal affect locus, i.e., perceived and induced affect, respectively, whereas most experiments examine only one locus in isolation. The affect locus comparison did not show qualitative differences in direction or category, but rather in the extent to which certain affects were perceived or felt, as was also shown by the few previous studies that compared affect loci (Evans & Schubert, 2008; Vieillard et al., 2008). The more subtle differences also contribute to our knowledge on affective response to timbre. Further extending McAdams et al.'s (2017) experiment, the results from Experiment 1 and 2 showed that the unpleasant affects that may be perceived in response to certain sounds are less strongly induced, consistent with previous findings (e.g., Zentner et al., 2008), and that this difference in affect loci is dependent on pitch register and instrument family. Tenderness was an exception here, which was overall more strongly perceived than induced as its concept may not translate clearly to induced affect. Similarly, Chapter 3 did not find any major differences in dimensionality between perceived and induced affect but did find more subtle differences, such as the fact that the dimensional affect model showed higher collinearity between the three affect dimensions for induced than perceived affect in response to single notes. Participants also agreed more strongly on their ratings and showed less frequent influence of individual differences in the perceived than the induced affect condition. This shows that findings on perceived affect are not directly translatable to induced affect on a one-to-one mapping, and that individual differences are an especially important subject of research when investigating induced affect.

One issue with isolating single notes in the analysis of affective response to timbre is that it is unclear how findings on single notes translate to full musical pieces. By conducting the second experiment with chromatic scales, which also only varied in pitch register and instrument family, the findings on single notes could be systematically compared to findings on longer musical excerpts which more closely approximate actual music. Indeed, the results showed that some findings are consistent with both single notes and chromatic scales (e.g., pitch and instrument family effects), but also that timbre descriptors predicting the affective response to single notes are less translatable to the descriptors predicting the affective response to chromatic scales. The affective response to chromatic scales does however appear to be slightly less susceptible to influence by individual differences. Differences between affect loci were also more apparent in response to chromatic scales than single notes. Future studies may take into account that findings are not always consistent with more ecologically valid examples of music, which should encourage them to further compare systematically the findings between isolated sounds and longer musical excerpts.

Both Chapters 3 and 4 investigated a relatively large number of potential sources of individual differences in the affective response to musical timbre. With this exploratory approach several important sources of individual differences can be identified for future research. In fact, all of the sources of individual differences that were included were in some way associated with the affective response: pre-existing mood, Big-Five personality traits, dispositional empathy, musical sophistication, and musical preferences. This is in agreement with previous findings that also found that these individual differences were associated with musical affect (Akkermans et al., 2019; Balteș & Miu, 2014; Pilgrim et al., 2017; Eerola et al., 2016; Garrido & Schubert, 2011; Juslin et al., 2008; Kawakami & Katahira, 2015; Ladinig & Schellenberg, 2012; Miu & Vuoskoski, 2017; Nusbaum & Silvia, 2011; Vuoskoski & Eerola, 2011), although to the best of our knowledge this is the only investigation that considers this many factors together in one musical affect study. As a results, we found somewhat surprisingly that empathy played the least frequent role in relation to affective response to timbre, even though this personality trait has been studied relatively frequently in relation to musical affect. Pre-existing mood proved to be more influential.

Finally, the online setting of the experiments proved to be a relatively new adventure that posed new challenges. However, as researchers conducting empirical studies, some of these challenges also deserve more consideration in the laboratory environment. For example, ensuring that participants are paying attention and not mindlessly clicking through an experiment is a challenge in both experimental settings. In the current experiments, there were no attention checks because there were no right or wrong answers, it was not possible to check whether participants had responded randomly. However, the exclusion criteria that determine how well the participants had read the instructions and characterized the uniformity of the rating behaviour of participants could, and perhaps should, also be applied in a laboratory setting. Particularly when the sample largely consists of students who are participating in experiments for compulsory credits and are lacking the intrinsic motivation to provide high quality data. The online setting furthermore allowed us to explore the various sources of individual differences by reaching a large, and less *WEIRD*, participant sample in a relatively short amount of time. Indeed, the results showed several consistencies with previous research, indicating that despite a lack of experimental control findings translate well from in-lab to on-line (Berinsky et al., 2012; Klein et al., 2014; Paolacci et al., 2010; Seow & Hauser, 2022).

6.3 Limitations and Future Directions

To continue on the topic of online experimentation, although it has proven to be highly beneficial for the current experiments, there were some limitations that may also be considered for future research. For example, the experiments did not include a test to ensure that the participants were wearing headphones, because in a trial phase at the Music Perception and Cognition Lab employing the test described by Milne et al. (2021), it appeared that several of us were able to "cheat" the test. Furthermore, although the Prolific participants all indicated that they had no hearing problems, an audiometry test could have further ensured that the participants were able to hear subtle differences in timbre. Future research is encouraged to further perfect tests that protect the auditory testing environment, because even now that laboratory experiments are possible again, there are many benefits to online experimentation.

The *WEIRD*-ness and diversity of the participant population should also be acknowledged. The age range (18–68 years) and average (M_{Exp1} = 29.0, SD_{Exp1} = 10.2; M_{Exp2} = 31.6, SD_{Exp2} = 10.4) goes beyond the expected age range and average of a student population. The education levels were somewhat varied, with most participants having a high-school, bachelor's, or master's degree. However, the majority of participants grew up in Europe and North America (Exp1, 88%; Exp. 2, 84%), falling under the *Western* part of WEIRD-ness. This is likely the result of the inclusion criterion to speak fluent English and perhaps an inherent bias in recruitment on the Prolific platform. Thus, although the participant sample can be considered less WEIRD than the average participant population, the current findings cannot be generalized to all socio-economic backgrounds and cultures. If experimenters are not able to reach a fully non-WEIRD population sample, then they should adequately report the demographics of the sample they do have. Beyond demographics, adequate reporting of the distribution of potential sources of individual differences, as was done in Table 2.4, also aid in the interpretation of results and the comparison of results between different experiments. Even if such results are not directly useful to an experimenter, they can help explain discrepancies in future results.

One major difference between Chapters 4 and 5 is in the role of pitch register. Whereas the polynomial mixed-effects models in Chapter 4 consistently showed that pitch register influenced affect ratings, the lasso and random forest regressions in Chapter 5 only found pitch as a predictor for energy arousal. That is, the quadratic register effects in Chapter 4 were not found by the analysis in Chapter 5, and only the linear register effect on energy arousal was consistent in both analysis approaches. Although random forest regression is a *nonlinear* analysis method, it was not specifically geared to detecting quadratic patterns. Employing a polynomial mixed-effects model in Chapter 5 was not possible, however, due to the large number of timbre descriptors as predictors. Perhaps the timbre descriptors that were consistently found to predict the affect ratings in Chapter 5, or another selection of timbre descriptors that differentiate the sounds, can be further explored by future research to examine the quadratic nature of pitch and pitch-related timbre descriptors.

