Examining Relationships Between Basic Emotion Perception and Musical Training in the Prosodic, Facial, and Lexical Channels of Communication and in Music

Jamie Twaite
The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Clinical Psychology Commons, Cognition and Perception Commons, Cognitive Neuroscience Commons, Music Education Commons, Music Performance Commons, Music Theory Commons, Music Therapy Commons, Other Neuroscience and Neurobiology Commons, and the Other Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/1435

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Examining Relationships between Basic Emotion Perception and Musical Training in the
Prosodic, Facial, and Lexical Channels of Communication and in Music

by

Jamie T. Twaite, M. Phil.

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
The City University of New York
2016
This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jamie T. Twaite

Date

Joan C. Borod, Ph.D.
Chair of Examining Committee

Date

Maureen O’Connor, Ph.D., J.D.
Executive Officer

Supervisory Committee
Justin Storbeck, Ph.D.
Howard Erlichman, Ph.D.

Readers
Jennifer Speilman, Ph.D.
Adam M. Brickman, Ph.D.

THE CITY UNIVERSITY OF NEW YORK
Abstract

EXAMINING RELATIONSHIPS BETWEEN BASIC EMOTION PERCEPTION AND MUSICAL TRAINING IN THE PROSODIC, FACIAL, AND LEXICAL CHANNELS OF COMMUNICATION AND IN MUSIC

By

Jamie T. Twaite

Dissertation chairperson: Joan C. Borod, Ph.D.

Research has suggested that intensive musical training may result in transfer effects from musical to non-musical domains. There is considerable research on perceptual and cognitive transfer effects associated with music, but, comparatively, fewer studies examined relationships between musical training and emotion processing. Preliminary findings, though equivocal, suggested that musical training is associated with enhanced perception of emotional prosody, consistent with a growing body of research demonstrating relationships between music and speech. In addition, few studies directly examined the relationship between musical training and the perception of emotions expressed in music, and no studies directly evaluated this relationship in the facial and lexical channels of emotion communication. In an effort to expand on prior findings, the current study characterized emotion perception differences between musicians and non-musicians in the prosodic, lexical, and facial channels of communication and in music.

A total of 119 healthy adults (18-40 years old) completed the study. Fifty-eight were musicians and 61 were controls. Participants were screened for neurological and psychiatric illness. They completed emotion perception tasks from the New York Emotion Battery (Borod,
Welkowitz, & Obler, 1992) and a music emotion perception task, created for this project, using stimuli developed by Eerola and Vuoskoski (2011). They also completed multiple non-emotional control measures, as well as neuropsychological and self-report measures, in order to control for any relevant participant group differences. Parametric and non-parametric statistical procedures were employed to evaluate for group differences in emotion perception accuracy for each of the emotional control tasks. Parametric and non-parametric procedures were also used to evaluate whether musicians and non-musicians differed with regard to their perception of basic emotions.

There was evidence for differences in emotion perception between musicians and non-musicians. Musicians were more accurate than non-musicians for the prosodic channel and for musical emotions. There were no group differences for the lexical or facial channels of emotion communication. When error patterns were examined, musicians and non-musicians were found to make similar patterns of misidentifications, suggesting that musicians and non-musicians were processing emotions similarly.

Results are discussed in the context of theories of music and speech, emotion perception processing, and learning transfer. This work serves to clarify and strengthen prior research demonstrating relationships between music and speech. It also has implications for understanding emotion perception as well as potential clinical implications, particularly for neurorehabilitation. Lastly, this work serves to guide future research on music and emotion processing.
Acknowledgements

This dissertation would not have been possible without the help, support, and guidance of numerous faculty members, colleagues, friends, and loved ones.

I would first and foremost like to thank my advisor, Dr. Joan C. Borod, who welcomed me into her lab and provided me the opportunity to pursue a career in clinical neuropsychology. I will always be incredibly grateful for her mentorship and guidance, and inspired by her passion for the study of emotion and neuropsychology.

I would also like to express my sincere appreciation to my committee members, Dr. Justin Storbeck and Dr. Howard Ehrlichman, for their support and helpful feedback throughout the dissertation process, as well as to my outside readers, Dr. Jennifer Stewart and Dr. Adam M. Brickman, for their thoughtful contributions.

I am also indebted to the many wonderful members of Dr. Borod’s Emotion Lab. I am particularly appreciative of Hira Bokari, Elina Aranova, Remington Stafford, and Ashba Hoda for their time and energy assisting on this project. I also cannot express how much the incredible members of my lab and cohort, especially Erica Meltzer, have helped me from the start all the way through those last few miles. I will always treasure their friendship.

I would like to thank my mother, Patricia, whose incredible love and sacrifices enabled me to attend college and set me on the path to a Ph.D. I would also have not been able to accomplish this task without the wonderful love, keen intelligence, and infinite patience of my amazing husband, Saar. Lastly, I would like to thank my good friends MacKay, Nicole, Molly, Phil, and Oliver for providing me with much needed distraction and emotional support and for keeping me grounded in the world outside of graduate school.

This work was supported by the CUNY Doctoral Student Research Grant program.
# Table of Contents

Introduction 1  
Research Question 1  
Literature Review 4  
Emotion Processing 4  
Decoding discrete emotions 5  
Facial 5  
Prosodic 6  
Lexical 7  
Postural/gestural 8  
Representation and perception of emotion in music 9  
What is the nature of musical emotion? 12  
Decoding discrete emotions in music 12  
Individual Differences in Emotion Perception 14  
Sex 14  
Ethnicity 14  
Age 16  
Emotion perception deficits and psychopathology 17  
Neural Basis of Emotion Perception 19  
Lateralization of emotion processing 19  
Facial channel 20  
Prosodic channel 20  
Lexical channel 21  
Postural/gestural channel 21  
Musical emotions 22  
Neuroscience and Music 23  
Neural basis of music processing 24  
Neurological differences between musicians and non-musicians 24  
Neuroanatomical differences 25  
Functional differences 25  
Critical periods for music training 27
Musical Training and Non-Emotional Mental Processes 27
Musical training and auditory processing 28
Musical Training and cognitive processes 29
Musical Training and Emotional Processes 32
Emotion Perception 32
Prosodic channel 32
Musical channel 38
Lexical, facial, and postural channels 41
Summary of perception findings 44
Emotional expression 45
Emotional experience 47
Emotional intelligence 49
Social functioning 51
Summary of musical training and emotion processes 53
Musical Training and the Aging Brain 53
Aims and Hypotheses 56
Aim I 56
Aim II 57
Aim III 59
Aim IV 60
Aim V 61
Methods 62
Participants 62
Inclusion and exclusion criteria 62
Procedures 64
Screening measures 65
Experimental tasks 65
Emotion identification tasks 65
Non-emotional control tasks 67
Prosodic channel 67
Lexical channel 68
Cognitive measures 68
Additional measures for group comparisons 69
Task administration 70
Statistical Analyses 71
  Data Inspection 72
  Data Normalization and Variance 72
  Participant demographics 72
  Group differences 72
  Data Analysis for Specific Aims and Hypotheses 72
    Aim I 73
    Aim II 73
    Aim III 74
    Aim IV 74
    Aim V 75
Results 76
  Data Inspection 76
  Tests of Normality and Homogeneity of Variance 76
  Demographic Characteristics 77
  Group Comparisons 78
  Non-emotional Control Measures 79
  Analyses Specific to Aims 79
    Aim I 79
    Aim II 80
    Aim III 80
    Aim IV 80
    Aim V 83
Discussion 84
  Summary of Results 84
    Musical training and emotion perception 85
      Musical training and prosodic emotion perception 85
      Musical training and musical emotion perception 87
Associations between length of training and accuracy 88
Musical training and lexical and facial emotion perception 89
Associations between ethnicity and emotion perception 90
Theoretical implications of emotion findings 91
Auditory processing 92
Domain-specific emotion perception processing 93
Domain-general emotion perception processing 94
Learning transfer 95
Theories of musical emotion 97
Musical training and cognitive abilities 98
Limitations 101
Clinical Implications 103
Directions for Future Research 104
Conclusions 106
References 162
## Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inclusion Criteria</td>
<td>107</td>
</tr>
<tr>
<td>2</td>
<td>Participant Demographic Characteristics by Group</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>Descriptives for Group Scores on Questionnaires</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td>Descriptives for Non-emotional control measures</td>
<td>108</td>
</tr>
<tr>
<td>5</td>
<td>Descriptives for Emotion Identification Tasks</td>
<td>109</td>
</tr>
<tr>
<td>6</td>
<td>Descriptives for Individual Musical Emotions</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>MID Distribution of Total Responses for Each Emotion in %</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>PID Distribution of Total Responses for Each Emotion in %</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>LID Distribution of Total Responses for Each Emotion in %</td>
<td>112</td>
</tr>
<tr>
<td>10</td>
<td>FID Distribution of Total Responses for Each Emotion in %</td>
<td>113</td>
</tr>
<tr>
<td>11</td>
<td>MID Distribution of Inaccurate Responses for Each Emotion</td>
<td>114</td>
</tr>
<tr>
<td>12</td>
<td>PID % Distribution of Inaccurate Responses for Each Emotion</td>
<td>115</td>
</tr>
<tr>
<td>13</td>
<td>Patterns of Misidentified Emotions for MID for Musicians &amp; Controls</td>
<td>116</td>
</tr>
<tr>
<td>14</td>
<td>Patterns of Misidentified Emotions for PID, FID, and LID for Musicians &amp; Controls</td>
<td>116</td>
</tr>
<tr>
<td>15</td>
<td>Descriptives for Cognitive measures</td>
<td>117</td>
</tr>
<tr>
<td>16</td>
<td>Group by Ethnicity by Emotion ANCOVA Statistics</td>
<td>117</td>
</tr>
</tbody>
</table>
**Figures**

Figure 1. Group Accuracy for Emotion Identification Tasks  
118

Figure 2. Group accuracy for Music Identification by Individual Emotion  
119

Figure 3. MID – Musicians - Distribution of Responses for Each Emotion  
120

Figure 4. MID – Controls – Distribution of Responses for Each Emotion  
121

Figure 5. PID – Musicians – Distribution of Responses for Each Emotion  
122

Figure 6. PID – Controls – Distribution of Responses for Each Emotion  
123

Figure 7. MID – Musicians – Distributions of Inaccurate Responses  
124

Figure 8. MID – Controls – Distributions of Inaccurate Responses  
125

Figure 9. PID – Musicians – Distributions of Inaccurate Responses  
126

Figure 10. PID – Controls - Distributions of Inaccurate Responses  
127
Appendices

Table 1: Power Analysis 128
Table 2: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Experimental Emotion Discrimination Tasks 128
Table 3: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Music Identification Individual Emotions 129
Table 4: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Non-emotional experimental tasks 129
Table 5: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Cognitive measures 130
Table 6: One-Way ANOVAs comparing White and Non-White participants’ emotion perception accuracy percentage scores 130
Table 7: Spearman Correlations between WTAR score and the MID or PID score for Musicians and Controls 131
Table 8: Descriptives PID, LID, and FID Individual Emotions (% correct) 131
Table 9: PID, LID, and FID ANCOVA Statistics 131
Table 10: Multiple Wald Test Statistics for LID and FID 132
Table 11: Relationships between Training Duration & Accuracy 132
Figure 1: Normal and detrended normal Q-Q plots for Music Identification 131
Figure 2: Normal and detrended normal Q-Q plots for Facial Identification 134
Figure 3: Normal and detrended normal Q-Q plots for Lexical Identification 136
Figure 4: Normal and detrended normal Q-Q plots for Prosodic Identification 138
Figure 5: Normal and detrended normal Q-Q plots for Music Identification Happy 140
Figure 6: Normal and detrended normal Q-Q plots for Music Identification Sad 142
Figure 7: Normal and detrended normal Q-Q plots for Music Identification Anger 144
Figure 8: Normal and detrended normal Q-Q plots for Music Identification Fear 146
Figure 9: Normal and detrended normal Q-Q plots for Music Identification Tender 148
Figure 10: Group Accuracy for Prosodic Identification by Individual Emotion 151
Figure 11: Group accuracy for Lexical Identification by Individual Emotion 152
Figure 12: Group accuracy for Facial Identification by Individual Emotion 153
Figure 13: LID Distribution of Responses for Each Emotion – Musicians 154
Figure 14: LID Distribution of Responses for Each Emotion – Controls 155
Figure 15: FID Distribution of Responses for Each Emotion – Musicians 156
Figure 16: FID Distribution of Responses for Each Emotion – Controls 157
Figure 17: LID Distributions of Inaccurate Responses – Musicians 158
Figure 18: LID Distributions of Inaccurate Responses – Controls 159
Figure 19: FID Distributions of Inaccurate Responses – Musicians 160
Figure 20: FID Distributions of Inaccurate Responses – Controls 161
Introduction

Research Question

Over the past century, an enormous volume of research literature has been generated regarding virtually all aspects of music as it relates to psychology. Topics range from fundamental questions about the evolutionary origins of musical ability to the application of music as a therapeutic intervention. One growing area of music psychology research seeks to examine the effects of musical training on non-musical mental processes, such as motor skills, auditory processes, and cognitive ability. Emotion researchers have also been drawn to the study of musical training, as musicians engage a broad range of neural networks during musical practice and performance, including systems overlapping with those used in the perception, experience, and expression of emotion.

Understanding and communicating emotions is key to successful human social interaction, from infancy through old age. In particular, the ability to identify emotionally salient information in the environment quickly and accurately enables us to respond accordingly and is crucial to survival. In daily life, accurate identification of others’ emotions is a fundamental skill in achieving successful interpersonal communication and allows us to successfully navigate everyday social interactions. Correspondingly, difficulty with emotion perception results in a wide range of impairments to social functioning. These include decreased social competence (Feldman, Philippot, & Custrini, 1991; Shimokawa, Yatomi, Anamizu, Torii, Isono, Sugai, & Kohno, 2001), impaired interpersonal communication (Spell & Frank, 2000), poorer quality interpersonal relationships and functioning (Carton, Kessler, & Pape, 1999; Ciarrochi, Chan, & Caputi, 2000), and increased psychopathology (Keltner & Kring, 1998).
Emotion processing is impaired in multiple psychiatric (e.g., schizophrenia, mood disorders, and personality disorders; Phillips, Drevets, Rauch, & Lane, 2003) and neurological (e.g., dementia and traumatic brain injury) conditions (e.g., Bornhofen & McDonald, 2008; Shimokawa et al., 2001). Emotion processing also changes as a function of age. Age-related declines in emotion perception have been reported for recognizing emotion in multiple channels of communication, including faces (i.e., facial channel), body language (i.e., postural/gestural channel), tone of voice (i.e., prosodic channel), and written language/speech content (i.e., lexical channel; e.g., Ruffman, Sullivan, & Dittrich, 2009). These declines appear most pronounced in the perception of affective prosody and in the perception of certain discrete emotions, including happiness, interest, and fear (Finley, Borod, Brickman, et al., 2008; Savage et al., 2013).

Emotion recognition is of particular relevance to the health and well-being of older adults. Loneliness and social isolation negatively impact both physical and cognitive health (Bath & Deeg, 2005; Cacioppo & Hawkley, 2009; Fry & Debats, 2006; Hawthorne, 2008; House, Landis, & Umberson, 1988; Wilson et al., 2007). It is beneficial to identify factors that influence emotion perception, as doing so not only enhances current understanding of emotion processes, but such factors could ultimately represent means to prevent or ameliorate deficits in emotion perception.

Preliminary research has found that individuals with musical training are more accurate in identifying emotions than non-musicians in the prosodic channel (e.g., Lima & Castro, 2011a; Thompson, Schellenberg, & Husain, 2004). However, results have not been consistent across all studies. This discrepancy is likely due to methodological differences, including variation in onset, duration, and intensity of participants’ musical training, and further clarification is needed. Furthermore, the relationship between musical training and emotion perception in non-auditory channels of communication, such as facial and lexical, has not been investigated. Evidence is
mixed as to if emotion perception is channel or modality specific (i.e., divided according to auditory versus visually expressed emotion, perhaps due to overlap between music and speech mechanisms) or there is an underlying “general processor,” such that emotion perception ability generalizes across channels due to a shared underlying mechanism (Borod, Pick, et al., 2000). Examination of both auditory and visual channels of emotion communication is therefore necessary to better characterize the nature of the relationship between musical training and emotion perception and to inform future research.

In an effort to replicate and extend previous work, the current project examined the perception of 8 discrete emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, fear, and anger) in the prosodic, facial, and lexical channels of communication and the perception of 5 discrete emotions (i.e., happiness, tenderness, sadness, fear, and anger) in an additional auditory modality, music. A total of 58 musically-trained and 61 musically-untrained individuals between the ages of 18-40 completed emotion identification tasks from the New York Emotion Battery (NYEB; Borod, Welkowitz, & Obler, 1992; Borod et al., 1998), a comprehensive test battery that examines emotion perception, expression, and experience in three channels of communication (i.e., prosodic, facial, and lexical). Participants also completed an analogous music emotion identification task, designed for this project using stimuli developed and validated by Eerola and Vuoskoski (2011). The primary aims of this project were to a) replicate and clarify prior research examining prosodic emotion perception in individuals with and without musical training and b) expand this research to assess emotion perception in additional areas, specifically an additional auditory modality (i.e., music) as well as in non-auditory channels of communication (i.e., facial and lexical). It was hypothesized that individuals with musical training would be more accurate than those without at identifying discrete emotions
in the prosodic channel. It was also predicted that this effect would be found in music emotion identification as well as in the lexical and facial channels. These findings have implications for understanding the mechanisms of emotion perception and, potentially, for the treatment or prevention of deficiencies of emotion perception.

**Literature Review**

**Emotion Processing**

Emotions can be defined as biologically-based responses to evocative stimuli that encompass multiple components, including cognitive appraisal/perception, expression, physiological arousal, subjective feeling, and goal-directed behavior (Borod, 1993; Plutchik, 1984). For the purpose of the present project, this review will focus primarily on emotion perception, defined as the “processing, appreciation, or comprehension of the emotional aspect of a stimulus” (Borod, 1992, p. 340).

Humans can express emotion through a number of what are termed channels of communication, including the prosodic channel (i.e., emotional speech prosody), lexical channel (i.e., word choice/language content), facial channel (i.e., facial expression; Borod et al., 1998), and the postural/gestural channel (i.e., body position and movement; Atkinson, Dittrich, Gemmell, & Young, 2004). It is also broadly accepted that emotions can be communicated through other mediums, such as music (Juslin & Laukka, 2003; Juslin & Västfjäll, 2008), although this position is not without controversy (Trainor & Schmidt, 2003).

Emotions can be characterized using multiple theoretical perspectives. One dominant theory is that emotions are natural kinds, that is, discrete entities with properties that can be objectively identified (e.g., surprise, happiness, sadness, disgust, fear, and anger; Barrett, 2006). Alternately, other theories have begun to challenge this model (Barrett, 2006; Barrett et al.,
and posit that emotions are better distinguished using dimensional models, such as according to valance (i.e., positive or negative emotionality) or behavioral function (i.e., if the emotion leads to approach or withdrawal behaviors; Davidson, 1995; Demaree, Everhart, Youngstrom, & Harrison, 2005). Within the range of discrete emotions, a subset of emotions has been identified that are thought to be universally expressed and perceived across cultures. These discrete emotions are referred to as “basic” emotions (Ekman, 1992; 1994). Although there is some variation as to which emotions are considered basic, there is a consensus that happiness, surprise, sadness, disgust, fear, and anger can be identified across cultures in multiple channels of communication (Ekman, 1994; Elfenbein & Ambady, 2002; Pell, Monetta, Paulmann, & Kotz, 2009). As the focus of the present project is on the perception of discrete emotions, a brief overview of the perception of discrete emotions, including cues that signal emotions and the neuroanatomy of emotion perception, will now be provided, focusing specifically on basic emotions.

**Decoding discrete emotions.** In order for basic emotions to be recognizable across cultures, they must in some way have universal signals or cues that denote particular emotions to the perceiver. The cues that convey discrete basic emotions across channels of communication will be briefly reviewed.

**Facial.** The facial channel, along with the prosodic channel, is perhaps the best characterized with regards to the cues that signal discrete emotions. Discrete facial emotions are signaled through specific combinations of facial musculature movements, which have been characterized in great detail using methods such as the Facial Action Coding System (FACS; Ekman, 1980; Ekman & Friesen, 1977). This objective method, which quantifies specific facial
muscle movements, has allowed researchers to examine the expression of basic emotions cross-culturally and validate the universality of basic facial emotions (Fasel & Luettin, 2003).

