UNDERSTANDING MUSICAL EMOTION: EXPLORING THE INTERACTION BETWEEN CUES, TRAINING, AND INTERPRETATION

UNDERSTANDING MUSICAL EMOTION: EXPLORING THE INTERACTION BETWEEN CUES, TRAINING, AND INTERPRETATION

By AIMEE BATTCOCK, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctorate of Philosophy (Ph.D.)

Lay Abstract

Musical performers and composers express emotions through the selection and use of various musical features, or cues. Studies exploring how listeners perceive emotion in music have identified several cues important to this process—often using tightly controlled (and constrained) tone sequences crafted for experimental purposes. More work is needed to examine how listeners decode communicated emotion in unaltered passages created by renowned composers—the kind of music routinely performed and enjoyed by audiences for generations. Here in three sets of experiments I apply a novel stimulus set and analysis to determine the relative importance of three musical features. Additionally, I explore the role of the listener's level of expertise as well as the importance of performers' interpretative decisions. My work offers a new way to understand the relationship between musical features and emotional messages, helping to clarify one of music's most mysterious and powerful capabilities.

Abstract

Previous work on conveyed musical emotion has often focused on experimentally composed and manipulated music, or multi-lined music selected to express overt emotions. This highly controlled approach may overlook some aspects of the complex relationship between composers, performers, and listeners in transmitting emotional messages. My PhD research focuses on how listeners perceive emotion in music, specifically, how listeners interpret musical features such as timing, mode and pitch in complex musical stimuli. I also demonstrate how listeners with musical expertise use cues differently to perceive emotion and the effect of performer interpretation on this communication process.

Throughout this dissertation I use a dimensional approach capturing perceived valence and arousal to assess complex musical stimuli. I adapted a technique used in other domains to music, affording an opportunity to explore nuanced relationships between cues and listener ratings of emotion. In Chapter 1 I show that musically untrained adults mainly use cues of timing and mode when rating emotional valence, mirroring previously reported. Additionally, I show that although pitch information emerges as a significant predictor of listener's valence ratings, listeners use it less than cues such as timing and mode. Further, I demonstrate that neither mode nor pitch information help listeners rate perceived arousal. Finally, in Chapter 4, I show differences in performer interpretation mediate the strength of individual cues, as well as the distribution of emotional ratings across each album. In Chapter 3, I demonstrate that listeners with musical

training use cues differently than untrained listeners, with more reliance on information communicated through mode when making judgements of emotional valence. Altogether these findings corroborate previous evidence suggesting timing and mode cues are of the greatest importance in conveying /perceiving emotion, this process is further mediated by individual differences in both pianist (interpretation) and listener (musical training)—underscoring the complex relationship between composer, performer, and audience.

Descriptive Note

McMaster University DOCTORATE OF PHILOSOPHY (2019) Hamilton, Ontario (Psychology, Neuroscience and Behaviour)

TITLE: Understanding musical emotion: Exploring the interaction between cues, training, and interpretation

AUTHOR: Aimee E. Battcock, B.Sc.

SUPERVISOR: Dr. Michael Schutz

PAGES: xvii, 268

Acknowledgements

First off, I would like to thank my PhD supervisor, Mike Schutz for the opportunity to spend the last five years doing research in an area I dreamt about as an undergrad. I would also like to express my deepest appreciation to the members of my supervisory committee, Laurel Trainor and Bruce Milliken, for their input and guidance as I progressed through this degree. To Karin Humphreys, thank you for being an additional source of guidance and support through all my regression problems. Your willingness to sit down and work through all my questions is a testament to how you truly care about students.

I would also like to thank all the lab-mates I've been lucky to work alongside of these past few years, for their feedback, their support and inspiring me become a better scientist. In addition, I'm extremely grateful for all of the wonderful undergraduates I've worked with— you have made my graduate mentor experience more rewarding, fun and overall better than I could have ever envisioned. Thank you to our most recent lab manager, my OG thesis student and one half of the #dreamteam, Annilee Baron. Collaborating with you in various academic capacities has been the best, and I don't want to stop doing it! I am so grateful you even considered coming back to ride out the last year to manage the lab, update holiday Bruce, and my Festive Vienna Sausages background.

The most important thing I've gained over the past few years has been the meaningful and long-lasting friendships within and outside of the department that have kept me going through this journey. It truly takes a village to help you through grad school— to get you through the tough times and celebrate in the good ones. First of all, a big thank you to all the kind, supportive grad students I've made as friends throughout my years and fabulous office admin in the PNB department for creating a welcoming atmosphere in the department. To Erica, my favourite 'lab-mate', thank you for providing me refuge in your office at a time when I needed it the most. Rachael Finnerty— Thank you for accepting me as a temporary office mate, but also for the numerous opportunities to guest lecture, and for your overall positivity and support.

To Ruth and Victoria S. — I couldn't have asked for two better people to be there with me from start to finish. Ruth, thank you for being my East Coast/Halifax loving, festive décor loving, and Barrett's Privateer singing other half. Victoria, thank you for being the most understanding sound board for all family issues, for bringing Normz into our lives and for your margz. To half of 'A&W' and the 'A Team', David Prete and Hector, you have empowered me to believe in myself, be kind to myself and simply to be myself. We here, we alive, we family.

Of course, a special mention is needed for two of the best lab-mates turned absolute best friends I could have ever asked for —Sharmila and Anna. Sharmila, your unwavering support and unbelievable kindness has been a much-needed addition to the lab and especially my life. I could not have finished my last two years without you, and I aspire to be as amazingly thoughtful as you are every day. Anna, best friends don't come better than you. Thank you for whole heartedly accepting me for the weird, anxious, grandma I am, for keeping me grounded, and taking me on a variety of adventures. Most importantly, thank you for being such a good best friend that you'd willingly visit me in one of your least favourite cities.

A big thank you goes out to my childhood best friends, parents and my large family, for the love and support over the many, many, many years I've spent away to pursue what seemed like an endless amount of education. I am all that I am because of you. Vicki, Bahiyyih, Corinne, you have and will always be 'home' to me. Thank you to my second family, the O'Kanes, for taking me in every Thanksgiving, for putting me on the mantle, and for being such great examples of the kindest people. It is an honour to be a part of your family. Emily — Thank you for the all the visits across the many places you've lived in the past 5 years, for the begrudging runs in the rain/heat/snow, and for all the movies and chats sitting on my bedroom floor from undergrad through grad school. I can't wait to continue that into old age.

To both of my grandmothers— your strength has fueled my strength, in all areas of life. To my grandpa(w) and Pop, your resourcefulness and generosity remind me that above all else, it is as important to be as kind as it is to be ambitious. Lastly, a special mention goes to my mom. You have been steadfast in your support for all my life choices (except maybe the tattoos). You've set the best example of what it is to be a hardworking, generous person, who finds or creates joy in the moments around her. Thank you for nurturing my desires to go

against the grain (like supporting my love of purple lipstick, even in our small town), for enabling all my crafting dreams, for your absolute selflessness and for supplying me with the framework of a person that can achieve what they set their mind to.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

ANOVA	Analysis of variance
bpm	Beats per minute
CI	Confidence interval
CV	Coefficient of variation
df	Degrees of freedom
η^2	Eta-squared
F	F ratio
M	Mean
MOE	Margin of error
n	Number of participants
p	p-value, probability
SD	Standard deviation
t	t-statistic
χ 2	Chi squared statistic
Z	z-score

Declaration of Academic Achievement

This thesis is comprised of three empirical that each examine how emotion is communicated to listeners. The studies examine different aspects of acoustic information in music that influence a communicated emotion. Chapter 2 is an article that has been published in a peer-reviewed journal. Chapters 3 and 4 consists of manuscripts that are currently in preparation for publishing. Each empirical chapter represents a single manuscript where I am the primary author.

The remainder of this preface is designed to clarify my contributions to the manuscripts that comprise the chapters of this thesis.

Chapter 1: Introduction

Author: Aimee E. Battcock

Chapter 2: Acoustically Expressing Affect

Authors: Aimee E. Battcock & Michael Schutz
Publication: Music Perception, 2019, volume 37, issue 1, pp. 66-91
Comments:
My role in this manuscript included experimental design, data collection of human participants, data analysis and interpretation of results. I am the primary author of this manuscript and the second author is my thesis supervisor. A number of undergraduate students assisted with data collection.

Chapter 3: The Influence of Musical Training on Perceived Emotion

Authors: Aimee E. Battcock & Michael Schutz
Publication: In Prep
Comments:
My role in this manuscript included experimental design, data collection of human participants, data analysis and interpretation of results. I am the primary author of this manuscript and the second author is my thesis supervisor. A number of undergraduate students assisted with data collection.

Chapter 4: Performer Differences and Communicated Emotion

Authors: Aimee E. Battcock & Michael Schutz
Publication: In Prep
Comments:

My role in this manuscript included experimental design, data collection of human participants, data analysis and interpretation of results. I am the primary author of this manuscript and the second author is my thesis supervisor. A number of undergraduate students assisted with data collection.

Chapter 5: Discussion

Author: Aimee E. Battcock

As a final note, each of Chapters 2 through 4 represents a manuscript that is intended to stand alone as a published article. As a result, there is a fair amount of redundancy present within the introductory sections to each of these chapters that the reader should be aware. However, the experiments presented within each manuscript are distinct works that are intended to address separate but complimentary objectives

Chapter 1: General Introduction

1.0 Introduction

As social species, communicating and perceiving emotional signals is crucial for daily interactions. Recognizing and responding appropriately to these signals helps us navigate relationships in the social world, promoting survival (Ekman, 1992; Izard, 1992). The importance of this ability is evident in human development, as sharing emotional information is crucial in infancy, where mothers use infant-directed speech to regulate the social relationship and emotional state with their children (Trainor, 1996; Trehub & Trainor, 1998). Auditory signals are one such medium through which emotions can be communicated. For example, humans can use vocal prosody in speech to help convey their emotions to each other, using the components, or features that comprise this acoustic medium to do so. Music is also one such medium through which emotional signals can be communicated to listeners, regardless of any verbal comprehension.

This chapter presents a brief overview of music and emotion research, focusing on factors that impact perceived emotion. I discuss the influence of features in notated music focusing on three specific cues (attack rate, mode, pitch). Beyond these specific cues, I will discuss influence of individual differences that emerge from listener differences in both (a) musical training and (b) performer interpretation. Additionally, this chapter discusses challenges with quantifying emotion conveyed through music. Chapters 2-4 explore these topics across a series of novel experiments that together shed insight into non-verbal emotional communication. These findings help clarify debates on the importance of musical cues such as mode and timing within music listening. Further, they contribute to the literature investigating which cues are helpful in the listening process, and how cues selected by the composer and manipulated by the performer influence perceived valence and arousal.

1.1 Expression of Emotion in Music

The ability for music to represent or convey emotions, distinct from evoked emotions or felt emotions, is thought to occur because musical structure contains characteristics that resemble expression in speech (Davies, 1980). Similar to communication in speech, emotion can be conveyed through information transmitted by a combination of acoustic cues or features in music. However, unlike speech, music can communicate emotion without semantic information. With music, whether an emotion is successfully conveyed is dependent on a number of elements that interact in complex ways as messages are transmitted from composer to performer to listener. This process can differ from everyday speech, as a composer who creates the musical composition does not always produce the auditory signal. Therefore, music offers a unique avenue to explore the complex relationship between structure and interpretation in communication of emotion.

Juslin's (1997) adaptation of Brunswik's (1956) visual perception Lens Model attempts to explain the communicative process of conveyed emotion in music. According to this adapted model, composers and performers can encode (e.g., express) emotion using a combination of probabilistic features, or cues (Figure 1). Upon exposure to this auditory information, listeners can use these cues to decode (e.g., perceive or recognize) the emotion that is expressed. Although the original adaptation of Lens Model helps unpack how cues can be used to transmit information in the auditory signal, it does not consider the impact of individual differences introduced by the signal's encoders (performers) and decoders (listeners). More recently, Juslin and Lindström (2010) extended this work to encompass this complex interaction that occurs between composer and performer. Using a factorial experimental design, the authors composed short sequences, varying cues of interest accordingly. They determined that different features appeared important for specific emotions; mode was important for 'sadness' and articulation emerged as significant to predictor listener ratings for 'fear'. In Chapters 3 and 4 of this dissertation, I expand on these aspects of the communication process using precomposed music to expand upon what we know of the Lens Model and how cues interact in perceived emotion.

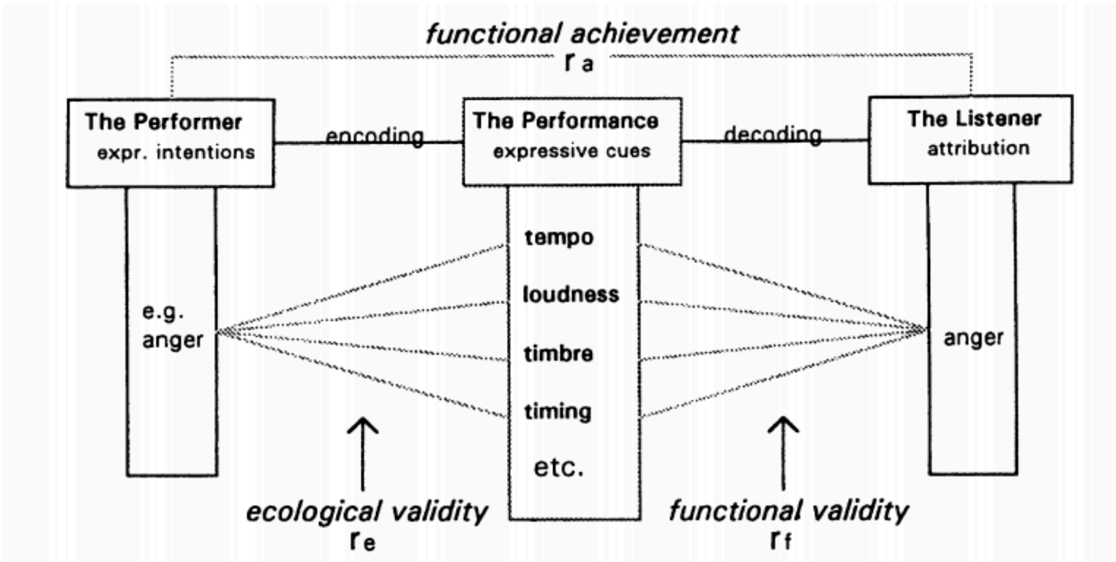


Figure 1.1 Juslin's (1997) adaptation of the Brunswik (1956) Lens Model illustrating communication of emotion to listeners using cues.

Several features are recognized as important in conveying musical emotion, including pitch (Hevner, 1937; Huron, 2008), articulation and dynamics (Eerola, Friberg, & Bresin, 2013), mode —major vs minor—(Costa, Fine, Enrico, & Bitti, 2004; Hevner, 1935), and timing (Balkwill & Thompson, 1999; Gagnon & Peretz, 2003). These features have been studied using a range of stimuli, including single lined melodies, composed or experimentally manipulated (Hailstone et al., 2009; Lindström, 2006; Quinto, Thompson, & Keating, 2013), with a smaller amount of studies using multi-lined excerpts (Eerola, 2011; Schubert, 2004). Three features that often emerge as effective features in conveyed emotion are cues of timing, mode and pitch (Gagnon & Peretz, 2003; Hevner, 1937; Ilie & Thompson, 2006). Cues like mode and timing are frequently found to be important predictors of listener assessments of perceived emotion

(Eerola et al., 2013). The influence of pitch cues have been found to transmit emotional information in a similar fashion to their functioning in speech (Bachorowski & Owren, 1995; Breitenstein, Lancker, & Daum, 2001; Hevner, 1937) however, the lack of consistency in results raises interesting questions. This dissertation focuses on these three cues as they function in the context of polyphonic, or multi-lined music, to help unpack how cues function to transmit emotion.

1.1.1 Timing

Features that convey temporal information are found to be one of the most important cues to convey emotion to listeners (Eerola et al., 2013). Timing, often considered tempo, reflects the rate at which the information or the music is presented. The informative nature of timing appears important for perception across modalities, including speech and music. Theories about the importance of timing emerge from its connection with biological cues, and behaviours such as speech and gait, as an indicator for energy and energy expenditure (Gomez & Danuser, 2007). Further, the relationship between temporal cues and energy may be related to physiological responses, where faster paced music can increase and individual's measured or perceived level of sympathetic arousal response (Dillman Carpentier & Potter, 2007; Husain, Thompson, & Schellenberg, 2002). Developmentally, the salience of temporal information in the form of musical tempo emerges earlier than cues such as mode. Children as young as four years are able to distinguish emotions in music based on tempo information (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). This early sensitivity to the temporal information of unfolding events in music may emerge due to innate and learned associations between speed and general behavioural responses (Mote, 2011).

In music specifically, timing information can be conveyed through a number of cues such as tempo, speed, or rhythm (Balkwill & Thompson, 1999; Juslin & Madison, 1999; Schellenberg, Krysciak, & Campbell, 2000). Timing is often discussed in terms of tempo (Balkwill & Thompson, 1999; Gagnon & Peretz, 2003; Scherer & Oshinsky, 1977), however this dissertation focuses on attack rate (i.e., note attacks per second) as a timing cue of interest. This measure of timing information considers both the number of attacks and the tempo and is comparable to articulation rate in speech. The benefit of examining attack rate, is the sensitivity to the use of fast tempos with longer attack durations, or slow tempos with short attack durations which may capture aspects of conveyed information that tempo does not.

1.1.2 Mode

This dissertation — grounded in music from a renowned composer who chose to naturally co-vary these musical cues — can help clarify conflicting views. Music theorists have argued that mode's role in expressed emotion is overly generalized and perhaps its role is misinterpreted as a result of the relation with other elements of musical structure (Hatten, 2004). However, some theorists have detailed common emotional qualities associated with each major and minor key (Freedberg, 2006). In contrast, perceptual evidence has continuously demonstrated a link between music written in major and minor modes and emotion. The major-minor distinction is found to be a strong predictor of emotion judgements, where major modes are found to associate with positively valenced emotions and moods such as happiness, and minor modes are associated with the more negatively valenced ones such as sadness (Costa et al., 2004; Hevner, 1935; Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005; Quinto et al., 2013; Webster & Weir, 2005). Exploring the connection between mode and emotion with more complex stimuli sheds light on the cue relationships while considering aspects like harmony, the natural co-variation of cues such as modality and timing, and performer interpretation.

1.1.3 Pitch

Pitch is an acoustic cue that refers to the perceptual property of sound that allows us to identify whether tones are 'high' or 'low'. Pitch communicates affective information to listeners in speech where the fundamental frequency (primary acoustic correlate of perceived pitch) is one of the acoustic features that exerts strong effects on expressed emotion (Banse & Scherer, 1996; Breitenstein, Lancker, & Daum, 2001; Scherer, 1995; Scherer & Oshinsky, 1977). In music, empirical evidence measuring various pitch-based cues such as pitch height, pitch range, pitch variability (Gabrielsson & Lindström, 2010; Ilie & Thompson, 2006) demonstrates the association with expressed emotion; however, the contribution of pitch does not appear to behave consistently. In general, higher octaves (higher frequencies of the tones) communicate more positive emotions such as happy and dreamy (Eitan & Timmers, 2010) and lower octaves, the more negative emotions such as sad and somber (Gundlach, 1935; Hevner, 1937; Scherer & Oshinsky, 1977; Watson, 1942; Wedin, 1972). Conversely, high and low pitches have also been associated with both positive and negative emotions (Ilie & Thompson, 2006; Scherer & Oshinsky, 1977). Given the probabilistic nature of cues to express emotion (Juslin, 1997), this may suggest other cues, aside from pitch, are more salient to express emotion. It may also be a result of the different pitch-based cues measured to examine expressed emotion, or the range in stimuli complexity in studies that use single-lined (Balkwill & Thompson, 1999) or multi-lined (Ilie & Thompson, 2006) music examples.

1.2. Individual Differences

A crucial component of the communication process resides in the listener (Figure 1.2). Within the perceptual process, individual backgrounds and experiences can influence the music listening experience (Juslin & Laukka, 2004), and may impact how one interprets the information around them. For example, personality traits can drive individual differences in emotional processing (Rusting, 1998) and given the subjective nature of music listening, can result in the same emotional signals being perceived differently. In fact, evidence suggests that personality traits can moderate mood-congruence with emotion ratings (Vuoskoski & Eerola, 2011).When examining expressed emotion, research often focuses on the influence of the composer or musician instead of the listener. Further, studies discussing individual characteristics including personality and musical training, present inconsistent findings that highlight the complicated nature of investigating the impact of individual differences on the perception of emotion.

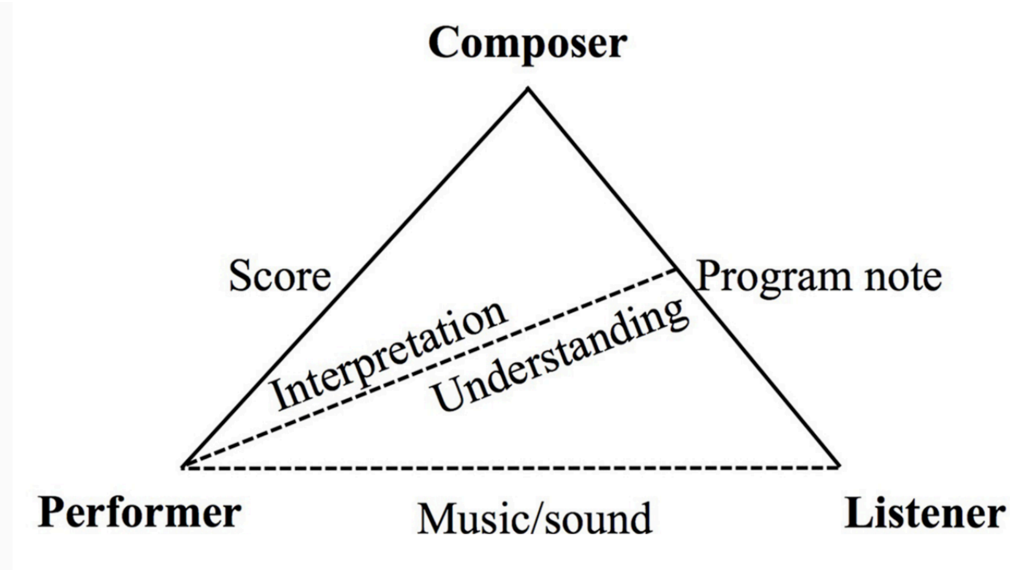


Figure 1.2. Blom, Bennett, & Stevenson (2016)*'s model of communication between composer, performer and listener.*

1.2.1. Individual Differences in Musical Training

Musical training involves acquiring knowledge on how to perform in an expressive manner to listeners, providing students with the tools to interpret musical compositions and convey intended emotions. Furthermore, individuals who partake in formal music lessons spend concentrated effort listening to and perceiving the auditory information of their chosen instrument. In general, individuals with musical training are found to process auditory stimuli differently (Ladinig & Schellenberg, 2012; Taruffi, Allen, Downing, & Heaton, 2017; Vuoskoski & Eerola, 2011). Differences in brain activity both in processing and detecting changes within the auditory signal suggest musical training can induce changes in the neural circuitry that affect how the brain process sounds. This raises questions about whether increased musical knowledge can affect perceptual processes regulating conveyed emotion.

Musical training is found to correlate with emotion recognition, where years of training can relate to recognition accuracy (Akkermans et al., 2018; Castro & Lima, 2014). These studies demonstrate that years of musical training correlate positively with accuracy, and differences emerge between the abilities of those musically trained and non-experts. Furthermore, it appears that with musical training, there is better identification accuracy for more complicated/often confused emotions, like anger (Castro & Lima, 2014). Investigation on the predictive weights of acoustic and musical cues also indicate differences in how musically trained listeners use cues. There, models of listener ratings show more explained variance for older musicians than non-musicians of the same age. These results suggest that musical training can influence how listeners use cues to perceive emotion.

Evidence suggesting potential advantages of musical training in perceiving musical emotions is mixed. Studies using forced choice paradigms for emotion recognition often find no difference in recognition accuracy between those with and without musical training (Juslin, 1997; Trimmer & Cuddy, 2008). What is apparent from these types of studies is that musical expertise is not necessarily a pre-requisite to perceiving emotion in music (Castro & Lima, 2014). However, this does not mean expertise is irrelevant. Music listeners continuously gain musical knowledge from the experience of listening to music, making them 'experienced' listeners (Bigand & Poulin-Charronnat, 2006). Thus, tasks that are overtly explicit can be completed with this basic knowledge. It also is possible that with experimentally composed or manipulated stimuli, a conveyed emotion is salient enough that there is no confusion regardless of musical knowledge/training. Investigating the presence of training effects with multi-lined, pre-composed music allows further insight to how musical knowledge or experience shapes the perceptual processing of communicated emotions in complex passages, with music frequently experienced in everyday situations.

Another issue or question that arises is the amount of musical training that is sufficient to develop differences within these recognition or identification tasks. The operational definition of a 'musical expert' varies from study to study on the amount of formal musical training required. A large sample of the studies that fail to find differences between trained and untrained participants often define experts as those having six years of musical training (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Trimmer & Cuddy, 2008) . This may suggest that benefits emerge after more long-term exposure/training. Additionally, what contributes to the distinction musical ‘expert’ is not agreed upon, where some argue years of musical training is not enough to define an ‘expert’.

The Goldsmith Musical Sophistication Index (Müllensiefen, Gingras, Musil, & Stewart, 2014) attempts to address challenges with a self-report measurement of expertise. Although this inventory does not directly test one’s musical ability, it is a valid and reliable instrument that can distinguish different aspects of musical sophistication: including self-reported perceptual and singing abilities along with active and emotional engagement with music and musical training. I use this measure in the methodology within Chapter 3 to confirm our selection of participants with musical training encompasses individuals who demonstrate a musical expertise above a normal population as compared to data norms from Müllensiefen, Gingras, Stewart, & Musil, (2013).

1.2.2 Individual Differences in Performance

Although composers can embed their emotional intention through musical structure, it is the performer who must interpret the notation and create the auditory signal. Thus, how a performer creates their performance can shape the listener’s experience. This helps explain why we may prefer one performer’s version of a piece over another. This interpretation, or performance expression often involves small and/or large variations of the expressive cues available

(Gabrielsson & Juslin, 1996; Kendall & Carterette, 1990; Palmer, 1997). Although some work explores the effects of performer interpretation, a large focus has been on measuring differences in expressive cues (Sloboda, 1985) or listener ability to identify different renditions (Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011) and on the features of musical structure involved in emotional expression and recognition (Eerola et al., 2013).

As mentioned above, listeners use cues that convey timing information (tempo, articulation) as well as dynamics when perceiving emotion (Dodson, 2011; Repp, 1992). Timing is the primary way in which performers are able to control the encoding of musical interpretation (Juslin, 1997; Repp, 1992). Music research explores how performers vary cues in their renditions of pieces, which often occur at the microstructural level for cues like tempo and dynamics (Macritchie, Eiholzer, & Italiana, 2012). However, less attention has centered around the influence of performer interpretation on perceived emotion for the same musical piece.

Listeners are sensitive to performance differences when rating both concurrently and retrospectively (Sloboda & Lehmann, 2001), even between similar performances. They can identify performance styles (Gingras et al., 2011) including differences in expressive phrasing (Macritchie et al., 2012). In addition, perceived emotionality of a performance is found to differ between performer interpretations (Sloboda & Lehmann, 2001). Although previous work has demonstrated similar associations between cues used by performers and decoded by listeners (Juslin, 2000), the effect of performer differences on cue use and the consequences on listener perception requires further understanding.

1.3. Challenges of Quantifying Emotion

Investigating emotional perception in music presents several significant experimental challenges, and differences in approaches to overcoming these challenges likely contribute to occasionally conflicting results. Stimuli selection is crucial when considering the reliability and generalizability of results. Further, based on the stimuli used, findings may overestimate the effect of cues or demonstrate conflicting evidence of their effect. Therefore, experimental results should be interpreted with respect to the stimuli used. Additionally, the push to incorporate ecologically valid music into experimental practices is important to capture the complex nature of how emotion can be communicated to listeners.

In order to analyze emotional responses, regression modelling of listener ratings is commonly used to determine the cues that can significantly predict responses. This can help identify which cues are important to the perception of specific emotions, or dimensions of emotion such as valence and arousal. Researchers employing this analysis find models are best at predicting variance associated with ratings of arousal in contrast to valence (Eerola, 2012; Eerola, Lartillot, & Toiviainen, 2009; Korhonen, Clausi, & Ed Jernigan, 2006; Vuoskoski & Eerola, 2011). Although it is not clear why this pattern emerges frequently in the literature, it may be a result of the selected cues used in these models.

Therefore, in addition to regression modelling, this dissertation employs commonality analysis to better understand the relations between cues which predict listener responses. Given the natural correlation between musical cues, this is a powerful tool for better understanding the complex ways in which inter-related cues combine (as well as act independently) in conveying emotional messages.

1.3.1 Stimuli

One of the major challenges with investigating emotion in music and the most important decision to make is the choice of stimuli. Originally, studies exploring the perception of emotion conveyed in music had focused on single-lined, experimentally designed and/or manipulated stimuli (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013). These studies have been instrumental in identifying important cues listeners use to decode emotion, however they may oversimply the complex relationship that occurs between cues and perception when listening to multi-lined, and more complex music. Some studies have extended on this work using selections of instrumental, classical or pop music (Eerola, 2011; Leman, Vermeulen, De Voogdt, Moelants, & Lesafre, 2005; Schubert, 2004; Yang & Chen, 2012). Those studies have modelled listener responses, adding to the literature on the relative weights of cues for perception, and whether these cue patterns can be seen across musical genres (Eerola, 2012). Although these speak more to the complexities in music of everyday listening, effects of familiarity, of multi-instrument/timbre highlight the need to continue exploration with naturalistic stimuli to best understand the listener experience.

The stimuli used in experiments designed to examine perceived emotions are often selected explicitly for their ability to convey specific basic emotions. Thus, what remains less understood is how cues are used with more emotionally ambiguous musical examples. This thesis investigates this topic using music from Bach's *Well-Tempered Clavier (WTC),* which offers numerous features desirable for experimental study. It affords some control over familiarity for untrained, non-musicians, as it is largely unfamiliar to this population. In addition, it consists of one instrumental timbre and is balanced with regards mode (a musical cues of interest) as it has 12 major and minor pieces in every key. Although it represents a set of musical compositions within one genre, it can be used to help build on our understanding of listener perceptual responses to music frequently heard in concert halls, comprised of cues that have been composed to co-vary together.

1.3.2 Measuring Perceptual Processes of Emotion

Two main methods frequently used to capture listener perception of emotion involve discrete or dimensional conceptions of emotion (Zentner & Eerola, 2010). Studies using these measures demonstrate listeners' abilities to perceive a range of emotion in music. Both methods appear successful in capturing conveyed emotions, however each provides subtle differences in capturing listener responses. Discrete methods involve providing specific examples of affective terms, and either ask participants to select the appropriate emotion conveyed or to rate the intensity to which each emotion is represented in the music (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013). This serves to measure accuracy rates and emotional intensity differences; however, it also has the potential to prime participants to recognize the conveyed emotion and can fail to account for more ambiguous or complex emotions.

Dimensional models conceptualize emotion into multiple dimensions or components. The most widely accepted and used model across research fields examining emotion is Russell's 2D circumplex model of affect (1980), which functions under the assumption that emotion can be broken down into two components: emotional valence (pleasure-displeasure) and arousal (activation-deactivation). This model has been used to measure emotional responses across fields, due to its reliability in capturing emotion in numerous contexts. Nonetheless, critics of dimensional models argue the difficulty in differentiation between emotions that fall close in the 2D space.

Although evidence suggests both models of emotion can capture similar amounts of variance in listener responses, the discrete model is found to be less reliable for ratings of ambiguous emotional musical examples (Vuoskoski & Eerola, 2011). This suggests dimensional models may best capture perception in response to more ambiguously expressed emotion given its ability to have a more fine-tuned scale of emotion. This is crucial in music listening experiences outside of the laboratory, where emotional messages are not explicitly dictated by a researcher and subsequently embedded by a composer.

1.4 Modelling Listener Responses of Perceived Emotion

1.4.1 Regression analysis

Regression analyses are frequently used in studies to assess how well cue predictors can explain variance in listener ratings of emotion. These models have revealed cues often function in an additive fashion (Eerola et al., 2013; Schubert, 2004) and can indicate how well cues predict listener ratings of valence and arousal, especially across different genres. Ratings of arousal are found to be better predicted by cues than valence, where upwards to 43-62% of variance is predicted for listener assessments of arousal and 16-43% of the variance of valence ratings (Eerola, 2011). The better model fit for ratings of arousal may reflect the selection of cues used as predictors in these models. A large portion of cues investigated often represent expressive, or performance cues such as tempo, loudness or articulation that demonstrate a stronger impact on ratings of emotional arousal than valence (Eerola et al., 2009; Schubert, 2004). Structural cues such as texture, melodic contour and harmony have been used as predictors and can help explain some variance; however, the degree to model fit for valence ratings using these cues appears to be piece dependent. This may help explain why modelling perceived valence across several pieces or excerpts results in less variance predicted (lower R^2 values) than for perceived arousal. Our methodology incorporates the cue of mode as a predictor, given its strong association with perceived emotion as mentioned above. Although major/minor distinction of mode with emotion is Western music specific, using mode as a predictor in listener ratings helps unpack the strong learned associations that become acculturated through exposure and can help explain how listeners hear emotions in music.

1.4.2 Commonality Analysis

Commonality analysis affords a method of assessing relative importance of predictors in a regression model. This technique partitions the explained variance accounted for by predictors within the regression model into components of unique and shared variance (Figure 1.3). Shared variance represents the variance two predictors have in common with the dependent variable and in contrast, the unique variance is what is independently predicted by a predictor (Ray-Mukherjee et al., 2014). Commonality analysis (or variance partitioning) can further identify the relative weight of predictors with respect to unique and shared variance. This is particularly important in situations with multicollinearity or correlations between the predictors. Correlations are frequently found between structural cues in Western tonal music. Therefore, using commonality analysis to assess the relative importance of quantified cues could be more beneficial than comparing beta value strengths. To the best of our knowledge, this project represents the first application of this statistical technique to music, allowing us to gain more insight into how key variables either uniquely or commonly predict responses of perceived emotion. This helps to have a better understanding of the relative weights of cues used in ratings of valence or arousal and helps us uncover how the relationship between timing and mode affect listener responses.

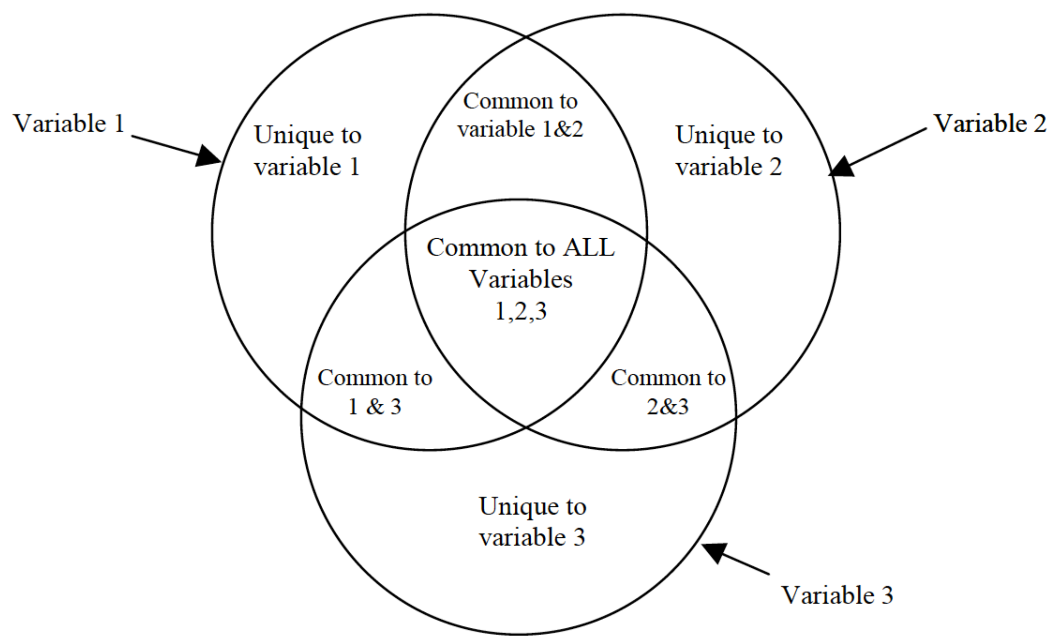


Figure 1.3. Capraro & Capraro's (2001) illustration of commonality analysis.

Although commonality analysis breaks down the components of unique and shared variance predicted, it does not provide a way to assess statistical differences between these components. As such, in Chapters 3 and 4 where I compare the relative weights of cues between ratings of different groups, we employ bootstrapping methods to calculate error bars to help quantify potential differences. These methods involve resampling with replacement to create simulated data distributions from an original dataset in order to estimate population statistics. Through the process of bootstrapping, we are able to simulate datasets of participant responses to determine measures of accuracy. With better understanding of the variability of each calculated commonality analysis coefficient we are able to make further comparisons between the relative weights of predictors in our models.

1.5. Thesis Overview

This thesis seeks to further explore the role of musical cues as they relate to listener perceptions of conveyed emotion in music. In the present work, I use an adaptation of Russell's 2D Circumplex Model of Affect (1980) to explore the relation between listener perception and musical structure. Further, I employ commonality analysis, a technique not previously used in this research field to uncover the relationships between cues of interest. Using this statistical analysis helps to untangle the influence of timing information and mode, two cues often linked in Western music, and aids in the argument that mode is a strong cue in the perception of emotion in music.

In Chapter 2, I discuss the importance of the information transmitted by select musical cues for the perception of emotions contained with complex, multi-lined music. Specifically, I demonstrate the importance of the information transmitted through cues of timing and mode. Additionally, in experiment 2 of

Chapter 2, I show cue importance shifts as a function of presenting musically 'resolved' excerpts to participant, where the cue of modality increases in its ability to predict for listener ratings. In Chapter 3, I show that listeners with musical training rely on information from cues differently than those without musical training, where their experience and intimate knowledge of musical structure and expression influence their perceptual processing. In particular, musical experts demonstrate a greater sensitivity to mode in experiment 2 (Chapter 3) than non-experts, as it is found to be more predictive of their ratings of emotional valence. In Chapter 4, I show that based on a performer's interpretative decisions on expressive cues, listeners rely on cues differently when making their judgements of communicated emotion. Finally, I will discuss how these studies advance the field of emotional perception in Chapter 5 and suggest some future directions for this research.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., … Frieler, K. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1–20.

Bachorowski, J. ., & Owren, M. (1995). Vocal expression of emotion: Intensity and context. *Psychological Science*, *6*(4), 219–224.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perceptionl*, *17*(1), 43–64.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*. *70*(3), 614.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130.

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 and Emotion*, *19*(8), 1113–1139.

Blom, D. M., Bennett, D., & Stevenson, I. (2016). The Composer’s Program Note for Newly Written Classical Music: Content and Intentions. *Frontiers in Psychology*, *7*, 1707.

Breitenstein, C., Lancker, D. Van, & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion, 15*(1), 57–79.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, California: University of California Press.

Castro, S. L., & Lima, C. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142.

Costa, M., Fine, P., Enrico, P., & Bitti, P. E. R. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception: An Interdisciplinary Journal*, *22*(1), 1–14.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Davies, S. (1980). The expression of emotion in music. *Mind*, *89*(353), 67–86.

Dillman Carpentier, F. R., & Potter, R. F. (2007). Effects of Music on Physiological Arousal: Explorations into Tempo and Genre. *Media Psychology*, *10*, 339–363.

Dodson, A. (2011). Expressive timing in expanded phrases: An empirical study of recordings of three Chopin preludes. *Music Performance Research*, *4*, 2–29.

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. *Journal of New Music Research*, *40*(4), 349–366.

Eerola, T. (2012). Modeling Listeners' Emotional Response to Music. *Topics in Cognitive Science*, *4*(4), 607–624.