Finally, whereas self-report measures were chosen here as an adequate measure of the subjective experience of perceived and induced affect, future research may explore how these measures relate to other behavioural and psychophysiological measures, particularly concerning induced affect. A further expansion of the *holistic* approach by simultaneously quantifying affect with different methods was beyond the scope of the dissertation research here. Although such different methods, effectively measuring different outlets or processes of the affective response will not necessarily be highly correlated or synchronous (see e.g., Cacioppo et al., 2000; Mauss & Robinson, 2009), a complete picture of the subjective experience cannot be assessed without considering the various measures together as a whole. Indeed, such a holistic approach, while also considering individual differences, is computationally very heavy, but there is much room for future research to explore large-scale experiments, considering online experimentation, gamification, or world-wide lab-collaborations.

6.4 Concluding Remarks

Indeed, the work for this dissertation was a challenge between the desire to take a holistic approach to researching affective response to timbre and the practicality of conducting suitable experiments that can answer these questions within the time frame of a PhD. The resulting two experiments have been able to investigate different affect quantification methods, study affective timbre at different levels of granularity, explore how the individual characteristics of the music listener inevitably influence the affective response, and demonstrate the challenges and possibilities of online research. While answering a few questions, I believe the findings also provide fruitful ground for many future research endeavors.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., Schlemmer, K., & Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1099–1118. https://doi.org/10.1080/02699931.2018.1541312

Alluri, V., & Toiviainen, P. (2012). Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. *Music Perception, 29*(3), 297–310. https://doi.org/10.1525/mp.2012.29.3.297

Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: A comparison of pilot data sets from lab, crowdsourced and convenience web samples. *Frontiers in Psychology, 10,* 2883. https://doi.org/10.3389/fpsyg.2019.02883

Balteș, F. R., & Miu, A. C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind, and Brain*, *24*(1), 580–65. https://doi.org/10.1037/pmu0000030

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. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, *12*(4), 579–599. https://doi.org/10.1080/026999398379574

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology*, *87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Barrett, L. F. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4), 292–297. https://doi.org/10.1177/1754073914534479

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*(1), 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Bartoń, K. (2022). *MuMIn: Multi-Model Inference* (1.47.1). https://CRAN.Rproject.org/package=MuMIn

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious mixed models* (arXiv:1506.04967). arXiv. http://arxiv.org/abs/1506.04967

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. https://doi.org/10.3758/s13428-011-0081-0

Ben-Shachar, M. S., Makowski, D., Lüdecke, D., Patil, I., Wiernik, B. M., Kelley, K., Stanley, D., Burnett, J., & Karreth, J. (2022). *effectsize: Indices of Effect Size* (0.8.2). https://CRAN.R-project.org/package=effectsize

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

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. https://doi.org/10.1080/02699930500204250

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, S., Martinez, M. J., & Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *NeuroReport*, *15*(13), 2033–2037.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlman, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In R. Lewis & J. M. Havilnad-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). Guilford.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*(1), 471–482.

Caetano, M., Saitis, C., & Siedenburg, K. (2019). Audio content descriptors of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, Perception, and Cognition* (pp. 297–333). Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_11

Camacho, A., & Harris, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America*, *124*(3), 1638–1652. https://doi.org/10.1121/1.2951592

Carr, N. R., Olsen, K. N., & Thompson, W. F. (2023). The perceptual and Emotional consequences of articulation in music. *Music Perception*, *40*(3), 202–219. https://doi.org/10.1525/mp.2023.40.3.202

Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2017). *Speaking about sounds: A tool for communication on sound features*. *15*(2), 85–109. https://doi.org/10.1504/jdr.2017.086749

Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion*, *4*(1), 3–22. https://doi.org/10.1037/1528-3542.4.1.3

Cespedes-Guevara, J., & Eerola, T. (2018). Music communicates affects, not basic emotions – A constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, *9*, 215. https://doi.org/10.3389/fpsyg.2018.00215

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods, 51*(5), 2022–2038. https://doi.org/10.3758/s13428-019-01273-7

Chau, C.-J., Wu, B., & Horner, A. (2015). The emotional characteristics and timbre of nonsustaining instrument sounds. *Journal of the Audio Engineering Society*, *63*(4), 228–244. http://dx.doi.org/10.17743/jaes.2015.0016

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://psycnet.apa.org/doi/10.1037/0022-3514.44.1.113

Diedenhofen, B., & Musch, J. (2015). Cocor: A Comprehensive solution for the statistical comparison of correlations. *PloS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Eerola, T., Armitage, J., Lavan, N., & Knight, S. (2021). Online data collection in auditory perception and cognition research: Recruitment, testing, data quality and ethical considerations. *Auditory Perception & Cognition*, *4*(3–4), 251–280. https://doi.org/10.1080/25742442.2021.2007718

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, *30*(1), 49–70. https://doi.org/10.1525/mp.2012.30.1.49

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 487. https://doi.org/10.3389/fpsyg.2013.00487

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. https://doi.org/10.1177/0305735610362821

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, *30*(3), 307–340. https://doi.org/10.1525/MP.2012.30.3.307

Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Frontiers in Psychology*, *7*, 1176. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01176

Eerola, T., Vuoskoski, J. K., Peltola, H.-R., Putkinen, V., & Schäfer, K. (2018). An integrative review of the enjoyment of sadness associated with music. *Physics of Life Reviews*, *25*, 100–121. https://doi.org/10.1016/j.plrev.2017.11.016

Egermann, H., & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception*, *31*(2), 139–156. https://doi.org/10.1525/mp.2013.31.2.139

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Wiley.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, *12*(1), 75–99. https://doi.org/10.1177/102986490801200105

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, *17*(3), 335–341. https://doi.org/10.3758/PBR.17.3.335

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Frego, R. J. (1999). Effects of aural and visual conditions on response to perceived artistic tension in music and dance. *Journal of Research in Music Education*, *47*(1), 31–43. https://doi.org/10.2307/3345826

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., & Yang, J. (2022). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models* (4.1-6). https://CRAN.R-project.org/package=glmnet

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*, 123–147. https://doi.org/10.1177/10298649020050S105

Gabrielsson, A., & Lindström, E. (1995). Emotional expression in synthesizer and sentograph performance. *Psychomusicology: A Journal of Research in Music Cognition*, *14*(1–2), 94. https://doi.org/10.1037/h0094089

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford University Press.