Regarding the perception of emotion, some research has suggested that particular emotions are also recognized using information gleaned from particular regions of the face, rather than through holistic processing of the entire face. The identification of happiness and disgust appears to be more dependent on information in the lower half of the face, including the mouth region, and the identification of sadness, fear, and anger on the upper half of the face, including the eye region, whereas surprise appears to be equally well identified using information from either the upper or lower portion of the face (Calder, Young, Keane, & Dean, 2000).

**Prosodic.** Scherer (1986) proposed that individual acoustic cues would, when combined, signal discrete emotions. Subsequently, research on the communication of discrete emotions in speech has been generally consistent with this prediction, and cues for discrete emotions have been characterized. Methods used span several types of speech productions (i.e., natural, induced, stimulated, and computer-generated emotional speech) and have aided in identifying the specific speech components that convey discrete emotions (Scherer, 2003). The physiological experience of emotion produces changes to respiration, articulation, and phonation, which in turn produce prosodic cues (Banse & Scherer, 1996). The cues most often identified as helpful in detecting specific emotions are fundamental frequency (e.g., pitch), timbre or tone quality (produced via phonation and articulation), loudness, intensity, and modifications to speech timing, such as rate, use and duration of pauses, and phase-duration (Banse & Scherer, 1996; Juslin & Laukka, 2001).
Emotions can be accurately recognized in emotional speech prosody at rates significantly higher than chance (Banse & Scherer, 1996; Juslin & Laukka, 2001). Based on these findings, it is argued that acoustic cues combine in relatively distinct patterns, thereby allowing the differentiation of discrete emotions (Belin, Fecteau, & Bédard, 2004; Pittam & Scherer, 1993). Patterns of acoustic cues have been identified for a number of basic emotions. For example, sadness is characterized by lower and less variable pitch (decreased fundamental frequency) and lower intensity, whereas happiness is characterized by higher and more variable pitch (increased fundamental frequency), increased intensity, and an accelerated rate of speech. However, not all emotions have well-characterized acoustic profiles (e.g., fear and anger). There is also considerable overlap in the cues associated with each emotion. For example, both fear and anger are associated with high, variable pitch and increased intensity, cue utilization very similar to that found in happiness. Yet, fear and anger are opposite in valence to happiness, and anger and happiness are both approach emotions, whereas fear is a withdrawal emotion. These findings suggest that perhaps there is either a wider range of vocal cues that have not yet been studied (Juslin & Laukka, 2001) or that a discrete emotion model may not be the best fit for vocally expressed emotions (Scherer, 2003).

**Lexical.** In contrast to the facial and prosodic channels, considerably less psychological research has been conducted to ascertain cues for decoding discrete emotions in the lexical channel. Much of the work characterizing this channel instead comes from computer science and linguistic researchers, who have analyzed rules for emotional language and built emotional word lexicons (Pennebaker, Mehl, & Niederhoffer, 2003). Although estimates indicate that there may be as many as 2,000 words for emotion within the English language (Wallace, 1973), these words can typically be categorized as subtypes of basic emotions. For example, joyful, delighted,
and elated are all words that are used to describe the emotion of happiness. However, similar to research on facial and prosodic emotion, arousal and valence appear to contribute significantly to the perception of emotional linguistic content, as does frequency of word use (Nakic, Smith, Busis, Vythilingam, & Blair, 2006; Scott, O'Donnell, Leuthold, & Sereno, 2009).

**Postural/Gestural.** It has long been thought that body acts and posture can communicate emotional information (Ekman & Friesen, 1967). However, in contrast to the robust body of literature and facial and prosodic emotion, the postural/gestural channel has only recently begun to be better characterized. Research using both static (i.e., fixed body postures) and dynamic stimuli (i.e., videos of emotional body movements) indicates that the emotions of happiness, anger, fear, and sadness can be readily identified from body posture in both posed static and dynamic stimuli, while disgust is more often detected at below-chance levels (Atkinson, Dittrich, Gemmell, & Young, 2004; Coulson, 2004; Gelder, 2006). The identification of anatomical cues signaling discrete emotions is still somewhat limited. For static body posture, it has been suggested that features such as facial orientation (toward or away from a stimulus), sagittal movement, spinal flexion, open/closed and forwards/backwards reaching, serve as cues for discrete emotions (Inouye, 1998). Head position (i.e., backwards, forwards, or upright) and arm position also convey information about discrete emotions. For example, happiness is associated with a backwards head position, no forward movement of the chest, and raised arms with straight elbows, and is more accurately identified when viewed from the front than from the back. In contrast, sadness is associated with a forwards head bend, arms down at the sides of the body, and a forward chest bend, with no twisting of the torso (Coulson, 2004). Additional research suggests that information about the velocity, acceleration, and smoothness/jerkiness of arm
movements is also used to help make decisions about discrete emotions, as does gait information such as arm swing, stride length, and walking speed (Montepare, Goldstein, & Clausen, 1987).

**Representation and perception of emotion in music.** The communication of emotion in music is distinct in several theoretical ways from that of the channels of communication reviewed above, and thus merits some additional review. It should be noted that music can arouse emotional feeling in the listener and that music listening is often used for emotion regulation purposes (Gabrielson, 2001; Juslin, 2000). However, the present focus is on the expression and perception of emotion in music. The idea that music can convey emotion is supported by a substantial body of empirical literature and is widely accepted among psychology researchers (Juslin, 2013; Swaminathan & Schellenberg, 2015). When asked to identify what, if anything, music expresses, 100% of survey participants identified emotion (Juslin & Laukka, 2004) and both children and adults can readily identify emotions in music (Lima & Castro, 2011a, 2011b; Terwogt & Grinsven, 1991; Thompson & Robitaille, 1992; Trainer & Trehub, 1992). Functional brain imaging research provides further evidence, as music listening activates regions also associated with emotional perception (Peretz & Zatorre, 2005). Although musical expression of emotion might be thought of as the emotional communications of the composer or musician, musical expression in the psychological literature is typically operationalized as synonymous with the perception of emotion in music. In other words, for psychologists, music expresses emotion because music is overwhelmingly perceived as expressing emotion (Juslin, 2013).

It should be noted that, outside of the psychological community, this conclusion is not nearly as straightforward (Trainor & Schmidt, 2003). Musicologists and philosophers have put forth a number of alternative perspectives about what, if anything, music can and does actually
express, and these positions can be broken down in several ways. “Absolutists” argue that meaning in music is exclusively the result of the perception of the music itself, whereas “referentialists” argue that music communicates meaning that is in some way specifically referring to the extramusical world, such as concepts or emotional states. Within the study of aesthetics, among other perspectives, there are “formalists” who argue that the value of a piece of art is derived exclusively from the perception of the physical properties of the artwork itself (e.g., colors, shapes, and lines, or, in the case of music, the acoustic properties) and “expressionists” who argue that the value of artwork is in its ability to effectively express inner states (e.g., emotions or the artist’s mental state; Meyer, 1961). These create a number of possible perspectives on musical emotion. For example, it has been argued that the meaning of music is exclusively in the perception and understanding of musical relationships, and, thus, can have intellectual and aesthetic meaning, but not emotional content (i.e., absolute formalists). However, more relevant to psychological research findings, is the position that music does convey emotion and that this expression is either 1) based exclusively on the structure and unfolding over time of musical sounds/acoustic elements, independent of extramusical concepts (i.e., absolute expressionists) or 2) dependent on understanding deliberate referential content, such as to a concept or emotion (i.e., referential expressionists; Meyer, 1961; Trainor & Schmidt, 2003). It should be noted that these positions are not mutually exclusive of one another (i.e., music may convey emotion through a combination of structure and reference).

Many psychologists, therefore, accept an expressionist position, that is, music does express emotion. However, there is not yet agreement as to precisely how and why music conveys emotion (i.e., is the emotional quality of music absolute, referential, or a combination of the two?). From an evolutionary perspective, emotions are adaptive reactions, primarily
nonverbal, that aid humans in navigating common obstacles and accomplishing shared goals that promote species survival (Juslin & Laukka, 2003; Trainor & Schmidt, 2003). Music, in contrast, is an art form, a product of human cognition, which does not have any immediately obvious survival value, and, on the surface, appears quite removed from the evolutionary origins of emotion. How is it, then, that music conveys emotion? Vocal expression appears to be the most phylogenetically continuous form of nonverbal expression and serves to promote a wide range of social functions across species (Juslin & Laukka, 2003). Although the origins of the development of music are not known, music making activity has been identified in preliterate cultures and has been suggested to promote emotional cohesion and greater social unity among groups (see Social Functioning on page 51). This similarity has led to the theory that music is perceived as expressing emotional content because it is, in many ways, analogous to speech (Juslin & Laukka, 2003), and perhaps even functions as a substitute for speech (such as to facilitate communication between nonverbal infants and their caregivers; Trainor & Schmidt, 2003). There is considerable empirical evidence that aligns with this theory. Musical emotions can be identified at accuracy levels similar to that of vocal emotions, are successfully identified cross culturally, can be identified at a very young age (perhaps in infancy), and appear to be communicated using acoustic cues that parallel the communication of emotion in speech (see Decoding discrete emotions in music on page 13). Additional factors do clearly contribute to music’s emotional expressiveness, such as individual associations with musical compositions, intellectual/aesthetic appreciation of music, and similarity between music and extramusical features (i.e., referential content), but these, either alone or in combination, cannot fully account for the evidence described above (Juslin, 2013).
What is the nature of musical emotion? If music does indeed communicate emotions, what is the nature of those emotions? There are two primary approaches to conceptualizing musical emotions: categorical models and dimensional models. Categorical models include, most prominently, the discrete emotion model, which encompasses the concept of basic emotions. However, there are other categorical models for music, such as Sherer’s (1984) component process approach, which states that there are as many musical emotion categories as there are possible outcomes of the appraisal process. There is also a “music specific” categorical model, which posits that musical emotions are distinct from the discrete emotions found in other channels of communication. This model proposes a set of emotions that are in fact unique to music and these emotions are aesthetic and reactive (versus the utilitarian and proactive emotions associated with voice or face) and includes emotional categories such as amazement, activation, sensuality, and dysphoria (Zentner, Grandjean, Scherer, 2008). Dimensional models, on the other hand, encompass a variety of approaches where musical emotions are placed on one or more continua such as valence or intensity. Of these, the most widely used is Russell’s (1980) circumplex model, which includes the two dimensions of arousal and pleasure (Eerola & Vuoskoski, 2011; Juslin, 2013).

There is a substantial body of literature investigating both categorical and dimensional approaches, and both have behavioral and neurophysiological support (Eerola & Vuoskoski, 2011). These two approaches are fundamentally opposed. However, there are recent efforts to integrate the two so as to develop a more sophisticated means of characterizing musical emotions that also better accounts for existing research findings. One such approach is Juslin’s (2013) “multiple layers” model. Juslin proposes that there is a core-layer of basic emotions coded via their similarity to vocal emotions. This core layer is in turn elaborated upon and, perhaps, even
modified by layers of coding built upon “intrinsic” responses (i.e., information from relationships within the music, akin to the absolute expressionist position) and “associative” responses (i.e., perceived associations between the music and some sort of extramusical entity, similar to the referential expressionist position).

Decoding discrete emotions in music. If discrete emotions can indeed be perceived in music, then, as with other channels of communication, it should be possible to identify cues that signal discrete emotions. The prevailing model for discrete emotion cues in music is based on the theory that the prosodic expression of emotion is highly related to the expression of musical emotions (as discussed above).

While this theory is still debated, there is nonetheless a substantial body of literature demonstrating that discrete emotions are conveyed in music by means of acoustic cues parallel to or derived from those used in emotional prosody. For example, speech rate in voice is analogous to tempo in music, intensity to sound level or loudness, F0 contours to pitch contour, and so on, while other acoustic features are identical in both (e.g., F0 variability). In a large meta-analysis of studies examining cues to vocal and musical emotion perception, Juslin and Laukka (2003) found that acoustic cues associated with the expression of emotion in speech overlapped almost completely for a number of discrete emotions (i.e., anger, tenderness, fear, happiness, and sadness). For example, in speech, anger is conveyed by a fast speech rate, high voice intensity, high variability of voice intensity, high F0 variability, ascending F0 contours, and microstructural irregularity, while, in music, anger is conveyed by a fast tempo, high sound level, high sound level variability, high F0 variability, ascending pitch contours, and microstructural irregularity (Juslin & Laukkam 2003). It should be noted that there are acoustic cues present in music that do not appear relevant to emotional communication in voice (e.g., staccato versus
legato articulation and vibrato), but add to the complexity and richness of emotion communication in music. However, there is nevertheless robust evidence that a number of basic emotions can be successfully identified at well-above chance levels in music and that these appear to be conveyed using cues largely parallel to those of speech (Eerola & Vuoskoski, 2011; Juslin, 2013; Juslin & Laukka, 2003).

**Individual Differences in Emotion Perception**

Research examining individual differences in emotional functioning has identified several demographic characteristics that are related to emotion perception ability, including sex, ethnicity, and age. These will be briefly reviewed below.

**Sex.** Emotion perception is frequently found to vary as a function of sex. In general, women are reported to be more accurate on measures of facial emotion perception than men (e.g., Campbell et al., 2002; Hampson, van Anders, & Mullin, 2006; Li, Yuan, & Yin, 2008; Miura, 1993; Montagne, Kessels, Frigerio, de Haan, & Perrett, 2005; Scholten, Aleman, Montagne, & Kahn, 2005). Although there is considerably less research examining other channels of communication, women have also been found to have an advantage in the lexical (Grunwald et al., 1999; Kimura & Hampson, 1993) and prosodic channels (Hall, 1978; Rymarczyk & Grabowska, 2007; Schirmer, Zysset, Kotz, & Yves von Cramon, 2004). In contrast, sex differences have not been observed in the postural/gestural channel (Montepare, Koff, Zaitchik, & Albert, 1999) or for musical emotions (Lima & Castro, 2011a, 2011b; Robazza, Macaluso, & D’urso, 1994), although very few studies were found that have examined this relationship.

**Ethnicity.** Although there are universal discrete emotions that are recognized across cultures (Ekman, 1972), there is nevertheless also evidence of what is termed an “in-group
advantage.” That is, members of a particular ethnic group are frequently found to be more accurate at recognizing emotions expressed by individuals from their own ethnic, cultural, and/or national group than they are for members of different groups (Elfenbein & Ambady, 2002; Markham & Wang, 1996; Pinkham et al., 2008; Scherer et al., 2001; Wickline, Bailey, & Nowicki, 2009). Although this phenomenon is most often examined in relation to the facial expression of emotion, there is also evidence of cultural/ethnic variation in the perception of emotional expression in the postural/gestural channels (Kleinsmith, De Silvia, Bianchi-Berthouze, 2006) although there is insufficient research to determine if these differences are in fact an in-group effect.

For music, cross-cultural research on musical emotions most often examines the extent to which individuals can perceive emotions in music composed in familiar versus unfamiliar tonal systems (e.g., is a Western listener able to accurately perceive emotion in both a Western symphony as well as a Hindustani raga?). Cross-cultural research on emotion perception on music is a relatively limited field, as the majority of studies have concentrated on predominantly Western music in a uni-cultural paradigm (i.e., perceiver and target music are of the same culture). That said, the majority of cross-cultural studies available have found that listeners are able to accurately perceive discrete emotions in tonally unfamiliar music at well-above chance levels (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004; Swaminathan & Schellenberg, 2015; Fritz et al., 2009). Limited available research assessing how the perceiver’s accuracy varies as a function of his or her and the expresser’s ethnicity further suggests universality of at least some basic emotions in music. For example, individuals native to Western Europe and Asia are able to accurately identify six basic emotions (happiness, sadness, fear, disgust, anger, and surprise) in music (Argstatter, 2015). It has been suggested that this apparent
universality is due to cross-cultural utilization of similar acoustic cues to convey musical emotions, such as consonance versus dissonance, volume, timbre, tempo, and complexity, regardless of the music’s tonal system or cultural origins (Balkwill, Thompson, & Matsunaga, 2004). Of note, this finding is consistent with the evolutionary approach to musical emotion arguing that music evolved to mimic vocal emotion cues discussed earlier (see Representation and perception of emotion in music on page 9). Similarly, patterns of physiological response and ratings of subjective feeling in response to music are also similar across cultures. In a study involving naive Canadian and Congolese Pygmy listeners (e.g., the Congolese Pygmies had no exposure to Western music or culture and vice versa), participants listened to musical excerpts of both Western and Congolese music. Both groups exhibited similar patterns of physiological response and reported similar subjective emotional experience across both types of music (Egermann, Fernando, Chuen, & McAdams, 2015). However, one study did find some subtle cultural differences in the perception of musical emotion (Gregory & Varney, 1996), which suggests that tonality may also play a culturally specific role in expressing emotion (e.g., the major key may signal happiness to a Western listener but not to a Japanese listener, due to cultural tradition).

**Age.** Age-related changes have been observed in the perception of emotion in different channels of communication, even when basic perceptual processes remain intact. As with sex and ethnicity, the facial channel is the best characterized with regard to age-related change. Specifically, declines in accuracy of facial emotion perception have been noted across the lifespan (Finley, Borod, Schmidt, et al., 2008; Isaacowitz et al., 2007; Prodan, Orbelo, & Ross, 2007; Ruffman, Henry, Livingstone, & Phillips, 2008; Savage et al., 2013; Slessor, Phillips, & Bull, 2010). A similar pattern of decline, independent of age-related hearing loss or cognitive
decline, is also observed for the prosodic channel (Dupuis & Pichora-Fuller, 2010; Finley, Borod, Schmidt, et al., 2008; Kiss & Ennis, 2001; Orbelo, Testa, & Ross, 2003; Orbelo, Grim, Talbott, & Ross, 2005; Savage et al., 2013). The lexical and postural/gestural channels are less well characterized, however, age-related declines have been found in the perception of single-words and sentence reading (Finley, Borod, Schmidt, et al., 2008; Grunwald et al., 1999; Isaacowitz et al., 2007; Savage et al., 2013) and in the perception of emotion in body movement and gesture (Montepare et al., 1999). It should be noted that, among studies that have investigated discrete emotions, different patterns of decline for discrete emotions have been found both within and across channels of communication. For example, in the facial channel, some studies found age-related decline in the perception of fear, anger, sadness, disgust, and happiness (Paulmann, Pell, & Kotz, 2008), whereas others have found declines only happiness, sadness, and anger (Ruffman et al., 2008), while different patterns have been found for the prosodic and lexical channels. This may be the result of methodological variation, but may also reflect different, channel-specific mechanisms underlying age-related declines in emotion perception. The perception of musical emotions also declines with age, and, consistent with literature demonstrating strong similarity between music and voice, the patterns of age-related changes for discrete musical and prosodic emotions are quite similar (Laukka & Juslin, 2007; Lima & Castro, 2011b).

**Emotion perception deficits and psychopathology**

A wide variety of psychiatric and neurological conditions are associated with deficits in emotion perception ability. These, in turn, can have broad impacts on daily social functioning, peer relationships, and overall quality of life (Cacioppo & Hawkley, 2009; Hawthorne, 2008; Wilson et al., 2007).
Decreased accuracy in the facial and prosodic channels is seen in several psychiatric disorders, including schizophrenia (Borod et al., 1990; Edwards, Jackson, & Pattison, 2002; Kohler, Walker, Martin, Healey, & Moberg, 2010; Weisgarber, Vermuelen, Peretz, Samson, Philippot, Maurage, … Constant., 2015) and mood disorders such as Major Depressive Disorder (Borod et al, 1990; Emerson, Harrison, & Everhart, 1999; Naranjo et al., 2011; Uekermann, Abdel-Hamid, Lehmkämper, Vollmoeller, & Daum, 2008) and bipolar disorder (Addington & Addington, 1998; Loughland, Williams, & Gordon, 2002; Vaskinn et al., 2007). Some personality disorders are also associated with impairments in facial emotion perception, such as psychopathy (Dawel, O’Kearney, McKone, & Palermo, 2012; Marsh & Blair, 2008) and borderline personality disorder (Bland, Williams, Scharer, & Manning, 2004).

Neurologically, deficits in facial and prosodic emotion perception have also been found in traumatic brain injury (Bornhofen and McDonald, 2008 and Shimokawa, et al., 2001) and in neurodegenerative disorders such as Parkinson’s disease (Borod et al., 1990; Smith, Smith, & Ellgring, 1996; for reviews, see McCabe, Borod, Meltzer, Spielman, & Ramig, 2010; Zgaljardic, Borod, Foldi, & Mattis, 2003) and frontotemporal dementia (Lavenu, Pasquier, Lebert, Petit, & Van der Linden, 1999; Lough et al., 2006; Rosen et al., 2004).