Eerola, T., Friberg, A., & Bresin, R. (2013a). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 487.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction Of Multidimensional Emotional Ratings in Music from Audio using Multivariate Regression Models. In K. Hirata, G. Tzanetakis, & K. Yoshii (Eds.), Proceedings of the *10th International Society for Music Information Retrieval. Kobe, Japan: International Society for Music Information Retrieval.*

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 162–200.

Freedberg, D. (2006). Composition and emotion. In M. Turner (Ed.), *The Artful Mind: Cognitive Science and the Riddle of Human Creativity* (pp. 73–89). Oxford University Press.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*, *24*, 68-91.

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

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, *40*(10), 1206–1220.

Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, *7*(2), 377–387.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *The American Journal of Psychology*, *47*(4), 624–643.

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. 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.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Bloomington, IN: Indiana University Press.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*, 103–118.

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*(4), 621–630.

Hoshino, E. (1996). The feeling of musical mode and its emotion in a melody. *Psychology of Music*, *24*(1), 29–46.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, *22*(2), 327–352.

Huron, D. (2008). A comparison of average pitch height and interval size in major- and minor-key themes : Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review*, *3*(2), 59–63.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception*, *20*(2), 151–171.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–330.

Izard, C. . (1992). Basic emotions, relations among emotions, and emotion-cognition. *Psychological Review*, *99*(3), 561–565.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, 14(4), 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1813.

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.

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364.

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception, 17*(2), 197–221.

Kendall, R. A., & Carterette, E. C. (1990). The Communication of Musical Expression. *Music Perception*, *8*(2), 129– 163.

Korhonen, M. D., Clausi, D. A., & Ed Jernigan, M. (2006). Modeling Emotional Content of Music Using System Identification. *Systems, Man and Cybernetics, Part B, IEEE Transactions on Systems, Man, and Cybernetics*, *36*(3), 588–599.

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.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesafre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, *34*(1), 39–67.

Lindström, E. (2006). Impact of melodic organization of melodic structure and emotional expression. *Musicae Scientiae*, *10*(1), 85–117.

Luck, G., Toiviainen, P., Errkkila, J., Lartillot, O., Riikkila, K., Makela, A., … Varri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, *36*(1), 25–45.

Macritchie, J., Eiholzer, H., & Italiana, S. (2012). Exploring the perceptual effects of performers ' interpretations. *Journal of Interdisciplinary Music Studies*, *6*(2), 177–200.

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*(3), 618–622.

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

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). *Goldsmiths Musical Sophistication Index (Gold-MSI) v1.0: Technical Report and Documentation Revision 0.3*. London.

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences*, *1060*, 450–453.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*(1), 115–138.

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4*, 1–8.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328.

Rigg, M. G. (1940). Speed as a determiner of musical mood. *Journal of Experimental Psychology*, *27*(5), 566–571.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception, 18*(2), 155–171.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology*, *39*(2), 273–293.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception*, *19*(1), 87–120.

Taruffi, L., Allen, R., Downing, J., & Heaton, P. (2017). Individual differences in music-perceived emotions: The influence of externally oriented thinking. *Music Perception*, *34*(3), 253–266.

Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development*, *19*(1), 83–92.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43–77.

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional Intelligence, Not Music Training, Predicts Recognition of Emotional Speech Prosody. *Emotion*, *8*(6), 838–849.

Vuoskoski, J. K., & Eerola, T. (2011). Special section on music in the brain: Research report The role of mood and personality in the perception of emotions represented by music. *Cortex*, *47*, 1099–1106.

Vuoskoski, J. k, & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173.

Watson, K. B. (1942). The nature and measurement of musical meanings. *Psychological Monographs*, *54*(2), i–43.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39.

Wedin, L. (1972). A multidimensional study of perceptual -emotional qualities in music. *Scandinavian Journal of Psychology*, *13*, 241–257.

Yang, Y.-H., & Chen, H. H. (2012). Machine Recognition of Music Emotion. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 1–30.

Zentner, M., & Eerola, T. (2010). Self-report based measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion* (pp. 187–221). Oxford: Oxford University Press.

Chapter 2

Battcock A. & Schutz. M. (2019). Acoustically Expressing Affect. *Music Perceptio*n, *33*(1), 66-91.

Preface

Listeners use musical features to decode musical emotions, however the roles of these cues in unaltered, multi-lined music are challenging to unravel. In particular, the role of mode in perceived emotion is a point of contention between psychologists and music theorists. In Chapter 2, university undergraduates are asked to rate perceived valence and arousal of excerpts from Bach's *Well-Tempered Clavier*. Attack rate, mode and pitch height emerge as the most important cues for ratings of valence, however attack rate appears as the only cue important for ratings of arousal. In a follow up experiment that gives more consideration to mode, using musically 'resolved' excerpts, participants again assess perceived valence and arousal of each excerpt. In this study, the cue of mode became more important for ratings of valence than other cues including attack rate and pitch. No changes are observed for cues used by listeners for ratings of arousal. Together, these studies show that although attack rate is consistently an important cue amongst dimensions of emotion, mode and pitch have a specialized function in communicating an emotion's valence. Further, these results suggest mode’s effect increases when considering the structural key changes of music designed to mix major and minor modes. Finally, to explore consequences of varying approaches to dimensional labelling, I had undergraduates rate perceived emotion using dimensional labels of valence and arousal in a third experiment to compare the effect of labelling on evaluations of the ‘energy’ dimension of emotion. Results confirm that the labelling of the intensity dimension as arousal leads to similar evaluations of emotion.

Chapter 2: Acoustically Expressing Affect

Abstract

Composers convey emotion through music by co-varying structural cues. Although the complex interplay provides a rich listening experience, this creates challenges for understanding the contributions of individual cues. Here we investigate how three specific cues (attack rate, mode and pitch height) work together to convey emotion in Bach's *Well Tempered-Clavier (WTC)*. In three experiments, we explore responses to (1) eight measure excerpts, (2) musically 'resolved' excerpts and (3) investigate the role of different standard dimensional scales of emotion. In each experiment, thirty non-musician participants rated perceived emotion along scales of valence and intensity (experiments 1 & 2) or valence and arousal (experiment 3) for 48 pieces in the *WTC*. Responses indicate listeners used attack rate, mode and pitch height to make judgements of valence, but only attack rate for intensity/arousal. Commonality analyses revealed mode predicted the most variance for valence ratings, followed by attack rate, with pitch height contributing minimally. In experiment 2 mode increased in predictive power compared to experiment 1. For experiment 3, using 'arousal' instead of 'intensity' showed similar results to experiment 1. We discuss how these results complement and extend previous findings of studies with tightly controlled stimuli, providing additional perspective on complex issues of interpersonal communication.

Keywords: emotion, perception, applied music cognition, valence, arousal

Introduction

"Music can serve as a way of capturing feelings, knowledge of feelings, or knowledge about the forms of feeling, communicating them from the performer or the creator to the listener" (Gardner, 1993, p. 124).

Music's relationship with emotion is one of the central reasons for our engagement with it (Juslin & Laukka, 2004) and continues to fascinate composers, listeners, psychologists, and neuroscientists alike. Similar to their use in vocal expression; listeners attend to and decode specific cues in lawful ways, with certain cues unique to music. Emotional communication is complex, governed by a multitude of factors both within the acoustic signal itself as well as from learned associations and experiences (i.e., national anthems, cultural conventions, etc.). The complexity and importance of this issue has generated sustained research interest (Hevner, 1936; Koelsch et al., 2004; Wiggins, 1998), finding consistent agreement in many aspects of its communicative abilities. Although some aspects are difficult to quantify precisely, a growing body of research on the relationship between psychophysical cues and their emotional associations has proven informative.

Timing as a cue for emotional expression

Timing is a powerful cue for emotional communication, however understanding its effect is complex as timing encompasses several distinct musical properties such as tempo and rhythm. (Balkwill & Thompson, 1999; Juslin & Madison, 1999; Schellenberg, Krysciak, & Campbell, 2000). Tempo, which describes the number of beats per minute, is of great importance for conveyed emotion (Balkwill & Thompson, 1999; Gagnon & Peretz, 2003; Scherer & Oshinsky, 1977). The role of musical tempo holds some parallels with articulation rate in speech, with fast and slow tempos associated with happiness and sadness respectively (Hevner, 1937; Juslin, 1997; Rigg, 1940). Sensitivity to tempo emerges at an early age, with children as young as four making affective judgments using tempo rather than familiarity (Mote, 2011). This develops earlier than their sensitivity to mode (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). Mote (2011) argues the dependency on tempo suggests children may generalize associations between speed and emotion in human behaviour — particularly speech — to music. This early sensitivity to timing may help explain why cues like tempo are found to have stronger effects than mode (Hevner, 1935, 1937).

Rhythm also plays a complex yet powerful role in emotional communication. The effect of rhythm is found to vary as a function of melody and intended emotion. In an experiment consisting of four measure melodies from unknown folksongs or experimentally composed melodies selected to express 'happy', 'sad', or 'scary', listeners rated melodies higher in the appropriately expressed emotion when excerpts contained rhythmic variation. In addition, the effect of rhythmic variation interacted with pitch (Schellenberg et al., 2000). The authors suggest their selection of emotional exemplars resulted in melodies that differed on a number of structural dimensions (number of contour changes, mean pitch level, as well as meter), which can explain why the effect of rhythm appeared context specific. Furthermore, the effect of rhythm can be so powerful it extends cross-culturally, correlating with emotions like joy, sadness and peace within Hindustani ragas presented to Western listeners (Balkwill & Thompson, 1999). In that study, participants rated pieces expressing joy to contain more simple rhythms in contrast to sad pieces, which participants judged to have more complex rhythms. In addition, these naïve Western listeners could accurately identify the intended emotions conveyed within the ragas. These findings demonstrated that despite unfamiliarity with the musical stimuli, the cue of rhythm remained a salient indicator of the conveyed emotion.

Mode as a cue for emotional expression

Unlike timing and pitch, the musical cue of mode is specific to music, referring to the structure of pitch information. Hevner's (1935) landmark work on mood associations with common Western modes (major and minor) illustrates that minor modes are associated with negatively valenced emotions such as 'sad' and 'melancholy,' whereas major melodies are described as 'cheerful' and 'gay'. In fact, mode is often a significant predictor of valence, with the major mode commonly associated with positively valenced emotions (Crowder, 1984; Costa et al., 2004).

The connection between emotion and musical mode is well established (Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005; Quinto, Thompson, & Keating, 2013; Webster & Weir, 2005), showing major-minor distinctions are useful predictors of emotions such as happiness and sadness (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990). The impact of mode is so strong it can shape emotional responses more so than pitch or timing (Hevner, 1935, 1937). However the relative contributes of mode and tempo are complex (Gagnon & Peretz, 2003; Juslin & Lindström, 2010).

Although powerful, mode is a culture-specific cue that requires learning (Corrigall & Trainor, 2014). Meyer's (1956) proposed theory of deviations highlights the idea that relationships between major and minor keys stem from expectations of regular and normative melodic progressions. In this regard, the associations and regularities must be internalized in order to form implicit and explicit musical expectations. Mode's power requires exposure: before the age of five, children are unable to identify this relationship between short melodies and emotional faces (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990).

Pitch as a cue for emotional expression

Emotion can also be conveyed through the perceptual property known as pitch — the subjective "highness" or "lowness" of a tone. Despite its clear role in speech (Bachorowski & Owren, 1995; Breitenstein, Lancker, & Daum, 2001; Scherer, 1995), it's musical role is less straight-forward. Pieces in higher octaves are generally found to be associated with more positive emotional adjectives such as happy, glad, and dreamy when assessing pairs of pitches (Eitan & Timmers, 2010), scales (Collier & Hubbard, 2001), commercially recorded works (Gundlach, 1935; Watson, 1942; Wedin, 1972), and transposed compositions (Hevner, 1937). Conversely, lower octaves are associated with negative emotions such as sad, agitated and somber (Gundlach, 1935; Hevner, 1937; Scherer & Oshinsky, 1977; Watson, 1942; Wedin, 1972).

However, research on discrete emotions provides a different perspective. For example, high pitches are in some cases associated with negative emotions, as well as low pitches with positive emotions (Ilie & Thompson, 2006; Scherer & Oshinsky, 1977). Secondly, pitch information (specifically pitch range) does not emerge as a strong predictor of listeners' ratings of target emotions across different musical cultures—although other cues do seem to translate (Balkwill & Thompson, 1999). Those authors suggest this may have occurred given pitch range plays an important role in expectancy, which can be generalized to emotional arousal, rather than specific emotions. Thus, the pitch information contained in Hindustani ragas did not provide useful information for listeners to interpret a specific, discrete emotion.

Research using the dimensional perspective of emotion also raises questions about pitch height's role. High-pitched music has been associated with both high and low-arousal emotional terms; listeners are found to associate high pitch with anger and fear (Scherer & Oshinsky, 1977; Wedin, 1972) in addition to affective adjectives representing low arousal states such as graceful and serene (Hevner, 1937). Musical stimuli lower in pitch have been associated with sadness and boredom (Hevner, 1937; Scherer & Oshinsky, 1977), but also with affective adjectives such as excitement and agitation (Hevner, 1937; Rigg, 1940).

Although a body of research suggest pitch height plays a role in musical emotion, its relationship appears less clear than cues such as tempo (Gabrielsson & Lindström, 2010) and mode. The varying effects of pitch may emerge in part, from the range of stimuli used within experiments. Differences may occur not only as a result of the increased complexity of polyphony (Ilie & Thompson, 2006) over monophony (Balkwill & Thompson, 1999), but also with respect to performed versus synthesized and manipulated (Scherer & Oshinsky, 1977) musical stimuli. Monophonic and experimental 'controlled' stimuli are often used for studies of the cue-response relationships, therefore more work on the natural use of cues will shed light on the complex relationships between cues and listener perceptions of musical emotion.

Measuring Emotional Communication

Assessments of musical emotion involve both discrete and dimensional models. Discrete models function as forced-choice paradigms based on the framework of Ekman's (1992) theory of basic emotions. These models assume a limited number of fundamental emotions such as anger, joy, sadness, fear, etc., derivative of biologically determined emotional responses (Borod, 2000). Experimental procedures utilizing discrete emotional models often require participants to select which discrete emotion is represented (Laukka et al., 2013).

Although discrete models facilitate paradigms involving recognition, they restrict the range of more complex, but recognizable emotions (Eerola & Vuoskoski, 2013).

In contrast, the dimensional model of emotion can offer more reliable measurement with emotionally ambiguous stimuli (Eerola & Vuoskoski, 2010). For example, Russell's (1980) popular circumplex model organizes emotional responses into two dimensions — valence and arousal. In this framework, valence represents the intrinsic positive or negative component of emotion and arousal represents the intensity or energy of the emotion. A number of studies have harnessed this view's utility in music (Wedin, 1969, 1972a, 1972b). In these studies, factor analyses on the semantic contents of adjectives or words listeners associated to musical excerpts indicated arousal and emotional valence emerge as the two main dimensions.

Two dimensional models can account for a large proportion of variance (Schubert, 1999), however the standard dimensions of valence and arousal alone fail to fully explain responses (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005) leading to interest in alternatives. For example, Schimmack and Grob (2000) argue that the ambiguous definition of arousal introduces confusion, which can be interpreted as either an awake-tired or tense-relaxed state. As such many studies explore variations from the standard dimensional model, using labels such as tension (Ilie & Thompson, 2006a), activity (Leman, Vermeulen, De Voogdt, Moelants, & Lesafre, 2005) and strength (Luck et al., 2008).

Consequently, here we assess emotion using different dimensional labels in order to contribute to ongoing discussions on this contested topic.

Reflections on stimuli used to explore emotion

Several studies use polyphonic musical examples, such as one drawing upon stimuli chosen to represent specific quadrants of the circumplex model (Dibben, 2004). Others focus on film soundtracks designed to stir up emotion (Vuoskoski & Eerola, 2011), offering insight into the processing of highly emotional musical experiences. However, the popularity and familiarity of this music introduce interesting challenges to interpreting results. Participants may be familiar with certain pieces of film music, having formed pre-existing associations with moments in the film, influencing their responses to the music. Furthermore, pieces from film soundtracks can contain sounds from multiple instruments in an orchestra or band, which introduce another layer of complexity (incorporating different timbres, pitch information, etc.).

The growing field of Music Information Retrieval (MIR) also extends the literature of perceived emotions in music by extracting features from stimuli to determine which predict emotion ratings. This approach has led to useful insight on a wide range of stimuli, such as polyphonic ringtones (Friberg, Schoonderwaldt, Hedblad, Fabiani, & Elowsson, 2014), film soundtracks (Eerola, 2011) and pop music (Yang & Chen, 2012). For example, Korhonen, Clausi, and Ed Jernigan (2006), used five excerpts of a Western art music styles, collecting the continual emotional appraisals for dimensions of valence and arousal. The authors used the overall median emotional appraisal across the response time-series to represent each piece in their analyses. They then created models of the emotional content for each piece as a function of time and musical features extracted from excerpts. As music is a time and potentially emotion varying stimuli, these time-series approaches can prove powerful tools for exploration. At the same time, requiring participants to provide continuous responses affects participants' cognitive load, potentially affecting their emotional responses. Additionally, although stimuli used in MIR research on this topic is rooted in naturalistic music listening, a large proportion focus on either pop music or soundtrack music containing multiple instruments. The sheer complexity of these naturalistic examples complicates efforts to draw strong conclusions about specific musical cues. Finally, the degree to which automated analyses accurately reflect the structural cues recognized as significant by music theorists is in itself an open question (Byrd & Crawford, 2002). Consequently, additional work is needed explore conveyed emotion in other musically polyphonic styles, and assessment of the effectiveness of feature extraction compared to score based cue quantification is crucial.

In order to provide a more focused perspective on the specific cues communicating emotional information, researchers often turn to monophonic (single-line) melodies affording rigorous quantification (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013). Others have turned to stimuli designed or composed to depict discrete emotions (Balkwill & Thompson, 1999; Hailstone et al., 2009). These approaches avoid the problems inherent with more naturalistic approaches such as studies of film music and/or MIR based analyses of large corpora of popular music, by offering precise control of multiple parameters. However, they are far removed from the types of music that so powerfully evoke strong emotions — such as the sounds heard in concert halls, home stereo systems, and personal listening devices. In addition, experimental designs independently manipulating cues such as pitch and timing to avoid confounds overlooks the powerful cumulative effects of the ways in which great composers chose to co-vary certain cues (Schutz, 2017).

Previous work has offered useful insight into musical emotion utilizing naturalistic stimuli with considerable variation on many dimensions, or tightly controlled stimuli with controlled manipulations. Here we aim to fill a gap between these approaches by exploring perceptual consequences of specific cues in unaltered renditions of widely performed and studied music. In order to identify the independent contributions of "natural" cues lacking independence, we drew on our team's previous extensive analysis and encoding of cues such as pitch timing and mode, as well as the technique of commonality analysis, or variance partitioning, on regression modeling. This provides novel insight into the unique and shared contributions of co-varying cues as manipulated by a renowned composer, offering useful new insight into how they work together to convey emotion.

The present study

Here we assess the relationship between musical structure and emotion perception of unaltered music written by a historically distinguished composer, J.S Bach (1685-1705). Building upon previous approaches manipulating cues such as pitch and timing, we explored the degree to which Bach's choices of mode, pitch and timing affect listeners' emotional responses to complex polyphonic music routinely performed and enjoyed in a wide variety of musical settings. Specifically, we used J.S Bach's well known *Well-Tempered Clavier (WTC) Book 1* as performed by Friedrich Gulda (Bach, 1973). Our approach complements and extends previous targeted explorations of manipulations to individual cues by exploring the perceptual consequences of the ways in which Bach naturally co-varied their use in a set of pieces still widely performed and studied. This preserves the musical complexity often experienced by listeners, offering on opportunity to assess generalizability of previous research on monophonic, or experimentally designed acoustic stimuli, as well as previous studies of emotional excerpts that likely came with extra-musical associations (i.e. film scores, popular music excerpts, etc.).

Given the significance of mode (Dalla Bella et al., 2001; Heinlein, 1928; Quinto et al., 2013; Juslin & Lindström, 2010), we wanted to base this exploration on a "balanced" set of major and minor key pieces. This proved surprisingly difficult, as western music is overwhelmingly written in major keys. Classical composers such as Haydn and Mozart display a bias towards the major mode

(Tan, Pfordresher, & Harré, 2010) which can also be found in both jazz (Broze & Shanahan, 2013) and rock (Temperley & de Clercq, 2013). As such, Bach's *WTC* is ideally suited for this exploration and offers a naturally balanced set of pieces with one Prelude and one Fugue in each major and minor key.

Emotional responses to the stimuli were encoded using a dimensional model in order to account for the complexity and richness of emotional affect within this set of pieces. We adapted Russell's (1980) circumplex model of emotion to represent the emotional space with scales of valence and arousal. For comparison, we tested two versions — one incorporating dimensions of valence and intensity (experiment 1 and 2), and another with dimensions of valence and arousal (experiment 3). In order to generalize our results broadly, we chose to use participants with minimal musical training. Although previous research indicates non-musicians and musicians may perceive emotional connotations in music similarly (Bigand et al., 2005; Juslin & Laukka, 2003), those of untrained participants allowed us to establish a consistent baseline which could be expanded upon in future research.

This study had two primary aims. First to determine the relationship between timing, mode, and pitch on the perception of emotion, as they naturally vary in an ecologically valid polyphonic stimulus. Second, to determine the validity of an alternative affective dimension, intensity, in lieu of Russell's (1980) dimension of arousal. Our hypotheses included predictions that (1) timing, mode and pitch cues will predict listener ratings of emotion (2) musical mode will increase in its importance within musically 'resolved' excerpts (excerpts cut to end in the piece's starting nominal key) and (3) cues will vary to the extent they are important across valence and intensity/arousal. For our second aim, we predicted that listener responses of emotional intensity (experiment 1) for perceived emotions would not be significantly different than ratings of perceived emotional arousal (experiment 2).

Experiment 1 (Intensity)

Method

Participants. We recruited thirty non-musicians (< 1 year of musical training) undergraduates (12 males, M= 19.7 years, SD=2.87, 18 females, M= 19.1 years, SD=3.00) from the McMaster University Psychology participant pool who reported normal hearing and normal or corrected-to-normal vision. The experiment met ethics standards according to the McMaster University Research Ethics Board. Participants received course credit in return for participation.

Musical Stimuli. Experimental stimuli consisted of audio recordings of J.S Bach's *WTC (Book 1)* as performed by Friedrich Gulda (n=48). Excerpts contained the first eight measures of each piece, with a two-second fade out starting at the ninth musical measure. Although faster and slower pieces varied in duration, this approach provided consistency in terms of musical units (measure length). Stimuli lasted 7-64 seconds in duration (M=30.2 seconds, SD=13.6). We prepared all excerpts using Amadeus Pro.

Cue Quantification. Beyond encoding the modality indicated in each piece's key signature, our analysis required quantifying two additional cues — pitch height and timing. We calculated pitch height using methods based upon Huron, Yim, and Chordia (2010), and later extended by Poon and Schutz (2015), to weight notes according to their duration (similar to other music, these pieces included both long and short notes). In this approach, pitch height is calculated by summing duration-weighted pitch values within each measure, then dividing by the sum of note durations within that measure. Previously, Poon and Schutz (2015) used this method to calculate theoretical averages for the first eight measures of each of the 48 pieces using tempi noted in a score. Here we used that approach as a point of departure, adjusting the tempi used in the calculations to reflect those in the stimuli played for participants. We also re-calculated information as needed for experiment 2, which involved excerpts of variable lengths rather than the 8 measure excerpts used in experiments 1 and 3. This ensured that all attack rate information used for comparison in each experiment corresponded to the stimuli heard by participants. Additional technical details on pitch and timing quantification methods are available in Poon and Schutz (2015) — including a figure annotating the exact pitch and timing values assigned to each note in the first measure of the C Major Prelude. We used this approach to calculate musical attack rate in part to allow for parallel comparisons of timing in speech, specifically with articulation rate (Johnstone & Scherer, 2000; Scherer, 2003).

Pitch height values varied from 33.13-53.00 (M=43.90, SD=4.03) corresponding ~F3 to ~ C#5, attack rate information for eight measure excerpts range 1.3-10.13 attacks per second (M=4.91, SD=2.18). We operationalized mode as the tonal center of the piece, as indicated by the denoted key signature of each score, coded dichotomously (0=minor, 1=Major). Admittedly, nominally minor excerpts in our experiment contained some major chords and vice-versa, making for a less controlled treatment of mode than monophonic excerpts created to be either unambiguously major or minor. Nonetheless, this is entirely in keeping with normative practice in musical composition, where harmonic progressions typically include both major and minor chords. As each of these pieces starts in the nominal key, we believe it is a reasonable way to explore mode as it is experienced in concert halls and on recordings—rather than the more controlled (but uncommon in natural practice) approaches found in psychological experiments.

Design and procedure. Participants first completed a consent form and musical experience survey (Appendix A)**,** then entered a sound-attenuating booth where the research assistant verbally instructed participants on the rating task. After each excerpt, participants rated two aspects of perceived emotion, using scales for valence and intensity. Instructions emphasized the full use of each scale displayed. Research assistants told participants they would be asked to provide ratings of emotion based on what the music conveys on two scales after listening to each piano excerpt. They described valence as how positive or negative the emotion sounds, ranging on a scale from 1 (negative) to 7 (positive). Intensity referred to the 'energy' of the emotion where high intensity pieces may sound excited or agitated, and low arousal pieces may sound dull or calm. The scale of intensity ranged from 1 (low intensity) to 100 (high intensity). We asked participants to make ratings based on emotion conveyed, rather than inquiring about emotions evoked (Gabrielsson, 2002). After hearing these instructions, participants completed 4 practice trials with alternate recordings not used in testing trials performed by Rosalyn Tureck (Bach, 1953), where they could ask the research assistant for procedural clarifications. We conducted the experiment using PsychoPy (Peirce, 2019), a Python-based psychology program and presented the experiment on a DELL monitor. Participants listened to the stimuli at a consistent and comfortable listening level through two Gateway 2000 speakers placed on either side of the computer monitor in a sound-attenuating booth (IAC Acoustics, Winchester, US). Each participant heard an individually randomized order of the 48 excerpts and provided responses via an Apple mouse connected to a 13-inch MacBook Pro located outside the booth.

Results

Visualizing participant data on Russell's the two-dimensional circumplex provides a useful first step to understanding emotional responses in these stimuli. Figure 1a shows ratings for the first experiment, illustrating minor key pieces received lower valence ratings than major for both Preludes (left column) and Fugues (right column). In fact, of the 24 Preludes, only one Major piece (B Major) fell in the lower half of valence ratings. Of the 24 Fugues, only one (C Major) clearly fell in the lower half of valence ratings (B Major and d minor Fugues tied for the 12th lowest valence rating). This is consistent with previous research indicating mode's strong effect on emotion and suggests our treatment of mode as a binary variable based on the nominal key of each piece provides a useful framework for understanding the emotional messages conveyed. However, as shown by Poon and Schutz (2015), composers co-vary cues in normative musical practice, making it difficult to understand the ultimate reason for this putative effect of mode. To explore this issue further, we turned to three separate statistical analyses. These both provide different perspectives on interpreting the data, as well as useful points of comparison with a rich literature on emotional communication in both speech and music.

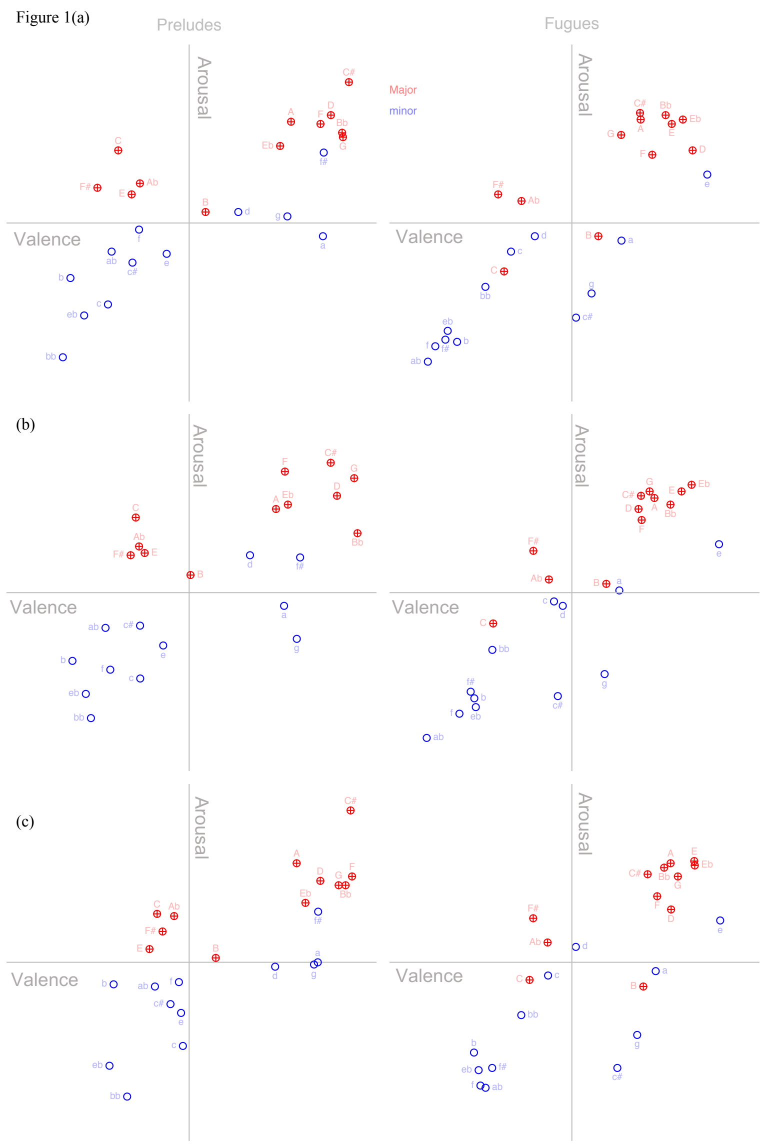


Figure 1. Mean ratings for all 48 pieces in the *WTC* (separated by preludes and fugues) across the 2D circumplex space for (a) Experiment 1, (b) Experiment 2 & (c) Experiment 3. Major key pieces are represented in red, minor key pieces in blue.

In order to clarify cue contributions, we assessed participant ratings from three perspectives. First, we examined Pearson product-moment and Pearson point bi-serial correlations between the three acoustic cues (attack rate, mode, pitch height) and the mean ratings from the two dimensions of response (valence, intensity). Second, we assessed the relationship between acoustic cues (attack rate, mode, pitch) and mean values of listener responses, as captured by a two-dimensional model of valence and intensity, using a least squares standard multiple linear regression on the mean ratings. Third, we further assessed relative cue contributions with commonality analyses to determine partitioned variance within the regression models. We determined Cronbach's alpha for listener ratings across 48 excerpts to be $\alpha=0.98$ for both valence and intensity ratings, suggesting participant ratings are highly consistent. Valence ratings ranged from 1.90 - 6.13 (M=4.12 SD=1.20) and intensity ratings ranged 21.00 – 82.20 (M=51.52, SD=19.69).

Correlations. Within the three acoustic cues, we found a significant correlation only between attack rate and mode ($r(46)=0.431$, $p<0.01$). Pitch height correlated significantly with neither attack rate ($r(46)=-0.138$, $p=0.350$) nor mode ($r(46)=0.142$, $p=0.334$). Independent-samples t-tests reveal significant differences in attack rate ($t(46)=-3.2419$, $p<0.05$), but not pitch height ($t(45)=-0.9758$, $p=0.33$) between major and minor key pieces. This is consistent with finding a significant correlation between mode and attack rate, but a lack of significant correlation between mode and pitch height. Within participant ratings, we found a positive correlation between the mean ratings of valence and intensity

($r(46)=0.795$, $p<0.001$), indicating the dimensions of the standard two dimensional model did not function independently in this context.

Exploring the relationship between acoustic cues and participant ratings, both attack rate ($r(46)=0.705$, $p<0.001$) and mode ($r(46)=0.762$, $p<0.001$) correlated with mean valence ratings. This is consistent with the visualization in Figure 1 suggesting mode plays a strong role in explaining valence ratings. Similarly, both attack rate ($r(46)=0.708$, $p<0.001$) and mode ($r(46)=0.442$, $p<0.002$) correlated with mean intensity ratings. In contrast pitch height did not play a meaningful role, as it correlated with neither valence ($r(46)=0.172$, $p=0.243$) nor intensity ($r(46)=-0.076$, $p=0.606$) ratings. This analysis suggests that emotional responses are affected by only by timing and mode, with minimal role of pitch height. This outcome is helpful in drawing contrasts between previous work on the perceptual consequences of pitch and timing on emotional speech (Breitenstein, Van Lancker, & Daum, 2001). However, correlations amongst the cues themselves (e.g., timing correlates with mode), which are likely common in music written for artistic purposes complicates interpretation of simple correlations between cues and ratings. Consequently, we turned to additional analyses to better understand what cues predict listener ratings of emotions, as well as the specific contributions of individual cues to participant responses.

Linear Regression Analysis. We ran a standard linear multiple regression analyses on normalized predictor values using the R Statistical Package to assess predictors of mean ratings of valence and intensity. We chose the Major mode as the reference level for mode, where the remaining level of the categorical variable (minor mode) is contrasted against it in analysis. The regression analysis revealed all three acoustic cues — attack rate, mode, and pitch height — significantly predicted ratings of valence (Table 1). In contrast, only attack rate predicted ratings of intensity (Table 1). This approach illustrates two important insights beyond those available from the correlations alone. First, when examined with this more nuanced assessment, mode does not predict intensity ratings. Although it correlated with intensity ratings in our first analysis, the linear regression suggests its contribution stems from its correlation with attack rate. Conversely, although we did not find a simple correlation between pitch height and valence ratings in our first analysis, it did serve as a significant predictor here.

Overall, the 3-cue predictor model accounted for 77% of the variance in ratings of valence (adjusted R^2=0.765) $F(3, 44) = 52.13$, p<0.001 and 49% of the variance in ratings of intensity, (adjusted R^2=0.492) $F(3, 44) = 16.2$, p<0.001. Tolerance and variance inflation factor (VIF) values indicate no issue of multicollinearity despite moderate correlation (r=0.431, p=0.002) between attack rate and mode (Attack rate, Tolerance=0.773, VIF=1.293; mode, Tolerance= .772, VIF= 1.295). The inclusion of interaction effects increased overall model predictability by a small amount for valence (adjusted R^2=0.771) $F(7,40)$=23.55, p<0.001) and intensity (adjusted R^2=0.494), $F(7,40)$= 7.553, p<0.001), suggesting cues functioned in an additive manner (See Appendix B).

Table 1.

Regression model for normalized attack rate, mode, pitch height and valence and intensity ratings (experiment 1). Beta values indicate strength and direction of relationship between each predictor variable and valence and intensity ratings. Default state for mode is Major.

	Valence				Intensity			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.5031	0.0435	6.264	$p<0.001$	0.6329	1.0479	5.356	$p<0.001$
Mode	-0.5212	0.1911	-6.485	$p<0.001$	-0.1708	4.6064	-1.445	$p=0.156$
Pitch Height	0.0215	0.0215	2.277	$p<0.01$	-0.0134	0. 5171	-0.124	$p=0.902$

Commonality Analysis. Finally, in order to more fully understand the overall contributions of each cue, we used commonality analysis to decompose the R^2 of each model. This technique affords examination of contributions of both unique and shared variance for each of our predictors[1] (Table 2 & 3). Here, 'shared' variance between predictors (overlapping regions in Figure 2) represent the variance those variables have in common with the dependent variable (Ray-Mukherjee et al., 2014). The presence of negative commonalities occurs when correlations among predictor variables have opposite signs (Pedhazur, 1997), or in

[1] Commonality analysis allows for reporting on the multivariate relationships between predictors beyond beta values, however does not address potential interaction effects within the model

the case that a variable confounds the explained variance of another variable in the model (Capraro & Capraro, 2001), such as a suppressor variable. Suppressor variables remove error variance in other predictors. As a result, the variable 'suppresses' irrelevant variance and increases the predictive ability of the other predictor and regression model overall (Cohen & Cohen, 1983; Capraro & Capraro, 2001).

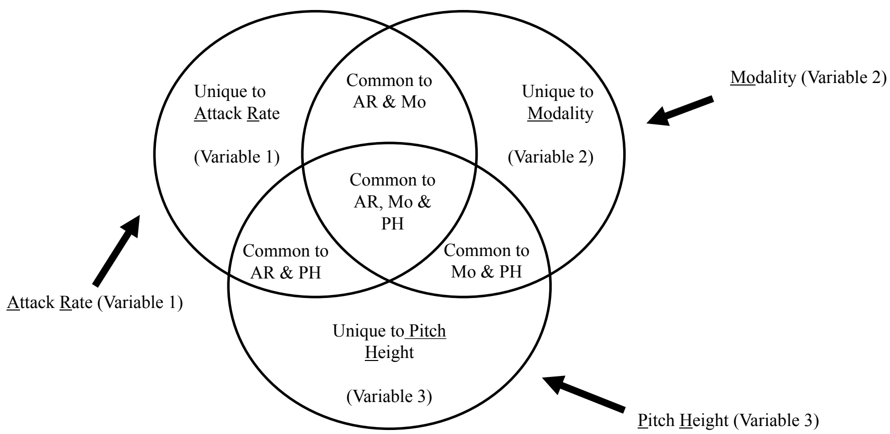


Figure 2. Visual representation of predictor relationships using commonality analysis as used here, adapted from original by Capraro & Capraro (2001).

Cue Contributions. To further explore the relative strengths of each cue, we examined their unique and shared contributions to predictions of participant response (Figures 3 & 4) using commonality analysis. Uniquely, attack rate accounted for the largest amount of variance within both valence (25%) and intensity (59%) ratings. Mode uniquely accounted for 27% of variance within valence ratings, but only 4% in intensity ratings. Pitch height uniquely accounted for 3% of variance in valence ratings but did not meaningfully contribute (<1%) to the intensity model.

Shared variance accounted for a total of 45% of valence and 37% of intensity ratings, with the largest contribution from the relationship between attack rate and mode (44% contributed to valence model, 36% to intensity model). Mode and pitch height contributed 6% of shared variance to the ratings of valence (Table 2) but did not to contribute to ratings of intensity (Table 3). Attack rate and pitch height accounted for -3% of shared valence variance in contrast to 4% of shared intensity variance. Variance common between all three cues explained -3% and -3% of variance in the valence and intensity models respectively. Some researchers interpret negative commonalities as indicating confounding suppression effects (Beaton 1973), whereas others postulate this suggests the predictor of interest has no influence (Frederick 1999). Capraro and Capraro (2001) caution the interpretation of negative values for variance common to all predictors: they argue a negative commonality value for all cues combined suggests an inverse relationship to the dependent variable, in contrast to the direct relationships found for individual predictors. As this represents the first application of commonality analysis to the study of music, for our purposes we believe it best to follow the latter approach and focus on cues with positive values.

Table 2.

Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 1)

		$R^2_{y.123}$ = .7655	% Explained Variance
Unique to X_1	Attack Rate	.1958	25.09%
Unique to X_2	Modality	.2099	26.89%
Unique to X_3	Pitch Height	.0259	3.31%
Common to X_1 and X_2	C (AR, Mo)	.3453	44.25%
Common to X_1 and X_3	C (AR, PH)	-.0218	-2.79%
Common to X_2 and X_3	C (Mo, PH)	.0477	6.12%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0224	-2.87%
	Totals	7655	100%

Table 3.