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (0.84.1). https://CRAN.R-project.org/package=irr

Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception*, *28*(3), 279–296. https://doi.org/10.1525/mp.2011.28.3.279

Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion, 7*(2), 377–387. https://doi.org/10.1037/1528-3542.7.2.377

González, I., & Déjean, S. (2021). *CCA: Canonical Correlation Analysis* (1.2.1). https://CRAN.R-project.org/package=CCA

Goodchild, M., Wild, J., & McAdams, S. (2019). Exploring emotional responses to orchestral gestures. *Musicae Scientiae*, *2*(1), 25–49. https://doi.org/10.1177/1029864917704033

Gosling, S. D., Sandy, C. J., John, O. P., & Potter, J. (2010). Wired but not WEIRD: The promise of the Internet in reaching more diverse samples. *Behavioral and Brain Sciences*, *33*(2–3), 94–95. https://doi.org/10.1017/S0140525X10000300

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. https://doi.org/10.1016/j.neuropsychologia.2006.07.012

Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., Baulac, M., & Peretz, I. (2006). Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. *Brain*, *129*(10), 2585–2592. https://doi.org/10.1093/brain/awl240

Grimaud, A. M., & Eerola, T. (2022). Emotional expression through musical cues: A comparison of production and perception approaches. *PLOS ONE*, *17*(12), e0279605. https://doi.org/10.1371/journal.pone.0279605

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception*, *24*(5), 473–484. https://doi.org/10.1525/mp.2007.24.5.473

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Michael, G., & Warren, J. D. (2009). It’s not what you play, it’s how you play it: Timbre affects perception of emotion in music. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2141–2155. https://doi.org/10.1080/17470210902765957

Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, *1*, 43–54.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

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), 47–56. https://doi.org/10.1037/a0016873

Huron, D., Anderson, N., & Shanahan, D. (2014). “You can’t play a sad song on the banjo": Acoustic factors in the judgment of instrument capacity to convey sadness. *Empirical Musicology Review*, *9*(1), 29–41. http://dx.doi.org/10.18061/emr.v9i1.4085

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–329. https://doi.org/10.1525/mp.2006.23.4.319

Järveläinen, H., Välimäki, V., & Karjalainen, M. (2001). Audibility of the timbral effects of inharmonicity in stringed instrument tones. *Acoustics Research Letters Online*, *2*(3), 79–84. https://doi.org/10.1121/1.1374756

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). Guilford Press.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

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. https://doi.org/10.1037/0033-2909.129.5.770

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. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion*, *8*(5), 668–683. https://doi.org/10.1037/a0013505

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 453–489). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.001.0001

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Bahavioral and Brain Sciences*, *31*, 559–621. https://doi.org/10.1017/S0140525X08005293

Kawakami, A., Furukawa, K., Katahira, K., Kamiyama, K., & Okanoya, K. (2013). Relations between musical structures and perceived and felt emotions. *Music Perception*, *30*(4), 407–417. https://doi.org/10.1525/mp.2013.30.4.407

Kawakami, A., & Katahira, K. (2015). Influence of trait empathy on the emotion evoked by sad music and on the preference for it. *Frontiers in Psychology*, *6*, 1541. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01541

Kazazis, S., Depalle, P., and McAdams, S. (2021). “The Timbre Toolbox User’s Manual,” https://github.com/MPCL-McGill/TimbreToolbox-R2021a (Last viewed 17 01 23).

Kendall, R. A., & Carterette, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck’s adjectives. *Music Perception*, *10*(4), 445–467. https://doi.org/10.2307/40285583

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, *68*(1), 17–26. https://doi.org/10.1016/j.ijpsycho.2007.12.001

Kivy, P. (1990). *Music alone: Philosophical reflections on the purely musical experience*. Cornell University Press.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Kreutz, G., Schubert, E., & Mitchell, L. A. (2008). Cognitive styles of music listening. *Music Perception*, *26*(1), 57–73. https://doi.org/10.1525/mp.2008.26.1.57

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*(4), 336–352. https://doi.org/10.1037/1196-1961.51.4.336

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2022). *caret: Classification and Regression Training* (6.0-93). https://CRAN.R-project.org/package=caret

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture: Relation of valence to arousal. *Journal of Personality*, *85*(4), 530–542. https://doi.org/10.1111/jopy.12258

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in linear mixed effects models* (3.1-3). https://CRAN.R-project.org/package=lmerTest

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 146–154. https://doi.org/10.1037/a0024671

Lahdelma, I., & Eerola, T. (2016). Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music*, *44*(1), 37–54. https://doi.org/10.1177/0305735614552006

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, *19*(5), 633–653. https://doi.org/10.1080/02699930441000445

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx)*, 237–244.

Lenth, R. V., Buerkner, P., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.8.0). https://CRAN.R-project.org/package=emmeans

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Lüdecke, D., Makowski, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., Wiernik, B. M., Arel-Bundock, V., Jullum, M., & gjo11. (2022). *performance: Assessment of regression models performance* (0.9.2). https://CRAN.R-project.org/package=performance

MacInnis, C. C., Boss, H. C., & Bourdage, J. S. (2020). More evidence of participant misrepresentation on MTurk and investigating who misrepresents. *Personality and Individual Differences*, *152*, 109603. https://doi.org/10.1016/j.paid.2019.109603

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping*, *30*, 267–275.

Marozeau, F., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*(5), 2946–2957. https://doi.org/10.1121/1.1618239

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198522577.003.0006

McAdams, S. (2019a). The perceptual representation of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 1–19). Springer International Publishing.

McAdams, S. (2019b). Timbre as a structuring force in music. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, Perception, and Cognition* (pp. 211–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_8

McAdams, S., Douglas, C., & Vempala, N. N. (2017). Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in Psychology*, *8*, 153. https://doi.org/10.3389/fpsyg.2017.00153

McAdams, S., & Goodchild, M. (2017a). *Perceptual evaluation of the recording and musical performance quality of OrchSim orchestral simulations compared to professional CD recordings* (MPCL Technical Report). McGill University. http://www.mcgill.ca/mpcl

McAdams, S., & Goodchild, M. (2017b). Musical structure: Sound and timbre. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 129–139). Routledge.

McAdams, S., & Siedenburg, K. (2019). Perception and cognition of musical timbre. In P. J. Rentfrow & D. J. Levitin (Eds.), *Foundations in music psychology: Theory and research* (pp. 71–120). MIT Press. https://doi.org/10.1037/e514592010-016

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. https://doi.org/10.1007/BF00419633

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Miu, A. C., & Vuoskoski, J. K. (2017). The social side of music listening: Empathy and contagion in music-induced emotions. In C. Waddington & E. King (Eds.), *Music and empathy* (pp. 124–138). Routledge.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Müllensiefen, D., Gingras, B., Stewart, L., & Ji, J. (2013). *Goldsmiths Musical Sophistication Index (Gold-MSI) v1.0: Technical report and documentation revision 0.3* (p. 69).

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nusbaum, E. C., & Silvia, P. J. (2011). Shivers and timbres: Personality and the experience of chills from music. *Social Psychological and Personality Science*, *2*(2), 199–204. https://doi.org/10.1177/1948550610386810

Nyklíček, I., Thayer, J. F., & Van Doornen, L. J. P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology*, *11*(4), 304–321.