Considerably less is known about how these disorders may impact perception in the postural and lexical channels and in musical emotion. However, deficits in the detection of emotion from posture have been documented in schizophrenia (Bigelow et al., 2006), and impaired musical emotion perception has been documented in schizophrenia (Weisgerber et al., 2015), Major Depressive Disorder (Naranjo et al., 2011), alcoholism (Kornreich et al., 2012), traumatic brain injury (e.g., Patel, Peretz, Tramo, & Labreque, 1998; Peretz, Gagnon, & Bouchard, 1998), and Alzheimer’s disease and semantic dementia (Hsieh, Hornberger, Piguet, &
Hodges, 2012). There is conflicting evidence regarding a relationship between autism and deficits in musical emotion perception (Heaton, Hermelin, & Pring, 1999).

**Neural Basis of Emotion Perception**

The study proposed hypothesizes that musical training, which engages neural circuitry also involved in emotion perception, is associated with greater emotion perception accuracy. Therefore, it will be helpful provide a brief overview of the basic neuroanatomical correlates of emotion perception. (For more comprehensive reviews of the neural circuitry of emotion processes, see Adolphs, 2002a, 2002b; Davidson & Irwin, 1999; Tamietto & de Gelder, 2010.)

Emotion processing can be examined at several neuroanatomical levels. These include, first, broad lateralization of emotional processes and second, the circuitry involved in the processing of individual channels or modalities (e.g., circuitry utilized in the processing of facial emotion). It has also been suggested that there may be specific, right hemisphere dominant systems for recognizing individual discrete emotions, particularly fear and disgust (Murphy, Nimmo-Smith, & Lawrence, 2003).

**Lateralization of emotion processing.** Aspects of emotion processing are thought to be differentiated by cerebral hemisphere. Regarding emotion processing, there are several theories that propose some form of hemispheric dominance for various aspects of emotion processing. The earliest theory is the right-hemisphere dominance hypothesis, which argues that emotion processes are globally lateralized to the right hemisphere (for a review, see Borod, 1992). There is extensive clinical documentation of changes to emotion processing following damage to the right hemisphere (for reviews, see Borod et al., 1998; Borod, Andelman, Obler, Tweedy, & Welkowitz, 1992; Borod, Bloom, Brickman, Nakhutina, & Curko, 2002; Heilman, Blonder, Bowers, & Crucian, 2000). Regarding emotion perception specifically, individuals with right
hemisphere damage are generally less accurate at identifying emotion than those with either left hemisphere damage or healthy controls (Adolphs, Damasio, Tranel, & Damasio, 1996; Borod et al., 1998; Kucharska-Pietura, Phillips, Gernand, & David, 2003; Mandal et al., 1999). An alternative model is the valence hypothesis, which argues that negative emotions are processed in the right hemisphere and positive emotions are processed in the left (Borod, Bloom, Brickman, Nakhutina, & Curko, 2002). This model originated with observations that left hemisphere damage can lead to a catastrophic depressive reaction (Goldstein, 1939). This model is supported by both behavioral studies using individuals with left and right hemisphere brain damage (e.g., Borod, 1993, Borod et al, 1997; Everhart & Harrison, 2000; Sackeim et al., 1982) and by EEG studies examining left versus right hemisphere activation in response to positive and negative stimuli (e.g., Davidson & Fox, 1982; Ekman and Davidson, 1993; Lee et al., 2004).

**Facial channel.** The perception of facial emotions does appear to be preferentially processed in the right hemisphere in behavioral studies (for review, see Borod et al., 2002), whereas evidence from imaging studies is more equivocal (for review, see Wager, Phan, Liberonz, & Taylor, 2003). Regarding specific neural circuitry, facial emotions are processed by a highly specialized network comprising multiple brain structures, including the fusiform gyrus, superior temporal sulcus, amygdala, prefrontal cortex (in particular, the orbitofrontal cortex), occipitotemporal cortex, and areas related to aspects of the somatosensory cortices (for review, see Adolphs, 2002).

**Prosodic channel.** The right hemisphere is dominant in mediating aspects of emotional language processing. Evidence for this relationship has been found in lesion studies (Borod, Cicero, et al., 1998), studies of dichotic listening (e.g., Dmitrieva, Gel’man, Zaitseva, & Orlov, 2006), and functional neuroimaging studies (e.g., Pihan, Altenmuller, & Ackermann, 1997;
Wildgruber, Ackermann, Kreifelts, & Ethofer, 2002). Schirmer and Kotz’s (2006) proposed model for vocal emotion perception suggests that this process involves integration of a distributed network of regions. These regions include initial recognition bilaterally at the level of the primary auditory cortex and superior temporal gyrus (where acoustic features, such as frequency and amplitude, are coded). Additional emotional features are then integrated at the level of the right superior temporal sulcus and right prefrontal cortex (Schirmer & Kotz, 2006). The amygdala and parietal lobe have also been implicated in prosodic emotion processing (Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002).

**Lexical channel.** Although general language processing is overwhelmingly associated with the left hemisphere, there is nevertheless evidence of a right hemisphere bias in processing emotional language specifically (e.g., Borod et al., 1992, 1998; Cicero, Borod, et al., 1999; Zgaljardic, Borod, & Sliwinski, 2002). Less is known about the neuroanatomical correlates of lexical emotion perception as compared to facial and prosodic perception. However, there is evidence that the left orbitofrontal gyrus and bilateral inferior frontal gyri are involved in the processing of emotional, but not neutral, words (Kuchinke et al., 2005). Neuronal activation may also vary as a function of valence, such that the perception of positive words is associated with activation in the bilateral middle temporal and superior frontal gyri while negative words are associated with amygdala, middle temporal cortex, and cingulate cortex activation (Kuchinke et al., 2005; Nakic et al., 2006).

**Postural/gestural channel.** Emotion communicated via bodily expression appears to be processed quite similarly to facial emotion, based, in part, on evidence demonstrating broad right hemisphere dominance in bodily expression processing (de Gelder, 2006). As with facial emotion, perception of bodily emotion engages neural circuitry that includes the amygdala,
fusiform gyrus (particularly in the identification of fear), cingulate cortex, rostralcaudal portion of right superior temporal sulcus, right parietal somatosensory cortex, and occipitotemoral regions (Myers, Twaite, Meltzer, Teague, Murray, & Borod, 2012).

**Musical emotions.** Evidence is limited and rather equivocal regarding hemispheric dominance for the processing of musical emotions. There is some indication of preferential processing of musical emotions in the right hemisphere (Park et al., 2014), but also evidence of left-hemisphere dominance in the processing of positive basic emotions (with no clear evidence of dominance for negative basic emotions; Khalfa, Delbe, Bigand, Reynaud, Chauvel, & Liegeois-Chauvel, 2008). In a study examining activation in response to aesthetic emotions, rather than basic emotions (for discussion of the various models of musical emotion, see *What is the nature of musical emotion?* on page 12), left hemisphere activation, particularly in striatum and insula, which has been observed for positive emotions (e.g., wonder and joy), whereas right hemisphere activation, particularly in the right striatum and orbitofrontal cortex, has been observed with low arousal emotions (e.g., nostalgia and tenderness; Trost, Ethofer, Zentner, Vuilleumier, 2012). This same study found activation in sensory and motor areas in response to high arousal emotions, activation of ventromedial prefrontal cortex and hippocampus for low-arousal emotions, and activation of the right parahippocampal gyrus for all but positive and high arousal emotions, indicating that there may be different patterns of activation and hemisphere dominance depending on both the level of arousal and the valence of the musical emotion perceived (Trost et al., 2012). Some studies also found differing patterns of activation in response to musical emotion depending on participant level of musicianship, perhaps because of the significant structural and functional changes associated with musical training (see *Neurological differences between musicians and non-musicians* on page 24), such as shifts of
hemisphere dominance in the processing of musical stimuli from the right (in non-musicians) to the left (in musicians; Dmitrevia et al., 2006).

As with the perception of prosodic emotion, there is auditory cortex activation during the processing of emotional musical stimuli, specifically bilateral superior temporal gyrus, right Rolandic operculum, and Heschl’s gyrus, for both musicians and non-musicians. A meta-analysis of functional neuroimaging studies also found activation in the amygdala, nucleus accumbens, hypothalamus, insula, cingulate cortex, orbitofrontal cortex, pre-supplementary motor area, and hippocampus in response to musical emotion (Koelsch, 2014), whereas another study concluded that recognition of musical and facial emotions depends on the right temporal pole, amygdala, and insula, and that musical emotion recognition (but not facial) was additionally associated with the left anterior and inferior temporal lobes (Hsieh et al., 2012). In a population of patients with frontotemporal dementia, patients were impaired on recognition of happiness, sadness, anger, and fear in music, faces, and voices, and the degree of impairment was associated with a lower amount of grey matter in the insula, orbitofrontal cortex, medial prefrontal cortex, anterior cingulated, anterior temporal cortex, posterior temporal cortex, parietal cortex, amygdala, and subcortical mesolimbic system (Omar et al., 2011). Studies contrasting pleasant versus unpleasant music have found involvement in the inferior frontal gyrus, anterior insula, parietal operculum, and ventral striatum for pleasant music, and amygdala, hippocampus, parahippocampal gyrus, and temporal poles for unpleasant music (Blood, Zatorre, Bermudez, & Evans, 1999).

**Neuroscience and Music**

It has been argued that music is an ideal candidate for neuroscientific inquiry (Zatorre, 2005). Listening to, composing, and playing music cumulatively recruit nearly every cognitive
function, including, but not limited to, auditory and motor learning, speech processing, emotion, memory, and executive functioning. Music listening, alone, utilizes skills in pitch discrimination, rhythm analysis, selective attention, working memory, and semantic analysis (Koelsch, 2011; Zatorre, 2005). Music training then adds not only motor learning but also learning a new, complex symbol system (music notation), which must then be converted into a complex procedural task. Music performance is, therefore, an astonishing example of the integration of multiple cognitive processes in order to simultaneously execute complex sensory-motor tasks with precise timing, while also monitoring and adjusting performance (Patel, 2013; Zatorre, 2005). Studying musical training then simultaneously allows us to advance our understanding of how the brain processes music while also providing insights into topics as varied as brain plasticity, speech and language, child development, aging, and emotion.

**Neural basis of music processing.** There has been considerable research examining the neural bases of music, and it is difficult to identify a non-musical process or brain region with which musical performance does not overlap in some way. It has also been proposed that, in addition to this overlap with non-musical processes, music may have its own discrete processing module. This music module would be responsible for the analysis of the music-specific aspects of pitch, rhythm, lexicon, and emotion. Lesion studies and studies of individuals with congenital amusia have lent some support to this hypothesis (Peretz & Coltheart, 2003). For a comprehensive review of the neural substrates of music processing, see Peretz and Zatorre (2005) and for the neural processing of music perception specifically, see Koelsch (2011).

**Neurological differences between musicians and non-musicians.** Given the host of neural processes engaged by musical training, it has been hypothesized that musicians’ brains will have increased plasticity. Indeed, the brains of individuals with musical training appear
distinctly different from those of untrained individuals, both neuroanatomically and functionally. These differences will now be briefly reviewed. For more comprehensive reviews of anatomical and functional differences associated with musical activity, see Barrett, Ashley, Strait, and Kraus (2013) and Herholz and Zatorre (2012).

**Neuroanatomical differences.** Neuroanatomical differences are observed between musicians and non-musicians, such that various brain structures appear to be larger in musicians than in non-musicians. White matter tracts have larger volume, including the corpus callosum (Schlaug, Jancke, Huang, Staiger, & Steinmetz, 1995; Schmithorst & Wilke, 2002). Cortical grey matter volume is also larger in multiple regions, including the primary motor cortex, including the intrasculcal length of the pre-central gyrus (a marker of cortical motor hand representation; Amunts, Jancke, Mohlberg, Steinmetz, & Zilles, 2000), and auditory areas, such as Heschl’s gyrus (Gaser & Schlaug, 2003). Enlargements of the planum temporale (Keenan, Thangaraj, Halpern, & Schlaug, 2001; Luders, Gaser, Jancke, & Schlaug, 2004; Schlaug et al., 1995b), the inferior frontal gyrus (Gaser & Schlaug, 2003; Luders et al., 2004), and the cerebellum (Hutchinson, Lee, Gaab, & Schlaug, 2003) are also observed.

**Functional differences.** There is considerable evidence that musicians also exhibit patterns of neural activity distinct from those of non-musicians. Musical training may exert an influence on hemisphere dominance (Dmitrevia et al., 2006), such that non-musicians appear to process melodic stimuli preferentially in the right hemisphere and musicians process them in the left hemisphere (Bever & Chiarello, 1974; Johnson, 1977), although some studies have found the opposite pattern (Gordon, 1980; Mazzuchhi, Parma, & Cattelani, 1981).

Cortical responses to auditory stimuli are larger in musicians (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Angulo-Perkins, Aubé, Peretz, Barrios, Armony, & Concha (2014),
which occurs even at the specific instrument level (Moreno et al., 2009). For example, piano players show greater evoked responses to the sounds of piano (Pantev et al., 1998) than non-musicians. Brainstem responses to auditory stimuli in musicians are stronger and faster than in non-musicians (Kraus & Chandraeskaran, 2010; Strait, Kraus, Skoe, & Ashley, 2009), even if the time of training was remote (Skoe & Kraus, 2012), and early and late ERP components of both sound and voice processing are enhanced in musicians as compared to non-musicians (Rigoulot, Pell, & Armony, 2015). Musicians also display increased connectivity in motor and sensory areas (Luo, Guo, Lai, Liao, Liu, Kendrick, Yao, & Li, 2012) and functional changes, such as increased activation, in the superior parietal cortex and inferior temporal gyrus, following training in reading musical notation (Stewart, 2008), among many other differences.

Notably, musical training is associated with functional changes in the processing of discrete musical emotions. Specifically, musicians show greater activation than non-musicians in the superior and middle frontal gyri of the right prefrontal cortex in response to sadness and greater activation in the supramarginal and inferior parietal gyri of the right parietal cortex in response to fear. The same study, however, found no differences for happiness (Park et al., 2014).

Although these differences are correlational, there is strong indication that they are the result of prolonged music training, rather than pre-existing. For one, functional enlargement of brain regions has been already demonstrated in animal models as a result of prolonged motor-skill learning (Stewart, 2008). In humans, longitudinal studies and careful comparison of individuals with various levels of musicianship have demonstrated strong associations between structural changes, such as increased grey matter volume, in auditory, visual, and motor regions and the duration and intensity of musical training (Bengtsson, Nagy, Skare, Forsman, Forssberg,
& Ullen, 2005; Gaser & Schlaug, 2003; Schlaug, Jancke, Huang, & Steinmetz, 1995). Furthermore, no neuroanatomical differences were found between a group of children about to begin piano or string lessons and a group not beginning music lessons (Norton, Winner, Cronin, Overy, Lee, & Schlaug, 2005), further supporting the notion that these differences are acquired as a result of musical training.

Critical periods for music training. There is also evidence to suggest that these differences are influenced not only by the duration and intensity of training but also by the age of training onset, which has resulted in the proposal of a music-learning “critical period” in early childhood. For example, when controlling for duration of training, children who began music training prior to age 7 demonstrate greater corpus callosum connectivity than children who began at a later age (Steele, Bailey, Zatorre, & Penhune, 2013) and greater right ventral premotor cortical volume, while also demonstrating corresponding behavioral advantages on auditory-motor tasks (Bailey, Zatorre, & Penhune, 2014). It is not clear, however, if this critical period is music-specific, an argument supported by evidence for a music-specific module, or if music learning at this age merely exploits the already plastic nature of the developing brain.

Musical Training and Non-Emotional Mental Processes

It is evident that extensive training in playing a musical instrument is associated with and likely causes pronounced changes to the brain anatomically and functionally. Furthermore, musical performance recruits neural circuitry that is also utilized in non-musical activities. This observation raises the possibility of transfer effects from musical training to non-musical domains. Transfer effects are generally defined as the likelihood of a skill in one domain to predict either a person’s skill level or ease of acquisition of skill in another domain. For example, if a person has had musical training, will that training predict better performance on tasks of
language than that of a person without musical training? Or will the training predict enhanced acquisition of language skills over time? Predictions regarding transfer effects may be based upon behavioral relationships (e.g., music performance promotes specific behaviors/skills that may also then be used in other domains) or from neuroanatomical evidence (e.g., learning to play music causes neuroanatomical and/or functional changes to specific brain regions, so then neural processes subserved by these regions may also be enhanced).

Although the term “transfer effect” is used broadly in literature discussing both correlational and causal studies of musical training, it does, nonetheless, connote directionality (skill transfers from one domain to another). Accordingly, “relationship” or “association” is used order to avoid the implication of casual effects. For the purposes of the proposed study, the focus is on training in playing a musical instrument, typically commencing at a relatively young age, and mental abilities. This line of inquiry is related to, but distinct, from lines of research that examine associations between cognitive abilities and innate musical talent or ability, and from lines of research that examine proposed performance benefits from listening to music (the so-called “Mozart effect,” for review, see Schellenberg & Weiss, 2013).

**Musical training and auditory processing.** Music learning utilizes and over-trains auditory processes, including both right-hemisphere pitch analysis and left-hemisphere temporal and rhythmic analysis (Baeck, 2002; Yuskatis, Parviz, Loui, Wan, & Pearl, 2015). Unsurprisingly, there is substantial evidence that musically trained individuals outperform untrained peers on a wide range of both musical and non-musical tasks. Music training is associated with superior processing of music harmony and key violations (Corrigall & Trainor, 2009), faster identification of transposed (Halpern, Kwak, Bartlett, & Dowling, 1996) or

Musicians have demonstrated superior frequency (Kraus & Chandrasekaran, 2010) and pitch (Marie, Kujala, & Besson, 2012) discrimination abilities, and longitudinal experimental studies have also shown musical training improves pitch processing, such as detection of small pitch differences (Besson, Schön, Moreno, Santos, & Magne, 2007), and sound discrimination ability (Francois, Chobert, Besson, & Schon, 2013). These differences also extend to tasks of speech perception (Jantzen & Scheurich, 2014; Kraus & Chandrasekaran, 2010; Patel & Iversen, 2007). Musicians demonstrate improved ability to detect speech in noise (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Parbery-Clark, Strait, & Kraus, 2011). It has also been shown that musicians are able to better utilize aspects of speech, such as pitch contours, to extract extralinguistic information, such as distinguishing between questions and statements (Kraus & Chandrasekaran, 2010).

**Musical training and cognitive processes.** Musical training is positively associated with enhanced cognitive abilities in a number of domains, including linguistic, memory, visuospatial, and executive functions. For more comprehensive reviews, see Miendlarzewska & Trost (2013) and Schellenberg & Weiss (2012).

More specifically, musical training is associated with superior phonological awareness, reading ability (Degé & Schwarzer, 2011; Gromko, 2005; Standley, 2008) and foreign language acquisition (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Marie, Magne, & Besson, 2010; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010; Sleve & Miyake, 2006). Musicians generally perform better than non-musicians on tasks of auditory (Cohen, Evans, Horotwitz, & Wolfe, 2011; Kilgour, Jakobson, & Cuddy, 2003) and verbal memory (Brandler & Rammsayer,
Musical training is also associated with enhanced visuospatial ability (Hetland, 2000; Sluming, Barrick, Howard, Cezayirli, Mayes, & Roberts, 2002; Patston & Tippett, 2011; Sluming, Brooks Howard, Downes, & Roberts, 2007; Stoesz, Jakobson, Kilgour, Lewycky, 2007), and musicians also appear to have less laterality bias than non-musicians (Patston, Corballis, Hogg, & Tippett, 2006; Patston, Hogg, & Tippett, 2011; Patston, Kirt, Rolfe, Corballis, & Tippett, 2007) as well as faster reaction times on visual tasks (Brochard, Dufour, & Deprés, 2004; Patston et al., 2007).

Lastly, musical training is associated with superior performance on many aspects of executive functioning (Degé et al., 2011), including response inhibition (Moreno et al., 2011, verbal fluency (Gibson, Folley, & Park 2009; Hassler, Birbaumer, & Feil, 1985), divergent thinking (Gibson, Folley, & Park 2009), and auditory working memory (Franklin et al., 2008; Hansen, Wallentin, & Vuust, 2012; Pallesen, Brattico, Bailey, Korvenoja, Koivisto, Gjedde & Carlson, 2010; Parbery-Clark et al., 2009; Posedel, Emery, Souza, & Fountain, 2012; Tierney, Bergeson-Dana, & Pisoni, 2008)

Interestingly, children with musical training do appear to have higher full-scale IQs (FSIQs), relative to untrained peers, when administered comprehensive tests of intelligence (Schellenberg, 2006, 2011a) and single measures of fluid intelligence, such as the Raven’s Progressive Matrices (Nutley et al., 2014) or Kaufman Brief Intelligence Test (Corrigall & Trainor, 2011), although not all studies have yielded similar results (e.g., Schellenberg & Moreno, 2010). A recent meta-analysis concluded that there is a positive association between
musical training and intelligence, regardless of the measure used (Jaschke, Eggermont, Honing, & Scherder, 2013).