Commonality Analysis for Variance in Listener Ratings of Intensity (Experiment 1)

		$R^2_{y.123}$ = .4939	% Explained Variance
Unique to X_1	Attack Rate	.3098	59.03%
Unique to X_2	Modality	.0225	4.30%
Unique to X_3	Pitch Height	.0002	0.02%
Common to X_1 and X_2	C (AR, Mo)	.1867	35.57%
Common to X_1 and X_3	C (AR, PH)	.0196	3.74%
Common to X_2 and X_3	C (Mo, PH)	. 0003	0.06%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0143	-2.72%
	Totals	.4939	100%

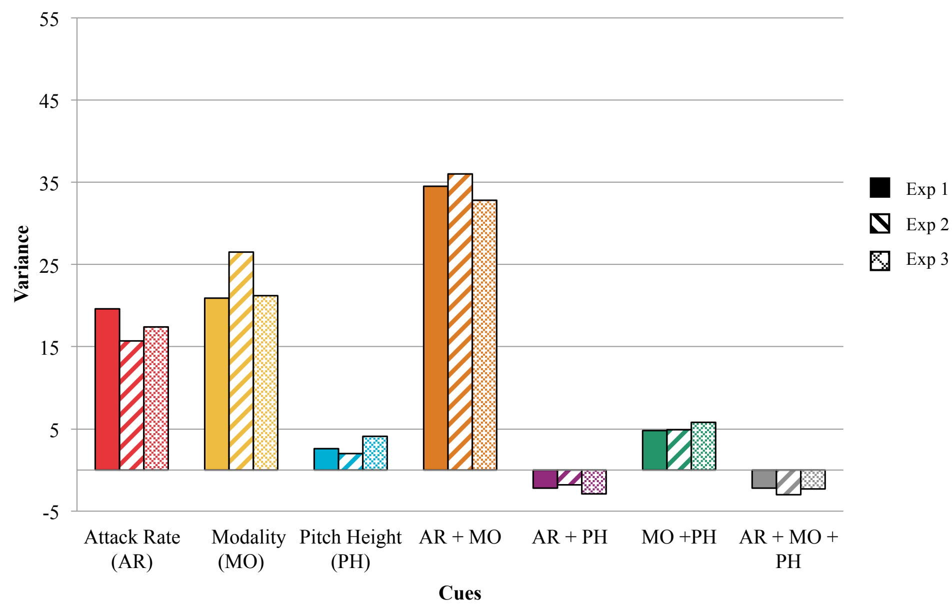


Figure 3. Unique and shared variance of valence ratings by musical cue. The unique and shared contributions of attack rate, and modality cues explained the vast majority of variance across three experiments. The three bars for each cue depict ratings made of both (1) 8 measure excerpts (experiment 1) and (2) variable length musically resolved excerpts using valence and intensity ratings (experiment 2), as well as (3) 8 measure excerpts using valence and arousal ratings (experiment 3).

Figure 4

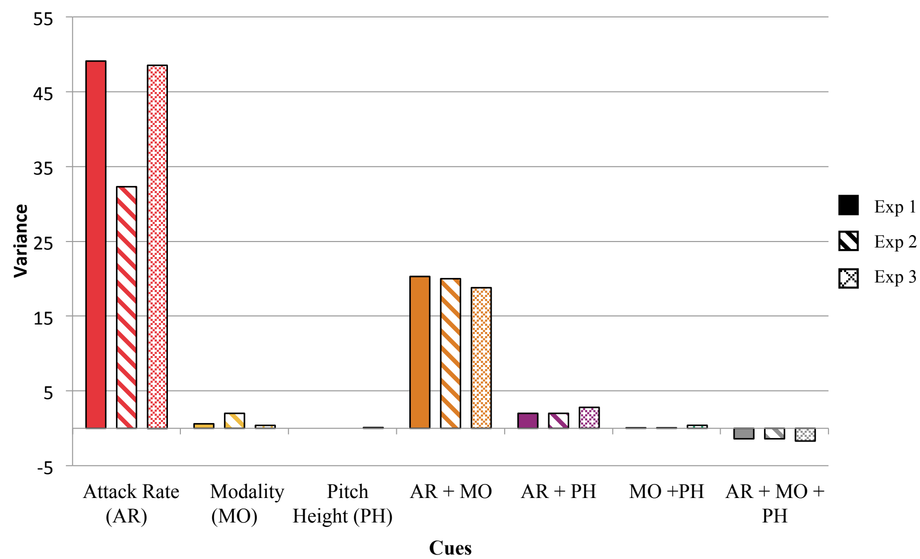


Figure 4. Unique and shared variance of Intensity/Arousal ratings by musical cue. Attack rate's unique and shared contribution with modality explained the majority of variance across perceived ratings of intensity/arousal. The three bars for each cue depict ratings made of both (1) 8 measure excerpts (experiment 1) and (2) variable length musically resolved excerpts using valence and intensity ratings (experiment 2), as well as (3) 8 measure excerpts using valence and arousal ratings (experiment 3).

Experiment 1 Discussion

Our results are consistent with previous findings in both music and speech that faster attack rates lead to higher judgments of valence and intensity, suggesting faster delivery of acoustic information may convey more positive emotions (Breitenstein et al., 2001; Juslin, 1997). In contrast to work from Ilie and Thompson (2006) and Scherer and Oshinsky (1977), we found pitch height did not correlate with valence, or intensity ratings, however appeared as a significant predictor within the three-cue regression model of valence ratings. Our analysis of the structural properties identified a correlation between mode and timing, consistent with previous findings that major key pieces tend to be faster than minor — both in these specific excerpts (Poon & Schutz, 2015) as well as more generally across a range of musical literature (Post & Huron, 2009). However, our results build on those outcomes by exploring perceptual evaluations of pieces varied in mode and timing. Additionally, they provide a useful converging measure to research using more constrained or systematically manipulated stimuli.

Attack rate significantly predicts listener ratings of both valence and intensity, indicating timing cues play an important role in both aspects of emotion. Both our linear regression and commonality analyses demonstrate timing as the most consistent predictor of emotional ratings. According to commonality analysis, attack rate uniquely predicted 25% of the total variance of valence ratings, and 59% of total intensity variance. Additionally, its shared contributions with mode predicted 44% of valence 35% of intensity variance. In contrast, pitch height contributed minimally (3% for valence, <1% for intensity). While attack rate remained the most valuable cue for ratings of intensity, mode uniquely predicted more variance of valence ratings than attack rate. This holds important implications for performer's interpretation of the musical score, for unlike pitch and mode, performers' decisions regarding tempo directly affect timing cues such as attack rate, and a review of well-known recordings of this music demonstrates considerable disagreement in tempo interpretation. For example, Palmer's (1994) review of tempi used in this set of pieces illustrates that Glenn Gould (1965) performed the E minor fugue (BWV 855) at twice the rate of Tureck (1953). Similarly, Newman (1973) performed the B minor prelude (BWV 869) at three times the rate of Gulda (1973). Finding that the cue most under control of performer interpretation plays a considerable role in emotion raises intriguing questions regarding the complex relationship between compositional structure and performer interpretation in shaping listeners' responses to musical passages.

Mode is typically regarded as an important cue for the perception of emotional valence (Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005). Our findings are to some degree consistent with this view, as depicted by plotting the mean rating of each piece across the circumplex space (Figure 1). Further, our statistical analyses illustrate that mode correlates with valence, with major key excerpts rated higher in valence and more intense. This is consistent with a large body of previous work, where major keys are commonly associated with positive valence in contrast to minor keys (Hevner 1935). Regression analyses converge with the correlational results by finding this cue significantly predicted valence ratings. However, they also illustrate that it played little role in predicting intensity ratings despite a significant correlational trend (likely a reflection of Bach's use of faster attack rates for major key pieces). According to our assessment of relative cue strength, mode functioned as the strongest cue for valence ratings, and second for intensity ratings. Uniquely, it predicted more variance associated with valence ratings (26.89%) than intensity (4.30%). This demonstrates that while mode is important for distinctions of valence, it may not be informative in the perception of emotional intensity. Furthermore, our results suggest mode's contribution to listener ratings, specifically for emotional intensity, may be a function of its relationship to cues more crucial to this emotional dimension, such as attack rate.

The selective effect of mode is particularly intriguing given disagreement over mode's significance in emotional evaluation. Although many studies of musical emotion have found it plays a powerful role (Dalla Bella et al., 2001; Hunter et al., 2008; Webster & Weir, 2005), prominent music theorists suggest its role is minimal and may be the result of its correlation with other cues aspects of musical structure (Hatten, 2004). Our results help to clarify some of the confusion over this important musical parameter by indicating within this corpus of music by a renowned composer, mode played an important role in listener perception of valence, but not intensity.

Research on music and speech suggest higher pitches correlate with positive valence (Hevner, 1937; Breitenstein et al., 2001). In contrast, here pitch height correlated with neither valence nor intensity. Furthermore, it had minimal predictive power in the commonality analysis. We suspect this difference may reflect the more complex role of pitch height in music with multiple voices and harmonic structure. Research on speech tokens often use a single voice for obvious reasons, and musical research exploring parallels often uses monophonic, or single-voiced stimuli (Hailstone et al., 2009; Lindström, 2006; Quinto, Thompson, & Keating, 2013). Although such simplified monophonic melodies provide a compelling parallel to speech, they share a tenuous connection to music that typically contains a great deal of pitch information beyond that of a single voice (i.e., polyphony, accompaniment, harmonic context, etc.).

Pitch height predicted valence ratings in the linear regression analysis (albeit to a lesser degree than other cues and cue combinations) but did not significantly predict intensity ratings. Although this might suggest some role for pitch height, our commonality analyses found it contributed minimally. Unique contributions of pitch height accounted for <1% (intensity) and 3% (valence) of listener ratings. Therefore, we conclude pitch height holds limited predictive value within this corpus of complex, polyphonic music created by a renowned composer for musical, rather than research, purposes.

In summary, our regression findings are somewhat consistent with previous work indicating the role of timing, mode and pitch in perceived emotion, however we found minimal contribution of pitch for both valence and arousal ratings. Our findings also suggest cue importance varies as a function of emotional dimension. All three cues predicted valence ratings, yet only attack rate predicted intensity ratings. Mode and pitch height served as better predictors of valence rather than intensity. These results inform previous debates on the importance of timing and mode (Gagnon & Peretz, 2003; Juslin & Lindström, 2010) suggesting timing cues (quantified as attack rate) contribute more to expressed emotion than mode. Finally, the commonality analyses suggest attack rate is the strongest contributor to both dimensions,

Previous research suggests mode and timing cues are of high importance for the perception of emotion (Eerola, Friberg, & Bresin, 2013; Gagnon & Peretz, 2003), where mode strongly predicts emotional valence. Therefore, the dominance of timing contributions in both dimensions of Experiment 1 raises an important issue: Would better control over musical key changes improve the weight of mode in listener judgements of valence? To assess this issue, we conducted an additional experiment with musically 'resolved' excerpts.

Experiment 2 (Musically resolved excerpts)

Our first experiment used eight measure excerpts for all 48 pieces. Although this approach has the benefit of consistency, some pieces modulated to different keys by the end of the excerpt (i.e. the eighth measure of the Cm Prelude outlines a C Major chord). In order to explore whether this affected mode's strength in experiment 1, we ran a second experiment using excerpts ending in the piece's nominal key (e.g. "C Major"). This required variability across stimuli length (in measures) but offers a useful complementary perspective to the strict eight measure durations of the first experiment, allowing for better insight into the relative strength of mode within this corpus of music. We then compared these responses to revised pitch and timing information corresponding to the segment evaluated—for excerpts longer than eight measures we calculated the pitch height and timing of the additional measures; for excerpts shorter than eight measures we removed the measure in question from the pitch and timing calculations used to predict responses. For example, as this experiment used an 11-measure segment of the D minor prelude, we calculated pitch and timing information for three additional measures beyond the eight calculated previously.

Method

We followed the same procedure as in the first experiment but used variable length (rather than eight measure) excerpts ending in the piece's nominal key. Participants included 30 non-musicians (<1 year musical training) undergraduate students (10 males, M=18.3 years, SD=0.67, 20 females, M=18.8 years, SD=1.02). They reported normal hearing and normal or corrected-to-normal vision. Musical stimuli ranged from 7-52 seconds (M=25.4 seconds, SD=11.0). Participants received course credit in return for participation.

Cue Quantification. Pitch and timing information corresponded the quantification of each cue within the specific number of measures required to reach a 'resolution' back to the home key for each excerpt. In these excerpts, pitch height values varied from 33.13- 53.13—corresponding ~F3 to ~ C#5—(M=43.87, SD=4.15), attack rate information ranged 1.30-10.13 attacks/second (M=4.87, SD=2.22). We mode the same way as in experiment 1 (0=minor, 1=Major).

Results

Visualizations of ratings on Russell's circumplex appear in Figure 1b for ease of comparison with previous results. Similar to the first experiment, only one Major Prelude (B Major) appeared in the lowest half of valence ratings, and only one Major key Fugue (C Major) appeared in the lowest half of valence ratings. Additionally, similar to experiment 1, the Cronbach's alpha for listener ratings appeared as α=0.97 for both valence and intensity ratings indicating high agreement across participants' ratings. Participants' valence rating ranged from 1.80 – 5.97 (M=4.12, SD=1.20) and intensity ratings ranged from 20.743- 83.93 (M=52.56, SD=18.20).

Correlations. As we recalculated pitch and timing information for these variable length excerpts, we re-ran our original analysis of the acoustic cues. Despite these changes we found a significant correlation between the cue of attack rate and mode ($r(46)=0.435$, $p<0.001$). Pitch height significantly correlated with neither attack rate ($r(46)=-0.165$, $p=0.261$) nor mode ($r(46)=0.126$, $p=0.392$). Similar to the first experiment, t-tests revealed a significant difference in attack rates ($t(46)=-3.2749$, $p<0.05$) between the major and minor key pieces, but no significant difference in pitch height ($t(45)=-0.8638$, $p=0.39$). Mean ratings of valence and intensity correlated significantly ($r(46)=0.780$, $p<0.001$), which suggests these dimensions functioned in a dependent manner.

Attack rate ($r(46)= 0.685$, $p<0.001$) and mode ($r(46)= 0.802$ $p<0.001$) significantly correlated with valence ratings. Attack rate ($r(46)=0.721$, $p<0.001$)

and mode ($r(46)=0.458$, $p<0.001$) also correlated significantly with the mean intensity ratings. In contrast, pitch height significantly correlated with neither mean ratings of valence ($r(46)=0.147$, $p=0.320$) nor intensity ($r(46)=-0.088$, $p=0.551$).

Regression analysis. All three cues significantly predicted participants' valence ratings. However, only attack rate predicted intensity ratings (Table 4). In contrast to the correlational results, mode did not predict intensity ratings. We found no significant simple correlation between pitch height and ratings of valence, however it significantly predicted listener judgements of valence in our regression model.

The three–cue predictor model accounted for 80% of the variance in valence ratings (adjusted R^2 =08029), $F(3, 44)$ = 59.73.9, $p<0.001$ and 54% of the variance in intensity ratings, (adjusted R^2 =0.5452), $F(3, 44)$ = 17.58, $p<0.001$. Regression models investigating interaction effects show similar predictability in variance prediction for valence (adjusted R^2 =0.790), $F(7,40)$=26.22, $p<0.001$ and intensity (adjusted R^2 =0.5127), $F(7,40)$=8.063 $p<0.001$ (See Appendix B).

Table 4.

The regression model for normalized attack rate, mode, pitch height and valence and intensity ratings (experiment 2). Beta values indicate strength and direction of relationship between each predictor variable and valence and intensity ratings.
.

	Valence				Intensity			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.5031	0.0418	5.920	$p<0.001$	0.4540	0.9521	5.513	$p<0.001$
Mode	-0.5212	0.1831	-7.688	$p<0.001$	-0.5863	4.1721	-1.546	$p=0.129$
Pitch Height	0.1667	0.0203	2.121	$p<0.05$	0.1457	0.4633	-0.044	$p=0.965$

Cue Contributions. As shown in Figures 3 and 4 (stripped bars), attack rate and mode accounted for the largest amount of unique variance within valence 20% and 33% respectively) ratings. Attack rate remained the only important contributor of intensity ratings (58%), and mode uniquely accounted for 5% of the model's variance. Pitch height uniquely accounted for 3% of variance for valence ratings, and none for intensity ratings. Shared variance explained 45% of total valence rating and 38% of total intensity rating variance. Attack rate and mode contributed the largest proportion of shared variance to both models (45% contributed to valence and 36% to intensity ratings). Variance shared between mode and pitch height contributed 6% to the valence model but contributed less than 1% to intensity ratings. In contrast, calculations for the relationship between attack rate and pitch height produced a -2% contribution to valence and 4% to intensity ratings. The shared variance common between all three cues accounted for approximately -3- -4% of the variance in valence and intensity models.

Table 5.

Commonality Analysis for Variance in Listener Ratings of Valence (experiment 2)

		$R^2_{y.123}$ = .8029	% Explained Variance
Unique to X_1	Attack Rate	.1570	19.56%
Unique to X_2	Modality	.2648	32.99%
Unique to X_3	Pitch Height	.0202	2.51%
Common to X_1 and X_2	C (AR, Mo)	.3595	44.78%
Common to X_1 and X_3	C (AR, PH)	-.0181	-2.25%
Common to X_2 and X_3	C (Mo, PH)	. 0493	6.14%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0299	-3.72%
	Totals	.8029	100%

Table 6.

Commonality Analysis for Variance in Listener Ratings of Intensity (Experiment 2)

		$R^2_{y.123}$ = .5452	% Explained Variance
Unique to X_1	Attack Rate	.3141	57.61%
Unique to X_2	Modality	.0247	4.53%
Unique to X_3	Pitch Height	0000	0.00%
Common to X_1 and X_2	C (AR, Mo)	.1987	36.44%
Common to X_1 and X_3	C (AR, PH)	.0217	3.97%
Common to X_2 and X_3	C (Mo, PH)	.0010	0.18%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0149	-2.73%
	Totals	.5452	100%

Discussion

Similar to experiment 1, experiment 2 highlights the relationship between attack rate (timing) information and mode within listener ratings of emotion. Correlation and regression results followed the same trends as reported in experiment 1: Attack rate and mode significantly correlated with valence and intensity, whereas pitch height significantly correlated with neither. Regression analyses indicated all three cues significantly predicted listener ratings of valence; however only attack rate predicted arousal ratings. As in experiment 1, this finding suggests the salience of cues as emotional indicators differ for the two dimensions.

Experiment 2 explored whether the influence of mode would increase when using excerpts starting and ending in the same nominal key. Although our findings here broadly mirrored those of the first experiment for valence ratings, the salience of musical mode increases when excerpts 'resolve' (i.e., end in the same nominal key in which they began), with mode increasing in its predictive power and on attack rate decreasing in predictability (see Table 3). This manipulation did not affect all cues, as pitch height's contribution remained small. As such, the results of experiment 2 suggest mode's predictive power is stronger when excerpts start and end in a consistent manner. This helps clarify mode's power in complex passages containing chords outside the target mode (i.e., major chords in nominally minor keys and vice-versa)—an approach that is common in actual musical practice although complicated to rigorously assess under controlled laboratory conditions.

Experiment 3 (Arousal)

The first two experiments quantified emotion employing an adaptation of Russell's (1980) 2D circumplex model of affect, using dimensions of valence and intensity. The literature contains some disagreement over the best label for the non-valence dimension. For example, Trainor and Schmidt (2001) use 'intensity', whereas 'arousal' is more common in other models (Russell 1980; Schubert, 2004). As 'intensity' is also used to describe the power, or physical characteristic of sound, it is possible participants might have confused emotional intensity with sound intensity in our first two experiments. Therefore, for the sake of thoroughness we ran a third experiment following the procedure and stimuli used in experiment 1 but labeling the ratings scales valence and arousal rather than valence and intensity. This afforded exploration of the consequences of different approaches to labeling the dimension representative of emotional 'energy', and to ensure listeners' understanding of the 'energy' dimension in the first experiment had not been conflated with sound intensity.

Method

We used an experimental procedure and cue quantification methods identical to the first (matching excerpt length at eight measures); however here participants rated perceived emotion on a scale of valence and a scale of arousal (rather than valence and intensity). In addition, cue quantification values remained identical to values calculated for experiment 1. Although we used the label of 'arousal' for the second dimension, the scale explanation given to participants remained identical to that of the 'intensity' scale in experiment 1. Participants included 30 undergraduate non-musicians (<1 year musical training) students (9 males, M=20.8 years, SD=2.96, 21 females, M=21.2 years, SD=3.96, with reported normal hearing and normal or corrected-to-normal vision. Musical stimuli ranged from 7-64 seconds (M=30.2 seconds, SD=13.6). Participants received course credit in return for participation.

Results

Visualizations of ratings on Russell's circumplex appear in Figure 1c for ease of comparison with previous results. Similar to the first experiment, only one Major Prelude (B Major) appeared in the lowest half of valence ratings, however two (C Major, B Major) of the Major key Fugues appeared in the lowest half of valence ratings. Cronbach's alpha for listener ratings demonstrated high internal consistency as in experiment 1 & 2 (α=0.97) for both valence and arousal ratings. Valence ratings from participants spanned from 1.97-6.30 (M=4.09, SD=1.12) and arousal ratings ranged 30.27-82.83 (M=56.99, SD=16.95).

Correlations. As the musical excerpts used in this experiment are identical to those of the first experiment, cue quantification analyses (inter-cue correlations and t-tests) are identical to those reported in experiment 1. In terms of their relationship to perceptual ratings, similar to previous experiments, attack rate ($r(46)=0.671, p<0.001$) and mode ($r(46)= 0.758$, $p<0.001$) correlated with mean valence ratings, with listeners giving higher rating to faster, major-key pieces. Attack rate ($r(46)= 0.702$, $p<0.001$) and mode ($r(46)= 0.418$, $p<0.01$) also correlated with mean intensity ratings, suggesting listeners' judgement of intensity to be higher when pieces had higher attack rates and major key structures. As in experiment 1 & 2, pitch height contrasted with the other cues, correlating with neither mean valence ($r(46)=0.219$, $p=0.134$) nor intensity ($r(46)=-0.112$, $p=0.450$) ratings.

Regression Analysis. Standard linear multiple regression analysis revealed attack rate, mode and pitch height contributed significantly towards mean valence ratings (Table 7). Analysis of arousal indicated attack rate as the only significant predictor (Table 7). Despite the correlation between mode and intensity, mode did not significantly predict mean intensity ratings in this regression analysis. However, although pitch height did not correlate significantly with valence, it significantly predicted valence ratings.

The 3-cue predictor model accounted for 48% of the variance in ratings of arousal adjusted R^2=0.4778, $F(3, 44) = 15.33$, p<0.001 and 75% of the variance in ratings of valence, adjusted R^2=0.7458, $F(3, 44) = 46.96$, p<0.001. Despite a moderate correlation between attack rate and mode (r=(46)0.445, p<0.01), tolerance and VIF values do not suggest multicollinearity (Attack rate, Tolerance=0.773, VIF=1.293; mode, Tolerance= .772,VIF= 1.295). Regression models investigating interaction effects show a small increase in variance prediction for valence (adjusted R^2 =0.7706), $F(7,40)= 23.55$, p<0.001 and arousal (adjusted R^2 =0.4847), F(7,40)= 7.317, p<0.001 models, however no interactions reached significance (See Appendix B3).

Table 7.

The regression model for normalized attack rate, mode, pitch height on valence and arousal ratings, as well as mode (Experiment 3). Beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings.

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.475	0.042	5.676	$p<0.001$	0.630	0.915	5.253	$p<0.001$
Mode	-0.524	0.186	-6.258	$p<0.001$	-0.585	4.092	-1.276	$p=0.272$
Pitch Height	0.210	0.021	2.759	$p<0.01$	0.156	0.452	-0.426	$p=0.672$

Cue Contributions. Attack rate accounted for the largest amount of unique variance within both valence (23%) and arousal (60%) ratings (Figure 3 & 4). Mode also contributed, accounting for 28% of unique valence variance, but only 4% of unique arousal variance. Uniquely, pitch height accounted for 5% of valence and less than 1% of arousal variance. The total shared variance across all cues accounted for 44 % of valence and 36% of arousal ratings, primarily from the variance shared between attack rate and mode (43% for valence ratings, 34% for intensity ratings). Mode and pitch height accounted for 8% of shared variance within the valence model and less than -1% within the arousal model. The shared variance contributed from attack rate and pitch height was -4% for valence ratings and 5% for ratings of arousal. Contribution from all three cues in variance contribution remained negative across both models, with -3 % contributed to valence and -3% to arousal.

Table 8.

Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 3)

		$R^2_{y.123}$ = .7620	% Explained Variance
Unique to X_1	Attack Rate	. 1742	22.87%
Unique to X_2	Modality	. 2119	27.80%
Unique to X_3	Pitch Height	.0412	5.40%
Common to X_1 and X_2	C (AR, Mo)	.3277	43.01%
Common to X_1 and X_3	C (AR, PH)	-.0285	-3.74%
Common to X_2 and X_3	C (Mo, PH)	.0581	7.62%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0226	-2.96%
	Totals	.7620	100%

Table 9.

Commonality Analysis for Variance in Listener Ratings of Arousal (Experiment 3)

		$R^2_{y.123}$ = .5111	% Explained Variance
Unique to X_1	Attack Rate	.3066	59.98%
Unique to X_2	Modality	.0181	3.54%
Unique to X_3	Pitch Height	.0279	0.39%
Common to X_1 and X_2	C (AR, Mo)	.1740	34.04%
Common to X_1 and X_3	C (AR, PH)	.0279	5.45%
Common to X_2 and X_3	C (Mo, PH)	-.0018	-0.35%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-.0156	-3.06%
	Totals	.5111	100%

Discussion

Our third experiment investigated the consequences of using different labels for the ‘energy’ dimension of the circumplex model of emotion. Regression (Table 7) and commonality analyses (Table 8 & 9) indicate minimal change from participant data collected in experiment 1, where participants rated the valence and intensity of perceived emotion. The intensity regression model in experiment 1 accounted for 49% (Table 3) whereas regression result for arousal ratings in experiment 3 accounted for approximately 48% (Table 9) of listener variance. This suggests both models similarly captured listener responses of perceived emotion within these musical excerpts.

Although this label change had little consequence, we felt it important to report for the sake of comparison with a wide range of existing research as both intensity (Trainor & Schmidt, 2001) and arousal (Russell 1980; Schubert, 2004) appear in the literature. We believe this is helpful in contextualizing our results, for although some studies question the effectiveness of alternative 2D models to quantify emotion (Eerola & Vuoskoski, 2013; Schimmack & Grob, 2000), models based on valence and arousal are considered standard despite disagreement over the specifics of dimensional labels. Therefore, we simply conclude that our approach captures similar aspects of the perceived emotional ‘energy’ in experiment 1 & 3 regardless of the label used for the non-valence dimension.

General Discussion

In this series of experiments, we explore the relationship between musical features and conveyed emotion using Bach's *Well Tempered Clavier (WTC)* — a prominent composition by a well-respected composer. Here we build upon previous corpus analysis of Bach's timing and pitch cues (Poon & Schutz, 2015) by empirically assessing their perceptual consequences. Complementing past work on highly emotive compositions such as film scores (Vuoskoski & Eerola, 2011) and familiar popular music (Yang & Chen, 2012), as well as tightly controlled manipulations to tone sequences (Hailstone et al., 2009; Lindström, 2006; Quinto et al., 2013), our results shed new light on the ways in which listeners respond to emotional cues when co-varying in a natural musical context. Cues such as attack rate and pitch height elicited affective consequences on listener judgements within musical stimuli in a manner complementing (though not always paralleling) those used in vocal expression. These findings are consistent with the view that music's power to communicate emotion may derive from our capacity to process parallel features in speech.

According to our model, attack rate, mode, and pitch height significantly predict ratings of valence, consistent with work documenting the effects of mode and articulation on valence (Gabrielsson & Lindström, 2001; Fabian & Schubert, 2003). Listeners also relied on attack rate (timing) cues to decode emotional intensity/arousal, common to results on speech and music (Ilie & Thompson, 2006; Schubert, 2004) in a manner previous with past findings—higher pitch heights (Bachorowski, 1999; Hevner 1937) and faster timings (Breitenstein, van Lancker, et al., 2001; Juslin, 1997) are linked with positively valenced emotions in both speech and music. Our finding that attack rates predict both intensity and arousal is also consistent with previous work on music (Vieillard et al., 2008). This demonstrates that the relationships between cues and responses within unaltered passages of ecologically valid music is in some ways consistent with research using composed monophonic and polyphonic music (Schubert 2004; Vieillard et al., 2008). Here we document how Bach wove acoustic cues such as attack rate (timing) and mode together to shape emotional messages within complex polyphonic music.

A linear model built using only three cues derived from score-based analysis accounted for approx. 49-79% of the variance in participants' ratings. Models incorporating more features such as loudness, tempo, melodic contour, texture and spectral centroid previously predicted 33-73% of perceived emotion within Romantic era music (Schubert, 2004). Our experiments employ music from a different era of musical style (Baroque), where relationships between cues such as mode and tempo differ from those in the Romantic era (Horn & Huron, 2015; Poon & Schutz, 2015). Despite differences in cue use across compositional styles, it is evident common cues such as attack rate (timing), and mode are pivotal in predicting participants' perception of emotion within music.

Our models of emotional valence predicted more variance across experiments (approx. 75-79%) than of intensity/arousal (48-51%). This contrasts with work done on modelling listener's perceived emotion, which predicts arousal better than valence (Eerola, 2012; Eerola, Lartillot, & Toiviainen, 2009; Korhonen et al., 2006; Vuoskoski & Eerola, 2011). Cross validation analyses used to compare models across various empirically tested datasets including classical, film, pop as well as mixed genre stimuli also show higher prediction rates for the perceived arousal than perceived valence both across (16% valence, 43% arousal) and within (43% valence, 62% arousal) genres (Eerola, 2011). There, systematic feature selection and principal components analysis selected nine orthogonal features covering dynamic, rhythmic, timbral and tonal aspects of the stimuli. Lower predictability for the intensity/arousal model may emerge in our results due to the lack of cues or features deemed 'expressive'. We chose to quantify only three specific cues, two of which represent structural features within the music. Previous literature has shown a number of cues to be associated with emotional arousal, such as tempo (Husain, Thompson, & Schellenberg, 2002), articulation, and loudness (Schubert, 2004) or sound intensity (Dean, Bailes, & Schubert, 2011). Perhaps with the inclusion of these additional cues, our model of intensity/arousal might be more predictive.

Although Eerola's (2011) analysis included more features, our models surprisingly demonstrated higher predictability — from essentially two cues. As mentioned above, the largest contribution occurred from the cue of attack rate, expressed as note attacks per second. Unlike that study, we extracted cues through score-based analysis, rather via the MIRtoolbox program. Thus, even for theoretically similar features such as event density — determined from the detection of onsets from the peaks evident in the amplitude envelope with respect to attack time and slope — attack rate may capture something different. Additionally, within the datasets used in Eerola (2011), 'Classical' stimuli encompassed a large mix of orchestral/ensemble recordings as well as range of Western musical styles including Baroque, Romantic, etc. Our findings reflect perceived emotion from a set of polyphonic musical examples performed on one instrument, derived from one style and one composer. Furthermore, it is important to point out model comparisons across datasets using polyphonic music indicated a genre specificity for how well features predict valence, although less so for arousal. This highlights the difference between how valence and arousal can be expressed in music, but also importance of exploring how cues function across styles of music, as core features of expressed emotion appear more effective for particular musical stimuli.

Our results also indicate the presence of interactions between musical features such as pitch height and attack rate made only small contributions (approx. 1% to models in experiment 1 & 3). Therefore consistent with previous research investigating cues in monophonic stimuli (Juslin & Lindström, 2010; Eerola & Vuoskoski, 2013)**,** the main driving effect of emotion perception appears driven by linear relationship between individual cues. However it is possible that inclusion of other features would improve predictive power for intensity/arousal.

Strength of Musical Cues

To assess relative cue strength, we used commonality analysis to calculate the unique and shared variance explained by three cues included in our model. Commonality analysis offers a powerful tool for picking apart contributions from the kinds of inter-related cues found in complex, composer created multi-voice stimuli. Variance partitioning of attack rate (timing), mode and pitch height ultimately allow us to statistically compare how much each cue contributes and gives a sense of their musical importance in this experimental context.

Timing. Attack rate remained the strongest predictor of explained variance across valence, intensity and arousal. This is consistent with research suggesting timing to be the most salient cue for emotion in music (Gagnon & Peretz, 2003), particularly for arousal (Schubert, 2004; Vieillard et al., 2008). The relationship observed between attack rate and arousal may stem from its general use in conveying information about energy. Attack rate describes the temporal rates of events, similar to rates of other behaviours such as speech, gait, etc. Thus as faster speech and walking pace suggests more energy and energy expenditure from an individual (Gomez & Danuser, 2007), the rate at which the musical structure unfolds can reflect the energy expenditure of a performer giving the performance, or the association between event rate with the other biologically important rate cues may provide listeners with information about communicated energy.

Unlike pitch and mode, attack rate represents a cue reflecting contributions from both composer and performer. It accounts for the structural decisions of the composer in the form of number of note attacks per measure, as well as the performer's choice of tempo for playing these rhythms. This suggests interesting future directions aimed at exploring the effect of different interpretation on musical communication. To some extent, the strength of timing here might reflect our use of musically untrained participants, who may have been less sensitive to mode—which requires specific musical knowledge or exposure to this type of music (Dalla Bella et al., 2001). Thus, it remains an open question whether cue weights would differ substantially amongst musically trained individuals.

Pitch height. In contrast to timing, pitch height played a smaller role (Figures 3 & 4), contributing minimally (0%-4.1% uniquely). It is possible that when hearing complex stimuli participants rely more on timing cues like attack rate than pitch height. Our use of "natural" stimuli admittedly poses challenges given the music complexity of polyphonic music created for artistic, rather than scientific purposes. However, this approach arguably assesses the role of mode in a more realistic manner, as audiences frequently encounter music with more complex uses of mode mixing chord qualities than is found in tone sequences artificially constructed to focus on one type of mode.

Music with high pitch has previously been linked to affective terms of both high and low arousal (Scherer & Oshinsky, 1977; Wedin, 1972a). Models of listener responses to Bach's *WTC* indicate pitch significantly predicted ratings of valence, but not arousal. This is consistent with cross-cultural work using monophonic Hindustani ragas, showing pitch information in the form of pitch range did not help listeners outside of the musical culture interpret specific, discrete emotions (Balkwill & Thompson, 1999), as well as work comparing studies using multi-genre polyphonic music, revealing pitch did not fall in the top ten features predicting either dimension of affective space (Eerola, 2011).

Our results provide an interesting counterpart to a previous study by Schellenberg et al. (2000) showing pitch manipulations to be more influential than rhythmic manipulations on affect perception. Their results indicate pitch as a more influential than timing when using novel, monophonic melodies performed by computers without harmonic context. Our contrary outcomes may reflect in part a different approach to timing; as our measure of attack rate considers the number of note onsets within the stimuli with respect to note durations, whereas they focus on manipulations of rhythmic structure. In addition, stimuli complexity may be a factor, as we employed the use of multi-voiced, polyphonic musical stimuli. Pitch height's importance is likely greater in the context of single-lined melodies, when there are fewer voices and musical features. Therefore, our experiment provides insight as to how pitch height functions within a natural musical context. It is possible that attack density is an important aspect within conveyed musical emotion and as such, this study produces insight into how listeners disentangle communicated emotion within non-manipulated, passages containing the natural co-variation of cues.

Mode. Across all three experiments, the use of musically resolved excerpts (experiment 2) led to the most accurate three cue model — perhaps in part due to increased predictive power of mode when ensuring excerpts started and ended in the same nominal key. Dalla Bella et al. (2001) reported that adults weighted mode more strongly than tempo, in contrast to children whose ratings seemed more reflective of tempo. Our results are inconsistent with those findings as here ratings by adults showed timing had a similar influence as mode in valence, and a much more powerful influence on arousal/intensity. However, our task differs from theirs in many ways, including the structure of stimuli. Our musical excerpts contained the kinds of complex harmonic progressions characteristic of classical music, in contrast to short melodies designed to clearly signal major vs. minor. Further research clarifying mode's role in harmonically complex passages similar to those written by great composers will help to clarify whether past findings on mode's effect may not generalize to passages of natural music with complex harmonic structure.

Most pieces in this set mixed both major and minor chords, and some begin to modulate (i.e. change their home key) within a few measures. Admittedly this makes analyses of mode difficult than in excerpts constructed to clearly articulate only major or minor keys. These distinctions matter — our resolved excerpts in experiment 2 attempted to control for these key changes, which resulted in an increase in mode's role. There are of course, limitations in our method as it is not a perfect example of a singular mode/key — modulations and/or shifts may have occurred throughout the excerpt. Additionally, our musically resolved excerpts varied in duration from excerpts used in experiment 1 & 2 (see Appendix C). As such, it is possible that stimuli duration played a role in the differences, and our conclusions should be interpreted in that light. Nonetheless, this trade-off is inevitable in evaluating music created for artistic, rather than scientific purposes. These "problematically complex" passages are more representative of the kinds of progressions moving listeners in concert halls and home stereos on a regular basis. Consequently, we see our work balancing realism and control as a helpful complement to previous research on highly controlled tone sequences. For although our findings are to some degree consistent with previous demonstrations of manipulations to single line melodies lacking harmonic context, they raise interesting questions surrounding mode's role in conveying affect within complex polyphonic compositions.

Measuring Emotion

Two-dimensional models are frequently used to quantify emotion in research (Rodà et al., 2014; Gomez & Danuser, 2004; Schubert, 1999, 2004). Using this method, emotions are broken down into components of varying degrees along the two dimensions, in contrast to discretely distinguishable categories. Russell's 2D circumplex model of affect (1980) is dominant in the field of emotion cognition and considered a standard in emotional quantification. Despite general agreement on utilizing valence to evaluate musical affect (Carroll, Yik, Russell & Barrett, 1999) researchers disagree over the best practice for additional dimension(s) (Schimmack & Grob, 2000; Vieillard et al., 2008). Previous studies use labels such as tension (Ilie & Thompson, 2006), activity (Leman et al., 2005) and/or strength (Luck et al., 2008).

To provide the most connection with the vast literature on musical emotion, our third experiment investigated the difference between various dimensions (arousal and intensity) as adapted from Russell's 2D model. Our results show each model accounted for similar amounts of variance across both dimensions. Small variations occurred between models where the valence/arousal model (Table 8 & 9) explained less of the variance within mean arousal and valence ratings than the valence/intensity model (Table 2 & 3). Overall these small differences suggest our experimental definition and use of 'intensity' captures the second or 'energetic' dimension of the 2D circumplex space, similar to 'arousal'. Given debate over the best measure of assessing emotion we believe this direct assessment of dimensional labeling is useful to note.

Concluding thoughts and broader implications

Together our experiments demonstrate the relative importance of attack rate (timing), mode, and pitch on emotional perception within a complex, composed musical corpus. This finding contributes to a growing literature on the relationships between cues and affective perception (Eerola, Friberg & Bresin, 2013**,** Dalla Bella, Peretz, Rousseau, & Gosselin, 2001), by assessing cue contributions in a corpus of renowned musical pieces performed and heard frequently in concert halls around the world. The *WTC* was developed as a teaching tool for classical musicians, and is pedagogically in frequent use helping to refine a performer's expressive skills involving aspects like articulation, tempo and phrasing (Paggioli de Carvalho 2016)**.** Thus, a selection like the *WTC* affords further opportunity to explore cues expressed in its performance.

As our study focused on a corpus of music by one particular composer, future work using a broader corpus will further explore generalizations of these findings. However, this focused exploration of such a prominent set of pieces offers a unique opportunity to explore the effects of three structural cues for emotion as encountered in a natural musical context. Applying empirical scientific methods to assess emotional encoding and decoding of acoustic cues in complex, naturalistic music contributes to understanding listener perception in an everyday context. Previous literature reinforces careful consideration of mode's role in listener perception. It's significance in conveying emotion is frequently reported (Gagnon & Peretz, 2003; Hevner, 1935; Hunter, Schellenberg, & Schimmack, 2010)**,** however our results indicate mode's role is perhaps less straight-forward in excerpts of natural music compared to musical cues such as attack rate. Intriguingly, this finding may help explain concerns voiced by music theorists that the view of major-as-happy is overly simplified and is essentially an abstraction ignoring actual compositional practice (Hatten, 2004).