OrchPlayMusic. (2022). OrchPlay Music. https://www.orchplaymusic.com/

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science*, *2*(3), 281–296. https://doi.org/10.1111/j.1745-6916.2007.00045.x

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. https://doi.org/10.1017/S1930297500002205

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902–2916. https://doi.org/10.1121/1.3642604

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. https://doi.org/10.1016/S0010-0277(98)00043-2

Pilgrim, L., Norris, J. I., & Hackathorn, J. (2017). Music is awesome: Influences of emotion, personality, and preference on experienced awe. *Journal of Consumer Behaviour*, *16*(5), 442-451. https://doi.org/10.1002/cb.1645

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. https://doi.org/10.1037/a0022406

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. https://doi.org/10.1037/0022-3514.84.6.1236

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388. https://doi.org/10.1177/0305735604046096

Risset, J.-C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *Psychology of Music* (2nd ed., pp. 113–169). Academic Press.

Robinson, K., & Patterson, R. (1995). The duration required to identify the instrument, the octave, or the pitch chroma of a musical note. *Music Perception*, 13(1), 1-15. https://doi.org/10.2307/40285682

Rodà, A., Canazza, S., & De Poli, G. (2014). Clustering affective qualities of classical music: Beyond the valence-arousal plane. *IEEE Transactions on Affective Computing*, *5*(4), 364–376. https://doi.org/10.1109/TAFFC.2014.2343222

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences, 10*(4), 251. https://doi.org/10. 3390/brainsci10040251

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon’s Mechanical Turk. *Communication Monographs, 85*(1), 140–156. https://doi.org/10.1080/03637751.2017. 1342043

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, *14*, 325–345. https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology, 4*, 837. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00837

Schubert, E., & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta acustica united with Acustica*, *92*(5), 820-825.

Schutz, M., Huron, D., Keeton, K., & Loewer, G. (2008). The happy xylophone: Acoustics affordances restrict an emotional palate. *Empirical Musicology Review*, *3*(3), 126–135. https://doi.org/10.18061/1811/34103

Seow, T. X. F., & Hauser, T. U. (2022). Reliability of web-based affective auditory stimulus presentation. *Behavior Research Methods*, *54*(1), 378–392. https://doi.org/10.3758/s13428-021-01643-0

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, *27*, 125–140.

Siedenburg, K., & McAdams, S. (2017). Four distinctions for the auditory “wastebasket” of timbre. *Frontiers in Psychology*, *8*, 1747. https://doi.org/10.3389/fpsyg.2017.01747

Siedenburg, K., Saitis, C., & McAdams, S. (2019). The present, past, and future of timbre research. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, R. R. Fay, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition*. Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_8

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*, 110–120.

Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *The Journal of the Acoustical Society of America*, *30*(5), 421–425. https://doi.org/10.1121/1.1909632

Sottile, G., Cilluffo, G., & Muggeo, V. M. (2022). *islasso: The Induced Smoothed Lasso* (1.4.3). https://CRAN.R-project.org/package=islasso

Strait, D. L., Chan, K., Ashley, R., & Kraus, N. (2012). Specialization among the specialized: Auditory brainstem function is tuned in to timbre. *Cortex*, *48*(3), 360–362. https://doi.org/10.1016/j.cortex.2011.03.015

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

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. https://doi.org/10.1080/02699930701503567

Vienna Symphonic Library GmbH. (2022). Vienna Symphonic Library. https://www.vsl.co.at

Vuoskoski, J. K., & Eerola, T. (2011a). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173. https://doi.org/10.1177/1029864911403367

Vuoskoski, J. K., & Eerola, T. (2011b). The role of mood and personality in the perception of emotions represented by music. *Cortex*, *47*(9), 1099–1106. https://doi.org/10.1016/j.cortex.2011.04.011

Warrenburg, L. A. (2020). People experience different emotions from melancholic and grieving music. *Music & Science, 3*, 205920432097738. https://doi.org/10.1177/2059204320977384

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded form*. Cedar Rapids: University of Iowa. https://ir.uiowa.edu/psychology_pubs/11

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weber, R. (1991). The continuous loudness judgment of temporally variable sounds with an “analog” category procedure. In A. Schick, J. Hellbruck, & R. Weber (Eds.), *Fifth Oldeburg symposium on psychological acoustics* (pp. 267–294). BIS.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., Freidank, M., Cai, J., & Protivinsky, T. (2021). *corrplot: Visualization of a Correlation Matrix* (0.92). https://CRAN.R-project.org/package=corrplot

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio. (2022). *ggplot2: Create elegant data visualisations using the grammar of graphics* (3.3.6). https://CRAN.R-project.org/package=ggplot2

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics, 79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361–2

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, *31*(4), 339–358. https://doi.org/10.1525/mp.2014.31.4.339

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An interlanguage unification of musical timbre. *Music Perception*, *32*(4), 394–412. https://doi.org/10.1525/mp.2015.32.4.394

Zentner, M., & Eerola, T. (2010). Self-report measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 187–221). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.001.0001

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. https://doi.org/10.1037/1528-3542.8.4.4

Appendix A

Table A1 *Type III Wald F-tests of Linear Mixed Models for Each Affect Scale*

	Valence					
	Exp. 1			Exp. 2		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register (R)	2, 52.9	4.79	.01	2, 33.7	30.70	<.0001
Family (F)	3, 47.0	13.87	<.0001	3, 20.0	11.89	.0001
Affect Locus (AL)	—	—	—	1, 4610.0	16.66	<.0001
R*F	—	—	—	6, 20.0	3.76	.01
R*AL	2, 175.6	17.57	<.0001	2, 4610.0	9.68	<.0001
F*AL	—	—	—	—	—	—
3-way	6, 7327.0	2.91	.008	—	—	—

	Tension					
	Exp. 1			Exp. 2		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 52.1	6.46	.003	2, 56.7	36.54	<.0001
Family	3, 47.0	9.95	<.0001	3, 31.3	16.09	<.0001
Affect Locus	—	—	—	1, 4606.6	33.52	<.0001
R*F	—	—	—	6, 31.3	8.65	<.0001
R*AL	2, 170.5	5.97	.003	—	—	—
F*AL	3, 7327.0	4.75	.003	—	—	—
3-way	6, 7327.0	2.22	.04	—	—	—

	Energy					
	Exp. 1			Exp. 2		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 67.1	36.21	<.0001	1, 85.8	32.76	<.0001
Family	—	—	—	—	—	—
Affect Locus	—	—	—	1, 4677.0	4.27	.04
R*F	6, 47.0	3.34	.008	—	—	—
R*AL	2, 186.0	7.57	.0007	—	—	—
F*AL	—	—	—	—	—	—
3-way	6, 7327.0	2.77	.01	—	—	—

Table A1 (Continued)

	Anger					
		Exp. 1			**Exp. 2**	
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 54.5	9.23	.0004	2, 28.0	33.51	<.0001
Family	3, 47.0	7.31	.0004	—	—	—
Affect Locus	—	—	—	1, 4601.0	219.61	<.0001
R*F	—	—	—	—	—	—
R*AL	2, 137.9	31.71	<.0001	2, 4610.0	74.87	<.0001
F*AL	3, 7277.0	17.69	<.0001	—	—	—
3-way	—	—	—	—	—	—