However, the directionality of the relationship between musical training and intelligence is key. Schellenberg (2008, 2011a) has argued that children with higher IQs are more likely to perform well on measures of cognitive ability and also more likely to take music lessons for any number of reasons (e.g., more likely to seek mentally engaging activities or to have more access to musical training). Nevertheless, Schellenberg (2004) conducted a randomized, experimental longitudinal study that demonstrated significant increases in FSIQ in 6-year olds as a result of one year of musical training and no FSIQ changes in children who had received either drama lessons or no lessons. Another experimental study by Moreno et al. (2009) also showed greater FSIQ increases in children who received music lessons than in those who received painting lessons.

In contrast, when adult musicians are compared to other matched adults (Bidelman, Hutka, & Moreno, 2013; Lima & Castro, 2011a) on measures of fluid intelligence (such as the Raven’s Progressive Matrices test) or when college students with various majors, including music, are compared on comprehensive measures of intelligence (Schellenberg, 2011a), differences in intelligence equivocal and, when present, typically minor. These observations may indicate that early benefits from music training for IQ are transient and simply accelerate the pace of early development. Alternately, individuals with high IQs may be predisposed to seek out musical training, but are not more likely to then become professional musicians.

The extent and precise nature of these relationships is still being teased apart, and conclusions specifically about transfer effects from music to cognitive abilities appear limited at this time to primarily phonological awareness (Corrigall & Trainor, 2011; Moreno et al., 2009),
visuospatial ability (Portowitz, Lichtenstein, Egorova, & Brand, 2009; Rauscher, 2002; Rauscher & Zupan, 2000), and executive functioning (Bugos, Pearlstein, McCrae, Brophy, & Bedenbaugh, 2007; Degé et al., 2011; Moreno et al., 2011). Jaschke et al. (2013) have called for greater uniformity in research methods and a focus on well-executed, longitudinal randomized control trials of at least 3 years in duration. This or any similar effort towards greater study uniformity would immensely improve the current state of the literature and enable firmer conclusions about transfer effects from music to cognition.

**Musical Training and Emotional Processes**

Relationships between musical training and emotion processes have received substantially less attention than those with auditory perception or cognition. However, many researchers take the position that the essence of music is the expression of emotions (Levitin & Tirovolas, 2009). Musical performance involves the experience, as well as expression, of emotion, typically through the postural/gestural channel of communication. Accordingly, a wide range of functional networks and brain regions recruited in musical performance also overlap with regions engaged in emotional processes (Peretz & Zatorre, 2005). Given this overlap, transfer effects may be possible not only from musical training to cognitive abilities but also to emotional processes. The existing literature examining on musical training and emotion perception will now be reviewed.

**Emotion perception.**

**Prosodic channel.** The rationale for assessing emotional prosody in musicians is typically based on evidence indicating relationships between music and speech and on the notion of auditory overtraining as a result of constant musical engagement. Musicians have highly developed auditory abilities, are thought to be superior to non-musicians at processing speech
MUSICAL TRAINING AND EMOTION PERCEPTION

(see Musical training and auditory processing on page 28), and, therefore, are anticipated to be better at the auditory processing skills necessary to detect emotion in speech. Furthermore, there is considerable support for the notion that emotions are communicated through speech in a manner analogous to music, including evidence of parallel use of acoustic cues to convey emotion in both speech and music (Patel, 2013; Scherer, 1995; Slevc, 2012; see, also, Juslin and Laukka, 2003 for a comprehensive review of similarities between emotion in speech and music).

Nilsonne and Sundberg (1985) provided early evidence that musicians may have a superior ability to ascertain emotional content in speech. In their study, conservatory students and law students listened to voice recordings made of psychiatric patients during periods of depression and periods of remission. Although both groups achieved above-chance levels of accuracy in differentiating depressed recordings from remission recordings, the conservatory students were significantly more accurate than the law students.

Thompson, Schellenberg, and Husain (2004) further explored this relationship in a three-part study. In the first part, musically trained adults were better able than untrained adults at identifying emotions conveyed in emotional tone sequences created by electronically transforming emotional speech. This effect was found across all four of the emotions tested (i.e., happy, sad, fearful, and angry). The second part replicated the first while also adding spoken sentences (spoken in an emotional tone of voice, yet neutral in lexical content) in both the participant’s native language of English and in an unfamiliar language, the Filipino language Tagalog. There was no overall effect of musical training; rather, musically trained individuals were better at identifying sad and fearful emotions, untrained were better at identifying happy emotions, and both groups were generally equivalent at identifying angry emotions. The authors theorized that untrained individuals had a bias towards choosing happy, but they could not
evaluate this because responses were recorded as only correct or incorrect (i.e., the emotion selected was not recorded). Notably, no effect of language was obtained in this analysis. When the Tagalog samples were examined separately, there was a main effect of training, such that participants with musical training were more accurate than those without, indicating that musical training is also associated with enhanced perception of prosodic emotion in a foreign language.

While both these studies show positive relationships between musical training and perception of emotion, both are correlational. In a third part, a longitudinal experiment, the authors randomly assigned 6-year-olds to one of four conditions: drama, singing, or keyboard lessons, or a no lesson group. After one year of training, prosody was measured through two prosodic discrimination tasks (i.e., happy-sad and fearful-angry). Results were mixed; the happy-sad task was uninformative due to ceiling effects, whereas the fearful-angry task revealed that both the keyboard and drama groups performed equivalently and significantly outperformed the no-lessons group. By contrast, the singing group was not significantly different from either the drama or no-training group (Thompson et al., 2004). These three studies, therefore, showed selective advantages of musical training in identifying emotional prosody, but results were not uniform across all conditions and emotions.

Another examination of musically trained and matched untrained children ages 7-17 revealed that the musically trained children were more accurate in identifying which of three emotions (joy, anger, and neutral) were expressed in speech. However, this study used a limited number of prosodic stimuli and was designed to assess left- versus right-hemisphere dominance for the processing of prosody in musicians, not to evaluate emotion perception accuracy. The authors also cited several articles that apparently demonstrate that adult vocalists are more sensitive to emotional speech content than are non-vocalists. However, these studies were
published in Russian and therefore were not able to be reviewed in this current paper (Dmitrieva et al., 2006).

Lima and Castro (2011a) compared trained adult musicians and matched non-musicians, all native Portuguese speakers. In contrast to prior studies, these authors used a much wider range of basic emotions and utilized neutral content sentences stated in emotional tones of voice rather than electronically manipulated stimuli. They found that musically trained participants were significantly more accurate in identifying emotions than were untrained participants in a forced-choice paradigm. Unlike in the Thompson et al. study (2004), this held true for all emotions included in the study (i.e., anger, fear, disgust, sadness, happiness, surprise, and neutral). Notably, the types of errors made by both groups were similar, suggesting that musicians are not perceiving the stimuli in a fundamentally different way than non-musicians. They also included a measurement of reaction time, and, interestingly, they did not find any group differences for the speed at which participants selected their answer, or any significant relationship, either positive or negative, between reaction time and accuracy.

A study by Young, Parsons, Stein, and Kringlebach (2012) demonstrated that non-depressed musicians were better able to discriminate which of two infant cries was distressed than were depressed musicians or non-depressed controls. Depressed musicians, on other hand, performed at about the same level as non-depressed controls. This study is notable as the stimuli used are the most naturalistic and thus arguably have the best external validity of the studies reviewed.

In an ERP study of emotional prosody processing, Pinheiro, Vasconcelos, Dias, Arrais, and Gonçalves (2015), 14 adult musicians and 14 controls listened to neutral sentences stated in either a neutral, happy, or angry tone of voice. Musicians were found to be more accurate at
identifying all three emotional conditions, but group differences were only significant for the angry condition. Functional differences were also observed between musicians and controls. Specifically, P50 and N100 ERP amplitudes were less negative for musicians across all conditions, which is thought to be suggestive of an influence of musical training on extraction of basic acoustic properties from speech.

Most recently, two-part study by Mualem and Lavidor (2015) yielded conflicting results. In part one, a short-term intervention study, 24 college students without prior musical training were assigned to one of two groups: a brief music intervention group or a brief art intervention group (designed to have parallel structure and levels of time commitment and engagement). The groups completed both a vocal emotion and a musical emotion identification task prior to the intervention and immediately following completion of the intervention. The music intervention group, but not the art intervention group, showed significant improvement in the identification of vocal emotions from pre to post. There was no significant improvement in music emotion identification. In part two, 24 college students without musical training and 23 music college students, all of whom had been trained in voice or an instrument for a minimum of 6 years, were compared on vocal and musical emotion identification tasks. In contrast to study one, no significant differences were found between groups for either task. It should be noted that this study was quite limited in sample size and, crucially, the interventions were designed specifically to emphasize learning and discussion about emotions in either music (the music intervention) or art (the art intervention), whereas these skills are typically not explicitly emphasized in traditional musical training of the type ostensibly undergone by their music school participants. These factors may therefore contribute to the conflicting findings.
Only one study was identified that did not find any support for improved prosodic perception in musicians. Trimmer and Cuddy (2008) reported an expansion of the Thompson et al. (2004) study, where they assessed both prosodic perception, via emotionally intoned sentences and tone analogues, and emotional intelligence, via the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Trimmer & Cuddy, 2008). The authors found no relationship between musical training and perception of emotional prosody, and instead found an association between accuracy and the experiential area score on the MSCEIT, which encompasses the ability to recognize emotion in faces and pictures and the ability to utilize emotions for cognitive processes.

It should be noted that this study differs from those previously discussed in several substantial ways. Rather than use matched groups of musicians and non-musicians, the authors used a single group of adults with an overall average of 6.5 years and range of 0-17 years of training, and examined relationships between musical training and both raw years of training as well as a music training score, generated for each participant based on responses to a questionnaire assessing various aspects of musical training. Although this approach might provide a more comprehensive way of assessing overall musical ability, the average of 6.5 years of training is low, and there may not have been enough participants with training sufficient to detect group differences as a function of training. Thompson et al. (2004), Lima and Castro (2011a), and Pinheiro et al. (2015) all required a minimum of 8 years of musical training for their musician groups, and in all three studies the actual averages obtained exceeded this minimum by several years. The exception is the study by Young et al. (2012) in which the cutoff was 4 years minimum of training. However, the emotion perception task used in by Trimmer and Cuddy (2008) also differs substantially from those discussed previously; studies that obtained effects of
musical training used forced choice tasks, such that participants must choose which emotion is represented in the stimuli from a predefined list of possible choices. Trimmer and Cuddy (2008) instead employed a paradigm whereby participants were asked to rate each prosodic stimuli for the prominence of each of four different emotions (angry, fearful, sad, and joyful), on a 0-10 scale, a considerably more complex and demanding task.

Factors that likely contribute to variability in findings across these studies include differences in the nature of the prosodic stimuli used (e.g., emotional speech or electronically modified sounds). Two of the studies used both gliding tone analogues, generated from electronically manipulating emotionally intoned sentences to produce pure pitch stimuli, and semantically neutral, emotionally intoned sentences (Thompson et al., 2004; Trimmer & Cuddy, 2008), one used readings of a “standard text,” the content of which was not specified, and filtered the speech to isolate fundamental frequency (Nilsonne & Sundberg, 1985), whereas two used only emotional speech (Dmitrieva et al., 2006; Lima & Castro, 2011a). These studies also take place across a variety of populations and in several languages (i.e., English, Russian, Portuguese, and Hebrew), although the latter strengthens the overall findings, as positive results were found in studies using each of these four languages. Lastly, although all studies employed discrete emotions, the number and type of emotions varied across all studies, further complicating comparison of results. Nevertheless, it does appear that musical training is associated with enhanced perception of prosody. While only two studies (Mualem & Lavidor, 2015; Thompson et al., 2004) contained longitudinal experiments, both did support transfer effects from musical training to prosodic emotional perception.

**Music.** The notion that musicians should be experts at the identification of emotion in music has considerable intuitive and theoretical appeal. It is argued that music can indeed
express emotions, perhaps was even created for the expression of emotion, and that this expressivity is what makes music so powerful to the listener (Eerola & Vuoskoski, 2011). The acoustic mechanisms that communicate emotion in music are similar, if not identical, to those utilized in emotional speech (Juslin, 2013), and as discussed above, musicians do appear to have an advantage in identifying emotional prosody. Musicians are also expert listeners with extensive exposure to music and training in music theory and performance, are skilled in communicating the emotions expressed in musical compositions (see Emotional Expression on page 45).

In an early study examining musical training and emotion recognition, Hevner (1935) examined the effects of musical training, musical talent, and intelligence on the ability to identify emotional valence (i.e., positive or negative emotion) in music. Although all participants were generally accurate, musical training was associated with slightly superior accuracy more so than was musical talent. Intelligence was not associated in any way with emotion identification accuracy. In a study of music perception over the lifespan (adults ages 17-84), Lima and Castro (2011b) found a positive correlation between music training and improved accuracy, regardless of age. Another study found that Chinese school children ages 6-10 with 6 or more months of music training were more accurate at identifying emotion in music (as well as in faces) than those without training (Yong & McBride-Chang, 2007). In a study where participants listened to a musical excerpt and then had to select other musical excerpts that elicited a similar emotional experience in the participant, slight differences were found between musicians and non-musicians, although there was general agreement as to how the emotions were classified with respect to type by both the non-musicians and musicians (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005).
However, several studies have failed to find emotion perception differences between children (Robazza et al., 1994) or adults (Mohn, Argstatter, Wilker, 2010; Mualem & Lavidor, 2015; Robazza et al., 1994; Ramos, Bueno, & Bigand, 2011) with and without musical training. In a related study, no difference was found between the length of time it took musicians or matched controls to make a decision about the valance of musical stimuli (Bachorik, Bangert, Loui, Larke, Berger, Rowe, & Schlaug, 2009). Interestingly, in a study of individuals with congenital amusia, participants were found to exhibit normal recognition of emotions (i.e., happiness, sadness, fear, and peacefulness) in music, even though they had demonstrated deficits in pitch perception. This finding suggests, potentially, a degree of dissociability between the perceptual and emotional processing of music, and that factors such as timbre and temporal characteristics (e.g., tempo) may play a greater role in the expression of emotion in music than previously thought (Gosselin, Paquette, & Peretz, 2015).

In sum, evidence is mixed as to whether or not musical training is associated with improved musical emotion perception, and none of the studies available is longitudinal, limiting conclusions about transfer effects. These studies all vary in a number of key ways that make comparison between them difficult. For one, they vary considerably as to participant age and to onset and duration of musical training, which makes comparison between these studies difficult. Training onset prior to age 7 is related to changes in the subcortical processing of timbre and pitch, whereas practice duration of 10 years or more is related to enhanced timing-related abilities (Strait et al., 2009), and these abilities are key for the perception of emotion in music. Of note, musicians do show greater preference for music that expresses mixed emotions (Ladinig & Schellenberg, 2012). Also, when musicians and non-musicians are asked to judge the duration of musical stimuli, non-musicians’ responses are influenced by the emotion (for example, they are
more likely to overestimate the length of sad musical segments), whereas musicians are more accurate and, therefore, appear less susceptible to the effect of the emotional content (Panagiotidi & Samartzi, 2013). There are, then, clearly differences between musicians and non-musicians with regard to interactions among perception, music, and emotion. However, children are able already to perceive emotion from an early age in a very consistent way (Kratus, 1993), and humans are, in general, already skilled listeners and identifiers of emotion (Nawrot, 2003).

These studies also vary considerably as to the nature of the musical stimuli and utilize a variety of theoretical approaches. There is no clear consensus yet as to whether music expresses emotions discretely, dimensionally, or through some other way. In the discrete model of musical emotion, emotions communicated in music are variations on basic emotions (e.g., happiness), whereas in the dimensional model, emotions are expressed as a combination of valence (i.e., positive or negative) and arousal (Eerola & Vuoskoski, 2011). Other models include music-specific theories of emotion in music or models based around intensity or preference. That there is no one clearly established theory of emotion in music makes it challenging to interpret these findings. Juslin (2013) has proposed a promising “multiple layer” theory of music emotion that unites basic emotions with additional levels of modification that enable the communication of more nuanced and complex emotions. This model may provide a means to better conceptualize musical emotion for future emotion perception studies.

**Lexical, facial, and postural channels.** Although lexical, facial, and prosodic channels of emotion communication do not directly overlap with auditory processing domains, as do music and prosody, Lima and Castro (2011a) nevertheless suggested that these channels might also be enhanced via transfer effects from musical training. Additionally, the notion of a right-hemisphere-based “general processor” that subserves emotion processes in all channels of
communication has been suggested and has some support for emotion perception in healthy adults (Borod, Pick, et al., 2000). Furthermore, musical training engages not only auditory emotion processes but also involves emotional experience, facial and postural expression (and, in the case of voice training, prosodic expression) during performance, and facial and postural perception during group performance. Therefore, theories of embodied cognition, which posit that the process of perceiving or thinking about emotion involves perceptual, somatovisceral, and motoric re-experiencing of the relevant emotion (Niedenthal, 2007), would suggest that musicians’ repeated experience and expression of emotion during musical performance may transfer to enhance perception of emotion.

Music has been shown to prime non-affective semantic processes, such as the identification of target words semantically related to the musical passage (e.g., the word “needle” paired with a musical passage representing sharp pain from a heart attack or the word “devotion” paired with a church anthem; Koelsch et al., 2004), and affective music has been shown to prime the identification of affective words (Sollberger, Reber, & Eckstein, 2003), demonstrating links between music and both non-emotional and emotional lexical content. It has also been observed that severity of impairment in music emotion recognition is associated with atrophy of areas dedicated to language processing, particularly verbal semantics, specifically the left anterior and inferior regions of the temporal lobe (Hsieh et al., 2012). This suggests a neuroanatomical relationship between music emotion processing and language processing, adding to the growing body of literature on language-music overlap. However, no studies have examined musicians’ perception of lexical emotion at this time.

Although the facial channel is perhaps theoretically and neuroanatomically the most distant from the musical channel, cross-modal affective priming has nevertheless been
demonstrated between the two channels. Affective music enhanced participants’ accuracy in perceiving the amount of emotion expressed in happy, sad, and, notably, neutral faces (Logeswaran & Bhattacharya, 2009), demonstrating the possibility of affective transfer from music to facial processing. There is also neuropsychological evidence of this relationship. Patients with semantic dementia exhibited difficulty recognizing both facial and musical emotions, particularly negative ones (Hsieh et al., 2012). Patients with schizophrenia show impaired recognition of emotion in face, voice, and music (Weisgarber et al., 2015). Recently detoxified, alcohol-dependent patients exhibited emotion perception deficits in face, voice, and music, suggesting a shared emotion recognition network damaged by alcohol use (Kornreich, Brevers, Canivet, Ermer, Naranjo, Constant,… Noël, 2012), and similar results are found in studies of depressed patients (Naranjo, Kornreich, Campanella, Noël, Vandriette, Gillian,…Constant, 2011). Furthermore, imaging supports considerable correspondence of brain activation during identification of musical and facial emotions, particularly the amygdala, insula, and right temporal pole. This supports the notion of a common neural substrate involved in the processing of both facial and musical emotion.

In a recent ERP study, participants were presented with emotional faces followed by emotional music stimuli and asked to evaluate if the two stimuli were emotionally congruent. Reaction time decreased when the facial emotion was congruent with the musical emotion, and ERP amplitude increased during the incongruent condition, indicating that facial emotion can prime the processing of musical emotion. However, no differences were found between participants with and without musical training (Kamiyama, Abla, Iwanaga, & Okanoya, 2013). In a study of non-verbal emotional expression in musical performance, both musicians and non-musicians assigned higher ratings of expressiveness and interest to audio-visual musical stimuli
(videotapes of a musician performing) that were performed in an expressive manner versus that performed in a "deadpan” manner. However, musicians assigned higher expressiveness ratings overall than non-musicians (Broughton, Stevens, & Malloch, 2006), and this may represent a greater sensitivity to non-verbal communication during musical performance on the part of musicians. In a study of 6-10 year-old Chinese children, children who had received at least 6 months of musical training were better at tasks of facial and musical emotion perception (Yong & McBride-Chang, 2007). These results are intriguing, however, as this was not a longitudinal study, it cannot be concluded if their superior ability was, in fact, due to musical training.