The results of the present study complement the relationships between perceived emotion and musical cues in the context of naturalistic musical stimuli. While the current study focused on Bach's *Well-Tempered Clavier*, future work could address music of other genres and time periods in order to determine whether these relationships change over centuries and continents. Insight into these changes can inform a deeper understanding between musical teaching practices and cognitive outcomes on a listener. In addition, the cues used within our models consist of predominantly composer-controlled features. Therefore, future studies should also consider performer-controlled cues and performer interpretation to further disentangle the connection between encoding and decoding within musical performances.

References

Bach, J. S. (1953). *Bach: The Well-Tempered Clavier Book I [Recorded by R. Tureck]*. [DC]. RFA-Neef, Wittingen, Germay : Deutsche Grammonphon (1953). (Original work published in 1722)

Bach, J. S. (1965). Bach: The Well-Tempered Clavier, Book I -Prelude #10 in E minor, BWV 855 [CD; Recorded by G. Gould]. New York: Sony Classical. (Original work published 1722)

Bach, J. S. (1973). *Bach: The Well-Tempered Clavier, Book I [Recorded by F. Gulda]*. *Bach: The Well-Tempered Clavier, Bk I*. [CD]. MPS-Tonstudio, Villingen, Germany: Decca (1973). (Original work published in 1722)

Bach, J. S. (1971). Bach: The Well-Tempered Clavier, Book I -Prelude #10 in E minor, BWV 869 [CD; Recorded by A.Newman]. New York: Columbia Records. (Original work published 1722)

Bachorowski, J. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, *8*(2), 53–57.

Bachorowski, J. ., & Owren, M. (1995). Vocal expression of emotion: Intensity and context. *Psychological Science*, *6*(4), 219–224.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64.

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.

Borod, J. C. (2000). *The Neuropsychology of Emotion* (1st ed.). Oxford University Press.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology*, *2*, 1–16.

Breitenstein, C., Lancker, D. Van, & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57–79.

Breitenstein, C., van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57–79.

Broze, Y., & Shanahan, D. (2013). Diachronic Changes in Jazz Harmony: A Cognitive Perspective. *Music Perception: An Interdisciplinary Journal*, *3*(1), 32–45.

Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. *Information Processing & Management*, *38*(2), 249–272.

Carroll, J. M., Yik, M. S., Yik, Russell, J.A. (1999). On the psychometric principles of affect. *Review of General Psychology*, *3*(1), 14–22.

Collier, W. G., & Hubbard, T. L. (2001). Musical scales and evaluations of happiness and awkwardness: effects of pitch, direction, and scale mode. *The American Journal of Psychology*, *114*(4), 355–375.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142–158.

Costa, M., Fine, P., Enrico, P., & Bitti, P. E. R. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception: An Interdisciplinary Journal*, *22*(1), 1–14.

Crowder, R. G. (1985). Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society*, *23*(4), 314–316.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation. *PLoS ONE*, *6*(4), e18591.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception: An Interdisciplinary Journal*, *22*(1), 79-115.

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. *Journal of New Music Research*, *40*(4), 349–366.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 1–12.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*.

Eerola, T., & Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49.

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies:

Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, *30*(3), 307–340.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 162–200.

Fabian, D., & Schubert, E. (2003). Expressive devices and perceived musical character in 34 performances of Variation-7 from Bach's "Goldberg Variations." *Musicae Scientiae*, 49–71.

Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., & Elowsson, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *The Journal of the Acoustical Society of America*, *136*, 1951–1963.

Gabrielsson, A. (2002). Perceived emotion and felt emotion: same or different? *Musicae Scientiae*, *6*(1; SPI), 123–148.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). New York, NY: Oxford University Press.

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

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Gardner, H. (1993). Multiple intelligences: The theory in practice. New York: Basic Books.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception: An Interdisciplinary Journal*, *12*(3), 279–290.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2), 91–103.

Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, *7*(2), 377–387.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *The American Journal of Psychology*, *47*(4), 624–643.

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. 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.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Bloomington, IN: Indiana University Press.

Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of Comparative Psychology*, *8*(2), 101–142.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*, 103–118.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, *48*(2), 246–268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*(4), 621–630.

Horn, K., & Huron, D. (2015). On the changing use of the major and minor modes 1750–1900. *Music Theory Online*, *21*(1), 1–11.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, *22*(2), 327–352.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47–56.

Huron, D., Yim, G., & Chordia, P. (2010). The effect of pitch exposure on sadness judgments: An association between sadness and lower than normal pitch. In S. M. Demorest, S. J. Morrison, & P.S Campbell (Eds), *Proceedings of the 11th International Conference on Music Perception and Cognition* (pp. 63–66). Seattle, WA: Causal Productions.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, *20*(2), 151–171.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception: An Interdisciplinary Journal*, *23*(4), 319–330.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotion* (Vol. 2, pp. 220–235). New York: Guilford.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*(4), 383–418.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217–238.

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364.

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception: An Interdisciplinary Journal*, *17*(2), 197–221.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189–202.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302–307.

Korhonen, M. D., Clausi, D. A., & Ed Jernigan, M. (2006). Modeling Emotional Content of Music Using System Identification. *Systems, Man and Cybernetics, Part B, IEEE Transactions on Systems, Man, and Cybernetics*, *36*(3), 588–599.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesafre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, *34*(1), 39–67.

Lindström, E. (2006). Impact of melodic organization of melodic structure and emotional expression. *Musicae Scientiae*, *10*(1), 85–117.

Luck, G., Toiviainen, P., Errkkila, J., Lartillot, O., Riikkila, K., Makela, A., … Varri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, *36*(1), 25–45.

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

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*(3), 618–622.

Paggioli de Carvalho, L. (2016). Bach's Well-Tempered Clavier: Pedagogical Approaches and the Different Styles of Preludes. *Per Musi*, 97–115.

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences*, *1060*, 450–453.

Palmer, W. A. (1994). *J.S. Bach: The Well-Tempered Clavier* (Third). USA: Alfred Music Publishing.

Pedhazur, E. J. (1997). Multiple regression in behavioral research: Ex- planation and prediction (3rd ed.). Fort Worth, TX: Harcourt Brace.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, *6*, 1–13.

Post, O., & Huron, D. (2009). Western Classical music in the minor mode is slower (Except in the Romantic period). *Empirical Musicology Review*, *4*(1), 2–10.

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4*, 1–8.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328.

Rigg, M. G. (1940). Speed as a determiner of musical mood. *Journal of Experimental Psychology*, *27*(5), 566–571.

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

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Schellenberg, E. G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune musical intervals. *Music Perception: An Interdisciplinary Journal*, *19*(2), 223–248.

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception: An Interdisciplinary Journal*, *18*(2), 155–171.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346.

Schimmack, U., & Grob, A. (2000). Dimensional model of core affect: A quantitative comparison by means of structural equation modeling. *The*

European Journal of Personality, *14*, 325–345.

Schubert, E. (1999). Measuring Emotion Continuously: Validity and Reliability of the Two-Dimensional Emotion-Space. *Australian Journal of Psychology*, *51*(3), 154–165.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Schutz, M. (2017). *Acoustic constraints and musical consequences: Exploring composers' use of cues for musical emotion*. *Frontiers in Psychology*, 8.

Tan, S.L., Pfordresher, P. Q., & Harré, R. (2010). *Psychology of music: From sound to significance* (1st ed.). New York, NY: Psychology Press.

Temperley, D., & de Clercq, T. (2013). Statistical Analysis of Harmony and Melody in Rock Music. *Journal of New Music Research*, *42*, 187–204.

Trainor, L. J., & Schmidt, L. A. (2001). Processing Emotions Induced by Music. *Cognition and Emotion*, *15*(4), 487–500.

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 and Emotion*, *22*(4), 720–752.

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

Watson, K. B. (1942). The nature and measurement of musical meanings. *Psychological Monographs*, *54*(2), i-43.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39.

Wedin, L. (1969). Dimension Analysis of Emotional Expression in Music. *Swedish Journal of Musicology*, *51*, 119–140.

Wedin, L. (1972a). A multidimensional study of perceptual -emotional qualities in music. *Scandinavian Journal of Psychology*, *13*, 241–257.

Wedin, L. (1972b). Evaluation of a Three-Dimensional Model of Emotional Expression in Music. *The Psychological Laboratories*, *54*(349), 1–17.

Wiggins, G. A. (1998). Music, syntax, and the meaning of "meaning." In *In Proceedings. of First Symposium on Music and Computers (pp.* 18–23). Corfu, Greece: Ionian University.

Yang, Y.H., & Chen, H. H. (2012). Machine recognition of music emotion. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 1–30.

Appendix A

Musical Training Survey

Date: _______ (dd/mm/yy) Experiment Num: ____ Participant Num: _____

1) What is your age in years at the time of this study: _____

2) Please list your year of study at McMaster (i.e. first year undergraduate student, second year graduate student, etc). ______________________________________

3) Gender - I am (circle one): male female transgendered

4) Do you consider yourself proficient on a musical instrument, and if so, which one?

____ Yes – if so, please list instrument(s): ____________________________________

____ No

5) Have you taken private musical lessons

____ Yes – if so, for how many years: _______________________________________

____ No

6) Approximately how many hours a week do you spend practicing/performing/jamming (as opposed to listening) to music?

____ (0-1) ______ (1-5) _______ (5-10) _______ (10-15) ______ (15-20) ____ (> 20)

7) Approximately how many hours a week do you spend listening to (as opposed to practicing or performing) music?

____ (0-1) ______ (1-5) _______ (5-10) _______ (10-15) ______ (15-20) ____ (> 20)

8) Do you own an iPod or personal music listening device?

____ Yes ____ No

8a) If you answered yes to question 8, what kind of device do you own:

____

8b) If you answered “yes” to question 8, approximately how many songs do you have on your device?

____

Appendix B

Experiment 1 Regression Tables with Interaction Terms

Table B1

	Valence				Intensity			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.4799	0.0459	10.45	p<0.001	0.6678	0.0747	8.94	p<0.001
Mode	-0.3283	0.0438	-7.50	p<0.001	-0.0363	0.0711	-0.5	p=0.613
Pitch Height	0.1387	0.0553	2.51	p<0.05	0.0201	0.0899	0.22	p=0.824
AR x Mo	0.0918	0.0460	2.00	p=0.053	0.0751	0.0747	1.01	p=0.321
AR x PH	0.0679	0.0443	1.53	p=0.133	0.1856	0.0720	2.58	p<0.05
Mo x PH	-0.0160	0.0553	-0.29	p=0.774	0.1040	0.0899	1.16	p=0.254
AR x PH x Mo	0.0096	0.0443	0.22	p=0.830	0.0420	0.072	0.58	P=0.564
Adjusted R^2	0.879				0.693			

Experiment 2 Regression Table with Interaction Terms

Table B2

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.4071	0.0742	5.13	$p<0.001$	0.6341	0.1181	5.37	$p<0.001$
Mode	0.7044	0.0988	9.40	$p<0.001$	0.3029	0.1116	2.71	$p<0.01$
Pitch Height	0.0702	0.0988	0.71	$p=0.48$	0.0272	0.1471	0.185	$p=0.854$
AR x Mo	0.0226	0.0802	0.28	$p=0.779$	0.0998	0.1194	0.836	$p=0.408$
AR x PH	-0.0457	0.1015	-0.45	$p=0.655$	-0.0266	0.1152	-0.176	$p=0.861$
Mo x PH	0.0901	0.0998	0.90	$p=0.372$	0.0588	0.1486	0.396	$p=0.694$
AR x PH x Mo	0.0575	0.1023	0.56	$p=0.578$	-0.1161	0.1528	-0.760	$p=0.412$
Adjusted R^2	0.7715				0.4934			

Experiment 3 Regression Table with Interaction Terms

Table B3

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.6440	0.0689	9.36	$p<0.001$	0.8954	0.0999	8.96	$p<0.001$
Mode	0.4399	0.0656	6.70	$p<0.001$	0.0265	0.0952	0.28	$p=0.782$
Pitch Height	0.2536	0.0829	3.06	$p<0.01$	0.0226	0.1203	0.19	$p=0.852$
AR x Mo	-0.1050	0.0695	-1.52	$p=0.139$	-0.1014	0.1009	-1.00	$p=0.321$
AR x PH	0.0938	0.0670	1.40	$p=0.169$	0.1982	0.0973	2.04	$p<0.05$
Mo x PH	0.0272	0.0838	0.33	$p=0.747$	-0.0389	0.0122	-0.32	$p=0.750$
AR x PH x Mo	-0.0382	0.0677	-0.56	$p=0.576$	-0.0673	0.0983	0.684	$p=0.498$
Adjusted R^2	0.8558				0.6962			

Appendix C

Experiment Stimuli Durations (in seconds)

Table C1

Piece	Exp 1 & 3 Durations	Exp 2 Durations
Fugue 1	00:54	00:35
Fugue 2	00:34	00:36
Fugue 3	00:21	00:17
Fugue 4	00:19	00:19
Fugue 5	00:36	00:20
Fugue 6	00:25	00:18
Fugue 7	00:22	00:21
Fugue 8	00:34	00:31
Fugue 9	00:18	00:17
Fugue 10	00:14	00:32
Fugue 11	00:10	00:14
Fugue 12	00:46	00:51
Fugue 13	00:35	00:28
Fugue 14	00:52	00:52
Fugue 15	00:16	00:28
Fugue 16	00:39	00:29
Fugue 17	00:41	00:43
Fugue 18	01:03	00:34
Fugue 19	00:17	00:17
Fugue 20	00:25	00:23
Fugue 21	00:15	00:20
Fugue 22	00:23	00:26
Fugue 23	00:49	00:49
Fugue 24	00:54	00:43
Prelude 1	00:30	00:14
Prelude 2	00:33	00:15
Prelude 3	00:08	00:08
Prelude 4	00:25	00:30
Prelude 5	00:17	00:24
Prelude 6	00:26	00:33
Prelude 7	00:25	00:30
Prelude 8	00:52	00:24
Prelude 9	00:30	00:27
Prelude 10	00:38	00:17
Prelude 11	00:27	00:07
Prelude 12	00:37	00:10
Prelude 13	00:29	00:29

Prelude 14	00:19	00:26
Prelude 15	00:24	00:30
Prelude 16	00:40	00:14
Prelude 17	00:20	00:21
Prelude 18	00:29	00:12
Prelude 19	00:19	00:18
Prelude 20	00:23	00:12
Prelude 21	00:24	00:28
Prelude 22	01:05	00:50
Prelude 23	00:27	00:20
Prelude 24	00:44	00:22

Correlations between Stimuli Durations and Listener Ratings

Table C2

	Valence Ratings		Intensity/Arousal Ratings	
Experiment	*r* Value	*p* Value	*r* Value	*p* Value
1	-0.64	<0.001	-0.61	<0.001
2	-0.42	<0.01	-0.38	<0.01
3	-0.62	<0.001	-0.63	<0.001

Author Note

Aimee Battcock, Department of Psychology, Neuroscience & Behaviour, McMaster University; Michael Schutz, Department of Music, McMaster University.

This research is supported in part by grants from Social Sciences and Humanities Research Council of Canada (SSHRC and the Canadian Foundation for Innovation (CFI)

Correspondence concerning this article should be addressed to Aimee Battcock, Department of Psychology, Neuroscience and Behaviour, Psychology Building (PC), Room 102, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada.
Email: battcoae@mcmaster.ca

Chapter 3

Battcock A. & Schutz. M. (in prep). Emotion and expertise: How listeners with formal music training use cues to perceived emotion.

Preface

In Chapter 2, I found that non-musician listeners use cues of attack rate, mode and pitch height to perceive conveyed valence and attack rate when rating perceived arousal. In the present chapter, I examine cue use in listeners with musical training. In experiment 1, I show mode predicts more variance for ratings of perceived valence for listeners with six or more years of formal music lessons. In experiment 2, I show the predictive strength of mode increases for musically 'resolved' excerpts for this population, as attack rates predictive power decreases. This shift in predictive weights appears stronger for listeners with musical training as compared to untrained listeners' data collected in Chapter 2. These experiments show individuals with musical training do use cues differently to perceive emotion, specifically with the cue of mode.

Abstract

Although studies often focus on the role of the composer and performer in the communication of musical emotion, the communicative process is also influenced by the listener's musical background or experience. As a result of conflicting evidence regarding the effects of musical training, questions surrounding the role of listener expertise in understanding conveyed musical emotion remain opaque. Here we examine emotional responses of musically trained listeners across two experiments using (1) eight measure excerpts and (2) musically resolved excerpts and compared them to responses collected from untrained listeners in Battcock and Schutz (2019). In each experiment thirty participants with six or more years of music training rated perceived emotion for 48 excerpts from Bach's *WTC* using scales of valence and arousal. Models of listener ratings predicted more variance for trained vs. untrained listeners (across both experiments). Using commonality analysis, Fischer Z score comparisons and margin of error calculations, mode explained more variance when listeners had music training. We found similar results in experiment 2, where mode contributed more to models of valence ratings. Additionally, mode also had a larger increase in predictive power compared to experiment 1 for trained listeners. These results clarify musical training's impact on the specific effects of cues in conveying musical emotion.

Keywords: emotion, perception, training, expertise, individual differences

Introduction

The communication of musical emotion is both powerful and personal. Listeners can bring their individual histories to the listening experience (Ladinig & Schellenberg, 2012; Taruffi, Allen, Downing, & Heaton, 2017; Vuoskoski & Eerola, 2011), responding in different ways to the same musical information due to interactions between the musical structure and personality traits, experience and expertise. Musical training can influence individual differences in conveyed emotion, as musical training can increase sensitivity to musical structure. This can be assessed in different ways, such as exploring differences in the perception of emotional speech, perceptual responses to musical stimuli, and the neural processing of sound. However, there is ongoing debate about whether musical training can be advantageous for listeners' perceptual processing of auditory stimuli like music and speech, where some evidence suggests untrained listeners perform similarly on certain behavioural tasks (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2006).

Evidence for Training's Effect on Emotion Perception in Auditory Stimuli

Speech. Exploration on the effects of formal music training has suggested that music lessons help promote sensitivity to emotions conveyed in speech (Thompson, Schellenberg & Husain 2004). Adult participants who had at least eight years of musical training performed better than untrained adults at identifying the emotions in tone sequences constructed to mimic prosody of spoken sentences conveying different emotions. A second experiment using sentences uttered in familiar and unfamiliar vocal languages as well as tone sequences indicated musicians had a superior ability to identify the emotion conveyed in sentences presented in an unfamiliar language. For the familiar sentences and tone sequences, differences in accuracy occurred only for specific emotions including sadness and fear. Their findings support the argument that one's ability to perceive emotion in auditory information, specifically vocal prosody, can be facilitated by training in music.

Musical expertise can even aid performance in vocal prosody emotion recognition tasks (Lima & Castro, 2011). In this study, participants heard sentences with emotionally neutral statements produced in seven emotional tones (anger, disgust, fear, happiness, sadness, surprise and neutrality), and performed a forced-choice identification for the emotional tone as well as an intensity judgement. Musicians identified the appropriate emotional tone more accurately across six different emotions. This effect appeared to be long lasting, as both younger and middle-aged musicians performed the recognition more accurately. However, the authors report musicians and controls exhibited similar acoustic profiles in the musical features that predicted their responses. They argue that lack of evidence for an effect of expertise could occur as a result of the discrepancy in criteria for level of music training across participants. Studies showing positive effects of musical training use participants with more extensive training (13+ years of musical training). It is also possible that with a dimensional approach to conveyed emotion, the larger range of rating scales can pick up on more nuanced differences between ratings of experts and non-experts.

Music. In an attempt to replicate findings from Gabrielsson and Juslin (1996) on musically expressed emotion and extend on the original study, Akkermans et al. (2018) investigated the effect of musicality, emotional intelligence, and emotional contagion on how listeners decode emotion. They assessed musical expertise using the Goldsmith Musical Sophistication Index, or Gold-MSI (Müllensiefen, Gingras, Musil, & Stewart, 2014). As in the original study, classically trained musicians recorded each of three different melodies according to seven different emotional expressions. Participants heard all seven expressions for one melody four times over 28 trials, and rated excerpts on Likert scales representing the seven affective adjectives. Using hierarchical models of participants' emotion ratings, the authors found that only the musical training predictor made a significant contribution to explaining participants' decoding accuracy. This supports the argument that musical training (as measured using the musical training scale from the Gold-MSI) affords some perceptual benefits when assessing communicated emotion. It is important to note however, the calculated models of listener ratings exhibited a moderate effect size (R^2 values ranging from .04 - .53) which suggests other factors play an important role in explaining individual differences.

Additional evidence exploring the effect of musical training and age on perceptual abilities demonstrates that training benefits emerge for older musicians than younger ones when compared to matched controls (Castro & Lima, 2014). Regardless of age, all participants in this study had acquired at least eight consecutive years of formal training on their instrument and played regularly at the moment of testing. Participants rated expressed emotion of each short polyphonic excerpts on four affective 10-point intensity scales. Years of music training correlated with emotion categorization accuracy, where the middle-aged (range 40-60 years) musicians performed more accurately than non-musicians. The authors also determined participants' responses for each emotion could be predicted by various combinations of measured structural cues including tempo, mode, pitch range, dissonance, and rhythmic irregularity. The model better predicted responses given by older musicians compared to non-musicians, which authors argue may be related to training advantages in recognition accuracy. Interestingly, differences emerged in the predictive strengths of some cues for the negatively valenced emotions, supporting the hypothesis that musicians may use cues differently to decode emotion than nonexperts.

Furthermore, changes in mode and tempo affect how listeners with musical training rate perceived valence and arousal differently than those without training (Ramos, Bueno, & Bigand, 2011). In this study authors compared participants with at least six years of formal training on least one instrument to those with no experience in any study of music in an emotion recognition task.

Participants heard excerpts consisting of different modes (seven possible Greek modes selected) and tempo (three possible tempos selected) combinations and had to select one of four emotion categories representing the excerpt. The effect of tempo manipulations on participants' valence ratings was felt more by musical experts and overall, the authors determined the effect of mode had been modulated by participants' musical background for both valence and arousal ratings. The authors point out they found only slight differences, where both groups exhibited high responsiveness to the experimental manipulations. As suggested with speech evidence, it is possible that with more than six years of musical training, musicians would become increasingly more sensitive to these differences in contrast to those without musical experience.

Musical training and brain processing evidence. Musical training affects how individuals process musical information. Evidence using EEG (Koelsch, Schmidt, & Kansok, 2002; Sherwin & Sajda, 2013) and MRI (Gaser & Schlaug, 2003) techniques suggest neural differences between musical experts and nonexperts. Those with musical experience are found to have earlier and longer peaks in EEG activity for anomalous music events (AMEs) like key changes or out-of-key pitches in an instrumental stimulus. Musical experts also appeared more accurate at reporting AMEs than non-experts demonstrating a link between brain and behavioural responses. In addition, those with musical training demonstrate superior processing of musical syntax, as demonstrated by early right anterior negativity (ERAN) responses to harmonically improper chords. Larger ERAN amplitudes occurred for those with musical training when presented with Neapolitan chords not in-key. The authors argue this suggests those with musical training had more specific representations of the musical regularities, leading to greater responses when violations are presented. Differences in processing auditory stimuli may further be evident in perceptual processes including decoding conveyed emotion in music.

Evidence for Training's Lack of Effect

However some conflicting evidence regarding the effect of musical training suggests trained participants perform similarly to untrained ones in tasks assessing accuracy and categorization of examples of musical or prosody (Juslin, 1997; Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Trimmer & Cuddy, 2008). As listeners gain musical knowledge from basic listening experience, it is possible that music listening alone is sufficient to create 'experienced' listeners (Bigand & Poulin-Charronnat, 2006). Although focused on induced emotions, work from Bigand et al. (2005) found emotional responses to music only weakly influenced by musical expertise. In that study, participants grouped excerpts of instrumental Western music inducing similar emotions similarly regardless of musical background.

Interestingly, these findings occurred even though the selected stimuli included excerpts of great complexity, suggesting non-musicians are able to process subtle musical structures in Western music to discern emotion. Given the relationship between induced and perceived emotion (Hunter, Schellenberg, & Schimmack, 2010), these results add to the body of research trying to assess potential processing benefits of musical training on behavioural measures. Bigand & Poulin-Charronnat's (2006) review highlights several studies covering a range of perceptual tasks including perceived tension, abilities to anticipate musical events which also fail to find a difference or advantage for those with musical training. However, it is unclear if there are additional, more recent studies finding a lack of training effect. This may reflect a potential publication bias to publish only significant findings (Mlinarić, Horvat, & Šupak Smolčić, 2017).

Given conflicting evidence regarding musical training's effect (Akkermans et al., 2018; Castro & Lima, 2014; Koelsch, Schmidt, & Kansok, 2002; Sherwin & Sajda, 2013), or lack thereof (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Trimmer & Cuddy, 2008), the effect of musical expertise remains opaque. Some studies fail to show advantages of musical training, which may occur due to the explicit nature of the tasks used. Although the current study follows an explicit protcol, asking participants to directly evaluate valence and arousal, unlike studies providing the possible discrete affect terms, we believe the dimensional measurement of emotion is a more reliable tool for rating excerpts that are less overt in their emotional message. This method is found to be more sensitive for ambiguous emotional content in music and shows higher inter-rater consistency for listener ratings of emotion (Eerola & Vuoskoski, 2011)

Present study

The primary aim of this study is to examine the relation between musical training perceived emotion extending on past work using polyphonic stimuli (Castro & Lima, 2014). Here we employ a dimensional approach to emotion (Di Mauro, Toffalini, Grassi, & Petrini, 2018) to investigate differences between musically trained and untrained individuals. Further, our study will help clarify ongoing debate surrounding effect of musical expertise on the perception of musical structure and emotion. Here we build off of previous work exploring the relationship between mode, pitch and timing and perceived emotion (Battcock & Schutz, 2019) by using Bach's *Well-Tempered Clavier (WTC)* to explore this topic. This polyphonic work contains 48 pieces with equal numbers of major and minor modes for each key, which provides an advantage for assessing a cue of interest, as it gives us a 'balanced' set of pieces to explore the effect of mode. Additionally, it allows us to assess the perceptual consequences of a musical work that remains widely performed and studied by musicians. Using this stimulus set, we found timing information to be of great importance across emotional valence and arousal, but mode and pitch height to only play a role in perceived valence. Additionally, employing commonality analysis allowed us to further determine the relationship between timing and mode, clarifying how these cues work together to convey emotional messages.

Research exploring the influence of musical training and the perception of emotion often use discrete models, where participants rate perceived emotion on different affective adjective scales (Akkermans et al. 2018; Castro & Lima, 2014; Gabrielsson & Juslin, 1996). Although that method offers precision for the intended affective terms selected, it may exert priming effects for listeners. Unlike discrete models of emotion, the dimensional approach affords ability to represent more variation in conveyed and perceived emotion (Eerola & Vuoskoski, 2013). The ability to measure components of emotion on a fine-grained scale make dimensional models better suited for detecting differences between trained and untrained listeners. Here we will compare these rating by trained musicians with ratings in the same paradigm by 'non musician' participants with less than one year of musical training (Battcock & Schutz, 2019). We will assess these differences in two separate contexts (1) with excerpts from Bach's *WTC* cut to be eight musical measures in length (2) using musically 'resolved' excerpts where each excerpt ends in the same nominal key as it started. Similar to Castro and Lima (2014) we use linear regression modelling with three quantified cues as predictors. The cues selected — attack rate (timing), mode and pitch height — represent three musical features empirically proven to have a role in communicated musical emotion (Balkwill & Thompson, 1999; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Hevner, 1935, 1937). Here we investigate the predictive weights of cues across participants with and without musical training, to determine how expertise affects how listeners decode emotion in music.

Experiment 1 (Eight measure excerpts)

Method

The following procedure and stimuli follow that of Battcock & Schutz (2019). As that paper provides full technical details, here we summarize that approach focusing on aspects of greatest relevance.

Participants. We recruited 30 participants with ≥6 years of formal musical training from McMaster University and attendees of the Ontario Music Educators Association's General Assembly held in Hamilton, Ontario (25 females, ages M=27.36, SD= 13.69, years of training M=6.73 SD=0.45). On average, participants scored in the 71st percentile of the overall General Sophistication score and in 79th percentile on the Musical Training subscale using the Goldsmiths Musical Sophistication Index (Gold-MSI) as based on norms reported by the Müllensiefen et al. (2013). Participants either received course credit, or compensation for their participation or participated as volunteers. The experiment met ethics standards according to the McMaster University Research Ethics Board.

Musical Stimuli. Our stimuli consisted of excerpts from all 48 pieces of Bach's *Well-Tempered Clavier* (Book 1) as recorded by Friedrich Gulda (J. S. Bach, 1973). Each excerpt contained the first eight musical measures of the pieces and featured a two-second fade out starting at the ninth measure. Excerpts lasted 7-64 seconds in duration (M=30.2 seconds, SD=13.6). All excerpts had been prepared using Amadeus Pro.

***Cue Quantification*.** Pitch height information is calculated using an approach used in Huron, Yim, & Chordia (2010) and Poon & Schutz, (2015) by summing duration-weighted pitch values within each measure, divided by the sum of note durations within that measure. Attack rate calculations are based on the tempi chosen by Friedrich Gulda's in his performance of the *WTC* — the recording used for this experiment. In addition, we re-calculated information as needed for Experiment 2 (for excerpts of variable length rather than eight measures). Pitch height values varied from 33.13-53.00 (M=43.90, SD=4.03) corresponding ~F3 to ~ C#5, attack rate information for eight measure excerpts range 1.3-10.13 attacks per second (M=4.91, SD=2.18). We operationalized modality as the tonal center of the piece, as indicated by the denoted key signature of each score, coded dichotomously (0=minor, 1=Major).

Design and procedure. Experiment occurred in two locations, the Ontario Music Educators Association (OMEA) general assembly held at the Sheraton in

Hamilton, Ontario and McMaster University. Participants from the OMEA event filled out a consent form and completed the experiment in an isolated room. Following the consent form, participants from McMaster University completed the experiment in a sound-attenuating booth (IAC Acoustics, Winchester, US). For both testing locations, the experiment ran on PsychoPy (Peirce et al., 2019), a Python-based program on either a 2014 MacBook Air (OS X 10.9.4). Participants heard stimuli at a consistent and comfortable listening level through Sennheiser HDA 200 headphones and provided responses using the MacBook's trackpad.

Research assistants verbally instructed each participant to rate the perceived emotion after each excerpt using two scales: valence and arousal. The instructions explained valence as referring to how positive or negative the expressed emotion sounded, as rated on a scale from 1 (negative) to 7 (positive), arousal represented the energy of the emotion which is to be rated on a scale from 1 (low) to 100 (high). Participants had been encouraged to use to the full range of the scales and reminded to rate the emotion they heard and not the emotion they felt. Participants completed four practice trials before beginning the experiment, using recordings of the same album performed by Angela Hewitt (Bach, 1998). Each participant listened to an individually randomized order of the 48 excerpts. Following the experiment, participants completed the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014).

Results

Participants' valence ratings (M=4.20, SD=1.57) ranged from 1 to 7 and arousal ratings (M=55.98, SD=25.04) ranged from 1 to 100. We calculated Cronbach's alpha for listener ratings across all 48 excerpts to be α= 0.84 for valence ratings and α=0.87 for arousal ratings, suggesting high internal consistency among listener responses. Mean ratings of valence and arousal are positively correlated ($r=.39$, $p<.001$), indicating our two dimensions did not function independently. Furthermore, there is a significant positive correlation between attack rate and mode ($r(46)=0.431$, $p<0.01$), demonstrating a relationship between faster attack rates in major modes. This relationship is also supported evidence from t-test analyses ($t(46)=-3.2419$, $p<0.05$).[2] Pitch height correlated significantly with neither attack rate ($r(46)=-0.138$, $p=0.350$) nor modality ($r(46)=0.142$, $p=0.334$).

Regression Analysis.

We assessed our cues as potential predictors for mean ratings of valence and arousal using standard linear multiple regression analysis from the R Statistical Package. The Major mode is chosen as the reference level for mode, meaning the remaining level of our categorical variable (minor) is contrasted against it in the analysis. For mean ratings of valence, all three cues, attack rate, mode and pitch height emerged as significant predictors (Table 1). The regression

[2] Correlations previously reported in Battcock & Schutz (in press)

model for mean arousal ratings indicated only attack rate as a significant predictor (Table 1).

The three-cue predictor models accounted for 81.2% of the variance in valence ratings (adjusted R^2=0.812), $F(3,44)$=68.68, p <.001 in contrast to 49.8% of variance in arousal ratings (adjusted R^2=0.498), $F(3,44)$=16.56, p<.001. Participants' predicted valence rating is equal to 0.549 + 0.248 (attack rate) – 0.933(mode) + 0.102 (pitch height). Valence ratings increased 0.248 for each note attack per second increase in attack rate, decreased 0.933 for the switch from major to minor mode and increased 0.102 for each increase in pitch. The predicted arousal rating is equal to 0.474 (attack rate), where arousal ratings increase 0.474 for each note attack per second increase in attack rate.

Table 1.

Regression model for normalized attack rate, mode, pitch height on valence and arousal ratings. Beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings. Default state for mode is Major.

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.248	0.049	5.023	*p*<.001	0.474	0.085	5.570	*p*<.001
Modality	-0.933	0.099	-9.384	*p*<.001	-0.235	0.171	-1.372	*p*=.177
Pitch Height	0.0102	0.0455	2.243	*p*<.05	0.050	0.078	0.634	*p*=.529

R^2	.812	.498
F	68.68	16.56

Commonality Analysis.

We used this technique to partition the R^2 of our models to reveal how much variance our predictors can explain independently of or in common with the other predictors in our model. Commonality analyses allows for a better understanding of regression models as it reveals relationships between the total, direct and indirect effects of regression predictors (Ray-Mukherjee et al., 2014). Negative values are possible using this method, as they can emerge as a consequence of correlations between predictors with opposite signs (Pedhazur, 1997) or due to a 'suppressor' variable included in the model that removes some amount of error variance of another predictor (Capraro & Capraro, 2001). Furthermore, some researchers argue this indicates a predictor has no to little influence (Frederick, 1999). We found negative commonality values in our application of commonality analysis for ratings of valence and arousal (Battcock & Schutz, 2019).

For this study we extend our previous approaches involving commonality analysis to use bootstrap methods providing confidence intervals for the estimations of cue. We then examined the cue contributions to the bootstrapped data from the participant response using commonality analysis to decompose the R^2 value into shared and unique variance of the model (Tables 2 & 3). Similar to findings from Battcock & Schutz (2019), mode accounted for the largest amount of explained variance (38.9%) of valence ratings, followed by attack rate (14.8%) and pitch height (3.1%). This indicates that even when partialling out any common or shared variance, mode remains the strongest predictor of valence ratings. The relation between attack rate and mode predicted the most shared variance (40.2%) compared to shared contributions of attack rate and pitch height (-2.15%) or mode and pitch height (8.21%) or all three cues combined (-2.96%). The larger variance amount common to mode and attack rate is reflective of the correlation we found between these two cues.

For the variance of arousal ratings, attack rate is the strongest predictor, accounting for 63.2%, followed by mode (3.5%) and pitch height (0.5%). As in our model for valence ratings, the shared contribution of attack rate and mode predicted the most variance (33.3%). Contributions of other cue combinations predicted less than 1% of the model variance (Table 2).

Table 2.

Commonality Analysis for Variance in Listener Ratings of Valence (experiment 1)

		$R^2_{y.123}$ = 0.8334	95% CIs	% Explained
Unique to X_1	Attack Rate	0.1233	[0.048, 0.158]	14.80%
Unique to X_2	Modality	0.3239	[0.257, 0.427]	38.86%
Unique to X_3	Pitch Height	0.0254	[0.011, 0.032]	3.05%
Common to X_1 and X_2	C (AR, Mo)	0.3350	[0.275, 0.342]	40.20%

Common to X_1 and X_3	C (AR, PH)	-0.0179	[-0.02, -0.010]	-2.15%
Common to X_2 and X_3	C (Mo, PH)	0.0684	[0.051, 0.077]	8.21%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	0.0247	[-0.026, -0.021]	-2.97%
	Totals	.8334		100

*The empirical 95% CIs were computed using the percentile method on bootstrapped samples.

Table 3.

Commonality Analysis for Variance in Listener Ratings of Arousal (experment 1)

		$R^2_{y.123}$ = 0.5429	95% CIs*	% Explained
Unique to X_1	Attack Rate	0.3428	[0.281, 0.371]	63.15%
Unique to X_2	Modality	0.0191	[0.010, 0.036]	3.53%
Unique to X_3	Pitch Height	0.0027	[0.000, 0.012]	0.49%
Common to X_1 and X_2	C (AR, Mo)	0.1809	[0.143, 0.213]	33.33%
Common to X_1 and X_3	C (AR, PH)	0.0029	[-0.012, 0.010]	0.54%
Common to X_2 and X_3	C (Mo, PH)	0.0050	[0.003, 0.011]	0.93%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-0.0106	[-0.013, -0.005]	-1.96%
	Totals	.5429		100

*The empirical 95% CIs were computed using the percentile method on bootstrapped samples.

Comparison to Non-Expert Data

Comparing ratings of these musically trained participants with ratings by those without training (Battcock & Schutz, 2019) allows for useful insight into differences that may emerge in how listeners perceive emotion. Overall, the model for valence ratings for listeners with music training accounted for proportionally more of the total variance (83.3%) than found for untrained listeners (76.2%) (Battcock & Schutz, 2019). We found a similar trend for models of arousal ratings where the model for listeners with music training explained proportionally more variance (54.3%) than for those with less than one year of training (51.1%), indicating our three-cue model to fit the perceptual ratings of musically trained listeners better.

In order to more directly compare cue weights between the two groups of listeners, we performed Fisher's Z-test to compare beta weights from trained and untrained listener models (Clogg, Petkova, & Haritou, 1995; Steiger, 1980). Analyses on the regression weights in models for ratings of valence show all cues have equivalent weights across the two groups for attack rate (Z=0.794, p=.785), mode (Z=-0.989, p=.184) and pitch height (Z=0.069, p=.755). However using this method on regression beta weights fails to address any correlations between the predictors (Ray-Mukherjee et al., 2014), which we found in both correlation and commonality analyses (See Appendix C and Figure 1). Therefore, commonality analysis is crucial to break down the relationship between unique and shared variance explained by our predictors.

Although commonality analysis is helpful in teasing apart the relative strength of different cues, it does not in and of itself provide a straightforward way to assess the significance of differences in cue strength. Therefore, we turned to bootstrapping in order to explore whether musical training meaningfully increased the strength of any particular cues. Bootstrapping involves repeatedly resampling from the original data set to create as multiple simulated data sets. These simulated data sets afford hypothesis testing and sample statistics in cases where these analytic solutions are not available (Mooney & Duval, 1993). For the purposes of our study, the bootstrapping method we used a resampling with replacement for 1000 runs with a sample of 30 (our actual sample included 30 participants). Descriptive information for the bootstrapped data can be found in Appendix A.

From the generated data sets we calculated confidence intervals (CIs) for each of the coefficients of the commonality analysis. With the bootstrapped CIs, we calculated the average margin of error (MOE) estimation for CI overlap for the coefficient representing the unique contribution of mode from our commonality analysis on the ratings of trained and untrained listeners. Using this estimation, an overlap of 'moderate' to 'small' of the confidence intervals can equate to a *p* value of $\leq .05$[3] (Cumming, 2012). In this case, moderate overlaps are calculated to be half of the average MOE of the two groups. For our data, the criterion value is

[3] Although this method is not standard in hypothesis testing, the benefit of using confidence intervals instead of p-values has been argued for across different fields (Ranstam, 2012; Rigby, 1999).

0.08 and the calculated overlap of confidence intervals is 0.02 (See Appendix C for details on the calculation), indicating that the coefficients for these two groups are likely to be significantly different from each other using an α level of .05.