	Fear					
		Exp. 1			**Exp. 2**	
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 60.6	23.61	<.0001	2, 33.4	31.99	<.0001
Family	3, 47.0	6.39	.001	3, 20.0	3.21	.05
Affect Locus	1, 132.2	20.01	<.0001	1, 4601.0	227.81	<.0001
R*F	6, 47.0	4.27	.002	6, 20.0	6.11	.0009
R*AL	—	—	—	—	—	—
F*AL	3, 7271.0	9.70	<.0001	3, 4601.0	2.67	.05
3-way	6, 7271.0	4.57	.0001	—	—	—

	Sadness					
		Exp. 1			**Exp. 2**	
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 61.8	13.35	<.0001	2, 42.2	4.31	.02
Family	3, 47.0	32.15	<.0001	3, 20.0	20.00	<.0001
Affect Locus	1, 131.8	22.64	<.0001	1, 4601.0	285.48	<.0001
R*F	—	—	—	—	—	—
R*AL	2, 201.8	9.46	.0001	—	—	—
F*AL	3, 7271.0	14.37	<.0001	3, 4601.0	5.15	.001
3-way	6, 7271.0	2.69	.01	—	—	—

Table A1 (Continued)

	Happiness					
	Exp. 1			**Exp. 2**		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 53.5	12.73	<.0001	2, 33.9	27.90	<.0001
Family	3, 47.0	11.52	<.0001	3, 20	11.44	.0001
Affect Locus	—	—	—	—	—	—
R*F	—	—	—	6, 20.0	4.91	.003
R*AL	2, 135.6	7.51	.0008	2, 4610.0	6.41	.002
F*AL	3, 7277.0	4.50	.004	—	—	—
3-way	—	—	—	—	—	—

	Tenderness					
	Exp. 1			**Exp. 2**		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 58.3	10.21	.0002	2, 48.5	31.22	<.0001
Family	3, 47.0	4.53	.007	3, 20.0	10.27	.0003
Affect Locus	1, 129.0	4.19	.04	1, 4610.0	104.41	<.0001
R*F	—	—	—	6, 20.0	8.49	.0001
R*AL	—	—	—	—	—	—
F*AL	—	—	—	—	—	—
3-way	—	—	—	—	—	—

	Preference					
	Exp. 1			**Exp. 2**		
Fixed Effect	***df***	***F***	***p***	***df***	***F***	***p***
Register	2, 51.9	4.11	.02	2, 34.5	13.04	<.0001
Family	3, 47.0	14.20	<.0001	3, 20.0	26.04	<.0001
Affect Locus	—	—	—	—	—	—
R*F	—	—	—	—	—	—
R*AL	—	—	—	—	—	—
F*AL	3, 7333.0	6.00	.0004	—	—	—
3-way	—	—	—	—	—	—

Table A2 *Bonferroni-Corrected Pairwise Comparisons of Instrument Families*

		Experiment 1				Experiment 2			
Scale	**Contrast**	**Estimate**	***SE***	***z***	***p***	**Estimate**	***SE***	***z***	***p***
Valence	**B–P**	–1.64	0.41	–3.96	.0005	–1.61	0.29	–5.49	<.0001
	B–S	0.03	0.40	0.07	1	–0.25	0.31	–0.79	1
	B–W	0.01	0.37	0.02	1	–0.46	0.28	–1.63	.62
	P–S	1.67	0.44	3.82	.0008	1.36	0.32	4.30	.0001
	P–W	1.65	0.40	4.08	.0003	1.14	0.29	3.96	.0005
	S–W	–0.02	0.39	–0.05	1	–0.22	0.31	–0.71	1
Tension	**B–P**	1.71	0.50	3.42	.004	1.92	0.28	6.96	<.0001
	B–S	–0.19	0.49	–0.40	1	0.37	0.29	1.27	1
	B–W	0.04	0.44	0.10	1	0.53	0.27	1.99	.28
	P–S	–1.90	0.53	–3.62	.002	–1.55	0.30	–5.20	<.0001
	P–W	–1.67	0.49	–3.43	.004	–1.38	0.27	–5.10	<.0001
	S–W	0.23	0.47	0.50	1	0.16	0.29	0.56	1
Anger	**B–P**	1.05	0.42	2.50	.07				
	B–S	–0.56	0.41	–1.36	1				
	B–W	0.02	0.37	0.06	1				
	P–S	–1.60	0.44	–3.64	.002				
	P–W	–1.03	0.41	–2.52	.07				
	S–W	0.57	0.39	1.46	.87				
Fear	**B–P**	0.53	0.31	1.70	.53	0.44	0.26	1.72	.51
	B–S	–0.40	0.30	–1.34	1	0.20	0.27	0.72	1
	B–W	–0.12	0.27	–0.43	1	0.50	0.25	2.04	.25
	P–S	–0.93	0.33	–2.86	.03	–0.24	0.28	–0.88	1
	P–W	–0.65	0.30	–2.15	.19	0.07	0.25	0.26	1
	S–W	0.28	0.29	0.97	1	0.31	0.27	1.15	1
Sadness	**B–P**	1.57	0.21	7.61	<.0001	0.90	0.19	4.73	<.0001
	B–S	–0.03	0.20	–0.16	1	–0.32	0.20	–1.57	.69
	B–W	–0.11	0.18	–0.59	1	–0.07	0.19	–0.38	1
	P–S	–1.60	0.22	–7.38	<.0001	–1.22	0.21	–5.93	<.0001
	P–W	–1.68	0.20	–8.37	<.0001	–0.97	0.19	–5.16	<.0001
	S–W	–0.08	0.19	–0.40	1	0.25	0.20	1.24	1
Happi-ness	**B–P**	–1.59	0.33	–4.89	<.0001	–1.15	0.21	–5.40	<.0001
	B–S	–0.11	0.32	–0.34	1	–0.03	0.23	–0.12	1
	B–W	0.11	0.29	0.39	1	–0.24	0.21	–1.14	1
	P–S	1.48	0.34	4.34	.0001	–0.24	0.21	–1.14	1
	P–W	1.70	0.32	5.39	<.0001	0.92	0.21	4.35	.0001
	S–W	0.22	0.31	0.72	1	–0.21	0.23	–0.93	1

Table A2 (continued)

		Experiment 1				Experiment 2			
Scale	**Contrast**	**Estimate**	***SE***	***z***	***p***	**Estimate**	***SE***	***z***	***p***
Tender-ness	**B–P**	–0.99	0.30	–3.28	.006	–1.04	0.16	–6.48	<.0001
	B–S	–0.35	0.29	–1.20	1	–0.36	0.17	–2.1	.20
	B–W	–0.11	0.27	–0.40	1	–0.35	0.16	–2.23	.15
	P–S	0.64	0.32	2.02	.26	0.67	0.17	3.90	.0006
	P–W	0.88	0.29	3.01	.02	0.69	0.16	4.37	.0001
	S–W	0.24	0.28	0.85	1	0.02	0.17	0.10	1
Pref-erence	**B–P**	–1.84	0.49	–3.73	.001	–1.87	0.31	–6.12	<.0001
	B–S	–0.02	0.48	–0.04	1	–0.15	0.33	–0.47	1
	B–W	–0.02	0.44	–0.04	1	–0.46	0.30	–1.55	.82
	P–S	1.82	0.52	3.51	.003	1.72	0.33	5.21	.0003
	P–W	1.83	0.48	3.81	.0008	1.41	0.30	4.68	.0009
	S–W	0.00	0.47	0.01	1	–0.31	0.32	–0.96	1

Note. Contrasts signify subtractions (B–P: Brass minus Percussion; B–S: Brass minus Strings; B–W: Brass minus Woodwinds; etc.), estimates signify results of those subtractions.