These were the only two studies identified that examine perception in the facial and/or postural/gestural channels, and no studies were identified that examine lexical perception and musical training. Exploration of these relationships would be quite beneficial for a number of reasons, including the possible identification of additional extra-musical processes that benefit from musical training. This line of research would also expand our current understanding of the relationships in perception ability between different emotion communication channels.

Furthermore, it has been proposed that language and music may be co-occurring natural abilities that represent subsets of a single shared cognitive processing domain or even that language is managed neuroanatomically and functionally as a specialized type of music (Koelsch, 2013). Therefore, exploration of the lexical channel would also be valuable for furthering our understanding of relationships between music and language processing.

*Summary of perception findings.* Musicians’ emotion perception in the facial, postural, and lexical channels is virtually unexamined in the literature. There is ample theoretical justification for examining these relationships in order to expand our understanding of both the possible benefits of musical training and underlying relationships between channels of emotion
communication. However, it cannot be ruled out that studies examining such relationships have yielded negative results and have therefore gone unpublished.

Examination of the associations between the prosodic and musical channels has yielded mixed results, including no relationship (e.g., Trimmer & Cuddy, 2008), relationships among all emotions examined (e.g., Lima & Castro, 2011a), and relationships among some emotions but not others (e.g., Thompson et al., 2004). Some variation is likely due to differences in the nature of the experimental task and the type, onset, and duration of musical training. Furthermore, healthy adults are already expert listeners and identifiers of emotion (Nawrot, 2003). There are also baseline differences in emotion perception in healthy adults, such that some emotions appear easier to detect than others. Corrigall & Schellenberg (2013) suggest that, per Sloboda (1985), musical training may enhance perception of subtle emotions, at least in music, and so stimuli using ambiguous emotions, emotion blends, and/or moderate intensity stimuli should help shed light on this hypothesis.

**Emotional expression.** Musical performance involves highly expressive facial and bodily movements, and it is evident that musicians intend to and are able to convey specific emotions to listeners (Juslin, 2000). Emotions are communicated during performance, through the postural and/or facial channels (Thompson, Graham, & Russo, 2005; Thompson, Russo, & Quinto, 2008), which are congruous with the emotion expressed by the piece being performed (Dahl & Friberg, 2007; Davidson, 2012). The movements and gestures used to communicate emotions are quite similar to those spontaneously evoked in music listeners, as well as in bodily communications in non-musical settings (Burger, Saarikallio, Luck, Thompson, & Toivianen, 2013). This additional level of communication has been shown to enhance viewer comprehension of emotion in music (Krahé, Hahn, & Whitney) and is intimately linked to
viewer engagement and enjoyment of music (Broughton et al., 2006). Given that musical performance is so expressive, are musicians better able to express emotion outside of musical performance than their untrained counterparts?

Unlike actors, musicians do not generally receive extensive, if any, explicit training in techniques for emotional expression. Instead, such skills are generally communicated implicitly in instrumental training, through modeling (Karlsson & Juslin, 2008). This teaching approach is based on an assumption that expressive ability is a function of innate musical talent and therefore cannot be taught (thus implying that musicians are naturally more expressive than non-musicians) (Hallam, 2006). However, there is evidence that expressivity in musical performance is in fact predominantly a function of their intensive training, and not innate (Sloboda, 1996).

Unfortunately, almost no studies could be identified that specifically or indirectly evaluated musicians’ emotional expressivity outside of performance settings, for any channel of communication. Livingstone, Thompson, and Russo (2009) did examine musician’s facial expressions in a laboratory setting via facial electromyography (EMG). Musicians viewed video clips of phrases sung in happy, sad, and neutral tones, and were then asked to imitate the stimuli. Recordings were obtained during viewing, prior to imitation, during imitation, and following imitation. Facial muscle movement analogous to the emotions expressed was observed at all stages. Notably, these results indicate that musicians had spontaneous facial expression when viewing and listening to the sung phrases, prior to initiation of imitation. However, while this study supports the involvement of facial expression in musical activities, no untrained participants were included, and so no conclusions can be drawn about the performance of musicians on the task relative to non-musicians. Regarding the lexical channel, as previously discussed, musical training may enhance vocabulary, but it is not known if there is any effect
specific to emotional lexical content. Similarly, the performance-monitoring and auditory training aspects of musical training and the associated enhanced perception of prosody suggest that musicians may have greater skill in effectively communicating emotional prosody, but this hypothesis has not been tested.

Within the music therapy literature, there is some evidence that music-based interventions facilitate aspects of emotional communication and expression (i.e., Clements-Cortés, 2004). However, the conclusions that can be drawn from this literature are limited for several reasons. First, in music therapy, music performance is frequently combined with listening to and composing music, and the contributions of each on overall outcomes cannot always be evaluated independently. Second, reports are also often in case study format or evaluate a very small number of participants, limiting the ability to draw conclusions about efficacy and generalizability. Finally, the musical training received in music therapy is significantly shorter in duration and less intense than that of professional or even highly engaged amateur musicians. That any effects have been observed after only brief training, however, would suggest that it is actually more likely that intensive musical training, particularly during critical periods in childhood, would lead to notable behavioral differences. The relationship between emotional expression and musical training, therefore, represents a promising but as yet unexplored area of emotion research.

Emotional experience. The subjective experience of emotion is intrinsically connected to musical experience. Music listening is done for pleasure, often to evoke specific emotions or as a means of emotion regulation (Juslin & Västfjäll, 2008). Although music training and performance include a focus on felt emotions (Karlsson & Juslin, 2008), music research is primarily concerned with questions about the experience of the listener, not the experience of the
performer. There is some indication that musicians may process and experience emotions in music differently from non-musicians, including slightly stronger responses to emotion in music (Bigand et al., 2005), but these findings are rather weak and may be better explained by the musician’s superior ability to analyze complex Western music, rather than by differences in emotional processing.

Effects of music training on more general mood state have received somewhat more inquiry. Again, as with expression, these results are found primarily in the clinical and music therapy literature. Active group musical participation appears to change overall mood state in preadolescents, reducing negative and enhancing positive emotions (Montello & Coons, 1998). Music therapy has also been successfully utilized to improve mood in individuals with affective disorders (Erkkilä et al., 2011; Koelsch, Offermanns, & Franzke, 2010) and in individuals with Alzheimer’s disease (Clément, Tonini, Khatir, Schiaratura, & Samson, 2012). In another study, a caregiver singing along with dementia patients, as compared to a no-music condition, helped the patients experience and express more positive emotion (Gövell, Brown, & Ekman, 2009). Music training has also been suggested as an intervention for alexithymia due to its ability to evoke emotional experience, which could then be associated with the cognitive correlates of emotional experience to enhance emotion understanding (Allen & Heaton, 2010).

In summary, evidence is limited as to whether musical training modifies subjective emotional experience, such as the nature or quality of the emotional experience evoked by exposure to emotional stimuli. Musical training does appear to be a promising intervention for improving mood across a wide variety of populations, but would benefit from more studies that control for different types of musical interventions so as to distinguish among benefits from socialization, music exposure, and music performance.
**Emotional intelligence.** It has been suggested that emotional intelligence may mediate relationships between musical training and emotional abilities such as emotion perception (Thompson et al., 2004). However, emotional intelligence can refer to any number of theoretical constructs, all of which have been operationalized in different ways. Two of the best-characterized approaches include “trait emotional intelligence” (trait EI), which is based on an individual’s self-perception and general dispositions (Petrides, Furnham, & Frederickson 2004), and “ability emotional intelligence” (ability EI), which is based on a cluster of factors related to how the individual uses emotions and engages in emotional functions in daily life (Petrides, Niven, & Mouskounti, 2006; Trimmer & Cuddy, 2008). Both of these approaches have been assessed in relation to musical training.

Petrides et al. (2006) examined musical training and trait EI in relationship to length of musical training and found a significant positive relationship, such that higher levels of trait EI corresponded with longer periods of musical training. However, although this supports a relationship between musical training and trait EI, directionality cannot be assessed. The authors note that the factor of “self-control” is positively associated with length of musical training and that this trait would be one likely to aid musicians to remain in training programs. However, other factors of trait EI, such as well-being and emotionality, were also positively associated with years of training, and these relationships merit further investigation.

Ability EI is often assessed using the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). The MSCEIT assesses ability EI as a function of two broad areas: 1) experiential, and 2) strategic. These two areas are in turn, divided into two more specific ability sets, such that the experiential set is comprised of a) perceiving emotions and b) using emotions to facilitate cognition, and the strategic set is comprised of a) understanding emotions and b)
managing emotions (Mayer, Caruso, & Salovey, 1999; Resnicow, Salovey, & Repp, 2004). Resnicow et al. (2004) examined ability EI in a small sample of 24 undergraduates whose musical training ranged from 0 to 15 years. They found virtually no correlation between either overall MSCEIT scores or the experiential subscore and years of musical training. Trimmer and Cuddy (2008) obtained similar results in a larger group of 100 participants whose training ranged from 0 to 17 years (mean years of training = 6.5). Schellenberg (2011b) compared the ability EI in musically trained (musical training mean of 10.2 years, SD = 1.9) and untrained participant groups. He found no relationship between either overall MSCEIT score or the perceiving emotions score and musical training. He did find a positive relationship between MSCEIT understanding emotions score and musical training, but this disappeared with controlling for IQ.

In order to assess the possibility that musical training may accelerate the development of emotion skills in childhood, Schellenberg and Mankarious (2012) assessed the ability to understand emotions in groups of musically trained and untrained 7-8 year olds, using the Test of Emotion Comprehension (TEC). Although they did find an association between enhanced understanding and musical training, it appears that this was mediated by differences in IQ. Given this mediating relationship, this finding indicates that the TEC is more of a cognitive task than an emotional one.

In sum, there is some support for an association between trait EI and musical training. However, the directionality is unclear, and it may be that individuals high in trait EI are more likely to seek out or persist with musical training. In contrast, there is no evidence that ability EI (as measured by the MSCEIT) or the related ability of emotion understanding (as measured by the TEC) is associated with musical training. These findings may be limited by the cognitive,
rather than emotional, demands of these measures and do not rule out associations between musical training and other conceptualizations of ability EI.

**Social functioning.** Musical performance in an ensemble requires frequent communication between musicians, including attention to non-verbal signals and response monitoring. Davidson (2012) noted that musicians’ body movements reflect the constant communication that takes place between members of an ensemble. Musical training also appears to be associated with improved emotional communication, in particular, the improved ability to perceive emotion communicated through speech (Dmitrieva et al., 2006; Lima & Castro, 2011a; Nilsonne & Sundberg, 1985; Thompson et al, 2004) and may be associated with other aspects of emotional functioning, as previously discussed. Lastly, musical training is associated with a broad range of cognitive benefits, including executive functioning and higher overall IQ, which may enhance social competence (Schellenberg & Weiss, 2013). It is not surprising that there has been attention given to the idea that musical training may improve social skills/socialization and might even serve as an intervention for populations with social skills deficits.

There is some evidence supporting transfer from musical training to social development. Gerry, Unrau, and Trainor (2012) found that 6 months of active musical training (i.e., active infant and parent participation in moving, singing, and playing percussion instruments), made infants more communicative and enhanced their social development, relative to controls with passive musical training (i.e., music listening while engaged in free play) or no musical training. Another study showed enhanced empathy in a group of 8-11 year olds relative to controls following one school year of a group-based music-training curriculum (Rabinowitch, Cross, & Burnard, 2013). Kirschner and Tomasello (2009) showed that children engaged in drumming had more synchronous performance in a social condition than when drumming to audio recordings or
a drum machine, indicating that musical activity may be fundamentally linked to socialization, and may actually facilitate social cooperativity.

In contrast, Schellenberg (2006) found no association between parents’ estimates of social skills and musical training in a correlational study with 6-11 year olds. Also in an experimental study, Schellenberg (2004) found no improvement in social skills after one year of musical training. Although self-esteem has been linked to improved social abilities it was not found to improve in a study examining musical training over 3 years in a group of fourth-graders (Costa-Giomi, 2004).

The variability in these studies may be accounted for, at least in part, by the method of musical instruction. Overy and Molnar-Szacks (2009) have proposed the Shared Affective Motion Experience (SAME) model. This model posits that music is perceived as an auditory signal but also as “intentional, hierarchically organized sequences of expressive motor acts behind the signal” made possible by the motor neuron system. Within this model, when listening to or engaged in musical activities, it is as though we are also experiencing the presence of another individual. On this basis, the authors argue that musical engagement has a wide range of therapeutic applications. However, the most effective type of musical engagement would be in a group setting and that the music-making activity should be as “naturally occurring” as possible and that the more abstracted or culturally specific the musical experience, the less generalized the effect will be. Within this framework, it is not surprising that the studies with positive results predominantly involved training in basic music skills taking place in a group setting.

Music therapy is a common intervention for disorders with social skills deficits, such as autism. Improvements in positive emotional experience, communication, emotional responsiveness, emotional synchronicity, and social skills have all been noted in multiple
reviews and meta-analyses following music interventions in individuals with autism (Gold, 2011; Kaplan & Steele, 2005; Simpson & Keen, 2011; Reschke-Hernández, 2011; Whipple, 2004; Wigram & Gold, 2006). However, as previously discussed, conclusions about the role of musical training specifically in these gains are limited.

**Summary of musical training and emotion processes.** In sum, our understanding of relationships between musical training and various aspects of emotional functioning is still quite limited. For emotional perception, the prosodic and musical channels are the best characterized. There is evidence that musicians are better at the identification of discrete emotions in both music and the voice. However, this may be true more for certain discrete emotions (e.g., sadness) or subtle emotions, and a causal relationship cannot be clearly determined. Musical training conveys social skills benefits, particularly if the type of training is highly social and does not involve highly abstracted forms of musical training, such as music learning from formal notation. Music therapy is associated with a host of emotional benefits, including more positive emotional experience and improved overall mood state, improved emotional communication and expressivity, and, in autistic children, improved socialization. It is not known if musical training is associated with enhanced perception in either the facial, postural, or lexical channels, and there is virtually no research examining emotional expressivity musicians outside of performance settings or in comparison to non-musicians in any channel of communication.

**Musical training and the aging brain.**

The focus on the transfer effects of musical training has been primarily driven by interest in the benefits of musical training in childhood, either as a way to enhance development in normally developing children or to serve as an intervention in a variety of populations, such as low SES and autism. However, the needs of an aging population have shifted interest from
childhood to the interaction between late-life development and musical training. Musical training has been proposed as an “antidote for aging” (Kraus & Anderson, 2013a), a means of cognitive reserve that may preserve or even enhance auditory and cognitive functions throughout the lifespan (Kraus & Anderson, 2013b; Kraus & Chandrasekaran, 2010; Patel, 2013). Indeed, older adults who receive musical training in early life show auditory and cognitive advantages over untrained older adults (Alain, Zendel, Hutka, & Bidelman, 2014; Goodling, Abner, Jicha, Krysico, & Schmitt, 2013; Hanna-Pladdy & MacKay, 2011).

These benefits may also not be restricted only to older adults who had training during childhood critical periods, but may also be conferred to adults who receive training later in life. Functional changes in hippocampal activation when detecting novel acoustic stimuli are apparent in adults after brief musical training (Herdener et al., 2010; Lappe, Herholz, Trainor, & Pantev, 2008). Brief musical training has been shown to enhance motor (MacLean, Brown, & Astell, 2014) and cognitive functioning, such as working memory and processing speed (Bugos et al., 2007), in older adults. Music-making is also a promising method for motor and cognitive rehabilitation in older adults (see Altenmüller & Schlaug, 2013, for review). Furthermore, as discussed earlier, aging is not only accompanied by declines in cognitive functioning, but also in aspects of emotional functioning, particularly the perception of emotion (Finley, Borod, Brickman, et al., 2008; Lima & Castro, 20011b; Savage et al., 2013), and many neurodegenerative diseases and diseases of aging are also associated with deficits in the perception and expression of emotion (see Emotion perception deficits and psychopathology on page 17). Music training may represent a promising intervention for addressing both age-related changes in emotion processing and emotion perception deficits in psychiatric and neurological
conditions, but first the relationships between musical training and emotion perception must be better characterized.
Aims and Hypotheses

On the basis of the literature reviewed above, the following aims and hypotheses of the current study are presented. For details regarding statistical procedures, please see the “Data Analysis” section to follow.

**Aim I:** To examine the relationship between musical training and emotion perception within the prosodic channel of communication.

The primary aim of this project was to replicate and expand on previous research examining prosodic emotion perception in individuals with and without musical training in light of promising, but still somewhat equivocal, prior findings. This study expands on prior research by evaluating this relationship in a diverse population of musically-trained and untrained adult participants while carefully controlling for pre-existing group or individual differences that may also contribute to emotion perception (e.g., perceptual factors).

**Hypothesis I:** It was predicted that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) within the prosodic channel of communication. This hypothesis is based primarily on two prior studies that have shown a positive association between musical training and prosodic emotion perception accuracy (Lima & Castro, 2011; Thompson et al., 2004), as well as limited behavioral and functional evidence from a number of other studies (e.g., Dmitrieva, 2006; Young et al., 2012). It has been suggested that both speech and music engage similar mechanisms on multiple levels. For one, auditory overtraining is a feature of intensive musical training and results in a superior ability to analyze acoustic material (see the *Musical Training and Auditory Processing* section on page 28). It would therefore follow that musically-trained individuals should be particularly
skilled in discerning emotional content from speech through acoustic analysis alone. Furthermore, it has been shown that both speech and music utilize similar acoustic mechanisms for the communication of emotion (i.e., the acoustic profiles associated with the communication of discrete emotions are identical in both speech and music), and individuals with music processing difficulties, both congenital and acquired, tend to also be less accurate at processing prosody in speech (Juslin & Laukka, 2003). Lastly, musical training and performance activate a range of neural regions and processes that overlap with those recruited in prosodic emotion processing, further supporting the possibility of transfer effects from musical to emotion processing networks.

**Aim II: To examine the relationship between musical training and emotion perception in non-auditory channels of emotion communication, specifically the facial and lexical channels.**

Lima and Castro (2011a) suggest that transfer effects from musical training to emotion perception may not be limited to auditory domains, based on the possibility of a general emotion processor that subserves emotion perception in all channels (Borod, 1993; Borod, Pick et al., 2000). At present, however, little to no research exists examining relationships between musical training and the perception of emotion in either the facial or lexical channel of emotion communication.

**Hypothesis I: It was predicted that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) in the lexical channel of communication.** There is a growing body of behavioral (e.g., Koelsch et al., 2004; Sollberger, Reber, & Eckstein, 2003), as well as functional and neuroanatomical (e.g., Hsieh et
MUSICAL TRAINING AND EMOTION PERCEPTION

al., 2012), evidence suggesting that affective and non-affective lexical and musical processes overlap significantly, perhaps the result of shared evolutionary origins. Additionally, musical performance and emotional processes involve reciprocal perceptual, somatovisceral, and motoric re-experiencing, and per *emotion embodiment theory* (Niedenthal, 2007), the bodily expression and experience of emotion can alter the way in which emotional information is processed and interpreted. Therefore, musicians’ repeated exposure to and engagement in emotional processes (via musical performance) create the possibility of transfer effects to music perception in all modalities, including non-auditory channels.

**Hypothesis II:** It was predicted that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) in the facial channel of communication. At present, only two studies have examined facial emotion perception in relation to musical training (Kamiyama, Abla, Iwanaga, & Okanoya, 2013; Yong & McBride-Chang, 2007), and they have yielded conflicting results. However, there is neuropsychological and neuroanatomical evidence that facial and musical emotion perception processes are subserved by highly overlapping neural systems (see the *Neural Basis of Emotion Perception* section on page 19), findings that are in line with the notion of a general emotion processor proposed by Borod and colleagues (1993, 2000). Additionally, as discussed in the prior hypothesis, theories of embodied emotion suggest that the repeated experience and expression of emotion in musical performance enhances other related processes, such as the perception of emotion. On this basis, we theorized that musicians would be more accurate in emotion perception ability for the lexical and facial channels of communication.
Aim III: To examine the relationship between musical training and emotion perception in music.

The perception of discrete emotions in music has previously been explored as a function of musical training in a handful of studies, with somewhat equivocal results. Furthermore, there is considerable debate as to the validity of using a discrete emotion model to characterize musical emotion, versus, for example, a dimensional model. This aim sought to better characterize the relationship between musical training and discrete musical emotion perception in adults with and without musical training. The inclusion of music also provides a second auditory modality of emotion expression, which is informative as to whether auditory overtraining (i.e., a high level of engagement in specialized tasks that involve listening to and analyzing sound, in this case, music) or relationships between speech and music may be related to enhanced prosodic emotion perception in musicians.