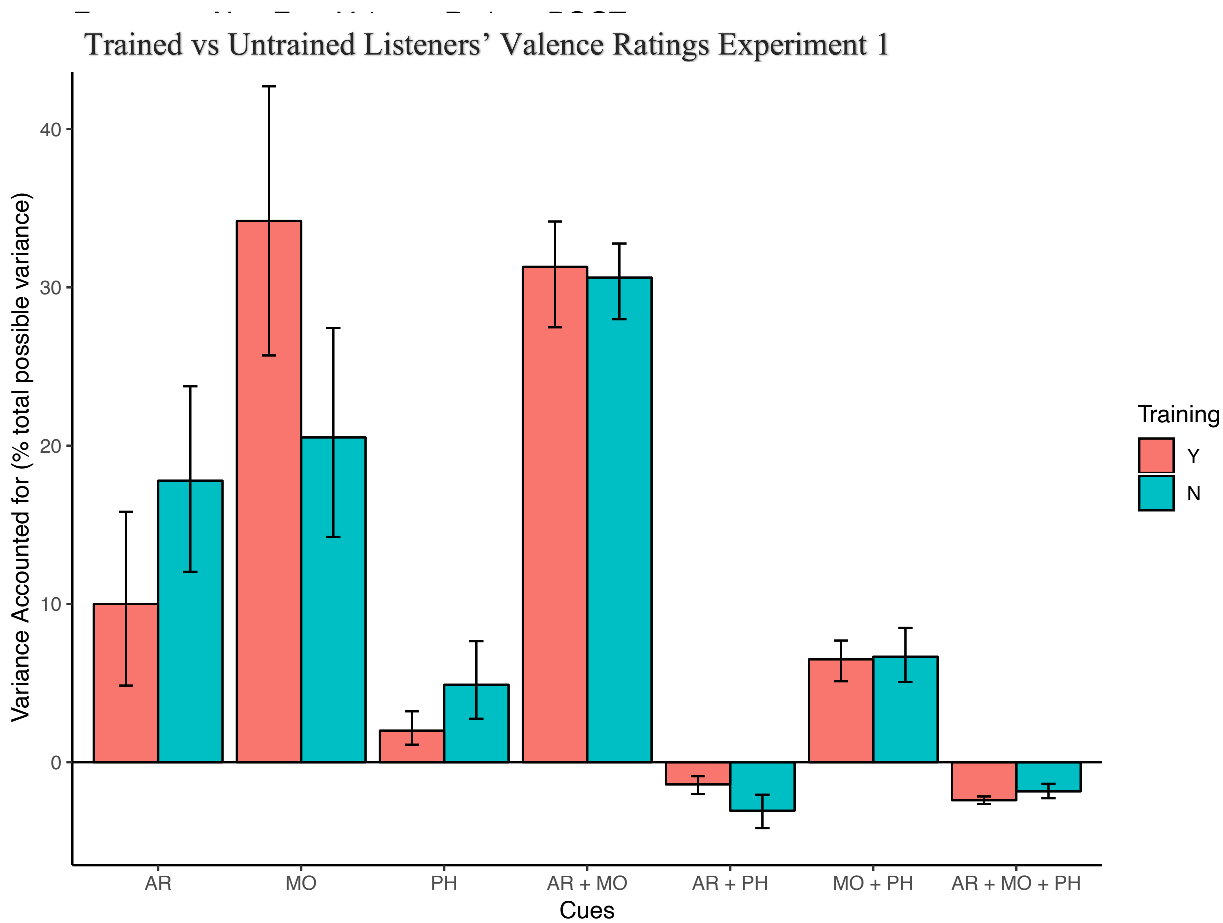


Figure 1. Unique and shared variance of valence ratings by musical cue. Individual bars depict cue weights calculated for each group of participants for Exp 1 (red=non musicians, blue=musicians). Error bars represent 95% confidence intervals. Attack rate uniquely explains more variance for those without musical training and modality explains a large majority of variance for those with musical training, although specific contributions vary.

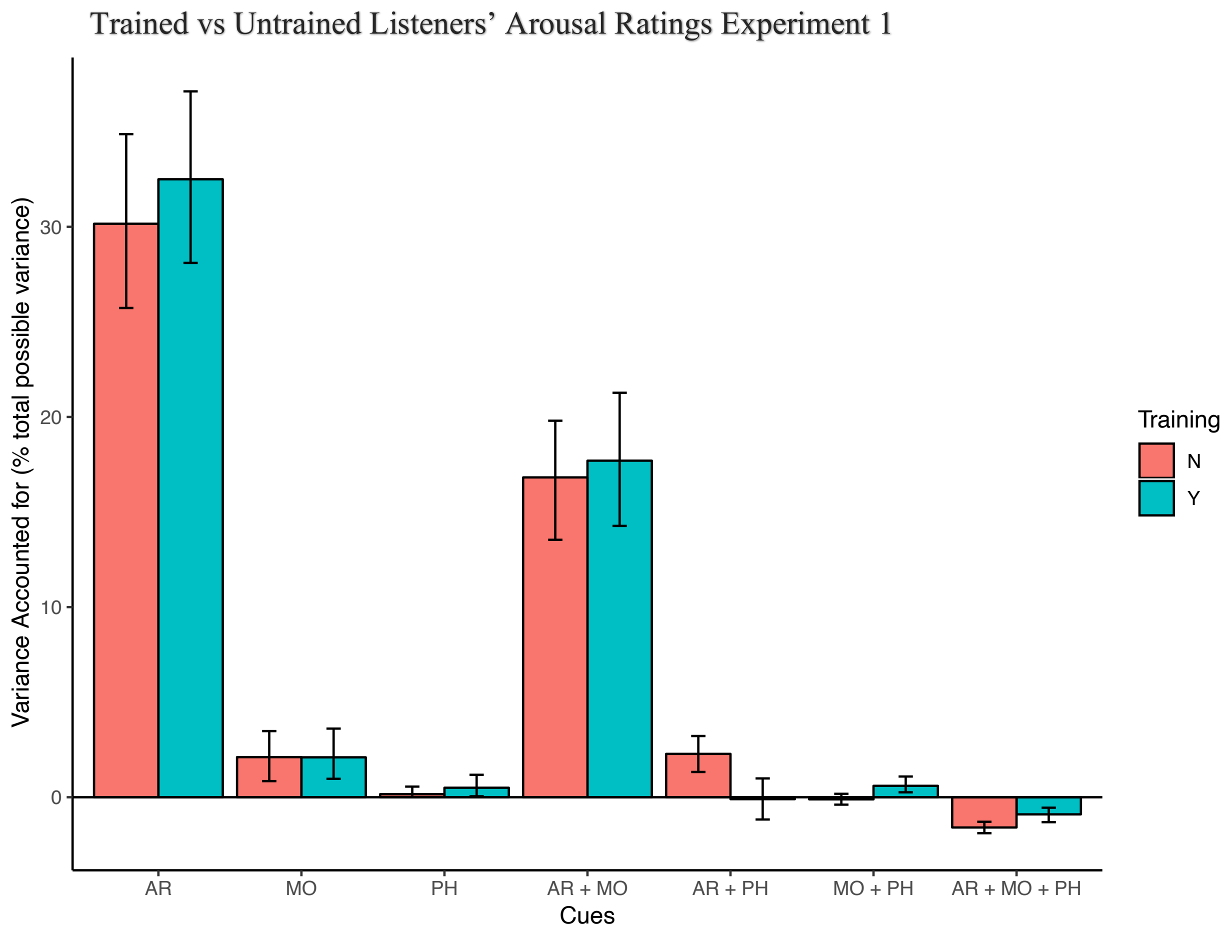


Figure 2. Unique and shared variance of arousal ratings by musical cue. Individual bars depict cue weights calculated for each group of participants for Exp 1 (red=non musicians, blue=musicians). Error bars represent 95% confidence intervals. Cue weights appear to explain variance similarly across participants with and without musical training.

Experiment 2 (Musically resolved excerpts)

Our first experiment assessed how listeners use cues of attack rate, mode and timing to perceive emotions in musical excerpts cut to be eight musical measures in length. One limitation of using precomposed stimuli such as the *WTC* is an inability to control for modulations or musical key changes that occur throughout the excerpts. These modulations result in changes in mode from major to minor keys or vice versa as the excerpt unfolds. Therefore, we ran a second experiment as in Battcock and Schutz (2019) ensuring excerpts ended to sound musically 'resolved', often ending in the piece's nominal key (e.g. the C minor excerpt for the experiment is cut at the point it returns to C minor). In many ways this offers a clearer assessment of modality's strength, although it by definition results in excerpts varying in duration. For this experiment we hypothesized (1) mode would increase in its importance for valence ratings based on ratings from those with musical training and (2) would be more important for trained compared to untrained listeners.

Method

We used an identical procedure for experiment 2, but with stimuli of variable length (in contrast to the eight measure excerpts in experiment 1) cut to be musically 'resolved', often ending in the piece's nominal key. As in experiment 1**,** participants included 30 individuals with ≥6 years of formal musical training from McMaster University and volunteers from the Ontario Music Educator's Association's General Assembly (21 females, ages M=25.07, SD=11.92, years of training M=6.57 SD=0.50). On average, participants scored in the 67th percentile on the General Sophistication scale and within the 79th percentile of the Gold-MSI Musical Training subscale. Undergraduate participants from McMaster University received course credit, or compensation for their participation. This experiment met McMaster University Research Ethics Board ethics standards. Musical stimuli ranged from 7-52 seconds (M=25.4 seconds, SD=11.0).

Cue Quantification. Pitch and timing information corresponded the quantification of each cue within the specific number of measures required to reach a 'resolution' back to the original mode for each excerpt. Pitch height values varied from 33.13-53.13 — corresponding ~F3 to ~ C#5 — (M=43.87, SD=4.15), attack rate information ranged 1.30-10.13 attacks/second (M=4.87, SD=2.22). We coded modality in the same way as in experiment 1 (0=minor, 1=Major).

Results

Valence ratings (M=3.94, SD=1.58) ranged from 1 to 7 and arousal ratings (M=53.78, SD=25.33) ranged from 1 to 100. Mean values of listener ratings of valence and arousal are significantly and positively correlated r=0.44, p<.001, indicating a lack of independence between our two dimensions as in experiment 1. The Cronbach's alpha values for ratings across our 48 excerpts are α= .79 for valence ratings and α=.95 for arousal ratings, suggesting less consistency among listener ratings of valence than arousal, however both values fall in the acceptable range. As in experiment 1, we calculated a significant positive correlation between the cues of attack rate and modality ($r(46)=0.435$, $p<.001$).[4] Pitch height significantly correlated with neither attack rate ($r(46)=-0.165$, $p=.261$) nor modality ($r(46)=0.126$, $p=.392$).

Regression analysis.

As with experiment 1, we ran linear regression analyses to assess predictors for listener ratings of emotion. All three cues significantly predicted participants' valence ratings, but only attack rate predicted arousal ratings (Table 4). The three-cue model for valence ratings accounted for 87% of variance (Adjusted R^2= 0.874), $F(3,44)=110$, $p<.001$). Predicted valence ratings are equal to 0.525 + 0.234 (attack rate) – 1.220 (mode) + 0.073 (pitch height), where valence ratings increase 0.234 for each increase in note attacks per second, decrease 1.220 from the switch to minor mode and increase 0.073 for each increase in pitch height. Our arousal rating model accounted for 52% of variance (Adjusted R^2=0.523), $F(3,44)=18.18$, $p<.001$, where predicted arousal ratings are equal to 0.534 (attack rate). As such, arousal ratings increased 0.534 for each increase in note attacks per second.

Across the two experiments, our models for valence ratings in experiment 2 (87.4%) accounted for proportionally more total variance than in experiment 1 (81.2%). The model for arousal ratings in experiment 2 (52.3%) also accounted for proportionally similar amounts of the total variance as seen in experiment 1

[4] Correlations reported in Battcock & Schutz (in press)

(49.38%). Comparing regression weights of cues between experiment 1 and 2 illustrates that mode's effect is significantly different (Z=-1.745, p <.05). This difference in mode's regression weight suggest mode is more predictive of valence ratings when individual pieces begin and end in the same mode. Attack rate and pitch height have equivalent regression weights in the two groups (Z=.115, p=.544 & Z=.156, p=.564, respectively), indicating no change in how listeners use these cues to make their emotion judgements.

Table 4.

Regression model for attack rate, mode, pitch height on valence and arousal ratings. Beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings. Default state for mode is Major.

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.234	0.048	4.892	p<.001	0.534	0.091	5.842	p<.001
Modality	-1.220	0.095	-12.802	p<.001	-0.024	0.182	-1.119	p=.269
Pitch Height	0.073	0.043	1.694	p=.097	-0.006	0.083	-0.070	p=.944
R^2			.874				.523	
F			110				18.18	

Commonality analysis. Uniquely, mode predicted the largest amount of variance associated with valence responses, accounting for 49.1% (Table 5). Attack rate and pitch height contributed 8.6% and 1.6% respectively. Together, attack rate and mode predicted the largest amount of shared variance in the model (37.3%), with a small amount predicted by the relation between mode and pitch height (7.7%). Values for the shared contributions between attack rate and pitch height and all three predictors remained below 0% (-1.2% and -3.1% respectively).

As in experiment 1, the R^2 breakdown of the model of arousal ratings (Table 6) indicates attack rate as the strongest predictor, uniquely accounting for 65.9% of the model variance. Mode and pitch height uniquely predicted only 2.6% and 0.5% of the variance from listener responses. With regards to shared contributions, only the relation between attack rate and mode predicted more than 1% of model variance (31.6%). Shared variance predicted by attack rate and pitch height accounted for 0.6%, and shared variance predicted by mode and pitch height 0.8%. The shared contribution of all three cues in the model predicted -1.9% of arousal rating variance.

Table 5.

Commonality Analysis for Variance in Listener Ratings of Valence (Experiment 2)

		$R^2_{y.123}$ = 0.8740	95% CIs*	% Explained
Unique to X_1	Attack Rate	0.0759	[0.039, 0.100]	8.56%
Unique to X_2	Modality	0.4358	[0.350, 0.500]	49.12%
Unique to X_3	Pitch Height	0.0014	[0.002, 0.016]	1.57%
Common to X_1 and X_2	C (AR, Mo)	0.3310	[0.329, 0.375]	37.31%
Common to X_1 and X_3	C (AR, PH)	-0.0103	[-0.012, -0.002]	-1.16%
Common to X_2 and X_3	C (Mo, PH)	0.0681	[0.038, 0.059]	7.68%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	-0.0273	[-0.308, -0.026]	-3.08%
	Totals	0.8740		100

*The empirical 95% CIs were computed using the percentile method on bootstrapped samples.

Table 6.

Commonality Analysis for Variance in Listener Ratings of Arousal (Experiment 2)

		$R^2_{y.123}$ = 0.5393	95% CI*	% Explained
Unique to X_1	Attack Rate	0.3553	[0.308, 0.380]	65.88%
Unique to X_2	Modality	0.0137	[0.007, 0.020]	2.55%
Unique to X_3	Pitch Height	0.0027	[0.00, 0.001]	0.49%
Common to X_1 and X_2	C (AR, Mo)	0.1703	[0.170, 0.204]	31.57%
Common to X_1 and X_3	C (AR, PH)	0.0033	[0.013, 0.029]	0.61%
Common to X_2 and X_3	C (Mo, PH)	0.0041	[-0.001, 0.002]	0.77%
Common to X_1, X_2 and X_3	C (AR, Mo, PH)	0.0101	[-0.015, -0.012]	-1.86%
	Totals	0.5393		100

*The empirical 95% CIs were computed using the percentile method on 1000 bootstrapped samples.

Comparison to Non-Expert Data

Further analyses on the regression models for valence ratings between trained and untrained listeners indicate our rating model of listeners with music training accounted for proportionally more of the total variance (83.3%) than for the untrained listeners (76.2%). We found a similar trend for arousal with more variance explained in the ratings made by those with music training (53.9%) than those with than one year of training (51.1%). We also calculated differences between predictors in experiment 1 and 2 for the valence ratings of untrained listeners using regression weights between experiments. This revealed mode has significantly different regression weights in experiment 1 and experiment 2 samples (Z=-1.745, p <.05).

Comparing beta coefficients from our regression models for trained and untrained listeners reveal mode to have a significantly different regression weight for the model of listener responses from those with and without musical training (Z=-1.854, p<.05). The cues of attack rate and pitch height have equivalent weights across the two groups (Z=0.373, p=.705 & Z=0.067, p=.749).

Comparison between experiments 1 and 2. Models of listener ratings for valence showed an increase in model fit for both trained (80% to 87%) and untrained listeners (76% to 81%) of 6-7% between experiment 1 and 2, where for both groups our three-cue model better predicted ratings in experiment 2. Regression models for the ratings of arousal demonstrated a different pattern: Model fit had a slight increase between experiment 1 and 2 for trained listeners (52% to 55%) however decreased in fit for untrained listeners (50% to 46%). Results of the commonality analysis on arousal ratings indicates a difference between how our listener groups use attack rate: attack rate predicts more variance in experiment 2 compared to experiment 1 for trained listeners and predicts less for untrained listeners. Overall, the model fit appeared better for ratings from musically trained listeners, suggesting listeners with music training may use the cues more systematically than untrained listeners.

The predictive weight of attack rate and pitch height did not change between experiments for neither group of listeners based on the Fisher Z test of beta weights. However, for trained listeners, the predictive weight of mode increased from experiment 1 to experiment 2 (Z=-1.745, p <.05). For untrained listeners the weight of mode did not change (Z=-1.0846, p=0.14). The results from this analysis suggests that the salience of mode increased for the excerpts that ended in the same nominal key — but only for musically trained listeners. Although the Fisher Z score emerges as nonsignificant, commonality analyses on the bootstrapped data shows mode's unique explanatory power increased from 20.5% to 27.2% (See Appendix B1 & B3) for untrained listeners when shared and unique contributions are taken into consideration. For musically trained listeners, mode's predictive weight increased from 32.4% to 43.6% (Tables 2 and 5).

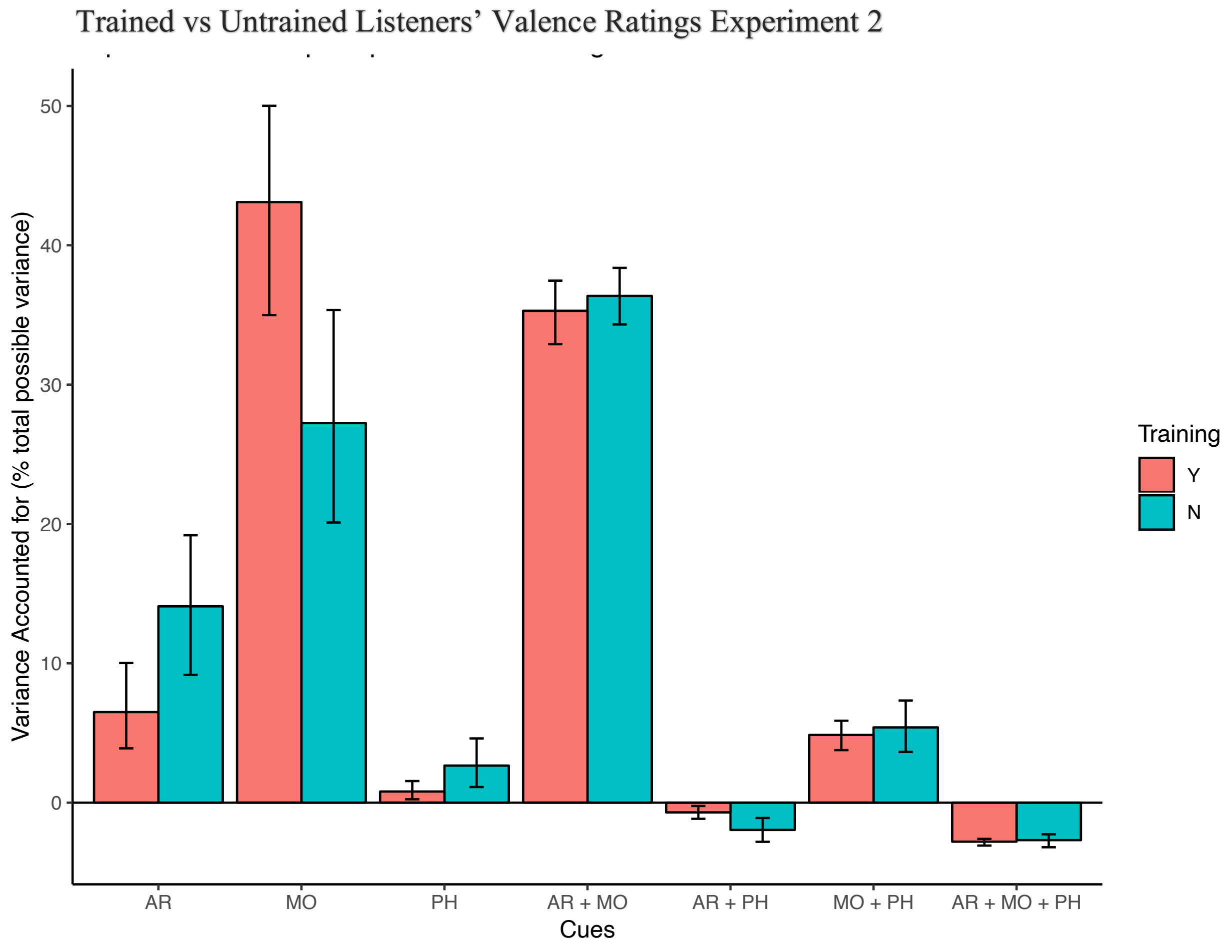


Figure 3. Unique and shared variance of valence ratings by musical cue. Individual bars depict cue weights calculated for each group of participants for experiment 2 (red=non musicians, blue=musicians). Error bars represent 95% confidence intervals. Attack rate uniquely explains more variance for those without musical training and modality explains a large majority of variance for those with musical training, although specific contributions vary.

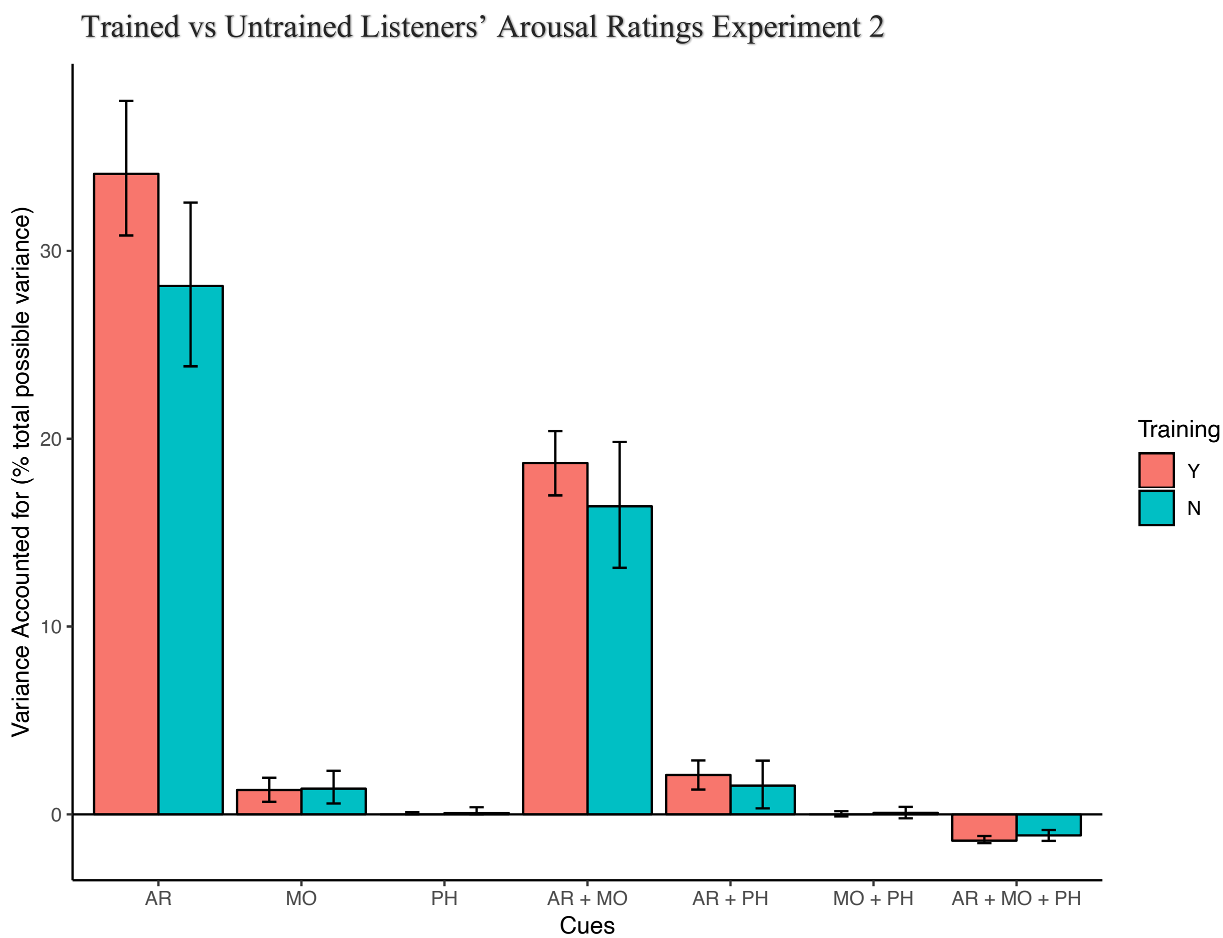


Figure 4. Unique and shared variance of arousal ratings by musical cue. Individual bars depict cue weights calculated for each group of participants for experiment 2 (red=non musicians, blue=musicians). Error bars represent 95% confidence intervals. Cue weights appear to explain variance similarly across participants with and without musical training.

Discussion

The results from our set of experiments demonstrate differences in how listeners with musical training use musical features to perceive emotion compared to responses collected from untrained listeners (Battcock & Schutz, 2019). Our data is consistent with the broad findings of Battcock & Schutz (2019) indicating the relation between mode, timing and pitch with perceived emotion in music. Yet here we complement and extend those results by exploring effect of musical training on the relative strength of cues used to communicate musical emotion. Our findings also further work done assessing the structural properties of music with listener ratings of perceived valence (Gagnon & Peretz, 2003) and arousal (Schubert, 2004; Vieillard et al., 2008), as we find mode more important for trained listeners in assessments of valence. Additionally, these results further illustrate that a model built on three cues derived from a score-based analysis can explain more variance for listeners with musical training.

Compared to the bootstrapped data from Battcock and Schutz (2019), models of ratings from trained listeners explain more variance for both valence and arousal ratings. This is consistent with the idea that trained listeners are more sensitive to particular cues than untrained listeners. Although Fisher's Z score analysis on predictor beta weights for valence ratings between the two populations indicated nonsignificant differences, further analyses using commonality analysis (Table 2) and MOE calculations on the bootstrapped CIs (Appendix C) revealed appreciable differences of the unique variance explained by mode (34.2% for trained listeners, 20.5% for untrained listeners). Mode's greater role for trained listeners is consistent with previous developmental work showing exposure or increased experience can change the relative weight given to mode when making assessments of emotion (Dalla Bella et al., 2001). This suggests that although structural cues generally affect listeners regardless of training, the specific mix of their effects is dependent upon the presence of musical training.

The acoustic profile generated from arousal ratings for trained participants appeared similar to the one calculated from bootstrapped ratings of untrained participants (Appendix B). The three cue models of arousal for each group only differed by 3% overall, which is a result of the increase of uniquely explained variance by attack rate for the trained listeners. Furthermore, visual inspection of the commonality analysis coefficients with confidence intervals do not suggest meaningful differences.

Grounding this study in well regarded music by Bach music offered an opportunity to explore naturally co-varying cues such as mode and timing, an issue difficult to explore when using more controlled stimuli (Schutz, 2017). Although we have used commonality analysis in an exploratory manner in previous studies (Battcock & Schutz, 2019), here our additional application of bootstrapping allowed us to directly assess differences in cue weights in a new way. This provides the novel insight that Bach's decision to co-vary cues such as mode and timing results in multiple pathways for listener detection of emotion to "converge"— whether they focused more on modality (experienced musicians) or timing (less experienced listeners). It is possible that part of the success of compositions such as the *WTC* lies in composers' innate ability to convey messages in redundant manners. Although future research is needed to explore this issue, this outcome is one of the benefits of using the *WTC* to balance issues of musical ecological validity with experimental control.

In experiment 2, we used musically 'resolved' stimuli to assess if the predictive strength of mode would increase when excerpts are cut to end in the same nominal key they began. Again, we see similar overarching trends as found in the untrained listeners used by Battcock and Schutz (2019) with respect to the relative weights of cues across ratings of valence and arousal. However, this exploration using trained listeners illustrates the greater emphasis placed upon mode as a result of listener experience. Additionally, further examination of emotion ratings between experiment 1 and 2 revealed stronger effects of mode in the musically 'resolved' excerpts for trained listeners than untrained listeners.

As in experiment 1, we see our models of listener ratings explained more variance in valence ratings for listeners with training (87.4%) than those without (81.1%). There did not appear to be any differences between models of arousal ratings for each population (54% explained for both). These results indicate the influence of individual differences, specifically differences in musical training, on the perception of emotional valence. In addition, we hypothesized mode would be more important for listeners with music training for valence ratings. Our results demonstrated that in contrast to bootstrapped responses from untrained participants (27.2%) collected by Battcock & Schutz (2019), mode uniquely predicted more variance for ratings of listeners who had formal music training (43.6%) (Figure 3). This is based on our MOE calculations of the bootstrapped confidence intervals (Appendix C) of the commonality analyses.

The breakdown of cue contributions for arousal ratings for participants with musical training showed similar values and trends as in experiment 1. These appeared to follow the cue profile for ratings from participants without musical training (See Appendix B). As we did not see any difference between listeners for cue profiles in experiment 1, we did not predict any differences in the calculated cue weights. In addition, the visual inspection of the commonality analysis coefficients with confidence intervals confirmed a lack of meaningful differences.

General Discussion

How does musical training affect the specific ways in which musical structure shapes emotional communication? We explored this question with two experiments, examining how listeners with musical training use cues of timing, mode and pitch height, compared to untrained listeners. Extending on our earlier work using Bach's *Well-Tempered Clavier* (Battcock & Schutz, 2019) as well as studies that employ excerpts recorded to convey basic emotions (Akkermans et al., 2018; Castro & Lima, 2014), here we show the effect of mode varies as a result of a listener's musical training. In light of counterevidence suggesting listeners with music training perform similarly to untrained listeners on behavioural tasks, these results underscore the presence of nuanced differences in how listeners process emotion that perhaps can often be difficult to capture based on the experimental methods and measures used.

Musical 'Expertise' and Perception/Perceptual Differences

Consistent with Lima and Castro (2011), we found similar trends in the cue profiles for features predicting listener responses to auditory stimuli for both trained and untrained participants. In their study the authors used discrete rating methods to gather emotional judgements on samples of vocal prosody and focused on regression analyses for each emotion to determine the cue profiles. Unlike their study, here we used commonality analyses in addition to regression modeling and found a difference in the strength of how mode predicted listener ratings of emotion. This novel approach illustrates that mode, a cue unique to music, predicted more variance for valence ratings for participants with musical training. Further it highlights the usefulness of using commonality analyses to tease apart the relationships between predictors and explained variance, demonstrating benefits of musical training with respect to specific cues conveying emotional information

Previous research exploring the effect of musical training and age using polyphonic, or single-lined instrumental excerpts demonstrated an influence of expertise for older participants, as years of musical training related to recognition accuracy (Castro & Lima, 2014). Their study focused on several acoustic cues such as tempo, mode and pitch range in their models of listener ratings and determined a range of explained variance dependent on the conveyed emotion as well as the significant predictors of listener ratings. There, regardless of music training participants could identify the intended emotions with high accuracy, however the authors found differences in how well the models of listener ratings fit for listeners with and without musical training. As such the authors suggest the effect of expertise might be small or difficult to detect. Similarly, our results indicated differences in how the models fit for trained (80% and 52% for valence and arousal in experiment 1, 87% and 55% for valence and arousal in experiment 2) and untrained (76% and 50% for experiment 1, 81 % and 54% for experiment 2 based on bootstrapped values) participants, particularly for ratings of valence. This suggests differences in how these groups of listeners are using cues of attack rate, mode and pitch height to make assessments of perceived emotion.

Further, Castro and Lima (2014) found variations in how cues predicted rating variance for negative emotions such as sad or scary, across younger and older musicians. The pattern of beta weights between trained and untrained listeners appeared similar, which the authors argue as suggesting listeners used similar inference rules in their perception of emotion. This had been determined using a multiple simultaneous regression analyses from collected intensity ratings for each of the four potential affect terms given for each excerpt. The results of our study, however, demonstrate a difference in the predictive weight of mode between trained and untrained listeners. In addition, we found the unique variance explained by mode increased more from experiment 1 to 2 for musically trained listeners than for untrained listeners, suggesting those with training were more sensitive to our resolved excerpts. As mentioned previously, differences may have emerged as a result of the stimuli used, as excerpts used in Castro and Lima

(2014) represented experimentally composed excerpts conveying specific intended emotions. Our stimuli came from a precomposed set by a widely recognized composers — crafted for artistic purposes rather than for a specific research aim. It is possible that with more ambiguous stimuli, differences in cue uses may emerge when emotional signal requires more attention or consideration in the decoding process.

Musical training and mode. The relationship between major and minor modes and perceived positive or negative emotions respectively, is hypothesized to develop through learned associations, or acculturation from exposure and experience with Western culture music. Understanding of the major/minor distinction with musical emotions emerges through development, as after five years of age children are found to use mode to match melodies to emotionally valenced faces (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990). Before this age, children predominately use timing information to understand expressed emotions (Dalla Bella et al., 2001). This pattern may emerge as children use similar performance cues to decode emotion in music as is used for nonverbal aspects in speech (Juslin & Laukka, 2003), consistent with findings that recognition of emotion in both music and speech develop in parallel (Vidas, Dingle, & Nelson, 2018). Given that the relationship between mode and perceived emotion becomes internalized through increased knowledge and familiarity with musical patterns of that culture, we might expect listeners with formal music training to use mode more than untrained listeners to decode conveyed emotion, particularly in more complex musical stimuli.

Although it has been suggested music listeners are themselves 'experienced listeners' (Bigand & Poulin-Charronnat, 2006), those with formal music training are often instructed to use cues to express emotion and therefore may use cues differently to decode expressed emotion. Our results demonstrate mode has a stronger effect on ratings of trained listeners than those with less than one year of musical training across both experiments. This could have occurred as a result of the complexity of the musical structure in our excerpts, leaving more 'naïve' listeners to use lower level cues like attack rate to understand what emotion is being transmitted, or cues commonly used to perceive emotion in vocal prosody such as timing, and loudness (Coutinho & Dibben, 2013).

Studies reporting no effect of musical training

Although some studies fail to find a training effect or advantage on perceptual tasks (Bigand & Poulin-Charronnat, 2006), evidence has shown musical training can be a significant predictor of listener ratings of emotion (Akkermans et al. 2018). In addition, conflicting findings on the effect of musical training and perception (Juslin, 1997; Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Trimmer & Cuddy, 2008) may emerge due to behavioural tasks used or the criteria for musical 'expertise' or training. Explicit and discrete methods often used in experimental tasks may produce ceiling effects in recognition accuracy and wash out any differences between trained and untrained listeners. Our study employs a dimensional approach with precomposed polyphonic stimuli which allows us to capture the complexity of perceiving emotion in music experienced in real world scenarios such as in concert halls, or from listening to album recordings.

Measuring musical expertise

The measured components considered as important in distinguishing a 'musician' or 'musical expertise' is an area that requires examination when considering conflicting results. The number of years of musical training is frequently used as the main qualifier when separating musicians or musical experts from non-musicians or nonexperts. Although years of training give a measure of how long individuals have spent learning musical theory and or musical performance, there may be other aspects that better reflect musicianship or expertise as encompassed by subscales included in Gold-MSI (active engagement, perceptual abilities, musical training, singing abilities and emotions). Regardless, the average criteria often used to qualify a musician has been variable across studies ranging from 6-13 years of formal training (Thompson, Schellenberg, & Ilie, 2004; Trimmer & Cuddy, 2008). Effects of expertise may be more pronounced with extensive training (13+ years) as suggested by Lima & Castro (2011), and supported by neural evidence indicating increased grey matter in several brain regions in 'expert' musicians (based on average daily practice and current profession) compared to non and amateur musicians (Gaser & Schlaug, 2013). Our study included individuals with six or more years of musical training, however on average they scored within the 67^{th} – 72^{nd} percentile on the General Sophistication scale and fell at the 79^{th} percentile on the Musical Ability scale of the Gold-MSI. It is possible with more training we could see a larger difference in the predictive weight of mode between trained and untrained listeners in emotion ratings. Future research should consider different aspects of expertise when exploring differences, as well as more extensive range of formal training.

Concluding thoughts

Our studies demonstrate how individuals with musical training are more affected by mode when perceiving conveyed emotion compared to untrained listeners. These results complement previous literature examining differences between behavioural and perceptual responses among musical experts and nonexperts, suggesting training can fine-tune the mechanisms used to decode musical emotions (Akkermans et al., 2018; Castro & Lima, 2014; Lima & Castro, 2011). In addition, our findings speak to literature exploring the role of individual differences and the effects of individual factors on emotion perception (Dibben, Coutinho, Vilar, & Estévez-Pérez, 2018; Taruffi et al., 2017; Vuoskoski & Eerola, 2011). Here we assess cue contributions, using regression analyses similar to Akkermans et al. (2018) and Eerola (2011) using to model of listener responses for valence and arousal, and incorporate commonality analysis to examine the unique and shared predicted variance to better break down cue contributions. Further, these experiments speak to conflicting behavioural results on perceptual differences between trained and untrained individuals, detecting a difference in the predictive weight of mode using a dimensional approach with participants who have six years or more of music training. Previous work indicates those with musical training respond to mode-emotion associations more reliably (Heinlein, 1928; Hevner, 1935), however evidence also suggests training is not necessary (Dalla Bella et al., 2001). In our studies, we demonstrate the degree mode's effect varies as a function of training, as mode holds more weight for trained listeners than those with less than one year of training. Thus, individual differences in perceiving emotion do emerge as result of formal music training.

The influence of mode in musically expressed emotion is one that faces some debate. Although evidence demonstrates it can be effective in conveyed positive or negative affect (Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005; Quinto & Thompson, 2013; Webster & Weir, 2005), music theorists argue its role is not as significant (Hatten, 2004). The argument is that results demonstrating mode's influence may emerge from its relationship or pairing with other structural cues such as timing, and not an inherent binary distinction between major equals 'happy' and minor equals 'sad'. Our results suggest mode can affect some aspects of emotion, like perceived valence, more than others such as perceived arousal. Additionally, it suggests that one potential explanation for this disagreement is that psychologists often use systematically varied stimuli offering a high degree of independent control over individual cues such as mode and timing. However, composers such as Bach essentially confounded these cues so that they co-varied. Consequently, disagreement over the role of modality in the communication of emotion could relate in part to different conceptions of how modality varies in passages created for scientific vs. artistic purposes.

The current study focuses on the perceptual responses of musically trained and untrained participants to one genre of musical stimuli within Western culture music. Even if they are unfamiliar with the *WTC* specifically (see Appendix D for familiarity responses), those with formal musical training can be familiar with the classical Western music styles due to training compared to those with minimal to no years of music training. Exploring expertise as well as familiarity effects using additional genres of music and incorporating commonality analysis can further extend our understanding of musical training on emotion perception and more broadly, the perceptual consequences of cue use and communicated emotion. Additionally, investigating familiarity or training in in non-Western cultures will help inform the relationship between cues and conveyed emotion with musical expertise in cross-cultural environments.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., … Frieler, K. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6) 1–20.

Bach, J. (1998). *Bach: The Well Tempered Clavier Book I* [CD; Recorded by A.

Hewitt]. London: Hyperion Records Ltd. (Original work published 1722).

Bach, J. S. (1973). *Bach: The Well-Tempered Clavier, Book I* - [CD; Recorded by F. Gulda]. MPS-Tonstudio, Villingen, Germany: Decca (Original work published in 1972).

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1), 43–64.

Battcock, A., & Schutz., M. (2019). Acoustically expressing affect. *Music Perception, 37*(1), 66-91.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and the duration of excerpts. *Cognition and Emotion*, *19*(8), 1113–1139.

Capraro, R. M., & Capraro, M. M. (2001). Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. *Multiple Linear Regression Viewpoints*, *27*(2), 16–23.