Table A3 *Affect Locus×Register Pairwise Comparisons*

		Experiment 1						Experiment 2					
		Trend		**I – P**	***SE***	***z***	***p***	**Trend**		**I – P**	***SE***	***z***	***p***
Scale		*I*	*P*					*I*	*P*				
Valence	**LN**	–0.16	0.16	–0.32	0.05	–5.93	<.0001	0.29	0.39	–0.10	0.02	–4.35	<.0001
Tension	**LN**	0.34	0.14	0.20	0.06	3.30	.001						
Energy	**LN**	0.32	0.56	–0.24	0.06	–3.85	.0001						
Anger	**LN**	0.17	–0.23	0.40	0.05	7.93	<.0001	–0.33	–0.63	0.30	0.03	11.61	<.0001
	QD	0.14	0.23	–0.09	0.03	–3.65	.0003	0.18	0.23	–0.06	0.02	–3.29	.001
Sadness	**LN**	–0.16	–0.32	0.16	0.05	3.41	.0006						
	QD	–0.00	–0.08	0.07	0.02	3.22	.0013						
Happi-ness	**LN**	0.17	0.33	–0.16	0.05	–3.22	.0013	0.27	0.35	–0.08	0.02	–3.52	.0004

Note. The LN row compares the linear register trends, the QD row compares the quadratic register trends. The *I – P* column shows the results of the estimated *induced* register trend minus the estimated *perceived* register trend.

Table A4 *Affect Locus×Family Pairwise Comparisons*

	Experiment 1					Experiment 2				
Scale	**Family**	**I – P**	***SE***	***z***	***p***					
Tension	B	0.27	0.15	1.80	.07					
	P	–0.30	0.15	–2.09	.04					
	S	–0.01	0.13	–0.10	.92					
	W	–0.17	0.14	–1.19	.23					
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***					
	B – P	0.57	0.15	3.84	.0007					
	B – W	0.44	0.15	3.02	.01					
	Family	**I – P**	***SE***	***z***	***p***					
Anger	B	–0.19	0.20	–0.93	.35					
	P	–0.95	0.20	–4.64	<.0001					
	S	–0.36	0.20	–1.76	.08					
	W	–0.40	0.20	–1.97	.05					
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***					
	B – P	0.75	0.11	6.71	<.0001					
	P – S	–0.59	0.11	–5.58	<.0001					
	P – W	–0.55	0.11	–5.14	<.0001					
	Family	**I – P**	***SE***	***z***	***p***	**Family**	**I – P**	***SE***	***z***	***p***
Fear	B	–0.98	0.23	–4.24	<.0001	B	–0.88	0.11	–8.18	<.0001
	P	–0.61	0.23	–2.71	.007	P	–0.55	0.11	–5.04	<.0001
	S	–1.15	0.22	–5.32	<.0001	S	–0.92	0.11	–8.42	<.0001
	W	–1.13	0.22	–5.05	<.0001	W	–0.91	0.11	–8.43	<.0001
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***	**AL × F Contrast**	**Estimate**	***SE***	***z***	***p***
	P – S	0.54	0.13	4.02	.0003	P – S	0.37	0.16	2.37	.08
	P – W	0.52	0.15	3.56	.002	P – W	0.36	0.15	2.38	.08
	Family	**I – P**	***SE***	***z***	***p***	**Family**	**I – P**	***SE***	***z***	***p***
Sadness	B	–1.57	0.25	–6.27	<.0001	B	–1.00	0.11	–8.80	<.0001
	P	–0.63	0.26	–2.44	.01	P	–0.52	0.11	–4.86	<.0001
	S	–1.3	0.26	–5.08	<.0001	S	–0.89	0.12	–7.36	<.0001
	W	–1.6	0.25	–6.49	<.0001	W	–0.98	0.11	–9.11	<.0001
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***	**AL × F Contrast**	**Estimate**	***SE***	***z***	***p***
	B – P	–0.94	0.16	–6.00	<.0001	B – P	–0.47	0.14	–3.42	.004
	P – S	0.67	0.16	4.05	.0003	P – S	0.36	0.14	2.61	.04
	P – W	0.97	0.15	6.41	<.0001	P – W	0.45	0.14	3.34	.005

Table A4 (continued)

	Family	I – P	*SE*	*z*	*p*
Happ-iness	B	–0.52	0.23	–2.30	.02
	P	–0.41	0.23	–1.81	.07
	S	–0.18	0.23	–0.76	.45
	W	–0.26	0.23	–1.17	.24
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***
	B – S	–0.35	0.11	–3.24	.007
	B – W	–0.26	0.10	–2.58	.05
	Family	**I – P**	***SE***	***z***	***p***
Pref-erence	B	–0.02	0.18	–0.13	.90
	P	–0.16	0.19	–0.88	.38
	S	–0.39	0.19	–1.61	.11
	W	0.08	0.18	0.46	.64
	AL × F Contrast	**Estimate**	***SE***	***z***	***p***
	B – S	0.27	0.10	2.81	.03
	P – W	–0.25	0.09	–2.62	.04
	S – W	–0.38	0.09	–4.09	.0002

Note. The *I – P* column shows the results of the average *induced* rating minus the average *perceived* rating, separated by instrument family. The *Estimate* column shows the results of the difference in *I – P* results between instrument families (*AL × F Contrast*).