Hypothesis I: It was predicted that both musicians and non-musicians would be able to identify basic emotions in music at above chance levels but that individuals with musical training would be significantly more accurate than the individuals without musical training at identifying basic emotions (i.e., happiness, tenderness, sadness, fear, and anger) in music.

The limited literature examining the relationship between musical emotion perception and musical training has yielded equivocal results. However, studies with a long duration and early onset of musical training have found a positive relationship (e.g., Lima & Castro, 2011b), and the musically trained population examined in the present study is comprised of adults with a long duration of musical training and a relatively early age of onset.

Furthermore, emotion is conveyed in music in ways that are highly analogous to speech, and prior evidence suggests that musical training is related to the accurate perception of
emotional prosody (see Aim 1, above). Onset and duration of musical training also affect processing of pitch and timing in music, respectively (Strait et al., 2009), both of which are key in perceiving emotion conveyed in music. Lastly, there is evidence that musical training causes not only neuroanatomical but also functional changes, specifically, increased neural activation when processing some discrete emotions (Park et al., 2014) and that the neuroanatomical regions activated during music listening and performance overlap significantly with those involved in the processing of musical emotion, creating the possibility of transfer effects between these two systems.

**Aim IV: To examine and assess if there are any differences in how musicians and non-musicians perceive individual basic emotions.**

Functional imaging studies have indicated that musical training may alter the way that basic emotions are processed. For example, musicians and non-musicians show different patterns of neural activation in response to fearful musical stimuli (Park et al., 2014; see the Neurological differences between musicians and non-musicians section on page 24). This raises the possibility that group differences in emotion perception accuracy may be the result of musicians processing basic emotions in a fundamentally different manner than non-musicians. Alternately, musicians may simply utilize the same processes as non-musicians to identify emotions (e.g., utilization of specific auditory cues in prosodic emotions), but are able to do so more proficiently as a result of their training. Assessing for different response patterns in the identification of individual emotions between groups can shed light on this area.

Furthermore, within the music emotion literature specifically, there is ongoing debate as to whether basic emotions are a valid means of conceptualizing musical emotion. It is also not clearly known what, if any, relationships there may be between musical training and the
perception of basic emotions in music specifically (see *Representation and perception of emotion in music* on page 9).

As behavioral and functional research in these areas is quite limited, this aim was exploratory in nature. There are no specific hypotheses as to which, if any, individual emotions would be more accurately identified by musicians or if patterns of emotion identification might differ between musicians and non-musicians.

**Aim V: To assess for cognitive differences between musicians and non-musicians.**

There is general agreement that musicians score higher on aggregate measures of general intelligence than non-musicians (see *Musical Training and cognitive processes* section on page 29.) However, regarding fluid intelligence and more specific cognitive abilities, there is a large, yet contradictory, body of literature investigating cognitive differences between musicians and non-musicians. This aim, therefore, seeks to compare the performance of musicians and non-musicians on measures of general intelligence and cognitive ability. As prior research regarding cognitive and intellectual differences between musicians and non-musicians is equivocal, these aims are exploratory in nature and no specific hypotheses have been generated.
Methods

Participants

Participants are 119 individuals between the ages of 18-40. Fifty-eight are musicians and 61 are healthy controls. All efforts were made to balance participants on key demographic variables, including age, gender, ethnicity, and education. Number of participants recruited was determined through a power analysis using G*Power 3.1 software (Faul, Erdfelder, Lang, & Buchner, 2007), using provided or calculated effect sizes from prior studies (see Table 1 of the Appendix).

Recruitment took place through the Queens College Subject Pool via a listing in the SONA system. Participants were also recruited from the Queens College campus and from the surrounding New York City community via fliers, through email lists maintained by the Aaron Copland School of Music at Queens College, and through internet postings in order to ensure outreach to individuals meeting criteria for inclusion in the musical training group. Participants were able to contact experimenters by either phone or email to express interest in the study. Participants were then asked to confirm that they meet the minimum study requirements listed on the recruitment materials prior coming into the laboratory to complete further screening measures.

Inclusion and exclusion criteria. All participants are between the ages of 18 and 40. A cutoff of 40 was selected for two reasons. First, prior research has shown that the ability to perceive emotion varies as a function of age, such that there is a decline in the ability to perceive affective prosody over the lifespan (Lima & Castro, 2011b). Second, to the best of our knowledge, only one study has examined the effect of musical training on prosodic emotion perception in individuals over the age of 40 (Lima & Castro, 2011a), and no studies have
examined the effect of musical training on communication channels other than prosody for any age range. Therefore, the age range was restricted to individuals under age 40 so as to first examine individuals with the highest level of emotion perception accuracy.

**Participant inclusion criteria.** In order to ensure sufficient English language ability to complete the study measures, all participants were either native English-speakers or learned English by age 7, a cutoff previously used in similar research in our laboratory. Participants who learned English after age 7 were excluded from the study. Participants were also excluded from the study if they have a history of neurological or psychiatric disorder, learning disability, substance abuse, or treatment with psychoactive medication, as these have been shown to potentially interfere with the ability to perceive emotion.

**Musician group criteria.** Based on prior research (Lima & Castro, 2011a; Thompson, Schellenberg, & Husain, 2004; Trimmer & Cuddy, 2008), inclusion criteria were selected in order to ensure a robust effect of musical training. Effects of musical training have not been obtained in studies where musically-trained participants had an average of 6.5 years of training (Trimmer & Cuddy, 2008), whereas effects have been found in a group with an average of 8 years of musical training (Thompson et al., 2004) and a group with an average of 12 years of musical training (Lima & Castro, 2011a). Based on this research, a minimum requirement of 8 years of musical training was selected for inclusion into this group.

Musical training was defined as formal lessons in playing a musical instrument, occurring on average once a week or more, and a distinction was made between training in a musical instrument and voice training. Lima and Castro (2011a) did not include vocalists in their study unless they also had instrumental training. Thompson, Schellenberg, and Husain (2004), in their experimental study examining emotion identification ability in children assigned to instrumental,
voice, or acting groups, found effects for the instrumental group but not for the voice-only group. Accordingly, vocalists were excluded from the study unless they are also meet inclusion criteria for one or more musical instruments.

Studies with significant effects have restricted participants to those with training onset in childhood, as this is a critical developmental period where musical training is likely to exert the most influence on non-musical domains. However, only one study provided an average age of onset of age 9 years old (SD=2.5; Lima & Castro, 2011a). Accordingly, inclusion criterion for training age of onset was set at a maximum of age 12. This age cutoff was selected to insure that participants have begun training during childhood, while not excessively restricting inclusion criteria in a way that might interfere with participant recruitment. Further, as musicians were recruited primarily from a university school of music, most musicians began training in early childhood.

Lastly, in order to ensure skill maintenance, all musician participants were engaged in at least 2 hours of weekly practice at the time of study participation. A minimum requirement of 2 hours per week reflects a balance between recruiting eligible participants and ensuring skill maintenance.

**Control group criteria.** In addition to meeting the overall group inclusion criteria detailed above, participants in the control group were screened to insure 3 or fewer years of musical training. Further, they were not currently involved in musical practice or performance, including instrumental or voice performance in informal settings (e.g., church choir). See Table 1 for musician versus control participant inclusion criteria.

**Procedures**

Prior to testing, participants provided informed written consent. The total administration
time of all questionnaires, screening tasks, and experimental and control tasks was approximately 3 hours, and all participants completed testing in a single session. Consent was continually assessed throughout the testing session, with breaks given as needed. Approval was obtained prior to the study start from the CUNY-wide Institutional Review Board (IRB). Queens College undergraduate students who participated through the SONA system received course credit in return for participation. Other participants were compensated $10 per hour/$5 per half-hour of participation in the study.

**Screening Measures.** All participants completed a brief screening questionnaire collecting participant demographic information, history of psychiatric or neurological disorder, history of substance abuse, history of learning disability, and level of musical training and musical exposure.

**Experimental Tasks.**

*Emotion identification tasks.* Participants completed a total of four emotion identification tasks. Three of these tasks were selected from the New York Emotion Battery (NYEB; Borod, Cicero, et al, 1998) in order to assess emotion identification ability in the facial, prosodic, and lexical channels of communication. Each of these tasks includes three positive (pleasant surprise, happiness, and interest) and five negative (fear, sadness, anger, disgust, and unpleasant surprise) discrete emotions. A fourth task was created to assess emotion identification ability in the musical channel of communication. This task includes two positive (happiness and tenderness) and three negative (anger, fear, and sadness) emotions. Stimuli for all four tasks were presented in a pseudorandomized order so as to prevent clusters of a particular emotional valence, and there were two forms of each task so as to minimize order effects within each task. All emotion stimuli were digitized and presented on a computer screen and
headphones via custom software written using the open source statistical software platform R (R Core Team, 2015) and Shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2015).

*Prosodic perception.* The NYEB prosodic identification task (PID) consists of two practice trials followed by 24 experimental trials. The stimuli consist of neutral-content sentences (e.g., “She put it on the table.”), spoken by actresses and actors (referred to as posers) in eight emotional tones. Each emotion appears three times, and poser gender is balanced. Each trial is presented once, with the option for a second presentation if needed, and participants then identified the emotion portrayed by selecting one of the eight possible emotion choices.

*Facial perception.* The NYEB facial identification task (FID) consists of two practice trials followed by 32 experimental trials. The stimuli consist of slides from Ekman and Friesen’s facial expression slides, which demonstrate happiness, fear, sadness, anger, and disgust, as well as additional slides created by Borod and colleagues to demonstrate interest, pleasant surprise, and unpleasant surprise. Each emotion is presented four times and poser gender is balanced. Slides are presented on a computer screen for 20 seconds, and participants then identified the emotion portrayed by one of the eight possible emotion choices.

*Lexical perception.* The NYEB lexical identification task (LID) consists of two practice trials followed by 24 experimental trials. The stimuli consist of emotional content sentences that represent happiness, fear, sadness, anger, disgust, pleasant surprise, and unpleasant surprise. Each emotion is represented three times in each task. The sentences are presented in black text in the center of a white computer screen for 20 seconds, and participants then identified the emotion portrayed in the sentence by selecting one of eight possible emotion choices.

*Music perception.* The musical stimuli for the music identification task (MID) are 15-second musical segments excerpted from film soundtracks, developed and validated by Eerola.
and Vuoskoski (2011). The music perception task consists of two practice trials and 50 experimental trials. Each of the 5 emotions is represented an equal number of times. Further, half of these segments have been judged to contain high levels of the target emotion, and half moderate levels, also equally distributed across the 5 emotions. The stimuli have been pseudorandomized in a manner analogous to that used with the NYEB stimuli in order to avoid clusters of particular emotional valence. Each trial was presented once, with the option for a second presentation if needed. Participants identified the emotion portrayed in the segment by selecting from one of five possible emotion choices (e.g., happy, tender, sad, anger, and fear).

**Non-emotional control tasks.** Three non-emotional control measures, analogous to the lexical and prosodic emotion perception tasks, and adapted from the NYEB, were administered. These tasks were selected from a larger set of NYEB control tasks in order to minimize the length of the study (and thus possible participant fatigue) while enabling control of key perceptual differences that may be present between the two groups.

**Prosodic channel.** Non-emotional control measures were administered for the prosodic channel for the following two reasons. First, it is possible that musicians are more accurate relative to controls in the perception of non-emotional prosody, given their specialized auditory training. In addition, although all participants learned English by the age of 7, it is still possible that non-native English speakers may be less accurate in the perception of English prosody. The non-emotional prosody measures allow for any possible group or individual differences to be controlled in subsequent statistical analyses. Two prosody control measures were used, the Benton Phoneme Discrimination task (BPID) and the Intonation Contours Perception task (ICID).

The BPID (Benton et al., 1983) is an auditory task where participants are presented with
30 tape-recorded pairs of nonsense words and are asked to judge whether the phonemes were the same or different from each other. ICID (Borod et al., 1992) is an auditory task where participants listen to 24 sentences with no meaning, one at a time, and identify if the sentence is voiced as a question, statement, or command.

*Lexical channel.* A control measure equivalent to the experimental lexical task was administered in order to control for possible participant differences in English language ability and to further ensure sufficient ability to complete the experimental emotion task. The Non-emotional Sentence Identification task (NSID; Borod, et al. 1992) consists of participants reading 24 sentences one at a time and then selecting which of 8 nonemotional human characteristics (i.e., body type, complexion, hair type, intelligence, personality, teeth, vision, and voice type) best fits the sentence.

*Cognitive measures.* Broad intellectual and cognitive ability were measured to assess for differences between the two groups, on the basis of prior literature suggesting that musicians outperform non-musicians on some measures of cognitive and intellectual ability (see *Literature review* section on page 4.) The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) is a brief cognitive screening measure and was administered to ensure adequate overall participant cognitive ability and, secondarily, to allow for comparison of groups on broad cognitive ability. It yields maximum score of 30 and a cutoff score of ≥ 26 is considered normal. A modified form of the Raven’s Advanced Progressive Matrices (RAPM; Raven, 1981; Bors & Stokes, 1998) was administered to estimate non-verbal intelligence, and consists of a series of patterns with a missing element that the participant must identify. The RAPM has 60 items, however, as a time-abbreviated version was administered, the anticipated maximum score obtainable is around 30. The Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) was
used to estimate verbal intelligence, and consists of a list of 50 irregular English words, which the participant must pronounce, in order of increasing difficulty. This yields a maximum raw score of 50, which is then converted to a standard score.

**Additional measures for group comparisons.** To evaluate and control for alexithymia, which has been suggested to interfere with an individual’s ability to perceive emotion, the Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994) was administered. The TAS-20 is a commonly used measure of alexithymia and assesses an individual’s ability to discern his/her feelings and express them. The TAS-20 yields an overall score and 3 subscales: TAS-1) Difficulty Describing Feelings subscale, TAS-2) Difficulty Identifying Feeling subscale, and TAS-3) Externally-Oriented Thinking subscale, used to measure the tendency of individuals to focus their attention externally. A typical statement on the TAS-20 is, “I am often confused about what emotion I am feeling.” The participant indicates to what extent he/she agrees with the statement on a 5-point scale, where 1 is "strongly disagree" and 5 is "strongly agree". An overall score of 52 or higher indicates possible alexithymia (Bagby et al., 1994).

As prior literature has suggested that there may be personality differences between musicians and non-musicians, all participants completed the Ten Item Personality Inventory (TIPI; Gosling et al., 2003), which evaluates personality according to the well-established Big Five personality model, consisting of 5 personality factors: openness, conscientiousness, extraversion, agreeableness, and neuroticism. This measure consists of 10 statements, a typical example of which is "I see myself as extraverted, enthusiastic." The participant must indicate to what extent he/she agrees with the statement on a 7-point scale where a score of 1 is "disagree strongly" and a score of 7 is "agree strongly." This measure yields 5 scores reflecting the extent to which a participant does or does not endorse the above personality traits, consisting of
extraversion (TIPI-E), agreeableness (TIPI-A), conscientiousness (TIPI-C), emotional stability (TIPI-ES), and openness (TIPI-O).

As the musical stimuli are excerpted from films and participant preference for or familiarity with particular musical genres could influence their evaluation of the stimuli, the Short Test of Music Preferences – Revised (STOMP-R; Rentfrow & Gosling, 2003) was administered to assess participant musical preferences. The STOMP-R consists of 23 different musical genres, which participants are asked to rate according to their preference on a 7-point scale where a score of 1 is "dislike strongly" and a score of 7 is "like strongly." The STOMP-R then yields overall scores reflecting participant preferences for music in four music preference dimensions. Reflective & Complex (STOMP-R&C) comprises bluegrass, blues, classical, folk, international/foreign, jazz, new age, and opera. Intense & Rebellious (STOMP-I&R) comprises alternative, heavy metal, punk, and rock. Upbeat & Conventional (STOMP-U&C) comprises country, gospel, oldies, pop, religious, and soundtracks/theme songs. Energetic & Rhythmic (STOMP-E&R) comprises dance/electronica, funk, rap/hip-hop, reggae, and soul/R&B. Groups were also be compared specifically for their preference for soundtracks/theme songs, as the music stimuli used on the music emotion identification task are taken from film soundtracks.

**Task administration.** The order of study tasks was carefully determined so as to reduce possible order effects as well as to minimize the impact of factors, such as fatigue or boredom, on the key experimental measures. Tasks were administered in a specific sequence, as follows:

1) The screening questionnaire was given first so as to ensure that participants met the study inclusion criteria.

2) The MoCA was administered, followed by the remaining two brief neuropsychological tests, the WTAR and Ravens, in randomized order.
3) The experimental tasks were then administered as follows:
   a. The three NYEB emotion identification tasks (LID, PID, and FID) were
      administered in randomized order across the emotion tasks as well as within tasks
      (such that either Form 1 or Form 2 of any of the tasks may be administered). The
      music identification task (MID), which was developed separately and contains a
different subset of basic emotions, was administered 50% of the time prior to the
NYEB tasks and 50% of the time following the NYEB tasks in order to control
for order effects and to control for the remote possibility that the music task could
have a priming effect.
   b. The NSID, ICID, and BPID were administered following the emotion
      identification tasks in a randomized order.
4) The TIPI, STOMP-R, and TAS-20 were administered at the end of the testing session,
as these measures are brief self-report inventories that are less likely to be affected by
fatigue than the experimental tasks. They were administered in a randomized order.

Statistical Analyses

Data scoring. Each response for the experimental emotion and non-emotional control
measures is scored as correct (1) or incorrect (0) and raw scores were converted to percent
correct to generate an overall accuracy score. In order to enable examination of perception of
discrete emotions in music, percent accuracy scores for each discrete emotion were also obtained
for the music identification task.

Data inspection. Statistical analyses were performed using SPSS (Version 22.0) and the
open source statistical software platform R (R Core Team, 2015). Prior to testing specific
hypotheses, data entry was double-checked and the data were inspected for outliers. To screen
for outliers, boxplots and scatterplots were created for each of the experimental emotional tasks for both the musician and control groups and these charts were visually inspected for outliers. Any scores 2.5 standard deviations above or below the mean for the participant’s group was inspected to determine if they represented a true outlier or normal variation within the sample.

**Data normalization and variance.** The Shapiro-Wilk test of normality (Maxwell & Delaney, 1990) was performed for all emotional and non-emotional perception tasks. Homogeneity of variance was also assessed, using Levene’s Test of Homogeneity of Variance (Levene, 1960). Efforts were made to transform any data with skewed distributions or unequal variance and then the Shapiro-Wilk and Levene’s tests were re-run for each variable.

**Participant demographics.** Groups were compared on key demographic variables (i.e., age, sex, handedness, and ethnicity) in order to evaluate group differences. For continuous variables (i.e., age and education) differences were examined using t-test. Chi-square tests were used to assess group differences for dichotomous variables, such as sex (1=male and 2=female), race/ethnicity (1=White and 2=non-White).

**Group differences.** In order assess for group differences on the non-emotional control measures (i.e., NSID, BPID, and ICID) and on other variables that may affect performance on the emotion identification tasks (e.g., alexithymia, personality, and musical preference), group scores were be compared using Mann Whitney-U tests or ANOVA, depending on data normality. Any significant group differences were included as covariates in the experimental data analyses.

**Data Analysis for Specific Aims and Hypotheses.**

*Aim I: To examine the relationship between musical training and basic emotion perception in the prosodic channel of communication.*
Hypothesis I proposes that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) in the prosodic channel of communication.

Aim II: To examine the relationship between musical training and emotion perception in non-auditory channels of emotion communication, specifically the facial and lexical channels.

Hypothesis I predicts that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) in the lexical channel of communication.

Hypothesis II predicts that individuals with musical training would be more accurate than those without musical training in identifying basic emotions (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, and anger) in the facial channel of communication.

The above hypotheses involve understanding group differences in performance on emotion tasks. As the test variables were all non-normally distributed, these hypotheses were initially tested with non-parametric procedures. Specifically, individual Mann-Whitney U Tests were conducted to initially compare group differences in overall emotion identification accuracy in the prosodic, lexical, and facial channels, respectively. However, there were significant differences between groups for relevant variables, and there is no standard way to control for these differences using non-parametric procedures. Therefore, for any channels with significant differences per Mann-Whitney tests, a one-way analysis of covariance was performed to assess for group differences when controlling for other relevant variables.
**Aim III:** To examine the relationship between musical training and basic emotion perception in music.

*Hypothesis I* predicts that both musicians and non-musicians would be able to identify basic emotions in music at above chance levels, but the individuals with musical training would be significantly more accurate than non-musicians at identifying basic emotions (i.e., happiness, tenderness, sadness, fear, and anger) in music. As test data were non-normally distributed, in order to address the above aim, a Mann-Whitney U Test was performed to initially compare differences between the two groups for accuracy at identification of musical emotion. However, as described above, it was necessary to covary for relevant group differences. Therefore, a one-way analysis of covariance was performed to compare group overall group accuracy on the music emotion identification task.