Castro, S. L., & Lima, C. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). *Statistical Methods for Comparing Regression Coefficients Between Models*. *Source: American Journal of Sociology*, 10, 1261-1293.

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition and Emotion*, *27*(4), 1–27.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.*. New York, NY, US: Routledge/Taylor & Francis Group.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation. *PLoS ONE*, *6*(4), e18591.

Di Mauro, M., Toffalini, E., Grassi, M., & Petrini, K. (2018). Effect of Long-Term Music Training on Emotion Perception From Drumming Improvisation. *Frontiers in Psychology*, 9, 2168.

Dibben, N., Coutinho, E., Vilar, J. A., & Estévez-Pérez, G. (2018). Do Individual Differences Influence Moment-by-Moment Reports of Emotion Perceived in Music and Speech Prosody? *Frontiers in Behavioral Neuroscience*, *12*(184).

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. *Journal of New Music Research*, *40*(4), 349–366.

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

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

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (pp. 305–318). Stamford, CT: JAI Press.

Gabrielsson, A., & Juslin, P. N. (1996). *Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience*. *Pyschology of Music,* 24, 68-91.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(27), 9240–9245.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Bloomington, IN: Indiana University Press.

Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of Comparative Psychology*, *8*(2), 101–142.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*, 103–118.

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*, 621–630.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, *22*(2), 327–352.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47–56.

Huron, D., Yim, G., & Chordia, P. (2010). The Effect of Pitch Exposure on Sadness Judgments: An Association Between Sadness and Lower Than Normal Pitch. In S. M Demorest, S.J Morisson, & P.S Campell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition* (pp. 63–66) Seattle, WA: Casual Productions.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*(4), 383–418.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception,* *8*(2), 189–202.

Koelsch, S., Schmidt, B.-H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychomusicology*, *39*, 657–663.

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.

Lima, C. F., & Castro, S. L. (2011). Speaking to the Trained Ear: Musical Expertise Enhances the Recognition of Emotions in Speech Prosody. *Emotion*, *11*(5), 1021–1031.

Lindström, E. (2006). Impact of melodic organization of melodic structure and emotional expression. *Musicae Scientiae*, *10*, 85–117.

Mlinarić, A., Horvat, M., & Šupak Smolčić, V. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia Medica*, *27*(3), 030201.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference.* Sage Publishing. Newbury Park, CA: Sage.

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).

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). *Goldsmiths Musical Sophistication Index (Gold-MSI) v1.0: Technical Report and Documentation Revision 0.3*. London: Goldsmiths, University of London.

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: a functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences*, *1060*, 450–453.

Pedhazur, E. . (1997). *Multiple regression in behavioural research: Explanation and prediction*. New York, NY: Thompson Learning.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, *6*, 1–13.

Quinto, L., & Thompson, W. F. (2013). Composers and performers have different capacities to manipulate arousal and valence. *Psychomusicology: Music, Mind, and Brain*, *23*, 137–150.

Ramos, D., Bueno, J. L. O., & Bigand, E. (2011). Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian Journal of Medical and Biological Research*, *44*(2), 165–172.

Ranstam, J. (2012). Why the P-value culture is bad and confidence intervals a better alternative. *Osteoarthritis and Cartilage*, *20*(8), 805–808.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328.

Rigby, A. S. (1999). Getting past the statistical referee: moving away from P-values and towards interval estimation. *Health Education Research*, *14*(6), 713–715.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Sherwin, J., & Sajda, P. (2013). Musical experts recruit action-related neural structures in harmonic anomaly detection: Evidence for embodied cognition in expertise. *Brain and Cognition*, *83*(2), 190–202.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251.

Taruffi, L., Allen, R., Downing, J., & Heaton, P. (2017). Individual differences in music-perceived emotions: The influence of externally oriented thinking. *Music Perception*, *34*(3), 253–266.

Thompson, W. F., Schellenberg, E. G., & Ilie, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64.

Trimmer, C. ., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*, 838–849.

Vidas, D., Dingle, G. A., & Nelson, N. L. (2018). Children's recognition of emotion in music and speech. *Music & Science*, 1–10.

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 and Emotion*, *22*(4), 720–752.

Vuoskoski, J. k, & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39.

Appendix A

Bootstrapped Commonality Analysis Distribution for Listener Ratings of Valence (Experiment 1)

Table A1.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.100	0.028	0.026	0.203	[0.048, 0.158]
Modality	1000	0.342	0.044	0.206	0.471	[0.257, 0.427]
Pitch Height	1000	0.020	0.005	0.007	0.040	[0.011, 0.032]
C (AR, Mo)	1000	0.313	0.017	0.233	0.354	[0.275, 0.342]
C (AR, PH)	1000	-0.014	0.003	-0.023	-0.006	[-0.02, -0.010]
C (Mo, PH)	1000	0.065	0.006	0.042	0.087	[0.051, 0.077]
C (AR, Mo, PH)	1000	-0.024	0.001	-0.028	-0.019	[-0.026, -0.021]

Bootstrapped Commonality Analysis Distribution for Listener Ratings of Arousal (Experiment 1)

Table A2.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.325	0.022	0.251	0.396	[0.281, 0.371]
Modality	1000	0.021	0.007	0.006	0.052	[0.010, 0.036]
Pitch Height	1000	0.005	0.003	0.000	0.019	[0.000, 0.012]
C (AR, Mo)	1000	0.177	0.018	0.124	0.236	[0.143, 0.213]
C (AR, PH)	1000	-0.001	0.005	-0.019	0.015	[-0.012, 0.010]
C (Mo, PH)	1000	0.006	0.002	0.001	0.015	[0.003, 0.011]
C (AR, Mo, PH)	1000	-0.009	0.002	-0.015	-0.003	[-0.013, -0.005]

Bootstrapped Commonality Analysis Distribution for Listener Ratings of Valence (Experiment 2)

Table A3.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.065	0.016	0.022	0.135	[0.039, 0.100]
Modality	1000	0.431	0.038	0.273	0.551	[0.350, 0.500]
Pitch Height	1000	0.008	0.003	0.001	0.021	[0.002, 0.016]
C (AR, Mo)	1000	0.353	0.012	0.307	0.385	[0.329, 0.375]
C (AR, PH)	1000	-0.007	0.002	-0.014	0.000	[-0.012, -0.002]
C (Mo, PH)	1000	0.0486	0.005	0.029	0.066	[0.038, 0.059]
C (AR, Mo, PH)	1000	-0.028	0.001	-0.033	-0.024	[-0.308, -0.026]

Bootstrapped Commonality Analysis Distribution for Listener Ratings of Arousal (Experiment 2)

Table A4.

	Bootstrapped Distribution					
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.341	0.018	0.292	0.415	[0.308, 0.380]
Modality	1000	0.013	0.003	0.005	0.024	[0.007, 0.020]
Pitch Height	1000	0.000	0.000	0.000	0.003	[0.00, 0.001]
C (AR, Mo)	1000	0.187	0.009	0.154	0.215	[0.170, 0.204]
C (AR, PH)	1000	0.021	0.004	0.009	0.033	[0.013, 0.029]
C (Mo, PH)	1000	0.000	0.001	-0.002	0.003	[-0.001, 0.002]
C (AR, Mo, PH)	1000	-0.014	0.001	-0.016	-0.010	[-0.015, -0.012]

Appendix B

Bootstrapped Commonality Analysis Distribution for Untrained Listener Ratings of Valence (Experiment 1)

Table B1.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	.179	.031	.081	.306	[.120, .238]
Modality	1000	.205	.034	.093	.332	[.142, .274]
Pitch Height	1000	.049	.013	.013	.091	[.027, .077]
C (AR, Mo)	1000	.306	.012	.266	.340	[.280, .328]
C (AR, PH)	1000	-.036	.006	-.049	-.012	[-.046, -.02,]
C (Mo, PH)	1000	.067	.009	.040	.096	[.051, .085]
C (AR, Mo, PH)	1000	-.184	.002	-.026	-.008	[-.023, -.014]

Bootstrapped Commonality Analysis Distribution for Untrained Listener Ratings of Arousal (Experiment 1)

Table B2.

	Bootstrapped Distribution					
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.302	0.024	0.219	0.387	[0.257, 0.349]
Modality	1000	0.021	0.007	0.036	0.043	[0.008, 0.035]
Pitch Height	1000	0.016	0.002	0.000	0.009	[0.000, 0.006]
C (AR, Mo)	1000	0.168	0.016	0.108	0.219	[0.135, 0.198]
C (AR, PH)	1000	0.023	0.005	0.067	0.038	[0.032, 0.010]
C (Mo, PH)	1000	-0.001	0.001	-0.005	0.004	[0.003, 0.011]
C (AR, Mo, PH)	1000	-0.016	0.002	-0.021	-0.011	[-0.013, -0.005]

Bootstrapped Commonality Analysis Distribution for Untrained Listener Ratings of Valence (Experiment 2)

Table B3.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	.141	0.026	.065	.214	[.092, .192]
Modality	1000	.272	0.040	.158	.405	[.201, .354]
Pitch Height	1000	.027	0.009	.006	.056	[.011, .046]
C (AR, Mo)	1000	.364	0.010	.322	.399	[.343, .384]
C (AR, PH)	1000	-.020	0.004	-.033	-.006	[-.028, -.011]
C (Mo, PH)	1000	.054	0.009	.029	.082	[.036, .073]
C (AR, Mo, PH)	1000	-.027	0.002	-.035	-.019	[-.032, -.023]

Bootstrapped Commonality Analysis Distribution for Untrained Listener Ratings of Arousal (Experiment 2)

Table B4.

		Bootstrapped Distribution				
Cue	N	Coeff	SD	Min	Max	95% CI
Attack Rate	1000	0.281	0.022	0.216	0.360	[0.239, 0.326]
Modality	1000	0.014	0.005	0.002	0.031	[0.006, 0.023]
Pitch Height	1000	0.073	0.001	0.000	0.008	[0.000, 0.004]
C (AR, Mo)	1000	0.164	0.018	0.106	0.218	[0.131, 0.198]
C (AR, PH)	1000	0.015	0.006	-0.006	0.036	[0.003, 0.029]
C (Mo, PH)	1000	0.008	0.002	-0.004	0.006	[-0.002, 0.004]
C (AR, Mo, PH)	1000	-0.011	0.001	-0.017	-0.006	[-0.014, -0.008]

Appendix C

Margin of Error Calculation for Valence Ratings Between Trained and Untrained Listeners (Experiment 1)

Table C1

Cue	Trained Listeners			Untrained Listeners			Avg MOE	Mod. Overlap	$LCI_{trained}$ - $UCI_{untrained}$
	U_{CI}	L_{CI}	Length	U_{CI}	L_{CI}	Leng.			
AR	0.158	0.048	0.110	0.238	0.128	0.117	0.114	0.057	-0.189
MO	0.427	0.257	0.170	0.274	0.104	0.132	0.151	0.075	-0.017
PH	0.032	0.011	0.021	0.077	0.055	0.049	0.035	0.018	-0.065
AR + MO	0.342	0.277	0.067	0.328	0.261	0.048	0.057	0.029	-0.053
AR + PH	-0.008	-0.020	0.011	-0.020	-0.032	0.021	0.016	0.008	0.000
MO + PH	0.077	0.051	0.026	0.085	0.059	0.034	0.030	0.015	-0.034
AR + MO + PH	-0.022	-0.026	0.005	-0.014	-0.018	0.009	0.004	0.003	-0.013

Note: CIs represent the 95% confidence interval arousal the mean; U_{CI}=Upper Confidence Interval; U_{LI}= Lower Confidence Interval; Length= length of the CI; Avg MOE= Average Margin of Error; Mod. Overlap= Calculated point of moderate overlap. $LCI_{trained}$- $UCI_{untrained}$ calculations represent the overlap calculation value.

Margin of Error Calculation for Valence Ratings Between Trained and Untrained Listeners (Experiment 2)

Table C2

Cue	Trained Listeners U_{CI}	Trained Listeners L_{CI}	Trained Listeners Leng.	Untrained Listeners U_{CI}	Untrained Listeners L_{CI}	Untrained Listeners Leng.	Avg MOE	Mod. Overlap	$LCI_{trained}$ - $UCI_{untrained}$
AR	0.100	0.039	0.061	0.192	0.092	0.100	0.0807	0.040	-0.153
MO	0.500	0.350	0.150	0.354	0.201	0.153	0.151	0.076	-0.004
PH	0.016	0.002	0.013	0.046	0.011	0.035	0.024	0.012	-0.044
AR + MO	0.375	0.329	0.046	0.384	0.343	0.041	0.043	0.022	-0.055
AR + PH	-0.002	-0.012	0.009	-0.011	-0.028	0.017	0.013	0.007	-0.001
MO + PH	0.059	0.038	0.021	0.073	0.036	0.037	0.029	0.0145	-0.036
AR + MO + PH	-0.026	-0.031	0.005	-0.023	-0.032	0.009	0.007	0.004	-0.008

Note: CIs represent the 95% confidence interval arousal the mean; U_{CI}=Upper Confidence Interval; U_{LI}= Lower Confidence Interval; Length= length of the CI; Avg MOE= Average Margin of Error; Mod. Overlap= Calculated point of moderate overlap. $LCI_{trained}$- $UCI_{untrained}$ calculations represent the overlap calculation value.

Appendix D

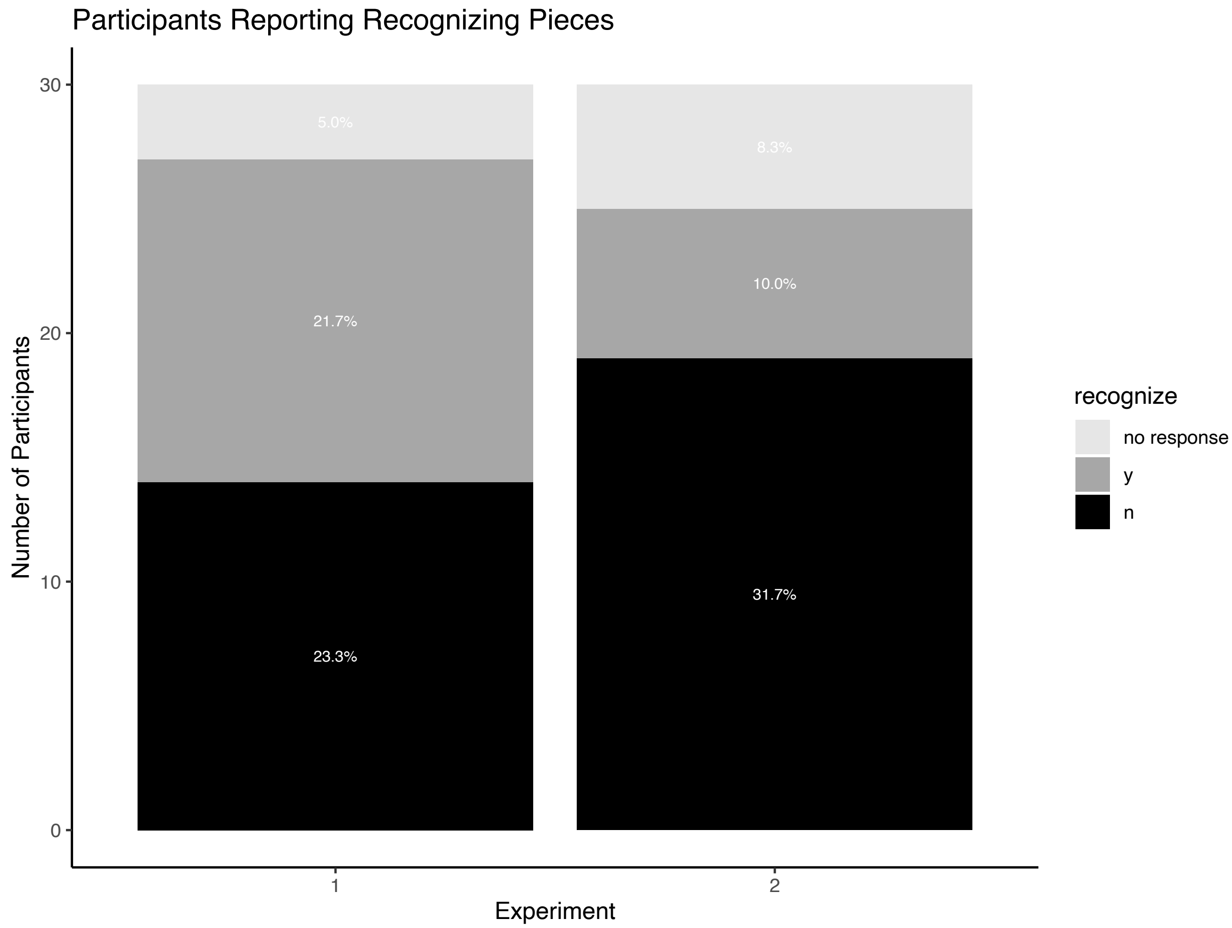


Figure D1. Familiarity responses from participants to "did you recognize any of the pieces" in debrief survey for experiment 1 and 2. Participants responded either yes or no, however there were a few missed responses due to RA error. Across both experiments a large majority of participants responded they had recognized some of the excerpts presented.

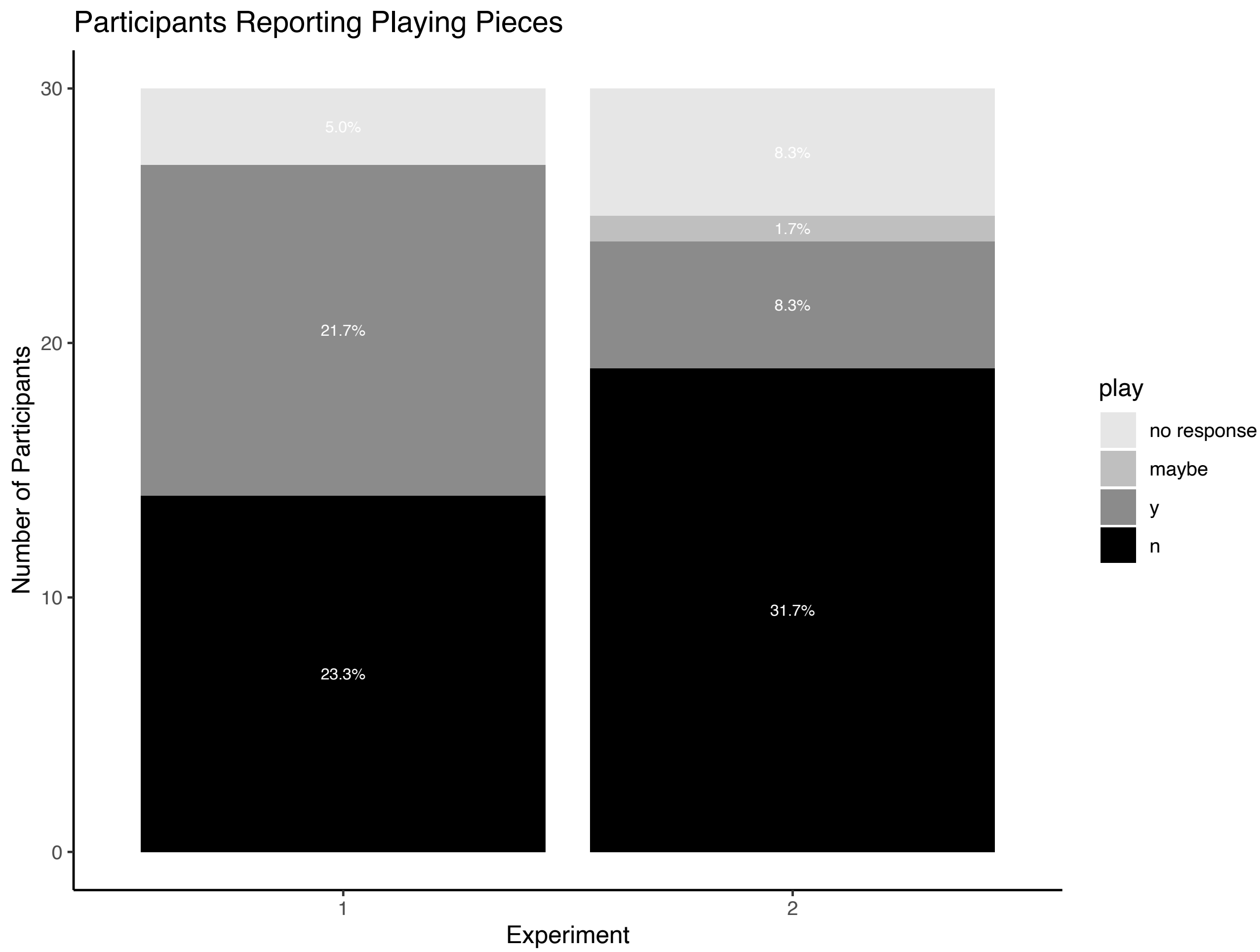


Figure D2. Participant responses across experiments 1 and 2 for follow up question “have you ever played any of the pieces recognized. Participants responded either yes or no, however there were a few missed responses due to RA error. Across both experiments the majority of responses is ‘no’, however in experiment 1 more participants reported playing some of the pieces than in experiment 2.

Author Note

Aimee Battcock, Department of Psychology, Neuroscience & Behaviour, McMaster University; Michael Schutz, Department of Music, McMaster University.

This research is supported in part by grants from Social Sciences and Humanities Research Council of Canada (SSHRC and the Canadian Foundation for Innovation (CFI)

Correspondence concerning this article should be addressed to Aimee Battcock, Department of Psychology, Neuroscience and Behaviour, Psychology Building (PC), Room 102, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada.
Email: battcoae@mcmaster.ca

Chapter 4

Battcock., A. & Schutz.,. M. (in prep). Performer differences and conveyed emotion: exploring interpretative decisions.

Preface

In the previous two chapters, I examined the influences of the musical structure (composition) and the listener, in the process of communicating emotion in music. In Chapter 4, I examined the influence of performer interpretation on cue used and conveyed emotion. Comparing listener ratings across seven interpretations in all 48 pieces of Bach's *Well-Tempered Clavier,* I observed notable differences in the emotional responses of different interpretations of the same piece. Further, I found greater variation in ratings of arousal for pieces compared to ratings of valence. In addition, there was greater variation across excerpts in minor modes rather than major modes. Finally, I show that models of listener ratings predicted varying amounts of rating variance dependant on performer. Together, these sets of experiment illustrate the impact of a performer's interpretative decisions on listener perception of conveyed valence and arousal.

Chapter 4: Performer Differences and Communicated Emotion

Abstract

In this series of exploratory experiments, we investigated differences in listeners' perceived emotion for various interpretations of pieces by JS Bach. For each study, we exposed thirty non-musician participants to 48 excerpts of the *Well-Tempered Clavier*, performed by one of seven pianists. After each excerpt, participants rated perceived emotion on scales of valence and arousal. From these ratings we (1) explore the degree to which the different interpretations of expert pianists affect the communication of emotion based on ratings of perceived arousal and valence, and (2) explore cue trade-offs in the between different performers.

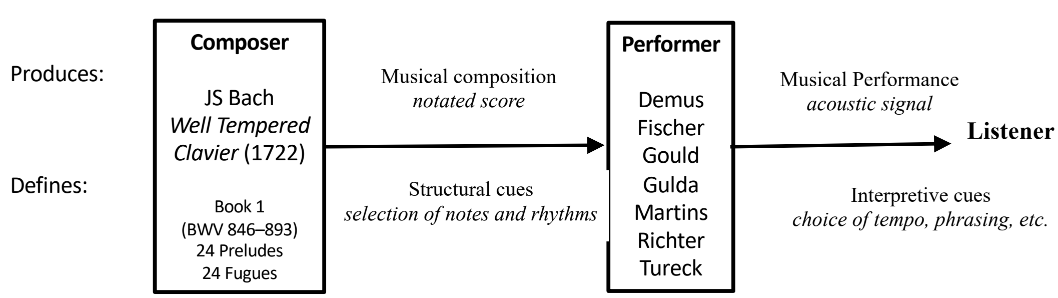
Our results indicate notable differences in the emotional responses of different interpretations of the same piece. Additionally, building on our past approaches we used multiple regression and commonality analysis to examine how listeners use select cues (attack rate, pitch height and mode) across various musical interpretations. Overall, we found similar trends across interpretation in the relationship between cues and perceived valence and arousal, with variation in the relative weight of attack rate across performers. Comparing the fit of our three-cue model for listener ratings of emotion across performers, we find the predictive value (R^2 values) of cues ranged from 51-78% for arousal ratings and 76-82% for valence ratings amongst performers. These results demonstrate that

performers' interpretative decisions lead to differences in listeners' perceived emotional experiences.

Keywords: emotion, perception, interpretation, performance, individual differences

Introduction

Music's ability to communicate emotional messages has fascinated generations of psychologists, musicologists, philosophers and anthropologists alike. Similar to speech, the structural aspects of music's acoustic signal can embody or convey emotional qualities to listeners. However, in the case of music, there is an extra layer of complexity in the transmission of emotion, as the person who creates the musical notation is not necessarily the person who produces the acoustic signal. This is particularly true for classical music, where so much of the standard repertoire is played by many different interpreters — who are generally trained and recognized for their performances, rather than their compositions. Therefore, music is an interesting avenue through which to investigate the process of emotional communication. Although this capability is widely recognized, as Levinson (1996) argues, a musical performance is not simply the presentation of the work's written sound structure (Macritchie, Eiholzer, & Italiana, 2012, p.179), but represents the combination of score and performers' interpretative



[Figure 1. Communication model between composer, performer and listener as applied to this project. Adapted from Blom, Bennett, & Stevenson (2016) and Juslin's Lens Model (1997)] decisions. Although a composer's intended emotional messages are conveyed in the structural cues codified in musical scores, performers shape these cues through their interpretation. Therefore, the ultimate listening experience of audiences is shaped by an intricate dialogue between composer and performer (Figure 1). This complex relationship is of great interest both theoretically and practically to a wide community of academics and researchers.

From a psychological perspective the interplay between structure and interpretation helps shed light on the complexities of the communication of coded emotional messages from multiple sources. Practically it is of great importance for performing musicians who are continually striving to put their unique stamp on frequently performed classics. Consequently, several instructional books help guide musicians on this important topic (Kramer, 2010; Silverman, 2008; Sinn, 2013; Thom, 2003). Interpretation is the lifeblood of musical performance — one of the most important dimensions along which performers are evaluated in public performances, recordings, juries and formal exams. Previous research exploring the complex relationships between cues and perceived emotion provides useful insight by examining how structural and interpretative aspects of music together shape emotional messages (Quinto, Thompson, & Taylor, 2014). Although the use of structural and interpretative cue combinations allows listeners to more accurately decode conveyed emotion on average, certain emotions are successfully transmitted with structural or composition cues alone, particularly in monophonic or single-lined melodies. As such, much is still needed to unpack the complex relationships between these two crucial aspects of musical listening. This is particularly challenging for the kinds of complex polyphonic music that defy easy deconstruction.

Performer Interpretation and Individuality

In the evaluation of the relative contributions of composers and performers to musical listening, Repp (1992), argues for a conceptual framework containing two basic aspects: normative and individual. The normative aspect is expected from performances across different artists, the individual aspect represents deviations from a single ideal norm of the musical score. This suggests performers can represent not only the written musical composition but include or add their interpretative expression.

Interpretation is considered an individualistic process based on a performer's musical intentions (Palmer, 1997). In a musical performance, performers make decisions regarding expressive cues to convey or emphasize emotional content. Studies using expert performers captured on commercially available recordings have explored these expressive differences across performers' use of timing cues showing variations in their interpretative decisions (Repp, 1992; Vines, Krumhansl, Wanderley, & Levitin, 2006; Dodson, 2011). These studies provide a useful starting point for empirically exploring differences in performer interpretation.

A performer's interpretative use of cues is embedded in a musical performance; thus, the auditory signal can transmit specific information regarding a performer's identity. In a study looking at 28 performances of Robert Schumann's "Träumerei" by 24 pianists, Repp (1992) analyzed interonset intervals of each performance to examine variations in global tempo choices in addition to overall performer timing profiles. Tempo varied across performers, however principal component analysis (PCA) on timing profiles revealed only a single factor that reflected some conformity in performances as all pianists observed the major ritardanadi of the piece. Variability in the PCA increased at lower levels of the structural hierarchy; as the analysis focused on fewer bars of music (i.e., smaller segments), more independent factors emerged. As such, patterns demonstrated performer differences for timing decisions evident at local levels of the musical structure. This is consistent with results from Dodson (2011) demonstrating heterogeneity in expressive timing within musical bars, despite a global similarity across many recordings of the same piece. An artist's individual interpretation can reflect variations of cues within the constraints "representing conscious or unconscious transgressions of the boundaries established by musical convention" (Repp, 1992). Although that paper explores acoustical difference in a signal based on performer interpretation, it does not address listeners' awareness of and sensitivity to these cues.

The current study builds upon that focus on performer interpretation by exploring the perceptual consequences of performer differences in performance. Additional work investigating influences of interpretation on performance cues indicate the most distinct differences of tempo and dynamic choices occur at a microstructural level, in contrast to a global one. For example, over the course of three performances of Chopin's Prelude in E minor Op.28 and the B Flat minor sonata Op.35, pianists produced similar overall tempo variations — but varied in the amount of variation as well as their consistency in shaping the timing of each bar (Macritchie et al., 2012). Dynamic choices also varied by performer, where the tempo and dynamic shaping across bars of the music appeared unique to each performer.

Following the performance analysis, participants notated their understanding of the phrasing, showing interpretive decisions by pianists affected listeners' understanding of the music's structure. Specifically, listeners identified major sectional boundaries differently within Chopin's Prelude in E minor based on performer. The pianists used expressive tempo and dynamic deviations to bring attention to specific elements, shaping listeners' experience of its structure. Thus, although interpretative differences of timing and dynamic cues occurred at the microstructural level, the placement of even small variations affect a listeners' understanding of the composition.

Listener Perception of Interpretative Decisions

The expressive choices made by a performer not only represent an artist's individuality but have the ability to shape a listener's perception of the musical experience. Performer interpretation and variation of cues such as dynamics (Nakamura, 1987) and ornamentation (Timmers & Ashley, 2007) are salient to listeners, and the individual decisions of specific performers play a key role in distinguishing performances of historically significant pieces — which have been played and recorded repeatedly. Timing information plays a particularly important role in such evaluations, as it helps listeners recognize a performer's expressive style. For example, cues such as tempo and articulation help listeners identify different performance styles (Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011). There, authors investigated whether participants could group together excerpts from the same performer and further explored the acoustical parameters listeners use to discriminate between interpretations. Regardless of the level of musical expertise, listeners correctly identified performers' individual style from excerpts as short as 10-14 seconds. Performer expertise appeared important as listeners more accurately categorized excerpts performed by those who had won prizes in musical competitions. Analysis of the acoustic information indicated mean tempo, articulation and onset synchrony varied as a function of performer, which presumably provided the acoustic basis on which listeners discriminated between different performance styles.

With respect to the listener's experience, performer interpretation influences how specific emotions are perceived by listeners. In a study from Juslin (2000), three professional guitar players recorded renditions of three selected songs with the intention of conveying 'anger', 'happiness', 'sadness' or fear. Focusing on three acoustical parameters of sound level, tempo and articulation, the authors reported large variability among performers for tempo, and differences in the use of articulation cues between interpretations. Listeners decoded these variations, which allowed them in many cases to accurately recognize the emotional intentions of the performer. Point biserial correlations of performers' intentions with listeners' affective ratings represented as achievement scores indicated variation based on to emotion and showed 70% of variance in listener's judgements could be accounted for by the performers' expressive intention. Individually, differences emerged for emotional expressions of happiness and fear, where performer 3 displayed significantly lower achievement scores than performer 1 or 2. Interestingly, other performers also displayed success in conveying the same emotions however using different performance strategies on cue use. This supports the proposed modified lens model from Juslin (1997), which hypothesizes that in the communication of emotion through music, cues function in a probabilistic and redundant way. The lens model helps to explain why the same emotion can be conveyed with different combinations of cues. Despite this general success, investigation and analysis on rating differences for varying interpretations needs further exploration to better understand how performer variations shape listener experiences. In addition, the use of explicit instructions to encode specific emotions into the musical stimuli — a method frequently used in the literature — may over represent how cues are frequently used to communicate emotion. The current study takes a naturalistic approach to account for the subtleties in performer's expressive decisions to convey emotion.

The present study

Here we complement and extend previous work exploring the effect of performer interpretation on listener perception. Our exploratory study examines the relationship between performer's interpretive decisions and listeners' evaluations of emotion — all within the context of performing the same set of widely-acclaimed pieces. Similar to the studies from Repp (1992) and Macritchie et al. (2012) above, we chose to use commercially available recordings of a renowned musical work to examine the natural variations in musical interpretation by professional musicians.

In order to investigate the effect of performer interpretation on listener perception, we used well respected recordings of Bach's *Well Tempered-Clavier* (*WTC*). We selected interpretations of seven notable pianists — chosen from a survey of important performances undertaken by Willard A. Palmer (Bach, 2004), a Baroque scholar and musicologist, who analyzed performance interpretation including tempo classifications for a particular set of albums. Using these landmark recordings, we explore how evaluations of perceived emotion differ based on performer, offering unique insight into the relationship between cue weights and individual interpretation. Many consider the *WTC* a teaching tool for musicians to gain skill in technique as well as decision making in performance, as

Bach did not include tempo markings in this set of works. The significance of this work in musical training is perhaps why the Royal Conservatory of Music includes many of its pieces in their exam requirements (Royal Conservatory of Music, 2015), and further why it may be an important set of pieces to use when exploring how musicians are taught to encode individuality and emotion in their performances.

Beyond its critical acceptance by music educators, the *WTC* is particularly well suited for exploration of emotional communication it contains 48 pieces balanced in modality across all chroma keys. This is important as it affords a type of control over the cue of modality, a cue empirically proven to be important in conveyed emotion in music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Eerola, Friberg, & Bresin, 2013; Quinto et al., 2014). This study thus builds on work of Macritchie et al., (2012), Repp (1992) exploring the effect of performer interpretation on the conveyed emotional content, and incorporates the analyses of composer and performer controlled quantified cues to break down the importance of cues for interpreting each performance.

Expanding on the analysis of the structural cues alone (Poon and Schutz, 2015), here we focus on their perceptual consequences — particularly with respect to interpretative choices by pianists. Evidence from our previous study on the perception of emotion suggested the timing cue of attack rate as an important predictor for listener perception. Cues of modality and pitch height information followed in importance respectively, for both emotional ratings of valence and arousal (Battcock & Schutz, 2019). One crucial aspect of interpretation exhibiting considerable difference of opinion is timing information (Repp 1992; Gingras et al., 2011), therefore we are interested in assessing how performers' choices in attack rate (note attacks per second) influence listener experience of the same pieces, and how those decision impact cue importance in a model containing both expressive and structural cues.

Given the exploratory nature of this investigation, we do not have specific hypotheses. Instead, our goals in undertaking this series of experiments are to (1) explore the degree to which the different interpretations of expert pianists affect the emotional messages of a highly regarded set of pieces by J.S Bach, (2) examine whether arousal or valence systematically differs more across different interpretations, and (3) explore trade-offs in the use of different cue for emotion between different performers. Together this novel exploration of how natural differences in interpretation affect listener responses will help connect previous work looking primarily at either (a) natural differences in the acoustic signal alone (i.e. Repp, 1992), or (b) whether systematic variations in performance are detected by listeners.

General Method

Overview

We used the following experimental design and procedures to test six new experiments using recordings from six different pianists. All apparatus, stimuli, and procedures remained identical to those used to assess responses to Friedrich Gulda in Battcock & Schutz (2019). Therefore, combining these data yields information for all seven pianists reviewed in Palmer's analyses (Bach, 2004).

Participants.

The sample comprised of a total of 180 (mean age = 18.8, SD=1.77) non-musician undergrads (n=30 for each experiment) from McMaster University's Psychology undergraduate participant pool. We considered 'non-musicians' to have <1 year of musical training. In return for their participation, participants received course credit. The experiment met ethical standards in accordance to the McMaster University Research Ethics Board.

Musical Stimuli.

Stimuli included 48 excerpts of J.S Bach's *WTC*'s Book 1 as performed by one of six notable performers: Edwin Fischer (Bach, 2007), Glenn Gould (Bach, 1993), Sviatoslav Richter (Bach, 1992), Rosalyn Tureck (Bach, 1953), Joao Carlos Martins (Bach, 1964) and Jöerg Demus (Bach, 1956). Excerpt durations ranged from 6 to 104 seconds in duration (M=28.79 seconds, SD=12.46), contained the first eight measures of each *WTC* piece, and included a two-second fade out starting at the ninth measure. Descriptive information on individual performer excerpts can be found in Appendix A. Although faster and slower pieces varied in duration, this provided consistency in terms of musical units (measure length). We used Amadeus Pro to cut and prep the stimuli for experimental testing.

Cue quantification.

As used in Battcock & Schutz (2019) we calculated values of pitch height and attack rate for all eight-measure excerpts across each performer. Pitch height values represented the average weighted pitch quantified from the eight measures. These values are obtained using the method in Poon & Schutz (2015), by summing duration-weighted pitch values of each measure, divided by the note duration sum of that measure. The calculated pitch height values varied from 33.13-53.00 (M=43.90, SD=4.03) corresponding ~F3 to ~ C#5. Attack rate (average note attacks per second for each eight measure excerpt) quantification used tempi values reflecting each performers performance. Overall attack rate values ranged from 0.87 to 12.80 note attacks per second (M=4.49, SD=2.46). Specific performer attack rate information can be found in Appendix A. In this study we defined mode as the tonal center of the piece, indicated by the denoted key signature of each score and coded it dichotomously (0= minor, 1= Major). Additionally, because we used commercial recordings of multiple performers, we quantified the root mean square (RMS) value of each audio file to measure the intensity of each excerpt to use as a covariate in our analyses. We obtained RMS values using Amadeus Pro software. RMS values of all excerpts ranged from -

40.6 to -16.6 (M=-27.7, SD=4.8). Details on RMS values for each excerpt can be found in Appendix A.

Design and procedure.

All experiments occurred in a sound-attenuating booth. Participants completed a consent and musical experience form (see Appendix B) before starting the experiment. Research assistants verbally explained the emotion rating tasks emphasizing full use of the scales. Instructions asked participants to provide ratings of the emotion conveyed through the music on scales of emotional valence (how positive or negative the emotion sounds) and arousal (the emotional energy communicated). Participants completed four practice trials using alternate recordings as performed by Angela Hewitt (Bach, 1998) with opportunity to ask procedural or clarification questions before the experiment began. Following presentation of each excerpt, participants rated perceived emotion on scales valence from 1 (negative) to 7 (positive), and arousal from 1 (low) to 100 (high). Each participant listened to an individually randomized order of the 48 excerpts.

The experiment ran on PsychoPy (Peirce et al., 2019), a Python-based program on either a 2014 MacBook Air (OS X 10.9.4) or a 2013 iMac (OS X 10.9.3) connected to a DELL monitor within the booth. Participants heard stimuli at a consistent and comfortable listening level through Sennheiser HDA 200 headphones and provided responses via either trackpad or Apple mouse connected to the iMac located outside the booth.

Results

Here we compare participants' ratings of emotional valence and arousal across seven experiments with seven performers — six newly tested and one previous from Battcock & Schutz (2019). Overall, mean valence ratings (M=4.07, SD=1.14) ranged from 1.41 to 6.3 and mean arousal ratings (M=55.16, SD=16.67) ranged from 19.57 to 87.17. Descriptive information on ratings for each performer's stimuli set can be found in Appendix C.

Analysis of Variance.

We conducted two, two-way between subjects' ANOVAs to compare the effect of excerpt and performer on ratings of valence and arousal with RMS values as a covariate. Equal variances cannot be assumed for our outcome variables as Levene's Test of Equality is violated for valence [$F(335,9695)=2.11$, $p<0.001$] and arousal [$F(335,9695)=2.46$, $p<0.001$], therefore all effects will be compared at the $p=0.001$ level. ANOVA results indicate there is a significant effect of excerpt [$F(47,9694)=121.12$ $p<.001$, $\eta_p^2=.370$] and performer [$F(6,9695)= 6.11$, $p<.001$, $\eta_p^2=.004$] (Figure 2) as well as an interaction effect between excerpt and performer [$F(281,9695)=3.08$, $p<0.001$, $\eta_p^2=.082$] on judgements of emotional valence conveyed (Table 1).