Table A5 *Three-Way Pairwise Comparisons (Exp. 1)*

Scale	**Register Trend**	**Family**	**I – P**	***SE***	***z***	***p***
Valence	QD	B	–0.08	0.03	–2.22	.03
		P	0.07	0.03	1.91	.06
		S	0.03	0.03	1.31	.19
		W	–0.02	0.03	–0.65	.52
		3-Way Contrast	**Estimate**	***SE***	***z***	***p***
		B – P	–0.14	0.04	–3.35	.005
		B – S	–0.11	0.04	–3.04	.01
	Register Trend	**Family**	**I – P**	***SE***	***z***	***p***
Tension	QD	B	0.07	0.04	1.79	.07
		P	–0.06	0.04	–1.61	.11
		S	–0.04	0.03	–1.45	.15
		W	–0.03	0.04	–0.81	.42
		3-Way Contrast	**Estimate**	***SE***	***z***	***p***
		B – P	0.13	0.05	2.78	.03
		B – S	0.11	0.04	2.76	.03
	Register Trend	**Family**	**I – P**	***SE***	***z***	***p***
Energy	LN	B	–0.34	0.09	–3.61	.0003
		P	–0.13	0.08	–1.75	.08
		S	–0.15	0.07	–2.14	.03
		W	–0.34	0.08	–4.17	<.0001
		3-Way Contrast	**Estimate**	***SE***	***z***	***p***
		S – W	0.20	0.07	2.67	.05
	Register Trend	**Family**	**I – P**	***SE***	***z***	***p***
Fear	LN	B	–0.02	0.09	–0.27	.79
		P	0.21	0.07	2.95	.003
		S	–0.04	0.06	–0.73	.47
		W	–0.09	0.08	–1.16	.25
		3-Way Contrast	**Estimate**	***SE***	***z***	***p***
		P – S	0.25	0.07	3.60	.002
		P – W	0.30	0.08	3.52	.003
	Register Trend	**Family**	**I – P**	***SE***	***z***	***p***
Sadness	QD	B	0.15	0.04	3.85	.0001
		P	0.01	0.04	0.30	.77
		S	0.03	0.03	1.07	.29
		W	0.10	0.04	2.71	.007
		3-Way Contrast	**Estimate**	***SE***	***z***	***p***
		B – P	0.14	0.05	2.75	.04
		B – S	0.12	0.04	2.71	.04

Note. The *I – P* column shows the difference in linear (LN) or quadratic (QD) register trend between affect loci. The *3-Way Contrast* column shows the difference in *I – P* between instrument families. For example, the estimate for the 3-way contrast of B – P shows the result of $(I – P)_{Brass} – (I – P)_{Percussion}$.

Table A6 *Moderation Effects of Individual Differences*

Moderator	Experiment	Valence	Tension	Energy	Anger	Fear	Sadness	Happiness	Tenderness	Preferences
BFI Agreeableness	1	–	–	–	–	–	–	F	–	–
	2	–	–	–	–	–	–	–	–	–
BFI Conscien-tiousness	1	–	F	–	–	–	–	–	–	–
	2	–	–	–	–	–	–	–	F	–
BFI Extraversion	1	–	–		–	–	–	–	–	–
	2	–	–	AL	–	AL	R	F	F	–
BFI Neuroticism	1	–	F	–	F×AL	–	–	–	–	–
	2	–	–	F	–	–	–	–	–	–
BFI Openness	1	F	–	F	–	–	F×AL	F×AL	F	F
	2	F	F	–	–	AL	–	–	–	–
Gold-MSI Emotions	1	F	F	–	–	–	–	–	F	–
	2	F	–	–	–	AL	–	–	–	–
Gold-MSI Engagement	1	–	F	F	F	F	–	–	–	F
	2	R×AL	–	–	–	AL	–	–	–	–
Gold-MSI General	1	F	F	F	–	–	3-way	–	–	F×AL
	2	F	F	–	–	AL	AL	–	F	–
Gold-MSI Perceptual	1	–	–	F	–	R×F	F×AL	F×AL	F	F×AL
	2	–	–	–	–	AL	AL	–	F	–
Gold-MSI Singing	1	F	–	F	–	R×F	–	–	–	F×AL
	2	R×F	R×F	–	–	AL	AL	–	F	–
Gold-MSI Training	1	F	F	–	–	AL	F	F×AL	–	F
	2	F	F	–	R×AL	–	AL	–	AL	F
IRI Empathic Concern	1	–	–	–	–	–	–	–	–	–
	2	–	–	AL	–	–	–	–	–	–
IRI Fantasy	1	–	–	–	–	3-way	–	–	F	–
	2	F	F	–	–	AL	–	–	–	–

Table A6 (continued)

Moderator	Experiment	Valence	Tension	Energy	Anger	Fear	Sadness	Happiness	Tenderness	Preferences
IRI Personal Distress	1	–	–	–	F	–	–	–	–	–
	2	R×AL	AL	–	–	–	–	–	–	–
IRI Perspect-ive Taking	1	–	–	F	F	–	–	–	–	–
	2	R×F	–	AL	–	–	–	–	–	–
IRI Total	1	–	–	–	–	–	–	–	–	–
	2	R×AL	–	AL	–	–	–	–	–	–
PANAS Positive	1	–	–	–	F	R×F	R	R	F	F
	2	–	F	F	R×AL	AL	–	–	–	–
PANAS Negative	1	–	–	–	–	–	–	–	F	–
	2	–	–	–	AL	AL	AL	–	–	–
STOMP-R Mellow	1	F	–	R×F	F×AL	–	–	–	–	F×AL
	2	R×AL	AL	AL	–	F	–	–	–	–
STOMP-R Unpretentious	1	–	–	–	–	–	F×AL	–	–	F×AL
	2	R×F	–	–	AL	AL	–	–	–	R
STOMP-R Sophisticated	1	F	F	–	F×AL	AL	–	–	–	F
	2	F	–	–	–	–	–	–	–	–
STOMP-R Intense	1	–	–	–	–	–	–	–	–	F
	2	–	–	–	–	–	–	–	–	–
STOMP-R Contemporary	1	–	–	–	–	–	–	–	–	–
	2	R×AL	R×AL	AL	–	–	–	–	–	–

Note. For each dependent variable (first row), we indicate the effects of register (R), family (F), affect locus (AL), or their interactions, that were moderated by each moderator (first column) with a significance of $p < .01$ and improving model performance by $\Delta\ AICc > 10$.

Appendix B

Table B1 *Model Performance for the Lasso and Random Forest Regressions and The Difference Between the Two Methods (%) in Predicting the Eerola12 Ratings, Mcadams17 Ratings, and Perceived Affect Ratings of Experiment 1*

		Eerola12			**McAdams17**			**Exp1–Per**		
		Lasso	**RF**	**%**	**Lasso**	**RF**	**%**	**Lasso**	**RF**	**%**
Valence	R^2	.58	.65	12.1	.71	.73	2.8	.86	.90	4.7
	Test-R^2	.69	.66	–4.3	.71	.72	1.4	.83	.90	8.4
	Avg NRMSE	.18	.15	–14.0	.16	.16	0.0	.10	.08	–23.1
Tension	R^2	.70	.71	1.4	.44	.64	45.5	.71	.81	14.1
	Test-R^2	.72	.67	–6.9	.44	.59	34.1	.67	.83	23.9
	Avg NRMSE	.14	.14	–2.6	.18	.15	–20.0	.15	.11	–25.6
Energy	R^2	.57	.58	1.8	.77	.83	7.8	.71	.79	11.3
	Test-R^2	.56	.59	5.4	.76	.85	11.8	.68	.76	11.8
	Avg NRMSE	.13	.12	–9.2	.14	.10	–25.0	.11	.09	–14.8

Note. R^2 was calculated on all data with fivefold cross-validation. *Test-*R^2 is a measure of predictive relevance, which is the averaged R^2 across the fivefold validation. The *NRMSE* is RMSE divided by mean y (valence/tension/energy), similarly averaged across the fivefold. The % column shows the percentage of increase in scores going from Lasso to RF.