**Aim IV:** To assess for differences in how musicians and non-musicians perceive individual basic emotions. In order to examine for differences in how musicians and non-musicians perceive basic emotions in music, a 2x5 mixed-design repeated measures ANCOVA (Group [Musicians and Controls] X Emotion [Happiness, Tender, Anger, Fear, and Sadness]) was performed to evaluate group accuracy for individual musical emotions.

In order to examine for differences in how musicians and controls identified individual basic emotions, the pattern of inaccurate responses for each emotion was briefly analyzed for all emotion identification tasks (i.e., MID, PID, FID, and LID). Additionally, an error analysis was performed to evaluate for significant differences in the patterns of musicians’ errors versus non-musicians’ errors for all emotion identification tasks found to have significant group differences. Specifically, Wald tests were used to examine the error distributions for both groups for each of the emotion identification tasks (i.e., MID, PID, FID, and LID) in order to determine if
musicians and controls differ significantly in how they misclassify individual emotions (for example, when musicians and controls are both incorrect about fear, do they then tend to both select the same emotion(s), instead, or do they select distinctly different emotions).

**Aim V: To assess for differences in cognitive ability between musicians and non-musicians.** To address the above aim, t-tests or Mann-Whitney U tests were used to compare overall group performances on the MoCA, RAPM, and WTAR, as appropriate based on data normality.
Results

Data Inspection

For the experimental tasks, 7 scores fell 2.5 standard deviations below the task means (4 musicians and 3 controls) and 1 score fell 2.5 standard deviations above the mean (one musician). For the nonemotional control tasks, 7 scores fell 2.5 standard deviations below the mean (4 musicians and 3 controls) and no scores fell 2.5 standard deviations above the mean. There were no instances in which one individual participant had more than one outlying score. For the cognitive measures, 10 scores fell 2.5 standard deviations below the mean (8 musicians and 2 controls), and one score fell 2.5 standard deviations above the mean (1 control). Each of these scores was investigated and it was determined that all outliers represented normal variability in performance.¹

Tests of Normality and Homogeneity of Variance

Shapiro-Wilk tests of normality (Shapiro & Wilk, 1965) were performed for all emotion identification tasks (i.e., FID, LID, MID, and PID) for total scores, for individual emotion scores for the music identification task (MID-IE), for the nonemotional control tasks (i.e., PDID, ICID, and NSID), and the cognitive measures (i.e., MoCA, RAPM, and WTAR). When the total score was used, results indicated that distributions were non-normal for musicians on FID and MID and non-normal for controls on FID, LID, and PID. When these tests were conducted for MID-IE, results indicated that distributions were non-normal for musicians on Happiness, Sadness, Anger, and Fear and non-normal for controls on Fear, Happiness, Sad, and Tender. When these tests were conducted for the non-emotional control tasks, results indicated that distributions were non-normal for musicians on Happiness, Sadness, Anger, and Fear and non-normal for controls on Fear, Happiness, Sad, and Tender. When these tests were conducted for the non-emotional control tasks, results indicated that distributions were non-normal for musicians on Happiness, Sadness, Anger, and Fear and non-normal for controls on Fear, Happiness, Sad, and Tender.

¹ Scores were also examined to rule out errors in data entry for computerized tasks, as well as scoring errors for neuropsychological measures and questionnaires. Furthermore, outlying scores were not systematic (i.e., no one participant was responsible for outlying scores for multiple variables). Lastly, all key analyses were run separately with outliers omitted, and results did not change when outliers were removed.
non-normal for both musicians and controls on ICID and BPID. When these tests were conducted for the cognitive measures, results indicated that distributions were non-normal for both musicians and controls on the MoCA and non-normal for musicians on the WTAR. Results of these analyses are presented on Tables 2, 3, 4, and 5 of the Appendix.

Levene’s Test of Homogeneity of Variance (Levene, 1960) was performed for all emotion identification tasks (i.e., FID, LID, MID, and PID) for total scores, for individual emotion scores for the music identification task (MID-IE), for the nonemotional control tasks (i.e., PDID, ICID, and NSID), and the cognitive measures (i.e., RAPM, MoCA, and WTAR). Results indicate that variances were equal for all measures except for the WTAR. Results of these analyses are presented on Tables 2, 3, 4, and 5 of the Appendix.

Several attempts were made to normalize the data using logarithmic, square root, arcsine, and reverse score transformations, but these results were largely unsuccessful. Non-parametric procedures were therefore conducted as appropriate. However, some of the primary hypotheses relate to understanding interactions between musical training, channel, and emotion, and these interactions cannot be tested using non-parametric procedures. Furthermore, if significant group differences are present on control tasks, covarying is necessary and cannot be performed using standard non-parametric procedures. Therefore, Q-Q plots were used to enable visual inspection of the distributions (see figures 1-9 in the Appendix for normal Q-Q and normal detrended Q-Q plots for each task). Visual inspection of these plots did not indicate gross departures from normality. As the ANOVA procedure is fairly robust to moderate violations of normality (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996; Ramsey, 1994), when necessary, parametric measures were used to test for group differences.

Demographic Characteristics
The final sample consisted of 119 healthy adults, 58 musicians and 61 non-musicians (i.e., controls). The demographic characteristics are listed in Table 2. Analyses comparing groups on demographic variables indicated that there were no significant (p < .05) differences for age ($t_{(117)} = 1.2, p = 0.226$), years of education ($t_{(117)} = .977, p = .331$), gender (1=male, 2=female; $\chi^2_{(1)} = 0.693, p = .405$), race/ethnicity (1=White, 2=non-White; $\chi^2_{(4)} = 2.96, p = 0.085$), or handedness (1=right, 2=left; $\chi^2_{(1)} = 0.434, p = 0.510$).

**Group Comparisons**

Group performances on questionnaires assessing alexithymia (i.e., TAS-20), personality (i.e., TIPI), and musical preference (i.e., STOMP) were compared using 2-tailed independent samples t-tests. For the TAS-20, there were no significant group differences for the TAS-1 ($t_{(117)} = -.79, p = 0.467$), TAS-2 ($t_{(117)} = -1.28, p = 0.203$), or TAS Overall scores ($t_{(117)} = -1.84, p = 0.068$). There was a significant group difference for the TAS-3 ($t_{(117)} = -2.56, p = .012$), such that controls reported greater externally-oriented thinking ($M=18.44$) than did musicians ($M=16.53$).

For the TIPI, there were no significant group differences for TIPI-EX ($t_{(117)} = .36, p = .719$), TIPI-A ($t_{(117)} = .146, p = .148$), TIPI-C ($t_{(117)} = .81, p = .421$), or TIPI-ES ($t_{(117)} = -1.67, p = 0.503$). There was a significant group difference for TIPI-O ($t_{(117)} = 3.55, p = 0.001$), such that musicians rated themselves as significantly more open to new experiences ($M=6.08$) than did controls ($M=5.34$). For the STOMP, there were no significant group differences for the STOMP-Intense & Rebellious ($t_{(117)} = 1.80, p = 0.074$), STOMP Upbeat & Conventional ($t_{(117)} = -0.49, p = .622$), or STOMP Energetic & Rhythmic ($t_{(117)} = -1.33, p = 0.186$). As the musical stimuli are

---

2Although group differences for ethnicity were not significant, there were, nonetheless, relatively more White participants in the musician group than in the control group (See Table 2). In order to evaluate if potential group differences for emotion perception might be the result of group differences in ethnicity, White and non-White participant mean % accuracy scores for the emotion perception tasks (e.g., MID, PID, LID, and FID) were compared within each of the groups (i.e., Musicians and Controls) using one-way ANOVAs. No significant differences were found either between White and non-White musicians or between White and non-White controls for any of the 4 emotion tasks. See Table 6 of the Appendix.
taken from soundtracks, overall group ratings for soundtracks/theme song preference alone were also examined. No significant group differences were found for rated soundtracks/theme song preferences \( (t_{(117)} = 0.61, p = 0.544) \). There was a significant group difference for STOMP Reflective & Complex \( (t_{(117)} = 4.89, p = 0.00) \), such that musicians preferred this musical dimension more than controls. See Table 3 for group descriptives.

**Non-emotional Control Measures**

Group performances on the non-emotional control measures (i.e., NSID, BPID, and ICID) were evaluated using non-parametric methods (i.e., Mann-Whitney U tests) as the data were not normally distributed. No significant differences between musicians and controls were found for NSID \( (U=1,702.00, p = .722) \), BPID \( (U=2,037.00, p = .151) \), or ICID \( (U=1,584.50, p = .321) \). See Table 4 for descriptives.

**Analyses Specific to Aims**

**Aim I: To examine the relationship between musical training and basic emotion perception in the prosodic channel of communication.**

**Hypothesis I.** To test the hypothesis that musicians would be more accurate than non-musicians in identifying emotions in the prosodic channel, the Mann-Whitney U-test was used to initially assess for overall group differences on the PID task. A significant group difference was found \( (U=1,332.00, p = 0.019) \), such that musicians (Mean rank=67.53) were more accurate than controls (Mean rank=52.84) at identifying prosodic emotions. In order to determine if this significance remained after accounting for relevant group differences, a one-way, between-subjects ANCOVA was performed to compare the two groups on PID with TAS-3 and TIPI-O scores as the covariates. A significant difference was again found between groups \( (F_{(1,115)} = ...) \)

---

3 Similarly, no differences were found using parametric measures (T-tests).
6.906, \( p = 0.010 \), such that musicians (\( M=59.41 \)) were more accurate than controls (\( M=53.75 \)) at identifying prosodic emotions. See Table 5 and Figure 1.

**Aim II: To examine the relationship between musical training and emotion perception in non-auditory channels of emotion communication, specifically facial and lexical.**

*Hypothesis I & II.* To test the hypothesis that musicians would be more accurate than non-musicians in identifying emotions in the facial and lexical channels of communication, Mann-Whitney U tests were used to initially assess for group differences. No significant group differences were found for either FID (\( U=1,615.50, p=0.412 \)) or LID (\( U=1,617.50, p=0.417 \)), and accordingly, no parametric procedures were performed. See Table 5 and Figure 1.

**Aim III: To examine the relationship between musical training and basic emotion perception in music.**

*Hypothesis I.* To test the hypothesis that musicians would be more accurate than controls at identifying emotions in music, Mann-Whitney U test was performed to assess for group differences in MID performance. A significant group difference was found (\( U=1,189.00, p=0.002 \)), such that musicians (Mean rank=70.00) were more accurate than controls (Mean rank=50.49) at identifying musical emotions. In order to determine if this significance remained after controlling for relevant variables, a one-way between-subjects ANCOVA was performed to compare the two groups on MID while covarying for TAS-3 and TIPI-O scores. Again, a significant difference was found between groups (\( F_{(1,115)} = 6.721, p = 0.011 \)), such that musicians (\( M=68.55 \)) were more accurate than controls (\( M=64.42 \)) at identifying musical emotions. See Table 5 and Figure 1.
Aim IV: To assess differences in how musicians and non-musicians perceive individual basic emotions. In order to compare group accuracy for individual basic musical emotions, a 2x5 mixed-design ANCOVA (Group [Musicians and Controls] X Emotion [Happiness, Tender, Anger, Fear, and Sadness]) was performed with TAS-3 and TIPI-O scores as covariates. The Bonferroni correction was used to control for Type 1 error. There was a significant main effect of group ($F_{(1)} = 6.721, p = 0.011$). There was no significant interaction effect of Group by Emotion ($F_{(4)} = 0.568, p = 0.686$). See Table 6 and Figure 2.

Overall, both musicians and controls identified the correct emotion the majority of the time for all tasks. The pattern of inaccurate responses for each emotion across both groups was briefly analyzed. With respect to MID (See Table 7 for % responses and Figures 3 & 4 for visualization), Happiness was misidentified mainly as Tender, Tender was misidentified mainly as Happiness or Sadness, Sadness as Tender, and Anger and Fear were mainly confused with one another. See Table 13 for a summary.

With respect to PID (see Table 8 for % responses and Figures 5 & 6 for visualization), Happiness was misidentified mainly as Pleasant Surprise, Pleasant Surprise as Interest or Unpleasant Surprise, Interest mainly as Unpleasant Surprise, Anger as Disgust, and Fear as Sadness. Sadness was mistaken for Disgust and Unpleasant Surprise, whereas Unpleasant Surprise and Disgust were both misidentified fairly equally with several emotions (Unpleasant Surprise with Pleasant Surprise, Interest, or Disgust, and Disgust with Anger, Interest, or Unpleasant Surprise, respectively). See Table 14 for summary.

Although no significant overall group differences were obtained for either LID or FID, patterns of responding were also briefly analyzed (see Tables 9 & 10). For both LID and FID, Happiness was misidentified mainly with Interest and Pleasant Surprise, Pleasant Surprise with...
Happiness, Unpleasant Surprise and Fear with one another, as well as Fear with Interest for LID only. Interest was misidentified mainly as Pleasant Surprise for LID and as Unpleasant Surprise for FID, Disgust for Unpleasant Surprise and Fear for LID and Disgust for Sadness for FID, and Anger as Disgust for LID and Interest for FID. For LID, Sadness was misidentified mainly as Unpleasant Surprise, while for FID, Sadness was mistaken for all emotions with the exception of those with positive valence (e.g., Happiness and Pleasant Surprise). See Table 14 for summary.

Error analyses were conducted to further examine for differences in basic emotion perception between musicians and controls for the tasks with significant group differences (i.e., MID and PID). Multiple Wald tests were used to test if there were significant differences between musicians and controls in their patterns of inaccurate responses for individual emotions (i.e., to assess for significant differences in the distributions of inaccurate responses). For MID, significant differences were found between the error response distributions of musicians and controls for Happiness ($\chi^2(4) = 10.33, p = 0.04$), Tender ($\chi^2(4) = 11.31, p = 0.02$), and Fear ($\chi^2(4) = 9.33, p = 0.05$). No significant differences were found for Sadness ($\chi^2(4) = 5.08, p = 0.28$) or Anger ($\chi^2(4) = 5.00, p = 0.29$). See Table 11 for % inaccurate responses. For PID, significant differences were found between the error response distributions of musicians and controls for Happiness ($\chi^2(7) = 32.69, p < 0.01$), Interest ($\chi^2(7) = 27.30, p < 0.01$), Unpleasant Surprise ($\chi^2(7) = 24.27, p < 0.01$), Disgust ($\chi^2(7) = 27.97, p < 0.01$), Anger ($\chi^2(7) = 27.64, p < 0.01$), and Fear ($\chi^2(7) = 25.29, p < 0.01$). No significant differences were found for Pleasant Surprise ($\chi^2(7) = 3.24, p = 0.86$) or Sadness ($\chi^2(7) = 0.55, p = 1.00$). See Table 12 for % inaccurate responses). The distributions were then visually inspected. Overall, even for those that were significantly different, the majority of musician and control responses, nonetheless, followed similar patterns of errors. The primary difference was typically that controls’ errors were more variable than
were those of musicians. In other words, when musicians were incorrect they selected the most commonly misidentified emotion slightly more often than controls, whereas controls selected a wider range of erroneous emotions overall. See Figures 7 and 8 for MID error distribution visualization and Figures 9 and 10 for PID error distribution visualization.

**Aim V: To assess for differences in cognitive ability between musicians and non-musicians.** Group performances on cognitive measures were compared using one-way ANOVAs. A significant group difference was found for the WTAR ($F_{(1,117)} = 19.538, p < 0.001$), such that musicians’ scores ($M=112.35$) were higher than those of controls ($M=103.39$). No significant differences were found for RAPM scores ($F_{(1,117)} = 1.738, p = 0.190$) or MoCA scores ($F_{(1,117)} = 1.561, p = .214$). See Table 15 for descriptive statistics.
Discussion

Summary of Results

The primary goal of this project was to expand on prior research examining the relationship between musical training and prosodic emotion perception and to extend this area of study to the lexical and facial channels of communication and musical emotions. Additional aims included examining the perception of individual basic emotions within these channels and a brief investigation into potential cognitive and intellectual differences between musicians and controls.

Using prosodic, lexical, and facial emotional stimuli from the NYEB (Borod, Welkowitz, et al., 1992, 1998) and musical emotional stimuli developed by Eerola and Vuoskoski (2011), we demonstrated that musicians are more accurate than untrained controls in the identification of basic emotions in the prosodic channel of communication and in music. In contrast, group differences were not found for the lexical and facial channels of communication. When patterns of responding were examined, we found that musicians and controls made similar patterns of errors in emotion identification within each of the emotion tasks. However, although both groups typically selected one or two emotions primarily when mistaken (e.g., those who made errors identifying happiness in music most often selected tender instead), controls tended to have a slightly wider and more variable distribution of misidentifications than did musicians. Finally, an examination of cognitive functioning indicated that musicians were better than controls on a measure of irregular word reading used to estimate verbal intelligence (i.e., WTAR), whereas the two groups did not differ on a brief cognitive screening measure (i.e., MoCA) or on a measure of non-verbal/fluid intelligence (i.e., RAPM).
Musical training and emotion perception. The primary aims of this project were to characterize the relationship between musical training and emotion perception in multiple modalities, specifically the prosodic, lexical, and facial channels of communication and emotions expressed in music. To examine these relationships, non-parametric methods, specifically the Mann-U Whitney Test, were first used to assess for group differences, as data were not normally distributed. When differences were found, parametric tests were then used to further examine these relationships. Specifically one-way ANCOVAs were used to assess for group differences in each task (i.e., prosodic, lexical, facial, and musical emotion identification) while also controlling for several group differences thought to potentially impact emotional processing (i.e., the Openness personality trait and the externally-oriented thinking component of the Toronto Alexithymia Scale).  

Musical training and prosodic emotion perception. On the basis of prior research suggesting that musical training may transfer positively to prosodic emotion perception, we hypothesized that individuals with musical training would be more accurate than untrained controls at the identification of emotions in the prosodic channel of communication. Consistent with this hypothesis, we found that musicians were significantly more accurate than controls on a measure of prosodic emotion perception. This finding is consistent with several prior studies, particularly ones by Lima and Castro (2011a) and Thompson and colleagues (2004), who also found that musicians outperformed controls on measures of prosodic emotion perception. Although similar in design, our study differs from that of Lima and Castro’s (2011a) in that we examined a highly heterogeneously diverse population of musicians and controls (e.g., with

---

4 Of note, group differences were also found for the WTAR, an estimate of verbal intelligence, such that musicians scored significantly higher than non-musicians. To determine whether verbal intelligence were a contributing factor to group differences for MID and PID, correlations were performed between each of these tasks and the WTAR score for each participant group (i.e., Musicians and Controls). No significant correlations were found. See Table 7 in the Appendix.
respect to race/ethnicity and native language status), whereas Lima and Castro’s study involved a more homogenous population, that is, all participants were Portuguese. Furthermore, in addition to examining non-auditory channels of communication, our study included two non-affective prosody tasks (i.e., BPID and ICID), which allowed us to ensure that both groups had adequate basic non-affective prosodic perception ability and, if present, to control for relevant group perceptual differences.

Our finding is in contrast to that of Trimmer and Cuddy (2008), who did not find a relationship between musical training and the perception of affective prosody. Notably, Trimmer and Cuddy used a large sample of native English-speaking undergraduates, but participants were not selected based specifically on musical training; rather, they evaluated a single group with a range of 0-17 years of musical training and a mean of 6.5 years of training. Our study, as well as other studies that found similar relationships, examined individuals with 8 or more years of training. We also recruited only individuals who began musical training prior to age 12 in order to ensure that musical training began during the proposed “critical period” for music skill acquisition (e.g., Bailey et al., 2014; Steele et al., 2013). As Trimmer and Cuddy (2008) did not discuss or report age of onset, it is uncertain at what age their participants initiated musical training.

In order to further explore relationships between musical training and emotion perception, we also examined the groups with respect to the eight individual emotions included in the prosodic identification task (i.e., happiness, pleasant surprise, interest, unpleasant surprise, sadness, disgust, anger, and fear). Wald tests were used to assess for significant differences between the two groups’ distributions of erroneous responses. Significant differences were found for all emotions except for pleasant surprise and sadness. These distributions were then visually
inspected. Results indicated that musicians do not appear to process prosodic emotions in a distinctly different way from non-musicians, as both musicians and controls tended to select the same or similar emotions when they made an error. See Table 10 of the Appendix for statistics.