For ratings of arousal, there is a significant effect of piece [$F(47,9695)=57.36$, $p<0.001$, $\eta_p^2=218$], performer [$F(6,9695)=34.21$, $p<0.001$, $\eta_p^2=0.02$] (Figure 3) and an interaction effect between excerpt and performer [$F(281,9695)=3.48$, $p<0.001$, $\eta_p^2=.09$] on participants emotional judgements (Table 2).

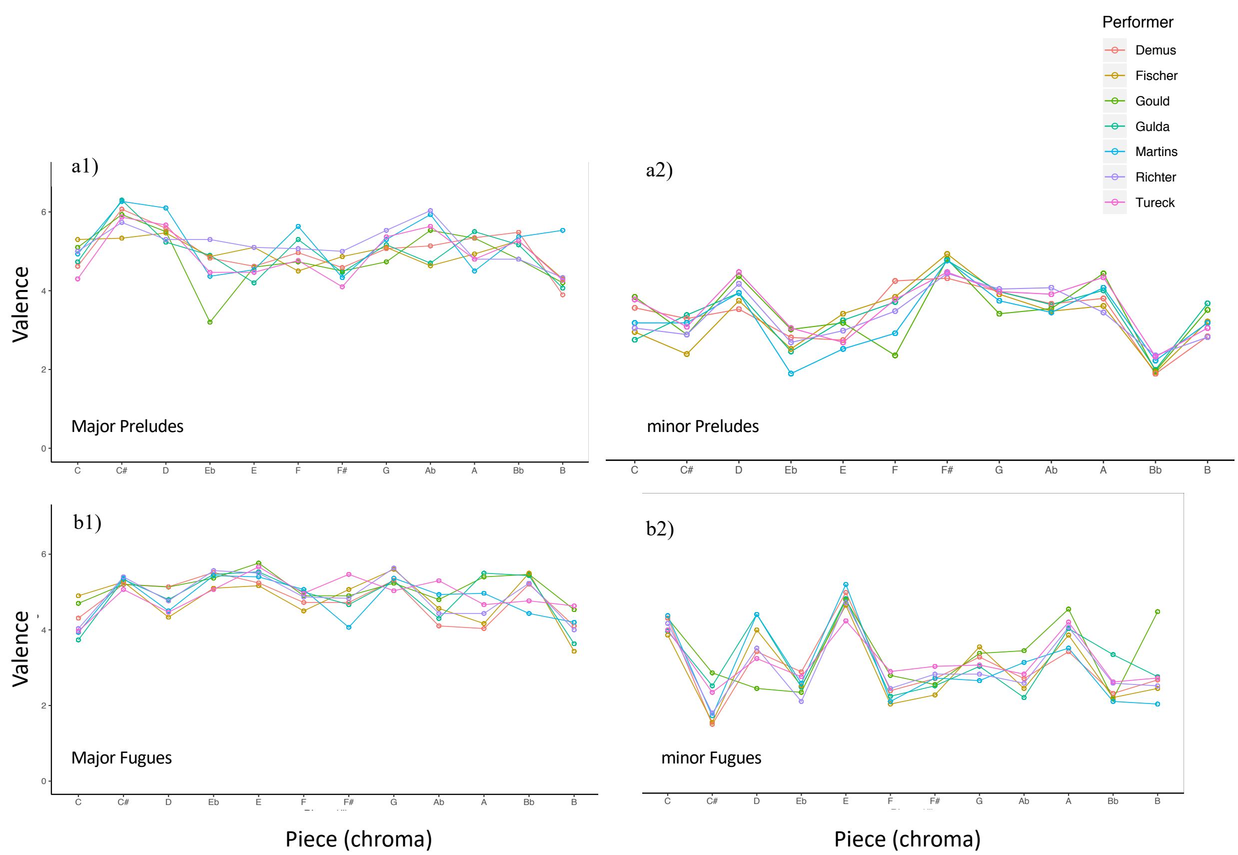


Figure 2. Mean valence ratings of all 48 pieces in the *WTC* across seven performers/seven experiments broken down into a1) Major preludes, a2) minor preludes, b1) Major fugues & b2) minor fugues. Each line represents mean ratings for one performer.

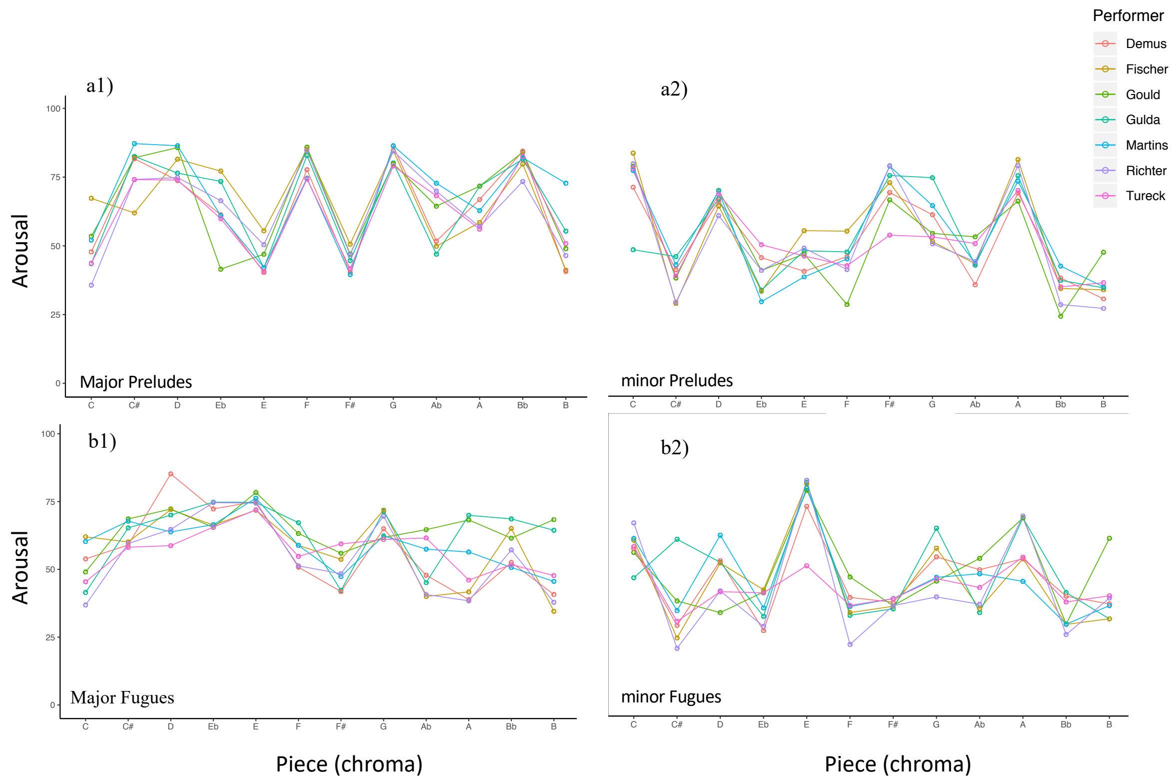


Figure 3. Mean arousal ratings of all 48 pieces in the *WTC* across seven performers/seven experiments broken down into a1) Major preludes, a2) minor preludes, b1) Major fugues & b2) minor fugues. Each line represents mean ratings for one performer.

Table 1

ANOVA results using valence rating as a criterion and RMS as a covariate

Predictor	Sum of Squares	df	Mean Square	F	*p*	Partial η^2
Intercept	6607.23	1	38.65	28.53	.001	.335
Performer	49.70	6	8.28	6.11	.001	.004
Piece	7713.81	47	164.12	121.12	.001	.370
Performer*Piece	1173.00	281	4.174	3.08	.001	.082

R squared= .496 (Adjusted *R* Squared=.479)

Table 2

ANOVA results using arousal rating as a criterion and RMS as a covariate

Predictor	Sum of Squares	df	Mean Square	F	*p*	Partial η^2
Intercept	1873797.84	1	8328.72	4977.48	.001	.339
Performer	77272.49	6	12878.74	34.21	.001	.021
Piece	1014838.40	47	21592.31	57.36	.001	.218
Performer*Piece	268282.31	281	1310.54	3.481	.001	.092

R squared= .433 (Adjusted *R* Squared=.414)

Variance of listener response.

As used in Quinto & Thompson (2013), we calculated the coefficient of variation (CV) to examine the range of valence and arousal ratings. This measure is calculated as the (standard deviation/mean) x 100 and represents a standardized measure of dispersion. According to listener ratings of valence and arousal, the CV for valence (CV= 39.56%) is lower than what is observed for arousal ratings (CV = 46.03%) when calculated across performers (Figure 4). These findings suggest that the influence of performer interpretation may have had a stronger effect on ratings of arousal, as ratings across performers are more variable for this dimension of perceived emotion than for ratings of valence.

We also calculated CVs for each performer across ratings of valence and arousal (Table 3). As with our overall calculation, the CV for arousal ratings remained higher than for valence ratings as seen across performers. The range of CVs for valence ratings across performers is approx. 4%, suggesting the level of consistency in valence ratings across pieces is similar across performers. For ratings of arousal, CVs across performers varied 10% from the highest to the lowest CV calculated. Using the R package 'cvequality' (Version 0.1.3; Marwick & Krishnamoorthy, 2019), we used the 'Modified signed-likelihood ratio test' (Krishnamoorthy & Lee, 2014) for equality of CVs of valence and arousal across performers. This test indicated variation in CVs across performers for valence ($\chi^2(6)=48.49$, $p<0.001$) and arousal ($\chi^2(6)=24.54$, $p<0.001$) is not due to chance. This indicates more variability in how some performers expressed arousal and valence across the pieces in the *WTC (Figure 5)*.

Table 3

Coefficients of Variation for each performer across valence and arousal ratings of all 48 excerpts.

Performer	CV Valence	CV Arousal
Demus	29.62	40.85
Fischer	33.01	43.50
Gulda	30.27	33.79
Gould	31.71	39.79
Martins	29.09	37.78
Richter	28.70	40.14
Tureck	30.62	38.91

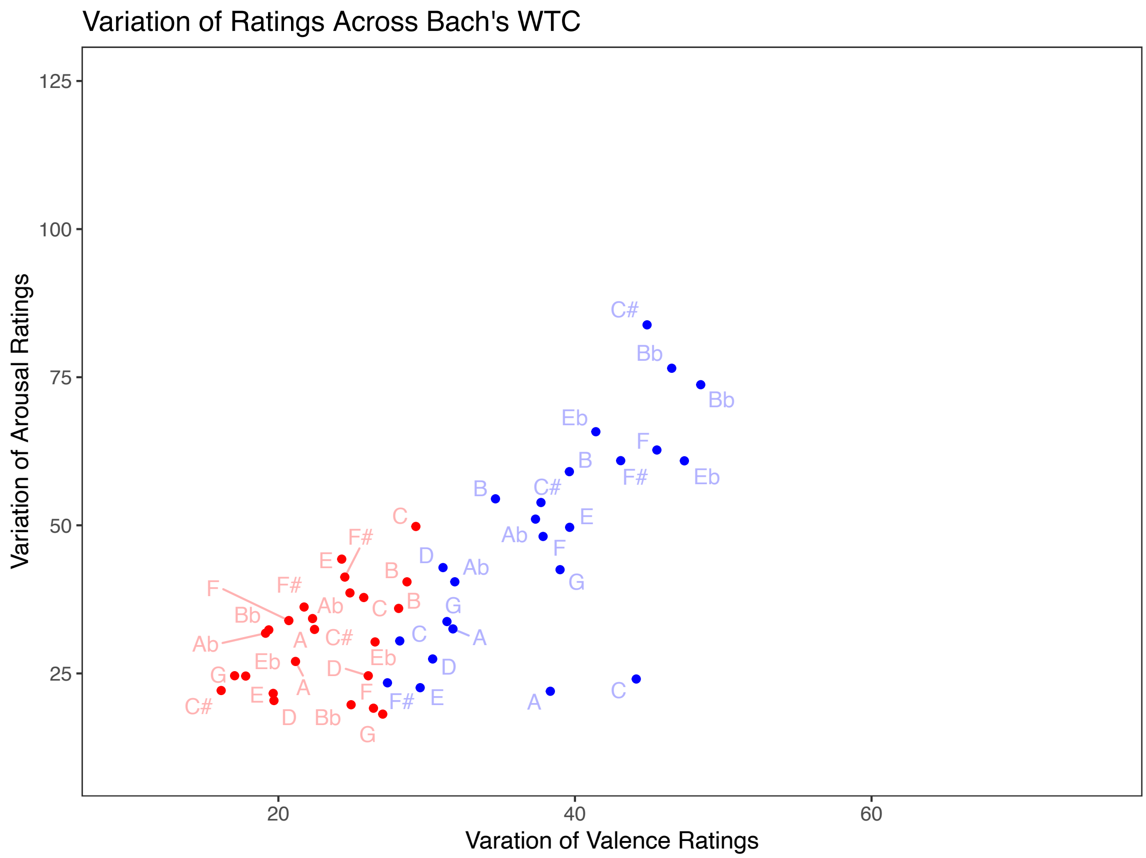


Figure 4. Coefficients of Variation for valence and arousal ratings across all performers for each excerpt. Major key excerpts are in red, minor key excerpts in blue. Each excerpt is denoted by their chroma key.

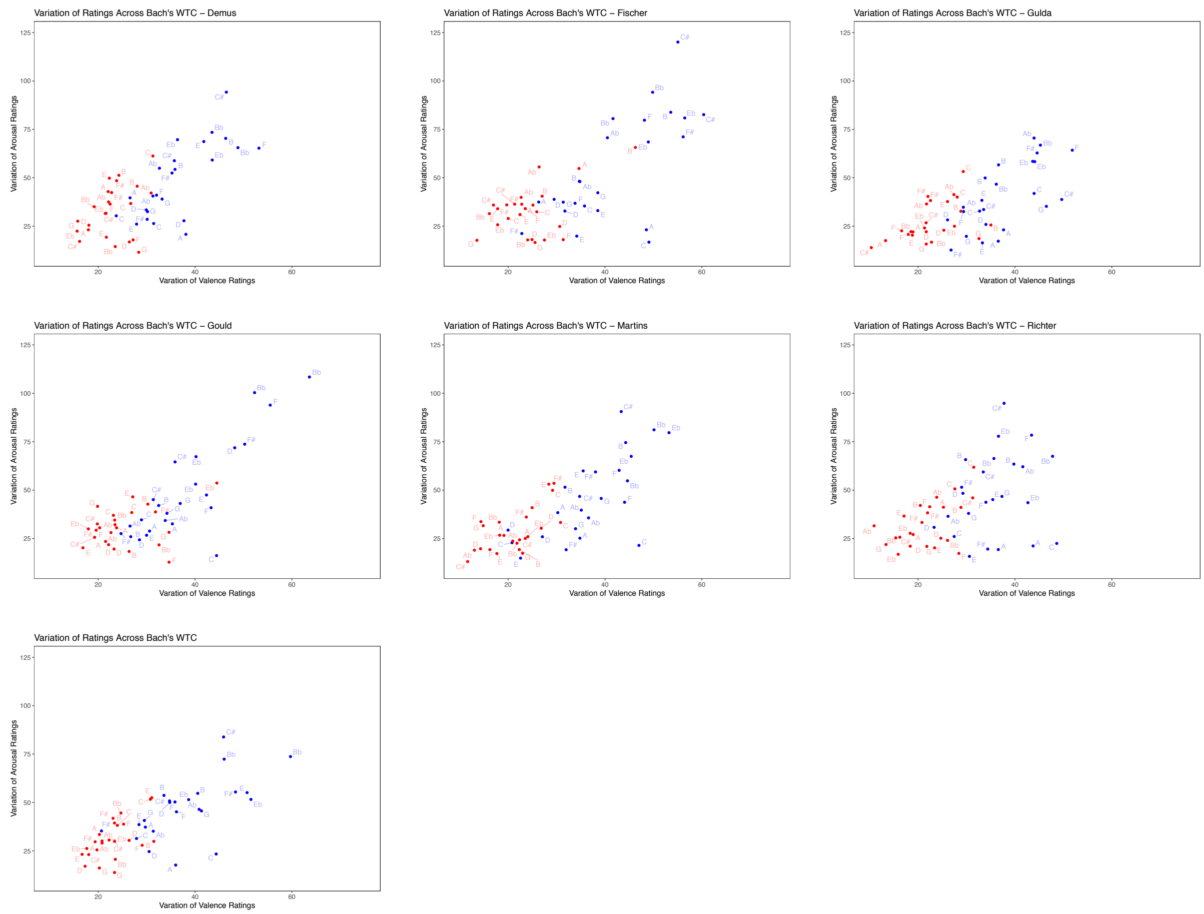


Figure 5. Coefficients of Variation for valence and arousal ratings across all excerpts for each performer. Major key excerpts are in red, minor key excerpts in blue. Each excerpt is denoted by their chroma key.

Linear regression analysis.

We used the R Statistical Package to run standard simultaneous multiple linear regression analysis to assess the overall influence of four predictors — attack rate, modality, pitch height and RMS — on mean ratings of valence and arousal. This helped to evaluate the impact of our selected musical cues and any differences in recording intensity in predicting ratings across experiments.

Further, we ran individual regression and commonality analyses for each experiment with individual performers in order to create a ‘cue profile’, displaying relative cue weights for each performer. For the dichotomous cue of modality, we chose the Major mode as the reference level for the analysis (thus the minor mode level of this factor is contrasted against it in analysis). The regression analysis on valence ratings indicated the cues of attack rate, modality, pitch height and RMS significantly predicted listener responses (Table 4). According to the model for arousal ratings, attack rate, pitch height and RMS emerged as significant predictor of listener responses.

Across the seven experiments, the three-cue predictor model accounted for 80.8% of the variance in listener ratings of valence (Table 4). In contrast, the same predictor model accounted for 78.8% of variance in ratings of arousal. Across individual regression analyses for each performer, three cue models of valence predicted 73.5 to 82.0% of rating variance and 50.7 to 78.8 % of variance in arousal ratings (See Table 5).

Table 4.

Regression model for attack rate, mode, pitch height and RMS on valence and arousal ratings across all performers. Beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings. Default state for mode is Major.

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.388	0.005	72.65	p<0.001	0.674	0.006	120.20	p<0.001
Modality	0.531	0.005	105.48	p<0.001	0.003	0.005	0.518	p=0.605
Pitch Height	0.173	0.005	37.61	p<0.001	0.076	0.005	15.73	p<0.005
RMS	0.198	0.005	39.48	p<0.001	0.356	0.005	67.52	p<0.01
R^2			.808				.788	
F			10520				9298	

Commonality Analysis Across Performer-Specific Models.

As used in Battcock & Schutz (2019) we calculated the overall contributions of each cue in participant ratings of emotion, using commonality analyses to decompose the R^2 of each performer's three cue predictor model (Table 5 & 6). Note here, regression models contain only three predictors (attack rate, mode and pitch height) as we are looking at the models of listener ratings for each individual experiment and are not comparing RMS values across performers. Contributions are broken down into shared and unique contributions to predict participant response. We also plotted this cue profile decomposition in Figure 6 and 7, where we used bootstrapping methods to calculate error bars for each cue. To obtain these error bar values we used a resampling with replacement for 1000 runs with a sample of 30 (our actual sample included 30 participants) for each performers' dataset.

Valence Ratings. Across performers we see similar trends of cue strengths (Figure 6), where mode (21-29%) accounted for the largest amount of variance uniquely. The effect of mode is evident in the circumplex visualization (Figure 8), where major key pieces fall predominately into the upper right quadrant. Uniquely, attack rate predicted 14.0-17.2% of variance in the model and pitch height uniquely contributed 1.4-4.3%. The largest contribution emerged from the relationship between attack rate and mode (33-37%). Attack rate and pitch height accounted for -1.4 – -3.3% of shared variance (Table 4), mode and pitch height accounted for 4.6-7.8% of variance and variance common between all three cues explained -2.2 - -5.3% of variance in valence ratings.

Arousal Ratings. Again, similar trends emerged across performers for cue weights predicting variance in arousal ratings (Figure 7). Attack rate uniquely predicted the largest amount of variance (30.1-64.1%). Cues of modality and pitch height uniquely predicted 0.01-1.8% and 0-2.1 of variance respectively. In terms of shared contributions, attack rate and mode remained the biggest predictor of shared variance (11.7-21.9%) compared to attack rate and pitch height (predicted -

1.9-3.0%) and mode and pitch height (predicted -0.30-0.73%). The shared contribution of all three cues predicted -2.26-0.22% of variance in arousal ratings (Table 6).

Table 5

Commonality Analysis for Variance in Listener Ratings of Valence Across Seven Performers

		Fischer	Gulda	Gould	Richter	Tureck	Martins	Demus
Unique X_1	Attack Rate	.1504	.1715	.1628	.1397	.1854	.1942	.1938
Unique X_2	Modality	.2390	.2114	.2226	.2867	.2579	.2564	.2427
Unique X_3	Pitch Height	.0380	.0438	.0225	.0233	.0163	.0139	.0314
Common to X_1 & X_2	C (AR, Mo)	.3523	.3282	.3213	.3559	.3676	.3294	.3494
Common to X_1 & X_3	C (AR, PH)	-.0325	-.0312	-.0186	-.0178	-.0159	-.0139	-.0263
Common to X_2 & X_3	C (Mo, PH)	.0781	.0626	.0463	.0545	.0607	.0466	.0586
Common to X_1, X_2 & X_3	C (AR, Mo, PH)	-.0491	-.0271	-.0219	-.0241	-.0525	-.0351	-.0306
	$R^2_{y.123}$=	.7762	.7593	.7350	.8181	.8195	.7916	.8190

Table 6

Commonality Analysis for Variance in Listener Ratings of Arousal Across Seven Performers

		Fischer	Gulda	Gould	Richter	Tureck	Martins	Demus
Unique X_1	Attack Rate	.5688	.3019	.4794	.6410	.5557	.5508	.5206
Unique X_2	Modality	.0001	.0181	.0168	.0042	.0069	.0096	.0006
Unique X_3	Pitch Height	.0205	.0012	.0000	.0196	.0131	.0019	.0001
Common to X_1 & X_2	C (AR, Mo)	.1622	.1740	.2342	.1165	.2187	.1982	.1488
Common to X_1 & X_3	C (AR, PH)	-.0097	.0287	.0230	-.0185	.0091	.0250	.0298
Common to X_2 & X_3	C (Mo, PH)	.0010	-.0012	.0012	-.0030	.0073	.0029	.0002
Common to X_1, X_2 & X_3	C (AR, Mo, PH)	-.0096	-.0162	-.0176	.0022	-.0226	-.0197	-.0158
	$R^2_{y.123}$=	.7333	.5065	.7371	.7620	.7881	.7687	.6843

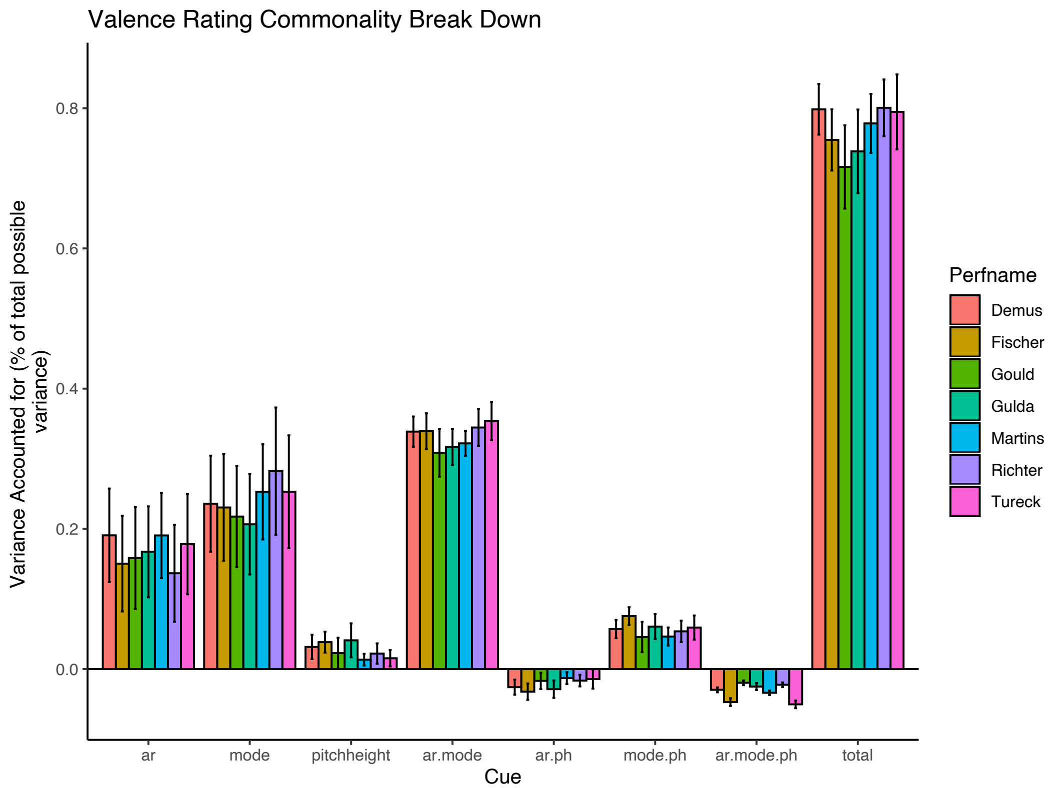


Figure 6. Unique and shared variance of valence ratings by musical cue. Individual bars depict cue weights for each performer. Attack rate and modality explain the vast majority of variance, although specific contributions vary. Error bars represent 95% confidence intervals on bootstrapped data.

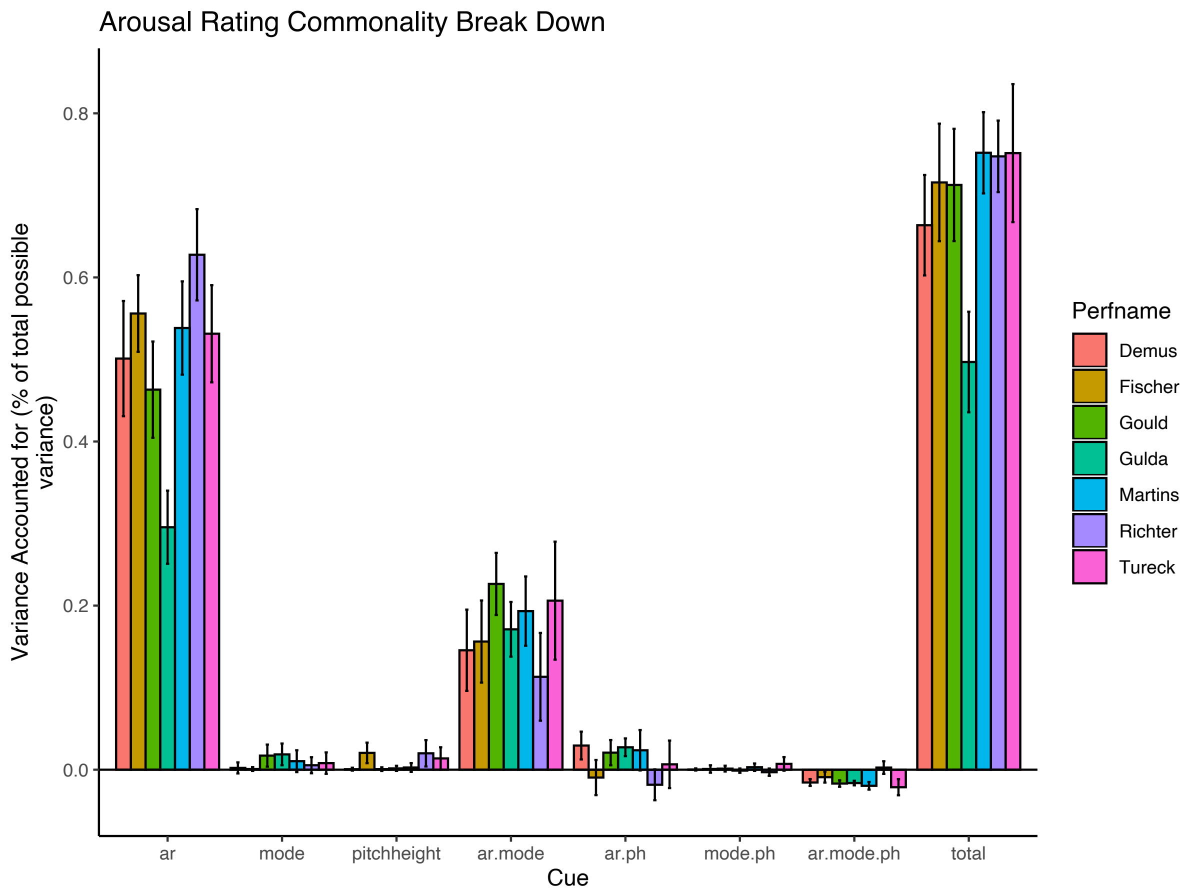


Figure 7. Unique and shared variance of arousal ratings, with individual bars illustrating cue weights for specific performers. Attack rate plays a crucial role, although the relative strength of its shared contribution with attack rate varies as a function of performer interpretation. Error bars represent 95% confidence intervals on bootstrapped data.

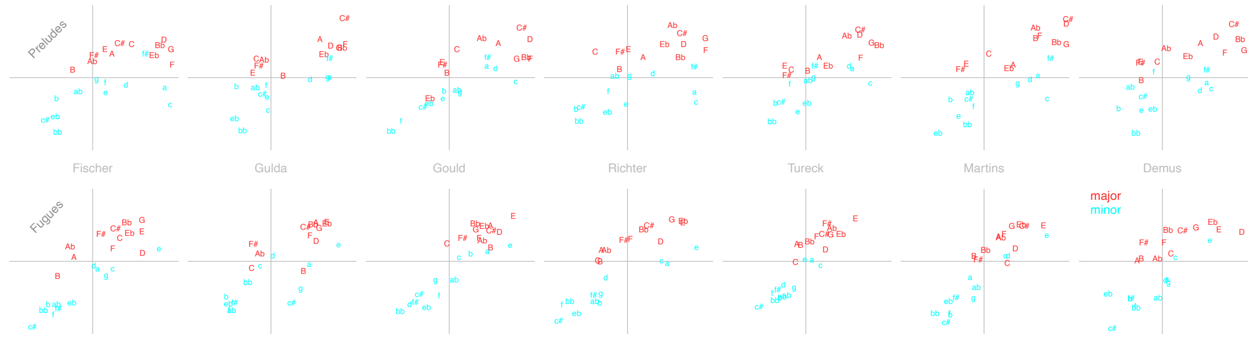


Figure 8. Mean ratings for all 48 pieces in the *WTC* (separated by preludes and fugues) plotted across the 2D circumplex space for several performers (a) Fischer, (b) Gulda, (c) Gould, (d) Richter, (e) Tureck, (f) Martins & (g) Demus. Major key pieces are represented in red, minor key pieces in blue.

Discussion

The significant effect of both excerpt and performer in our ANOVA are consistent with previous findings that musical emotion reflects a combination of structural and interpretive cues (Eerola et al., 2013; Schubert, 2004). Although the effect of performer interpretation is small (η_p^2=.004), the significant interaction between performer and piece as well as the performer profiles (Figure 2 & 3) illustrate interpretive differences vary by piece (for example, Eb Major Prelude and the B Major Prelude). We suspect this reflects a complex interaction between structure and interpretation, with the complexity of some pieces leaving little room for large variations in interpretation. Similarly, structural considerations might lead to greater or lesser agreement across performers interpretations. Although we cannot speak to the direction of the interaction from our results, we can see from further visual inspection the effect of interpretation and piece on conveyed emotion is a nuanced relationship involving performer interpretation that can be executed with the musical structure given.

These results raise interesting questions about the complex communicative collaboration between composer and performer in conveying emotion — an issue previously explored by Quinto, Thompson, & Taylor (2014). The coefficients of variation for valence (39.6%) and arousal (46.0%) indicated significantly more variation in participants' rating of arousal than valence. This finding is similar to that of Quinto and Thompson (2013) who found higher variation in judgements of perceived emotion in excerpts containing only salient performance cues (excerpts had been composed to be emotionally ambiguous). There, the authors argue this provided evidence performance cues had greater influence on the communication of arousal. That study demonstrated when both compositional cues and performance cues are available to listeners, the CV for each dimension appeared similar (29.94% valence, 25.78% arousal). In contrast, our findings indicate a difference in valence and arousal CVs for excerpts containing a combination of compositional and performance cues. Unlike Quinto and Thompson (2013) who used monophonic, or single lined melodies composed to represent specific emotions, we used precomposed music from a renowned composer often used as a teaching tool for performers. It's possible the use of such complex, polyphonic music with untrained listeners caused them to have a greater reliance on performance cues than structural ones which led to greater sensitivity in interpretative choices across pieces.

Interestingly, when we visually examine CVs for valence and arousal for each of the 48 excerpts (Figure 4 & 5), we see a bifurcation of mode where minor key excerpts appear to have more variation across both dimensions ($\chi^2(1)=31.69$, $p<0.001$ for valence ratings and $\chi^2(1)=20.01$, $p<0.001$ for arousal ratings) than those in major keys. From these results, the question emerges as to whether performers' choice of attack rates is more consistent when interpreting major key pieces than minor. Here we find minor key excerpts (CV=61.5%) are found to be significantly ($\chi^2(1)=13.66$, $p<0.001$) more variable in chosen attack rates than major key excerpts (CV=42.8%), which may support the hypothesis of consistency in interpreting major key pieces. Furthermore, we see this division across all performers to varying degrees, which may reflect an effect of the compositional structure of major vs minor key pieces on the expressive choices in the *WTC* (Table 3). Given our results show minor key pieces are indeed rated lower in perceived valence than major key pieces ($t(10079)=-256$, $p<0.001$), this may reflect the complexity and range in expressing or perceiving (Hunter, Glenn Schellenberg, & Stalinski, 2011; Laukka & Juslin, 2007) emotions that are low arousal, and negatively valenced. This is consistent with previous work on the perception of negatively valence emotions like sadness, nostalgia, or longing that are often confused or are less accurately identified than positively valenced emotions like happiness (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013).

We see this division of mode occur across CVs calculated for ratings of each performer's recordings. This may have occurred as result of stimuli differences, where Quinto and Thompson (2013) used melodic fragments consisting of seven to nine notes, each composed to communicate a specific emotion. In contrast the excerpts in our study are more reflective of passages in actual music with more complex polyphony, emotional ambiguity, and mixture of major and minor chords. Given this complexity, our listeners with little to no musical training may have fixated more on performance cues when judging emotion. Additionally, we saw that CVs for arousal varied more across performers than for valence, further suggesting effects of expressive choices as a stronger influence for perceived arousal than perceived valence.

We ran regression and commonality analyses for each performer experiment to compare cue weights across interpretations. Generally, cue profiles followed similar patterns to Battcock and Schutz (2019), with all three cues predicting listener ratings of valence, and only attack rate emerging as a predictor of ratings of arousal. However, the fit of each regression model varied across experiments, suggesting our three cues are more predictive for some performers than others. Models of valence ratings explained from 73.5% to 82.0% and models of arousal ratings explained from 50.7% to 78.8% depending on the performer recording used. The variation of model fits are greater for arousal ratings than valence ratings, supporting the hypothesis that compositional cues such as mode and pitch are more important for valence than arousal (Quinto & Thompson, 2013).

The commonality analyses allowed for a detailed breakdown of the relative cue weights with respect to unique and shared variance in our model. This is important as there is a relationship between attack rate and mode, where major key pieces of this set are faster than minor key pieces ($t(46)=-3.2419$, $p<0.05$) as presented in Battcock and Schutz (2019). The cue profiles calculated from the commonality analyses indicate that for ratings of arousal, performer's expressive choices of attack rate are more salient to listeners for specific performers, where the predictive power of attack rate to explain variance uniquely and independently of other predictors varies based on the performer recording used (30.2% to 56.8%), which resulted in the variation of model fit. As a consequence of the interpretive choices of attack rate, the shared variance explained by attack rate and mode also differed across performer recording (14.9% to 21.9%). It is important to note that it is possible that given our limited number of expressive cues in the model, performers' manipulation of other expressive cues like that of dynamics, articulation or timing fluctuations at a micro level have additional effects on perceived arousal that explain additional variance and might account for the variation in attack rate cue strength.

In addition, commonality analysis allows us to add insight in the debate on the role of mode in perceived emotion between music psychologists and music theorists. Empirical evidence using single lined pre-composed or experimentally designed excerpts demonstrate listeners perceive more positively valenced affective terms with the major mode, and negatively valenced affective terms when presented with minor modes (Dalla Bella et al., 2001; Hunter et al., 2008; Webster & Weir, 2005). Music theorists however reason mode's apparent influence is due largely to its pairing with other structural cues (Hatten, 2004). With our analysis we assessed the unique and common variance explained by mode and its relationship to timing information across seven difference performances. Here we show mode affects perception of valence, but not arousal. Further, with commonality analysis we separate out the relationship between mode and timing, observing variance commonly predicted by mode and attack rate, as well as uniquely predicted by mode. For ratings of valence, we see mode uniquely predicts a notable amount (21%-27%) of variance, even when the variance commonly predicted with attack rate (32%-37%) is considered.

General Discussion

The importance of expressive interpretation for a musician's career is paramount. Performers strive to differentiate themselves, often with the goal of creating emotionally captivating performances that speak to their audiences. The present data sheds light on the complex relationship between performers' interpretative decisions on a musical composition and the expressive outcomes judged by listeners. Focus of the perceptual consequences of performer differences often centers on the identification of interpretations or expressive aspects of a performance. Our work, similar to Juslin (1997), considers how performer interpretation leads to variations in the communicated emotional messages. Previous research has focused on the recognition or grouping of performer interpretations (Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011), or identifying the expressive phrasing used (Macritchie et al., 2012). Here we explore how performer differences influence the perception of emotion using a widely performed set of pieces that have been interpreted from many well-known and respected musicians. Rather than emotion recognition accuracy, our work focuses on how listeners ratings vary on dimensions of valence and arousal. Further, we discuss how our results touch on the role composer controlled (structural) or performer controlled (performance) cues play in communicating emotion through music.

Composer vs performer-controlled cues in conveyed emotion

Plotting the ratings of each piece across performers, we observed an influence of mode on perceived emotion for all interpretations (Figure 8). This raises interesting questions about the different contributions of composers and performers to emotional communication. Although composers' emotional intentions are encoded through the notation of structural cues — pitches, rhythms, notated tempi, etc., performers then study, interpret, and ultimately perform these notes and rhythms, adding additional layers to the emotional signal. Therefore, the relative weight of composer controlled versus performer-controlled cues is of broad interest with respect to emotional expression.

Studies that have investigated how listeners perceive emotion in music show composer cues are also influential to perceived emotion. In a more recent study from Juslin and Lindström (2010), 75-80% of variance associated with listener ratings of emotion could be explained from a linear combination of both performer and composer cues. Using synthesized musical stimuli composed for experimental purposes, the authors varied eight features (pitch, mode, melodic progression, rhythm, tempo, sound level, articulation and timbre) according to a factorial design. Musically trained participants rated musical stimuli along five affective adjective scales ('Happy', 'Sad', 'Angry', 'Fearful' and 'Tender'). Their regression analysis indicated composer-controlled cues contributed to perceived emotion, however they may be less influential than cues involved in performer expression. Our work demonstrates variance associated with valence and arousal respectively could be attributed to the combination of both composer-controlled and performer manipulated cues, where weights of these cues differed according to the dimension of emotion measured. Our findings are therefore consistent with Juslin and Lindstrom (2010), even with the use of stimuli that has not been systematically manipulated for each cue of interest. However, unlike their study, here we see that for ratings of valence the composer-controlled cue of mode appears more important than the performer-manipulated cue of attack rate which suggests a potential specialized function for each type of cue on perceived emotion.