Table B2 *Model Performance for the Lasso and Random Forest Regressions and the Difference Between the Two Methods (%) in Predicting the Perceived and Induced Affect Ratings of Experiment 1*

		Perceived			**Induced**		
		Lasso	**RF**	**%**	**Lasso**	**RF**	**%**
Valence	R^2	.86	.90	4.7	.70	.84	20.0
	Test-R^2	.83	.90	8.4	.68	.86	26.5
	Avg NRMSE	.10	.08	–23.1	.13	.09	–30.8
Tension	R^2	.71	.81	14.1	.61	.82	34.4
	Test-R^2	.67	.83	23.9	.58	.80	37.9
	Avg NRMSE	.15	.11	–25.6	.16	.11	–31.3
Energy	R^2	.71	.79	11.3	.63	.62	–1.6
	Test-R^2	.68	.76	11.8	.64	.77	20.3
	Avg NRMSE	.11	.09	–14.8	.07	.06	–14.3
Mean	R^2	.76	.83	10.0	.65	.76	17.6
	Test-R^2	.73	.83	14.7	.63	.81	28.2
	Avg NRMSE	.12	.09	–21.2	.12	.09	–25.5

Note. See Table B1. The mean rows show the R^2, *Test*-R^2, and *Avg NRMSE* averaged over valence, tension, and energy arousal.

Table B3 *Model Performance for the Lasso and Random Forest Regressions and the Difference Between the Two Methods (%) in Predicting the Perceived and Induced Affect Ratings on Discrete Scales in Experiment 1*

		Perceived			Induced			Mean		
		Lasso	**RF**	**%**	**Lasso**	**RF**	**%**	**Lasso**	**RF**	**%**
Anger	R^2	.71	.86	21.1	.63	.82	30.2	.67	.84	25.7
	Test-R^2	.73	.82	12.3	.63	.84	33.3	.68	.83	22.8
	Avg NRMSE	.24	.20	–16.7	.26	.17	–34.6	.25	.19	–25.7
Fear	R^2	.65	.76	16.9	.75	.71	–5.3	.70	.74	5.8
	Test-R^2	.67	.77	14.9	.75	.76	1.3	.71	.77	8.1
	Avg NRMSE	.22	.19	–13.6	.17	.20	17.6	.20	.20	2.0
Sadness	R^2	.81	.85	4.9	.79	.81	2.5	.80	.83	3.7
	Test-R^2	.86	.85	–1.2	.78	.81	3.8	.82	.83	1.3
	Avg NRMSE	.11	.11	0.0	.11	.10	–9.1	.11	.11	–4.6
Happiness	R^2	.72	.76	5.6	.83	.86	3.6	.78	.81	4.6
	Test-R^2	.77	.81	5.2	.87	.86	–1.1	.82	.84	2.1
	Avg NRMSE	.19	.18	–5.3	.16	.15	–6.3	.18	.17	–5.8
Tendernes s	R^2	.77	.81	5.2	.75	.82	9.3	.76	.82	7.3
	Test-R^2	.75	.82	9.3	.70	.81	15.7	.73	.82	12.5
	Avg NRMSE	.15	.14	–6.7	.13	.11	–15.4	.14	.13	–11.1
Mean	R^2	.73	.81	10.7	.75	.80	8.1	.74	.81	9.4
	Test-R^2	.76	.81	8.1	.75	.82	10.6	.75	.82	9.4
	Avg NRMSE	.18	.16	–8.5	.17	.15	–9.6	.17	.16	–9.0

Note. See Table B1. The mean rows show the R^2, *Test-*R^2, and *Avg NRMSE* averaged over valence, tension, and energy arousal. Similarly, the mean columns show the R^2, *Test-*R^2, and *Avg NRMSE* averaged over perceived and induced affect.

Table B4 *Model Performance for the Lasso and Random Forest Regressions and the Difference Between the Two Methods (%) in Predicting the Perceived and Induced Affect Ratings of Experiment 2*

		Perceived			Induced		
		Lasso	**RF**	**%**	**Lasso**	**RF**	**%**
Valence	R^2	.81	.71	–12.3	.71	.67	–5.6
	Test-R^2	.80	.72	–10.0	.71	.66	–7.0
	Avg NRMSE	.10	.14	33.3	.11	.14	24.1
Tension	R^2	.61	.65	6.6	.46	.69	50.0
	Test-R^2	.51	.61	19.6	.44	.65	47.7
	Avg NRMSE	.14	.15	7.6	.15	.16	7.8
Energy	R^2	.77	.79	2.6	.77	.68	–11.7
	Test-R^2	.79	.73	–7.6	.84	.75	–10.7
	Avg NRMSE	.09	.09	11.1	.07	.09	15.0
Mean	R^2	.73	.72	–1.1	.65	.68	10.9
	Test-R^2	.70	.69	0.7	.66	.69	10.0
	Avg NRMSE	.11	.13	17.3	.11	.13	15.6

Note. See Table B1. The mean rows show the R^2, *Test*-R^2, and *Avg NRMSE* averaged over valence, tension, and energy arousal.

Table B5 *Model Performance for the Lasso and Random Forest Regressions and the Difference Between the Two Methods (%) in Predicting the Perceived and Induced Affect Ratings on Discrete Scales in Experiment 2*

		Perceived			**Induced**		
		Lasso	**RF**	**%**	**Lasso**	**RF**	**%**
Anger	R^2	.90	.91	1.1	.82	.80	–2.4
	Test-R^2	.89	.90	1.1	.80	.77	–3.8
	Avg NRMSE	.18	.17	–10.5	.17	.18	7.3
Fear	R^2	.69	.75	8.7	.75	.75	0.0
	Test-R^2	.72	.59	–18.1	.75	.63	–16.0
	Avg NRMSE	.12	.15	20.0	.13	.14	9.3
Sadness	R^2	.86	.65	–24.4	.60	.57	–5.0
	Test-R^2	.83	.78	–6.0	.59	.57	–3.4
	Avg NRMSE	.07	.12	72.0	.09	.12	36.0
Happiness	R^2	.81	.74	–8.6	.85	.69	–18.8
	Test-R^2	.83	.79	–4.8	.88	.85	–3.4
	Avg NRMSE	.15	.20	32.4	.10	.17	64.0
Tenderness	R^2	.76	.61	–19.7	.81	.70	–13.6
	Test-R^2	.80	.66	–17.5	.83	.76	–8.4
	Avg NRMSE	.11	.14	33.3	.10	.11	17.9
Mean	R^2	.80	.73	–8.6	.77	.70	–8.0
	Test-R^2	.81	.74	–9.1	.77	.72	–7.0
	Avg NRMSE	.13	.15	29.4	.12	.14	26.9

Note. See Table B1. The mean rows show the R^2, *Test*-R^2, and *Avg NRMSE* averaged over valence, tension, and energy arousal. Similarly, the mean columns show the R^2, *Test*-R^2, and *Avg NRMSE* averaged over perceived and induced affect.