In our study all listeners heard the same compositional structure — with different performers varying interpretative dimensions such as tempo, articulation, and dynamics. From our results, we found these interpretive cues led to more variation in perceived arousal than valence across performers and pieces. This is consistent with the idea that both composer (structural) and performer controlled (performance) features contribute to emotional responses (Quinto and Thompson, 2014). In that series of studies, listeners rated emotional valence and arousal for musician-composed excerpts either performed with a specific intended emotion or rendered neutral and expressionless through MIDI software. In order to quantify the composer and performer-controlled cues, musical excerpts were subject to acoustic analysis using Praat and MIRToolbox software. Overall, regression analyses on listener ratings indicated the cues predicted 58-59% of valence and arousal variance. The model included compositional cues such as mode, mean fundamental frequency, range (number of semitones between the lowest and highest frequencies), mean interval size, as well as performance cues of articulation, mean intensity level, intensity variability, high-frequency energy and tempo. Researchers found cues affected valence and arousal ratings differently; compositional cues such as mode had a greater influence on valence ratings whereas performance cues more strongly influenced arousal ratings. In addition, regardless of the intended emotion use of performance expression enhanced listeners' ratings of emotional valence and arousal. As such, a performer's interpretative decisions on expressive musical cues elicits a strong influence on the communicated emotion with a potentially larger influence over emotional arousal in music. We found similar results to those findings in this study as well as Battcock and Schutz (2019), as we see attack rate emerges as the only significant predictor of arousal ratings, where all three cues appeared significant for valence ratings. Our results had been found using polyphonic, or multi-lined precomposed music, which helps strengthen the argument that composer or performer manipulated cues have varying influences on the dimensions perceived emotion in music.

Examining the main goals of our studies

Although we did not have specific hypotheses, we had three areas of primary interest motiving this work. Firstly, we wanted to explore how different interpretations of pieces from J.S Bach, as performed by highly trained musicians would affect communicated emotions. Using a dimensional approach to conveyed emotions, we are able to compare how listeners rated perceived valence and arousal for each performer across all 48 excerpts of the *WTC*. Our second focus involved assessing what aspect of emotion — arousal or valence — is more affected by differences in interpretation. We hoped to extend on previous work of Quinto and Thompson (2013), which investigated the impact of structural or performance cues affected these dimensions of emotion. Finally, we wanted to look at the effect of interpretation on the relative weights of cues used by listeners to make assessments of emotion. The application of commonality analysis is beneficial to this goal as allows us to tease apart the cue relationships and thus compare their importance across each performer.

The interaction between structural and interpretive cues

Our first research interest focused on the question about how different interpretations of expert pianists affect emotion messages. Here our experiments bridged work measuring differences in perceptual expressivity or emotionality and work measuring perceived emotion as it compares listener ratings of communicated emotion across different performer recordings. Similar to Sloboda and Lehman's (2001) study, we observed differences in how listeners experienced performer interpretations. In their study, authors reported that in both post performance and continuous intensity judgements of a Chopin piece recorded by ten pianists, listeners evaluated the 20 performances differently based on expressed emotionality. Although their study did not focus on emotion conveyed in a performance, it demonstrated that local divergences between performers resulted in variations of how listeners rated the emotionality. Our study focused on the consequences of interpretation in the emotional information transmitted to listeners, where we saw that with both visual inspection from plotting each excerpt on the circumplex (Figure 4) as well as the performer profiles (Figures 2 & 3) that performances from one performer can express different levels of arousal and/or valence than other renditions. This helps shed light on how the same musical structure can communicate varying emotional information based on how a performer chooses to manipulate the available expressive cues. We did find that any specific performer remaining consistently different in perceptual ratings, consistent with the idea that certain compositional structures afford more individualized interpretations. The specific compositional features giving rise to this greater performance flexibility remains an open question that would benefit from future research attention.

The relationship between valence and arousal across different interpretations

Our second interest was to examine whether arousal or valence systematically differed more across performances. We found that more variation on average in ratings of arousal than valence. This suggests performer interpretations has a stronger influence on perceived arousal than valence. The musical features or cues available for performers to manipulate are often found to be important to perceived arousal (Quinto et al., 2014). Here we only included one cue under the performer's control, attack rate. Results from our regression models of listener's arousal ratings indicated a large range in how much variance each performer model could explain for (50% – 79%). This may suggest there are other expressive cues outside of attack rate that vary in importance across performers that may help to explain the remaining variance left unexplained. Further, the larger amount of variation in arousal ratings across performers speaks to work done from Quinto and Thompson (2014) examining composer versus performer-controlled cues and their influence on perceived ratings of emotion.

Interestingly, our results show a strong effect of the composer dictated cue of mode, where regardless of interpretation, minor is more variable in the ratings of valence and arousal than major keys (Figure 4). The measure of variation of these rating differences varies across performers (Figure 5), which suggests that listeners perceived certain performers' interpretations more consistently. Using Bach's *WTC* affords us the unique opportunity to assess the influence of mode across a balanced set of keys. Often times stimuli in similar perceptual experiments consists of either a handful of precomposed pieces from different composers, or musical excerpts that have been experimentally composed or manipulated for the directives of the study.

Cue trade-offs in interpretation

Finally, our third goal of the study was to explore any cue trade-off that occurred across performances. The use of commonality analysis in our models of listener ratings allowed us to tease apart cue contributions based on unique and shared variance of our three predictors. However, this analysis does not offer a statistical method to directly compare differences of these cue contributions. As such, we employed bootstrapping to determine appropriate error bars to give us a better idea about any potential significant differences. Using bootstrapping we repeatedly resampled from our original data sets to create multiple simulated data sets, which provides a distribution from which to calculate our error bars. Based on those analyses for bootstrapped valence and arousal ratings, we did see shifts in how much cues could account for variance of that respective model. Specifically, for the unique contributions for our cues of attack rate, mode and pitch height predicted varying amounts of variance based on performance. It is possible performers used expressive cues differently to emphasize the musical structure, which led listener responses to rely on cues differently for different performances.

Although we are not able to directly comment further on this relationship given this is a comparison across different listener groups, and our limited number of predictors—we can observe the relative weights of our measured cues to identify how listeners are using cues for each performance to perceive the conveyed emotion. This speaks to Juslin's (1997) Lens Model that explains how the redundancy of cues allows for emotions to be successfully communicated with different cue combinations. Performers are still able to convey similar emotions without having to use or manipulate the same cues across performances. This is particularly useful in instances when the cues accessible to musicians are limited, such as when performing on instruments with restricted acoustic affordances (i.e., harpsichord does not allow for the same dynamic range as the piano).

Limitations and Future Directions

The present exploratory set of studies is subject to some limitations that can be addressed with future studies. Firstly, this work compares results across seven experiments, where participants are exposed one performer's set of recordings. Although this mitigated the effect of differences in recording quality across performers, it did not allow for direct comparisons of specific pieces across performers as participants heard performances by only a single pianist. Future work should more directly explore this issue, by testing select pieces as performed by all performers to compare emotion ratings. In addition, our studies focused only on three cues, two of which represent strictly composer controlled musical features. Future research should expand this list to include additional performance cues to assess how the relative weights of these cues change according to interpretation. To that effect it would also be beneficial to further examine the relationship between differences in these performance cues and the impact on the variation in emotion ratings.

Our selection of Bach's *WTC* as stimuli allowed for us to study interpretation effects form highly trained performers. Although this work is not known for being overtly emotional, it offers useful insight to how and how performers vary emotional expression when given greater flexibility. Nonetheless, choosing a more emotionally expressive work in future studies would offer useful complementary insight into the range of the potential effect of interpretation on perceived emotion. Another interesting aspect not addressed in this work is the influence of individual factors or differences added to the communication process from the listener. For example personality traits (Taruffi, Allen, Downing, & Heaton, 2017; Vuoskoski & Eerola, 2011), and musical experience (Akkermans et al., 2018; MacGregor, & Müllensiefen, 2017; Castro & Lima, 2014) affect the emotional messages received by listeners. Therefore, future work should also address the interaction of differences in listeners' backgrounds and differences in performers' interpretations.

Limitations aside, the results demonstrate regardless of interpretation, listeners are more variable in emotion assessments of minor key pieces. Further, our results indicate minor key pieces are rated more negatively than major key pieces. Taken together these findings highlight the difficulty in communicating negatively valenced emotions. In addition, even with the presence of structural and performance cues, there is a greater variation in perceived arousal than valence. This corroborates findings that suggest structural or composer-controlled cues have a greater influence of conveyed valence than variations in performance cues (Quinto and Thompson, 2014).

The present study also contributes to literature exploring expressive consequences of interpretation by examining how those choices affect listener perception. Although these differences may not necessarily reflect changes in the emotional category conveyed (i.e. 'happiness', or 'sadness'), variations for levels of conveyed emotional arousal. Thus, the strength or effectiveness of a conveyed emotion may vary as a result of these interpretive choices. As written by Daniel Leech-Wilkinson (2012) "Performance is not simply a reproduction, a performance of something, but a process, created by performers and mentally constructed (uniquely and temporarily by each listener)". When investigating the process of conveyed emotion through musical signals, it is important to consider aspects of the composer, performer and listener. This novel exploratory study explores two of these vital components, helping to shed light onto the complex process of conveying emotions through music.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., … Frieler, K. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, 3*3*(6), 1–20.

Bach, J. S. (1956). *Bach: The Well-Tempered Clavier, Book I [CD; Recorded by J. Demus]*. Universal City: California: MCA Records. (Original work published 1722)

Bach, J.S. (1964). *Bach: The Well-Tempered Clavier, Book I [CD; Recorded by J.M. Martins]*. New York: Connoisseur Society. (Original work published 1722)

Bach, J.S. (1992). *Bach: The Well-Tempered Clavier, Book I [CD;Recorded by S. Richter]*. Germany: RCA Victor Gold Seal (1973). (Original work published 1722)

Bach, J.S. (1993). *Bach: The Well-Tempered Clavier Book I [CD;Recorded by G. Gould]*. New York: New York: Sony Music Entertainment. (Original work published 1722)

Bach, J.S. (1998). *Bach: The Well Tempered Clavier Book I [CD; Recorded by A. Hewitt]*. London: Hyperion Records Ltd. (Original work published 1722)

Bach, J.S. (2007). *Bach: The Well-Tempered Clavier, Book I [CD; Recorded by E. Fischer]*. Abbey Road: London: EMI Records. (Original work published 1722)

Bach, J.S. (1953). *Bach: The Well-Tempered Clavier Book I [CD; Recorded by R. Tureck]*. RFA-Neef, Wittingen, Germany: Deutsche Grammonphon (1953). (Original work published 1722)

Bach, J.S. (2004). *The Well-Tempered Clavier, Volume I*. (W. A. Palmer, Ed.) (3rd ed.). Los Angeles, CA: Alfred Music Publishing.

Battcock, A., & Schutz, M. (2019). Acoustically Expressing Affect. *Music Perception*, *37(1)*, 66–91.

Blom, D. M., Bennett, D., & Stevenson, I. (2016). The Composer's Program Note for Newly Written Classical Music: Content and Intentions. *Frontiers in Psychology*, *7*, 1707.

Castro, S. L., & Lima, C. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Dodson, A. (2011). Expressive timing in expanded phrases: An empirical study of recordings of three Chopin preludes. *Music Performance Research*, *4*, 2–29.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 487.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, *40*(10), 1206–1220.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Bloomington, IN: Indiana University Press.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, *22*, 327–352.

Hunter, P. G., Glenn Schellenberg, E., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology*, *110*, 80–93.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, 14(4), 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1813.

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1-3), 334–364.

Kramer, L. (2010). *Interpreting Music*. London, England: University of California Press.

Krishnamoorthy, K., & Lee, M. (2014). Improved tests for the equality of normal coefficients of variation. *Comput Stat*, *29*, 215–232.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449.

Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, *31*, 182–191.

Leech-Wilkinson, D. (2012). Compositions, Scores, Performances, Meanings. *Music Theory Online*, *18*.

Levinson, J. (1996). Performative versus critical interpretation in music. In J. Levinson (Ed.), *The pleasures of aesthetics : philosophical essays*. Ithaca, NY : Cornell University Press.

MacGregor, C.S, & Müllensiefen, D. (2017). Factors Influencing Discrimination of Emotional Expression Conveyed Through Music Performance. In M. C. Harrison, Peter (Ed.), *Proceedings of the 10th International Conference of Students of Systematic Musicology* (pp. 1–7). London, United Kingdom: Zenodo.

Macritchie, J., Eiholzer, H., & Italiana, S. (2012). Exploring the perceptual effects of performers ' interpretations. *Journal of Interdisciplinary Music Studies*, *6*(2), 177–200.

Marwick, B., & Krishnamoorthy, K. (2019). cvequality: Tests for the Equality of

Coefficients of Variation from Multiple Groups.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, *41*(6), 525–533.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, *6*, 1–13.

Quinto, L., & Thompson, W. F. (2013). Composers and Performers Have Different Capacities to Manipulate Arousal and Valence. *Psychomusicology: Music, Mind, and Brain*, *23*(3), 137–150.

Quinto, L., Thompson, W. F., & Taylor, A. (2014). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music*, *42*(4), 503–524.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's '"Träumerei."' *The Journal of the Acoustical Society of America*, *92*, 2546-2568.

Royal Conservatory of Music. (2015). Piano Syllabus. Toronto, ON: Frederick Harris Music Co., Ltd.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Silverman, M. (2008). A performer's creative processes: implications for teaching and learning musical interpretation. *Music Education Research*, *10*(2), 249–269.

Sinn, D. R. (2013). *Playing Beyond the Notes: A Pianist's Guide to Interpretation*. Oxford University Press.

Sloboda, J., & Lehmann, A. (2001). Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a

Chopin Piano Prelude. *Music Perception*, *19*, 87–120.

Taruffi, L., Allen, R., Downing, J., & Heaton, P. (2017). Individual differences in music-perceived emotions: The influence of externally oriented thinking. *Music Perception*, *34*(3), 253–266.

Thom, P. (2003). The interpretation of music in performance. *The British Journal of Aesthetics*, *43*(2), 126–137.

Timmers, R., & Ashley, R. (2007). Emotional Ornamentation in Performances of Handel Sonata. *Music Perception*, *25*, 117–134.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113.

Vuoskoski, J. K., & Eerola, T. (2011). Special section on music in the brain: Research report The role of mood and personality in the perception of emotions represented by music. *Cortex*, *47*, 1099–1106.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*, 19–39.

Appendix A

Performer Excerpt Durations (seconds)

Table A1

Performer	M	SD	Min	Max
Fischer	28.88	11.72	9	65
Gulda	30.69	13.46	8	65
Gould	27.71	17.35	6	104
Richter	28.29	11.64	8	53
Tureck	31.75	11.57	7	58
Martins	30.88	13.35	7	64
Demus	29.63	10.81	7	58

Performer Attack Rate Descriptive Information

Table A2

Performer	M	SD	Min	Max
Fischer	4.67	2.57	1.2	12.8
Gulda	4.9	2.25	1.3	11.13
Gould	4.75	2.69	.96	11.73
Richter	4.61	2.57	.94	11.73
Tureck	3.91	2.09	.87	10.67
Martins	4.3	2.63	.9	11.2
Demus	4.28	2.23	1.2	10.13

Performer Excerpt RMS Descriptive Values

Table A3

Performer	M	SD	Min	Max
Fischer	-26.4	4.5	-37.5	-18.8
Gulda	-29.6	6.1	-40.6	-20.5
Gould	-22.3	3.9	-36.3	-22.3
Richter	-26.8	4.8	-36.9	-16.8
Tureck	-26.8	5.1	-39.3	-16.6
Martins	-29.4	4.1	-39.9	-22.9
Demus	-27.5	3.49	-35.9	-19.9

Appendix B

Musical Training Survey

Date: _______ (dd/mm/yy) Experiment Num: ____ Participant Num: _____

1) What is your age in years at the time of this study: _____

2) Please list your year of study at McMaster (i.e. first year undergraduate student, second year graduate student, etc). ____

3) Gender - I am (circle one): male female transgendered

4) Do you consider yourself proficient on a musical instrument, and if so, which one?

____ Yes – if so, please list instrument(s): ______________________________________

____ No

5) Have you taken private musical lessons

____ Yes – if so, for how many years: ____

____ No

6) Approximately how many hours a week do you spend practicing/performing/jamming (as opposed to listening) to music?

____ (0-1) ______ (1-5) _______ (5-10) _______ (10-15) ______ (15-20) ____ (> 20)

7) Approximately how many hours a week do you spend listening to (as opposed to practicing or performing) music?

____ (0-1) ______ (1-5) _______ (5-10) _______ (10-15) ______ (15-20) ____ (> 20)

8) Do you own an iPod or personal music listening device?

____ Yes ____ No

8a) If you answered yes to question 8, what kind of device do you own:

____

8b) If you answered "yes" to question 8, approximately how many songs do you have on your device?

____

Appendix C

Descriptives on Valence and Arousal Ratings Across Performers

Table C1

	Valence				Arousal			
Performer	M	SD	Min	Max	M	SD	Min	Max
Fischer	3.98	1.16	1.47	5.6	55.26	17.49	23.33	85.1
Gulda	4.09	1.11	1.97	6.3	56.99	16.78	30.27	82.82
Gould	4.15	1.11	1.93	5.93	57.5	16.24	24.33	85.87
Richter	4.07	1,17	1,7	6.03	52.57	18.72	19.57	84.57
Tureck	4.11	1.02	2.23	5.87	53.26	13.5	29.37	82.47
Martins	4.05	1,26	1,63	6.27	56.56	16.73	28.33	67.17
Demus	4.03	1.12	1.41	6.06	53.91	16.6	26	86.21

Author Note

Aimee Battcock, Department of Psychology, Neuroscience & Behaviour, McMaster University; Michael Schutz, Department of Music, McMaster University.

This research is supported in part by grants from Social Sciences and Humanities Research Council of Canada (SSHRC and the Canadian Foundation for Innovation (CFI)

Correspondence concerning this article should be addressed to Aimee Battcock, Department of Psychology, Neuroscience and Behaviour, Psychology Building (PC), Room 102, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada.
Email: battcoae@mcmaster.ca

Chapter 5: General Discussion

5.1 Thesis Findings and Theoretical Contributions

The foundation of work on musically communicated emotions has centered around tightly controlled or manipulated stimuli (Hailstone et al., 2009; Lindström, 2006; Quinto, Thompson, & Keating, 2013). Therefore, further insight into how cues function across genres in precomposed, multi-lined music is beneficial to understanding the way music can transmit emotional messages. Studies have incorporated more naturalistic methods, using film soundtracks (Eerola, 2011;Vuoskoski & Eerola, 2011), pop music (Yang & Chen, 2012) or excerpts selected to represent specific emotions (Dibben, 2004) to explore emotional perception, however, selection of these stimuli introduces challenges due to familiarity effects of such popular works. For example, pre-existing associations between film scenes and musical excerpts can impact perceived emotion, and the use of multiple timbres included in film music add further complexity in understanding the processing of conveyed emotional content. As such, this dissertation extends this work by examining pre-composed/recorded music unfamiliar to untrained listeners, containing one musical timbre. In addition, investigating the transmission of emotional signals through music not selected for its overt expression of an emotion can illustrate differences in cue use that may emerge when the emotional signal requires more consideration to decode. Overall, this dissertation offers three main contributions to the literature: (1) Insight into the individual and joint contributions of cues in a set of polyphonic pieces by a well-known composer; (2) The influence of musical training on the importance of cues in perceiving emotion; and (3) How performer interpretation and cue use interact to impact listener perception.

In the previous chapters I applied a statistical technique not previously used in the music cognition literature — commonality analysis (i.e., variance partitioning). I use this method on listener ratings of perceived emotion in unaltered music excerpts to investigate the phenomena of communicated emotion. Paired with more conventional techniques such as regression and correlation analyses, it provides a novel and powerful way to address the importance of predictors as it teases apart unique and shared contributions. This is particularly useful in contexts such as music, where predictors or cues may be correlated. For example, within Western culture music, timing and mode are often correlated, where major key pieces are found to be faster than minor key pieces (Poon & Schutz, 2015; Post & Huron, 2009). Therefore, commonality analysis allows us to extract useful information from perceptual ratings of such musical stimuli, in order to uncover how listeners use these cues.

5.1.1 Emotion perception in natural music

Chapter 2 focused on examining the roles of musical structure on perception. In that chapter, I showed that cues of attack rate, mode and pitch contribute to the perception of conveyed valence, but only attack rate emerged as an important cue for the perception of arousal. Further, I found that when applying commonality analysis to models of listener ratings of valence, attack rate and mode predicted similar amounts of rating variance. In contrast to studies that focus on the cue of tempo as an important cue to communicate emotion (Balkwill & Thompson, 1999a; P.N Juslin, 1997; Vieillard et al., 2008), my research uses attack rate to represent conveyed timing information. This cue captures temporal information; however, it is also sensitive to the number of note attacks presented within the selected musical excerpt. As such, it is a good representation of how much musical information is transmitted given the speed or rate at which the notes are performed. In Chapter 1, experiment 2, I found that the predictive weight of mode increased for listener ratings of emotion using musically 'resolved' excerpts. Some excerpts used as stimuli in the first experiment contained modulations or key changes, therefore we used 'resolved' excerpts in experiment 2 to provide a more representative sample of each mode. There, mode did indeed take on greater prominence as a result of this treatment—despite the fact that participants in this experiment had less than one year of formal training. This finding motivated Chapter 3's exploration directly assessing the role of musical training on cue prominence.

5.1.2 Individual differences (listeners)

The individual experiences and characteristics of the listener can influence how musical cues are decoded (Castro & Lima, 2014; Lima & Castro, 2011; Taruffi, Allen, Downing, & Heaton, 2017; Vuoskoski & Eerola, 2011). The level of musical training is one such characteristic that can impact how auditory information is processed and perceived. In Chapter 3, I compared ratings of perceived emotion from listeners with musical training to the data collected and presented in Chapter 2 that involved untrained listeners. From this comparison, I demonstrated differences in the relative weights of cues between individuals with and without musical training. Specifically, I showed that musical mode is more predictive of valence ratings for individuals with more musical training.

Regardless of training, listeners are found to use similar patterns of cues, or cue profiles, implying these populations use similar inference rules in their perception of emotion (Castro & Lima, 2014). Although the initial regression results in Chapter 2 showed comparable findings between groups, the commonality analysis illustrated differences between cue weights of these groups when considering the unique contributions of mode. In Chapter 3, experiment 2, as found in Chapter 1, the strength of mode's independent prediction of ratings of valence increased for trained listeners. In addition, I showed that the regression modelling of listener ratings indicates the three-cue model predicts proportionally more variance for listeners with training than untrained listeners. This result suggests formally trained listeners use these cues more systematically when making assessments of conveyed emotion. Together with the data presented in Chapter 2, this illustrates that the ending tonal area of excerpts affects modality's role in emotional communication—regardless of musical training.

5.1.3 Individual differences (performers)

A performer's interpretative decisions impact how cues are used to communicate intended emotions of a musical piece (Gabrielsson & Juslin, 1996; Juslin, 1997). In Chapter 4, I explored the perceptual consequences of performer interpretation by comparing listeners' emotion ratings across seven experiments, utilizing seven full recordings from well-established and notable performers. Consistent with Quinto and Thompson (2013, 2014), I demonstrated larger variations in listener ratings for perceived arousal in contrast to perceived valence. This result highlights the influence of performer interpretation and expressive cues in conveyed emotion with respect to perceived arousal. Further, these results suggest the importance of compositional cues in perceived valence within multi-lined, precomposed music. Overall, findings in this thesis indicate the specialized nature of compositional and expressive cues applies to music beyond short excerpts composed for experimental purposes.

In addition, I found the effects of piece to be stronger than performer effects in predicting emotion ratings, however the interaction between piece and performer indicated that the effect of performer differed across pieces. From this set of experiments, I showed that listener ratings are more variable for minor key excerpts across performers, demonstrating performers convey a wider range of emotion for pieces expressing more negatively valenced emotions. This result supports work hypothesizing difficulty in conveying negatively valenced emotions (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013). Further, it demonstrates consistency in listener ratings for more positively valenced emotions in musical excerpts, not chosen to explicitly express any basic intended emotions. Lastly, using commonality analysis, I examined performer differences by calculating the 'cue profiles' for each performer, indicating the weight of the unique and shared contributions of cues for predicted variance. These calculated values along with bootstrapped error bars revealed significant differences between performers with respect to attack rate's ability to predict arousal ratings. This result highlights the complex relationship between structural and interpretive cues in emotional communication.

5.2 Limitations and Implications

5.2.1 Limitations

To the best of my knowledge, this thesis is the first to use regression modelling in combination with commonality analyses to explore the relationship between musical cues (i.e., attack rate, mode and pitch height) and perceived emotion in precomposed music grounded in Western tradition. Although this approach offers insight into the complex communicative process of conveyed emotion, it brings some inherent limitations. Overall, the experiments in Chapters 2-4 focus mainly on three musical cues. Selected for their importance reported within the literature (Balkwill & Thompson, 1999; Hunter, Schellenberg, & Schimmack, 2008; Ilie & Thompson, 2006; Quinto, Thompson, & Keating, 2013), this nonetheless represents a limited set of all features that can transmit musical emotion. Work on music and emotion has identified other cues that influence perceived emotion (Bresin & Friberg, 2011; Eerola, Friberg, & Bresin, 2013; Schubert, 2004), and therefore future work could quantify additional cues to better compare the importance of individual musical features.

Furthermore, of the cues studied in this thesis, attack rate is the only expressive cue explored. Across all experimental chapters, models of listener ratings predicted valence better than arousal. This finding is in contrast to research using instrumental film musical excerpts (Eerola, 2012; Eerola, Lartillot, & Toiviainen, 2009; Vuoskoski & Eerola, 2011). Thus, results from this dissertation are limited in explaining whether models of listener ratings of emotion could be better predicted by other expressive cues than attack rate. Future studies aiming to assess the cues important to conveyed arousal could quantify expressive cues known to play a role, such as dynamics (Eerola et al., 2013; Nakamura, 1987) and articulation (Bresin & Friberg, 2011; Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011). In addition, quantification of cues in this thesis used only an average value per-piece—yet, music is a dynamic stimulus with events continuously unfolding and changing over time. As emotion can fluctuate throughout a musical excerpt, a retrospective evaluation might not fully capture the dynamic unfolding of emotional experiences. Future studies mapping cue use to emotional fluctuations using continuous rating measures (particularly when comparing performer interpretations) could prove invaluable to enhancing our understanding of how perceived emotion varies on a moment to moment basis.

In Chapter 3, I reported that results across both experiments 1 and 2 demonstrated the effect of mode is stronger on ratings from listeners with more years of musical training. Compared to the findings in Chapter 2, this result suggests musical training leads to a greater sensitivity to mode with respect to conveyed emotion. This is further supported by the results in the second experiment of Chapter 3, where the variance predicted by mode is larger for trained listeners than for those with less than one year of musical training. Mode is known to be an acculturated cue, as knowledge of the association between mode and emotion comes with exposure or experience (Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1990). Taken with the results of Chapter 3, this suggests those with formal years of training have more exposure or experience in learning how mode is used to convey emotion. However, these results cannot answer whether the effect of mode on the perceptual ratings of trained listeners is the result of the assessment of the excerpt as a whole, or if knowledge of the starting mode determines their ratings of valence. Listeners with musical training are often aware of emotional connotations of major/minor modes and have the ability to explicitly identify mode by listening. This raises the possibility of top-down biases in their assessments. Although experiment 2 uses musically 'resolved' excerpts to ensure they end in the same mode as they began, the question about mode's influence as the excerpt unfolds still remains. As such, continuous ratings would be beneficial to determine how mode influences perceived emotion moment-by-moment.

Chapter 4 examined the influence of performers' interpretative decisions regarding cue use and listener perception. Listeners are successful in grouping performances by performer (Gingras et al., 2011) and can identify the use of expressive cues (Macritchie, Eiholzer, & Italiana, 2012). Additionally, work from Juslin (2000) suggests performers can vary the emotion communicated to listeners. The methodology in Chapter 4 does not directly compare renditions of the same piece across different performers but evaluated rating differences between each experiment containing a performer's full recording of the *Well-Tempered Clavier*. Thus, it remains unclear which cues performers varied that may explain differences in listener ratings. Future work directly contrasting expressive performances of the same piece can elucidate the relationship between cue variations and listener responses of emotion.

Lastly, findings in all chapters of this dissertation emerged from the use of stimuli from a specific musical era in Western tradition. Although the results complemented and extended what has been found in the literature using tightly controlled or manipulated stimuli (Balkwill & Thompson, 1999; Hailstone et al., 2009; Lindström, 2006), they do not always generalize across all musical genres and eras. Even within music composed within the same genre (e.g. classical music), composers can use cues differently to express emotion. For example, the relationship between timing and mode cues appear to follow different trends or patterns in music from the Romantic era, compared to those works created by composers of Baroque music (Horn & Huron, 2015; Poon & Schutz, 2015; Post &

Huron, 2009). As such future experiments should test the generalizability of cue effects using this methodology with a wider range of stimuli.

5.2.2 Implications for music theorists

Some music theorists are skeptical of perceptual evidence demonstrating a strong effect of mode (particularly the major/minor distinction) on perceived emotion. Although the data presented here suggest mode is a strong cue for emotion, some music theorists argue such results speak more to associations between mode and structural cues other than mode itself (i.e., as mentioned in section 5.2.1, the correlation between mode and timing). My dissertation contributes to this debate by showing how mode has a greater influence over perceived valence than arousal. Using a dimensional modelling approach, I have shown that mode significantly predicts how listeners assess the communicated emotion as sounding positive or negative. Contrastingly, mode does not predict the conveyed emotional intensity, as evidenced by its minimal influence on listener ratings of arousal. This is consistent with other perceptual evidence suggesting a difference in how structural versus performance cues influence these two dimensions of emotion (Quinto & Thompson, 2013; Quinto, Thompson, & Taylor, 2014).

Additionally, my results illustrate the impact of the relationship between cues of mode and timing. Music theorists argue mode's proposed effect on perceived emotion is largely a function of its association with other cues. Using commonality analyses I show that a large proportion of variance in our regression models is commonly predicted by mode and attack rate. This shared predicted variance is a result of the correlation between mode and attack rate within the stimuli, where excerpts in major modes have faster attack rates than excerpts in minor modes. However, through the analyses, I identified the effect of this relationship on predicted variance and could tease apart the unique effects of each cue, revealing their independent effects on listener ratings. This demonstrates that for valence ratings, mode and attack rate appeared similar in their predictive weight in the first experiment of Chapter 2. In experiment 2 of that same chapter, the use of resolved excerpts created a shift in cue weights, as mode predicted proportionally more variance than attack rate. Thus, this dissertation adds some clarity to the debate on the importance of mode as an emotionally expressive cue.

5.2.3 Implications for developing clinical applications

Although the findings presented in this thesis do not directly lend themselves to clinical applications, they help in the understanding of interpersonal communication, specifically with non-verbal emotional communication. Musical therapeutic strategies are often used to facilitate emotional expression (Clements-Cortés, 2004; Darrow, 2006). These strategies rely on understanding the expressed emotions within performances and are crucial, given the importance for therapists to establish a meaningful relationship with their client through shared music making, often known as joint clinical improvisation (Alvin & Warwick, 1991). Recent clinical studies reported benefits of emotional expression in music. For example, many reports show music therapy can improve social skills in children with autism spectrum disorder (ASD), promoting non-verbal communication through musical improvisation (Novenia, 2019).

These findings illustrate the relationship between cues and dimensions of communicated emotion. Understanding how musical features can express valence and/or arousal allows therapists to use cues appropriately in practices where they are required to respond musically to clients expressing emotion. In particular, the 'iso-principle' is a technique in music therapy that requires therapists to meet the mood of the client with music, and gradually alters the musical expression to reach a desired mood state (Davis, Gfeller, & Thaut, 2008). Difficulty in recognizing a client's expressed emotions, and how that is achieved in music, can affect the success of this technique. In addition, knowledge of the relationships between musical cues and expressed emotion is crucial to respond appropriately to continue the musical dialogue and achieve a desired mood state in a client. Specifically, relationships between cues and dimensions of emotion can be key to communicating appropriate levels of emotional arousal and valence that that will eventually lead to effective changes in mood. This technique is applicable to a range of clinical populations who have difficulty with verbal expression of emotion ranging from those with ASD, to those experiencing depression (Bodner et al., 2007), or those unable to communicate verbally. This highlights the importance of not only using music to express emotion, but to have the tools to understand what is being communicated.

5.2.4 Implications for understanding the role of musical training

Neural evidence suggests musical experts process auditory information differently compared to listeners without training (Koelsch, Schmidt, & Kansok, 2002; Sherwin & Sajda, 2013). However behavioural evidence regarding the effect of musical training is more equivocal. In some cases listeners perform similarly regardless of levels of training (Bigand & Poulin-Charronnat, 2006), however more recent work suggests musical training may fine-tune perceptual processes which make trained listeners more accurate in tasks involving emotion recognition (Akkermans et al., 2018; Castro & Lima, 2014; Lima & Castro, 2011). In this thesis, I showed that listeners with six or more years of formal music training are more sensitive to the cue of mode than untrained listeners, which is consistent with views that training can affect how listeners process cues to perceive emotion.

Differences in performer expression have often been studied, noting variations in how expressive cues are used (Gingras et al., 2011; Macritchie et al., 2012; Repp, 1992). These variations are purposely selected to encode a performer's intended emotion, which as a result can change how the listener experiences the music (Sloboda & Lehmann, 2001). Examining the effect of interpretative choices on perceived valence and arousal, I showed models predicting listener ratings vary in the amount of variance accounted for by attack rate, mode and pitch height across both dimensions. The variation however among models of perceived arousal is greater than for perceived valence, again demonstrating expressive cues used by the performer have greater influence over conveyed arousal than valence.

5.3 Summary

With novel application of this technique in a unique stimulus set, my work offers three important contributions to existing research on communicated emotion in music: (1) Insight on the individual and joint contributions of correlated structural cues on varying dimensions of emotion; (2) The influence of individual differences such as musical training on cue weights in perceiving emotion; and (3) Interactions between performer interpretations and cue use and how this may impact perception.

The primary goal of this thesis was to determine trade-offs in the weighting of cues conveying emotion in widely studied and heard music, taking into account a wide variety of naturally and important, but difficult to assess factors—such as individual differences in performers' interpretations as well as listeners' training. Together, these results provide novel insight into issues of broad relevance. Specifically, Chapters 2-4 demonstrate how cues of timing, mode and pitch contribute both uniquely and jointly listener perception of emotion. This extends previous work using regression modelling to explore the importance of timing and mode cues on perceived emotion (Eerola 2011; Eerola 2012; Gagnon & Peretz, 2003; Schubert, 2004). In addition, this body of work emphasizes how cues can impact dimensions of valence and arousal differently, where mode and timing were the strongest cues for listener ratings of valence, but only attack rate appeared important for ratings of arousal. These results add to debates from music theorists on the utility of mode in perceived emotion, and support research indicating cues selected for by the composer (structural cues) have greater impact on perceived valence and cues manipulated by the performer (expressive cues) are more important for perceived arousal (Quinto & Thompson, 2013). Further, the present results broaden such findings to unaltered multi-lined music, composed of covaried cues. This illustrates the specialized effect of structural and expressive cues occurs in more complex music frequently experienced in concert halls or with personal listening devices.

Overall, this work extends research exploring how music can be a tool for emotional communication, focusing on the factors that influence the chain of communication. Insight into the function of musical features, listener differences and variations in performer interpretations helps us unpack how emotional signals can be decoded and better understand the emotional power of music. Further it provides a new analysis to investigate the relationship between cues and listener ratings in polyphonic, or multi-line musical works where cues are co-varied by the composer.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., … Frieler, K. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1–20.

Alvin, J., & Warwick, A. (1991). *Music Therapy for the Autistic Child*. Oxford: Oxford University Press.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Perception*, *17*, 43–64.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we 'experienced listeners';? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130.

Bodner, E., Iancu, I., Gilboa, A., Sarel, A., Mazor, A., & Amir, D. (2007). Finding words for emotions: The reactions of patients with major depressive disorder towards various musical excerpts. *The Arts in Psychotherapy*, *34*(2), 142–150.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex*, *47*(9), 1068–1081.

Castro, S. L., & Lima, C. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142.

Clements-Cortés. (2004). The use of music in facilitating emotional expression in the terminally ill. *American Journal of Hospice & Palliative Medicine*, *21*(4), 255–260.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Darrow, A. A. (2006). The role of music in deaf culture: Deaf students' perception of emotion in music. *Journal of Music Therapy*, *43*(1), 2–15.

Davis, W. ., Gfeller, K. ., & Thaut, M. H. (2008). *An introduction to music therapy theory and practice* (3rd ed.). The American Music Therapy Association.

Dibben, N. (2004). *The Role of Peripheral Feedback in Emotional Experience With Music*. *Music Perception, 22*(1), 79-115.

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. *Journal of New Music Research*, *40*(4), 349–366.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in*

Psychology, *4*, 487.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction Of Multidimensional Emotional Ratings in Music from Audio using Multivariate Regression Models. In K. Hirata, G. Tzanetakis, & K. Yoshii (Eds.), Proceedings of the *10th International Society for Music Information Retrieval. Kobe, Japan: International Society for Music Information Retrieval.*

Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience,. *Psychology of Music,* 24, 68-91.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(27), 9240–9245.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception, 12*(3), 279–290.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, *40*(10), 1206–1220.

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. 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.

Horn, K., & Huron, D. (2015). On the changing use of the major and minor modes 1750–1900. *Music Theory Online*, *21*(1), 1–11.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, *22*(2), 327–352.

Juslin, P.N. (1997). Emotional communication in music performance: A functionalist perspective. *Music Perception*, *14*(4), 383–418.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1813.

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189–202.

Koelsch, S., Schmidt, B.-H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychomusicology*, *39*, 657–663.

Korhonen, M. D., Clausi, D. A., & Ed Jernigan, M. (2006). Modeling Emotional Content of Music Using System Identification. *Systems, Man and Cybernetics, Part B, IEEE Transactions on Systems, Man, and Cybernetics*, *36*(3), 588–599.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449.

Lima, C. F., & Castro, S. L. (2011). Speaking to the Trained Ear: Musical Expertise Enhances the Recognition of Emotions in Speech Prosody. *Emotion*, *11*(5), 1021–1031.

Lindström, E. (2006). Impact of melodic organization of melodic structure and emotional expression. *Musicae Scientiae*, *10*(1), 85–117.

Macritchie, J., Eiholzer, H., & Italiana, S. (2012). Exploring the perceptual effects of performers ' interpretations. *Journal of Interdisciplinary Music Studies*, *6*(2), 177–200.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, *41*(6), 525–533.

Novenia, M. V. (2019). The use of music therapy to improve non-verbal communication skills for children with autism. *International Journal of*

Humanities and Social Sciences, *13*(3), 243–246.

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences*, *1060*, 450–453.

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, *6*, 1–13.

Post, O., & Huron, D. (2009). Western Classical music in the minor mode is slower (Except in the Romantic period). *Empirical Musicology Review*, *4*(1), 2–10.

Quinto, L., & Thompson, W. F. (2013). Composers and Performers Have Different Capacities to Manipulate Arousal and Valence. *Psychomusicology: Music, Mind, and Brain*, *23*(3), 137–150.

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4*, 1–8.

Quinto, L., Thompson, W. F., & Taylor, A. (2014). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music*, *42*(4), 503–524.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's '"Träumerei."' *The Journal of the Acoustical Society of America*. *92*, 2546-2568.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Sherwin, J., & Sajda, P. (2013). Musical experts recruit action-related neural structures in harmonic anomaly detection: Evidence for embodied cognition in expertise. *Brain and Cognition*, *83*(2), 190–202.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception l*, *19*(1), 87–120.

Taruffi, L., Allen, R., Downing, J., & Heaton, P. (2017). Individual differences in music-perceived emotions: The influence of externally oriented thinking.

Music Perception, *34*(3), 253–266.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., Bouchard, B., … Sabourin, C. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Emotion and Cognition*, *22*(4), 720–752.

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

Vuoskoski, J. k, & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39.

Yang, Y.-H., & Chen, H. H. (2012). Machine Recognition of Music Emotion. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 1–30.