On an exploratory basis, we performed a $2 \times 8$ ANCOVA (Group [Musicians and Controls] by Emotion [Happiness, Pleasant Surprise, Interest, Unpleasant Surprise, Sadness, Disgust, Anger, and Fear]), with TIP-I-O and TAS-3 as covariates, to further examine individual emotions. A main effect of Group remained for PID, such that musicians were more accurate than controls. However, no significant interaction effects were obtained. See Tables 8 and 9 of the Appendix for descriptive statistics. Musicians were more accurate than Controls on six of the eight emotions tested, specifically, for happiness, interest, unpleasant surprise, sadness, disgust, and fear, but not pleasant surprise or anger, though not all of these comparisons were significant. Prior studies have obtained similar results for happiness, surprise, neutrality, disgust, sadness, anger, and/or fear (e.g., Lima & Castro, 2011a, 2011b, 2012; Pinheiro et al., 2015, Thompson et al., 2004). The relationship between prosodic perception and musical training, therefore, appears to be general across emotions, rather than, for example, isolated to a single emotion or emotion-type (e.g., negatively-valenced emotions). Overall, findings suggest shared mechanisms for the processing of music and speech prosody. Implications of these findings are discussed in more detail below.

Musical training and musical emotion perception. Another primary aim of the current project was to evaluate the perception of basic musical emotions. Consistent with our hypothesis, we found that musicians and controls were both generally able to identify the target musical emotion, but musicians were, nonetheless, significantly more accurate. Next, we examined relationships between musical training and individual basic musical emotions by performing a 2
x 5 mixed-design ANCOVA (Group [Musicians and Controls] X Emotion [Happiness, Tender, Anger, Fear, and Sadness]). An overall effect of Group was present, such that musicians were significantly more accurate than non-musicians, and no interaction effects were obtained. Patterns of erroneous responding were then examined. Significant differences were found between groups for Happiness, Tender, and Fear, but not for Sadness, or Anger. As with prosodic perception, when visually inspected, distributions revealed that both musicians and controls made similar types of errors, but controls were more likely to misidentify an emotion as one of several different emotions, whereas musicians more often selected the most commonly misidentified emotion. These similar patterns of identification indicate that musicians and non-musicians are likely identifying basic musical emotions using the same mechanisms. Overall, these findings have implications for models of emotion processing and theories of musical emotion and will be discussed further.

Associations between length of training and accuracy. The current study design is limited in that it does not allow for conclusions regarding causality or directionality with respect to the relationship between musical training and emotion perception. Our rationale in positing these relationships was partially causal; specifically, that musical training may result in enhanced emotion perception. Alternately, it is possible that those individuals with strong innate acoustic and/or emotion abilities are drawn to music as a pursuit or that common factor(s) give rise to both. In order to further explore this issue, we carried out Spearman correlations between duration of training and MID and PID task accuracy scores, respectively (see Table 11 of the Appendix). Significant, but modest, positive correlations were found for both tasks such that, as duration of training increases, so does task accuracy. This suggests a “dose-dependent”-type relationship between years of training and accuracy. Nevertheless, the possibility of other
Musical training and lexical and facial emotion perception. To our knowledge, this is the first study to directly examine emotion perception in the lexical and facial channels of communication in adult musicians. On the basis of preliminary emotion perception findings in musicians, as well as related behavioral and neuroanatomical research, we hypothesized that musicians would be more accurate than controls in the perception of emotion in both the lexical and facial channels of communication. Contrary to these hypotheses, however, no significant group differences were obtained for either task.

Given the possibility of an association between musical training and a subset of emotions, we examined individual emotions in LID and FID on an exploratory basis. A 2 x 8 ANCOVA (Group [Musicians and Controls] by Emotion [Happiness, Pleasant Surprise, Interest, Unpleasant Surprise, Sadness, Disgust, Anger, and Fear]) was performed with TIPI-O and TAS-3 as covariates. However, no significant main effect of Group or interaction effects were found for either task (see Tables 7 & 8). In sum, no associations were found between musical training and emotion perception in the lexical and facial channels. Implications for theories of learning transfer and general versus domain-specific models of emotion perception processing will be discussed below.

It should be noted that our study examined overall group differences between musically trained and untrained individuals and that all of our musically trained participants were carefully selected to ensure a high level of training (≥ 8 years). Nevertheless, it is possible that a greater duration of training might still be associated with enhanced accuracy in the lexical and facial channels. In order to further investigate this possibility, Spearman correlations were performed.
between years of training and LID or FID task accuracy (see Table 11 of the Appendix). No significant correlations were found, however, indicating that accuracy for these channels does not significantly increase commensurate with years of training.

**The effect of ethnicity on emotion perception.** Although it was not significant, when ethnicity (i.e., White versus Non-White) was examined as a function of participant group (i.e., Musicians or Controls), there was a trend finding (p = .085) such that there were relatively more White subjects in the Musician group and relatively more non-White subjects in the Control group. In light of research regarding an “in-group advantage” (IGA) for perceiving emotional stimuli (e.g., Bell, 2008), the variable of ethnicity is important to consider. IGA refers to the possibility that “individuals can more easily and accurately understand emotional expressions originating from members of their own cultural group rather than expressions originating from members of a different cultural group” (Elfenbein, 2007, p. 51). In the current study, White subjects might achieve greater accuracy in emotion perception since the stimuli for the three NYEB tasks were posed or created by White individuals.

To explore this possibility, we conducted a 3-way ANCOVA for Group (2), Ethnicity (2), and Emotion (5 or 8), covarying for the TAS-3 and TIPI-O, on accuracy for each of the emotion perception tasks (i.e., MID, PID, FID, and LID). See Table 16 for statistics. For Group, there continued to be significant findings for MID (p = .003) and PID (p = .026), with Musicians more accurate than Controls. For Ethnicity, only one of the 4 effects was significant, that is, for PID, where White participants scored higher than non-Whites, consistent with the IGA theory. Importantly, in terms of the findings reported in the Results section, there were no significant interactions involving Group and Ethnicity for any of the tasks. There were significant main effects of Emotion for each task. Generally, for MID, Fear was the most accurately perceived
emotion, and Tenderness the least accurately perceived. For PID, Sadness and Disgust were the most accurately perceived, and Fear the least. For FID, Interest and Unpleasant Surprise were less accurately perceived than the other emotions. Finally, for LID, sadness was more accurately perceived than the other emotions.

In sum, our study demonstrates that musicians are more accurate than non-musicians in the identification of prosodic and musical emotion, but this was not the case for lexical or facial emotion, suggesting that music and emotional prosody may be processed via some overlapping mechanisms. Implications for models of learning and transfer, emotion perception, and musical emotion will now be discussed.

**Theoretical Implications of Emotion Findings**

The current study found a positive relationship between musical training and the perception of basic emotions in both intonation/speech output (i.e., emotional prosody) and music, but not in the lexical or facial channel. From a neurocognitive perspective, these findings are consistent with the hypothesis that music and spoken language engage shared mechanisms, and adds to a growing body of behavioral, functional, and neuroanatomical evidence showing overlap between music and speech at multiple levels of processing (e.g., Koelsch, et al. 2005, 2011; McMullen & Saffran, 2004; Moreno et al., 2009; Parbery-Clark et al., 2009, 2011). As both musicians and controls were able to accurately identify basic emotions in music, findings also offer further validation for a basic emotion model in music.

With regards to emotion processing specifically, it is not yet known which stage(s) of processing is shared with music or at what point(s) there might be domain transfer between music training and emotion perception. Areas of possible overlap and/or transfer include, but are not limited to: the level of initial auditory feature extraction (e.g., location, pitch, timbre,
louderness; processed by the brainstem and the primary auditory cortices), the level of auditory gestalt formation (e.g., superior temporal gyrus), the level of vitalization (i.e., physiological/body response, involving autonomic and endocrine systems and multimodal association cortices), the meaning-making/semantic level for speech (e.g., inferior frontal gyrus, middle temporal gyrus, and fusiform gyrus; (Koelsch, 2011; Peretz & Coltheart, 2003; Peretz & Zattore, 2005) and/or the conceptual system of emotions (subserved by multiple brain regions, particularly limbic regions; Barrett, 2006; Koelsch, 2011; Niedenthal, 2007).

Given the number of possible loci at which overlap may occur, it is beyond the scope of the current study to provide a satisfactory account for the relationship between music and emotion perception. However, as our study is the first to examine musical training and emotion perception in multiple communication channels, we are able to make suggestions about the most likely points of overlap and/or transfer that can serve as a guide for future research.

**Auditory processing.** Given our finding of significant differences for auditory emotion perception only, group differences in auditory perceptual abilities remain a clear possibility. There is a large body of literature documenting relationships between music training and auditory abilities such as pitch processing (e.g., Besson et al., 2007), and pitch is known to contribute to the differentiation of emotions in both speech and music. As speech intonation/prosody stems partially from pitch variation, our study included two control measures of non-emotional prosodic perception (i.e., BPID and BCID). Interestingly, we did not find group differences on either control measure. As both groups had a high degree of accuracy on these tasks, it is possible that these measures were simply not sensitive enough to differentiate group differences in pitch or non-emotional prosody perception (see Table 4). However, it is also evident that discrete emotions are conveyed in speech and music not only via pitch but also via...
perceptual qualities such as timbre, loudness, intensity, and variation in rate and timing (e.g., Banse & Scherer, 1996). Musicians may, therefore, have superior ability not in simply processing pitch or individual acoustic cues. Instead, they may have an enhanced ability to integrate acoustic information in order to extract complex semantic information, such as emotional salience.

**Domain-specific emotion perception processing.** It should be noted, that, consistent with prior research, we found similar patterns in the identification of prosodic and musical emotions across both groups. Furthermore, Lima and Castro (2011a) found that acoustic cues were equally predictive of emotion categorization for both musicians and controls. These findings raise the possibility that overlap between music and emotion processing occurs after the level of basic auditory feature extraction and integration.

It has been argued that music was developed to mimic speech and, therefore, that music has the ability to elicit emotional responses in a manner analogous to speech (e.g., Juslin, 2003). If this were the case, then music listening would engage neural processes devoted to the processing of emotional prosody and, in turn, elicit an emotional experience. This experience would then be interpreted by the listener as an intentional expression of emotion in music (e.g., Juslin, 2013; Juslin & Västfjäll, 2008). Such a process would constitute overlap at the domain-specific level of prosodic emotion perception.

Alternately, it may be the case that refinement of musical emotion perception transfers to prosodic perception. Barrett (2006) has proposed that there are various levels at which individuals are able to refine and specify emotional states, called granularity. Individuals with lower emotional granularity tend to perceive emotional states as broad and non-specific, whereas those with high granularity tend to perceive more specific, differentiated emotional states (e.g.,
Barrett, 2006; 2009). It is possible that musical training, because it involves constant perception and expression of emotion in a musical context, refines emotional granularity for music, which, in turn, is then transferred to prosodic emotion perception because of shared acoustic cues.

Both of these approaches would account for our findings that musicians tended to be narrower in their error patterns and generally more accurate than controls across most of the basic emotions examined, including basic emotions that have *not* been well validated in music (e.g., disgust; Erøla & Vuoskoski, 2011). Either music evokes emotions based on their similarity to emotions in speech, increased granularity for musical emotion perception enhances the ability to discriminate and categorize prosodic emotion (thus allowing individuals to perform more accurately on identification tasks even when the emotion is not one typically represented in music), *or* both mechanisms may work in conjunction to mutually enhance auditory emotion perception.

**Domain-general emotion perception processing.** On the basis of significant correlations among emotion identification communication channels (i.e., facial, prosodic, and lexical), Borod et al. (2000) proposed that there is a general emotion processor for the perception of emotion (see also Finley, Borod, Brickman, et al., 2008). The authors noted that, although the information for each channel is processed through generally separate sensory systems, for perception to be strongly related across channels, affective content is likely mapped onto perceptual stimuli, rather than processed only within its respective sensory modality. Similarly, Trimmer and Cuddy (2008) argued that the recognition of musical and prosodic emotions might not be a function of acoustic sensitivity but, rather, “the operation of a cross-modal emotional processing system” such as that proposed by Scherer and Ellgring (2007).
By including the lexical and facial channels, we, therefore, investigated the possibility that a general emotion perception processing system represents a point of overlap through which transfer can occur from musical training to emotion perception, but we did not find support for this hypothesis. Our findings were, instead, more consistent with literature demonstrating dissociations between perception in the various channels of communication. For example, impairment in facial emotion perception is well documented in patients with Alzheimer’s disease, whereas patients’ ability to identify musical emotions appears generally preserved (e.g., Drapeau, Gosselin, Gagnon, Peretz, & Lorrain, 2009). Similarly, Gosselin, Peretz, Hasboun, Baulac, and Samson (2011) demonstrated a dissociation between the perception of scary music and fearful faces in individuals following resection of the anteromedial temporal lobe, similar to findings previously reported by Adolphs, Tranel, and Damasio (2001) regarding dissociation between facial and prosodic perception. In related findings, Hsieh et al. (2012), in their examinations of patients with semantic dementia, determined that the recognition of musical emotions uses greater cognitive and neural resources than does the recognition of facial emotions. Banziger, Grandjean, and Scherer (2009), using a principal component analysis to examine facial, prosodic, and postural emotion perception, found separate auditory modality and visual modality factors. In sum, our findings are more consistent with modality-specific emotion perception processors than with a general emotion perception-processing model. It would be interesting, in future research with this data set, to examine relationships between the channels in order to further explore these findings.

Learning transfer. Another possible explanation for the absence of lexical and facial group differences is that, even with the presence of a general emotion processor, musical training is not sufficient to cause a transfer effect. The current study hypothesized positive associations
between emotion perception ability and musical training partially on the basis of two possible pathways through which learning transfer might occur. The first is the possibility of a transfer from musical training to prosodic and musical emotion perception specifically, given evidence of shared mechanisms, and, the second, transfer of musical training to overall emotion perception ability. Based on our findings, implications for learning transfer models will now be briefly discussed.

In their proposed taxonomy for learning transfer, Barnett and Ceci (2002) argued that transfer is more likely to occur when there are multiple shared dimensions (e.g., temporal, physical, functional, and social contexts; knowledge domains; and modalities) between the training and transfer contexts. Near transfer effects (i.e., the training and transfer contexts are highly similar with respect to the above dimensions) are well established for a wide range of perceptual, motor, and cognitive skills (e.g., Schellenberg & Weiss, 2013; Moreno et al., 2012). Far transfer effects (i.e., training and transfer contexts share fewer of the above dimensions) are typically more difficult to ascertain, and the frequency at which they occur is highly debated (e.g., Zelinski, 2009).

Using Barnett and Ceci’s (2002) framework, transfer from musical training to prosodic and musical emotion perception would therefore be considered a near transfer. There is repeated demonstration of superior auditory processing in musicians, explicit overlap between the auditory processes used in auditory emotion perception and in music learning/performance (e.g., pitch, rhythm, timbre, and timing), and overlapping neural networks recruited during both tasks, as discussed previously. In contrast, there are fewer shared dimensions between music performance and facial and lexical emotion perception, and those dimensions are less direct (e.g., facial/postural expression and perception of emotion during performance), making this a
proposed far transfer. Our findings suggest the possibility of a near transfer effect, but do not support far transfer. This is consistent with Barnett and Ceci’s (2002) proposed model and with the position that far transfer effects tend to be both rare and often involve explicit cuing regarding the transferred principle (Detterman, 1993). This may also clarify prior inconsistencies with regards to findings on prosody perception. For example, Mualem and Lavidor (2015) showed increased emotion perception accuracy in participants who completed a very brief music intervention, but not in individuals with long-term musical training. However, their intervention design included an explicit focus on emotion in music that is not common in standard music training models (Karlsson & Juslin, 2008). The findings from the current study add to our understanding of models of learning transfer, a particularly relevant area of inquiry given increasing interest in skill transfer as a means of cognitive rehabilitation and psychotherapeutic intervention (e.g., Wykes & Spaulding, 2011).

Theories of musical emotion. The current study evaluated the perception of emotion in music, an area that has received relatively little attention in comparison to other channels (Lima & Castro, 2012). Our study demonstrated that musicians and controls are both able to identify basic emotions in music, but that musicians are more accurate than controls. As prior findings regarding both general musical emotion perception and perception in musicians specifically have been equivocal, our study may provide clarification regarding key theoretical issues regarding emotion and music. In particular, there is considerable debate as to the most appropriate model to use when characterizing musical emotions, given the complex, extra-linguistic nature of emotional responses to music. Our findings demonstrated successful identification of discrete basic emotions in music, supporting the validity of a basic emotion model. Additionally, musicians demonstrated the same patterns of identification for three of the four overlapping
musical and prosodic emotions (i.e., happiness, sadness, and fear, but not anger), further suggesting that music and prosody communicate emotions in an analogous fashion using specific acoustic cues. That said, it is evident that some emotions (e.g., disgust and surprise; Eerola & Vuoskoski, 2011) are present in the prosodic, facial, and lexical channels but are not well identified in music. This discrepancy may be reconciled through the development of newer models of musical emotion, such as Juslin’s (2013) “multiple layers” model, which incorporates both a vocal-similarity model and appraisal.

Bigand and Poulin-Charronnat (2006), among others, have argued that we are “experienced listeners,” such that many musical capacities (e.g., the ability to process underlying musical structures) are automatically acquired, to a high degree of sophistication, through passive exposure to music and do not necessitate any form of explicit training. It has been argued that this, in turn, extends to emotional responses to music, such that all listeners are able to discern emotional content from music and that musical training is unlikely to modify emotional responses to music. In contrast, our findings suggest that, while there does appear to be an innate ability to discern emotion from music, this ability is nevertheless positively related to musical training, and as such is either an innate faculty that can vary across individuals or is modifiable through experience.

Musical Training and Cognitive Abilities

An additional, purely exploratory, aim of the current study was to assess for the possibility of cognitive differences between musicians and controls. In order ensure that the data collection time was feasible for participants, measures were selected that were brief and could also function as screening measures (e.g., although participants were screened for neurological disorders, these measures provided an additional means by which to verify adequate cognitive
ability to complete the study tasks). It is, therefore, important to emphasize that these measures were not intended to provide a comprehensive assessment of cognitive functioning.

Group performances were compared using one-way ANOVAs for each of the cognitive measures (i.e., MoCA, RAPM, and WTAR). No significant Group differences were obtained for either the MoCA or the RAPM. A significant group difference was obtained for the WTAR, such that musicians were significantly more accurate than non-musicians.

Regarding these findings, the MoCA does assess specific cognitive domains in which musicians are thought to have an advantage (e.g., verbal fluency and verbal memory; Schellenberg & Weiss, 2013). However, it should be noted that the MoCA is a very broad cognitive screening measure originally designed to assess for mild cognitive impairment (although it is now widely used as a basic assessment tool across a variety of clinical settings). Therefore, it is possible that significant group differences were not obtained due to a lack of sufficient sensitivity. It would also be expected, as we found, that the majority of healthy individuals would score at or above the cutoff, creating a ceiling effect in the scores for both groups. Similarly, the MoCA yields an aggregate score that encompasses multiple cognitive domains, a number of which are not associated with an advantage of musical training (e.g., visual confrontation naming). It is therefore possible that there may be differences between groups with respect to individual domains or tasks that are not reflected in the aggregate score.

The remaining measures given are both means of assessing general intelligence. Our finding of a significant group difference, favoring musicians, for the WTAR, but no difference for the RAPM, is consistent with some prior research evaluating intelligence and musical training. In general, musicians have been found to outperform non-musicians on measures of general intelligence (e.g., Schellenberg, 2004; Schellenberg, 2006). However, when differences
are found, they are typically found for intelligence test batteries (e.g., Wechsler test full scale IQ scores) or measures of academic achievement. In contrast, the majority of studies that have failed to find group differences have used measures of fluid/non-verbal intelligence, such as the RAPM or Matrix Reasoning subtest from the Wechsler intelligence tests (e.g., Schellenberg & Moreno, 2010). This discrepancy may also be explained by findings of transfer effects from musical training to language learning. That is, musicians may have a selective advantage on verbally-mediated, but not non-verbal, measures. The WTAR assesses acquired word knowledge through the reading of irregularly spelled words that cannot be phonetically decoded. This finding may therefore reflect superior early language learning on the part of musicians, as prior research suggests musical training is associated with increased vocabulary and overall superior language learning ability (e.g., Degé & Schwarzer, 2011). However, as mentioned previously, our study design does not allow for conclusions regarding causality. The possibility should therefore be raised that this finding is, in fact, not directly related to musical training. Alternately, there is the possibility that high functioning children are more likely to seek out novel and mentally engaging activities such as music lessons. Similarly, there also may be shared factors that contribute to both language ability and propensity to engage in musical training in childhood, such as parent level of education, overall socioeconomic status, and similar interrelated factors. Lastly, there may instead be non-music factors that explain participants’ language acquisition and/or performances on the WTAR specifically, as it is a measure developed and normed on a native English-speaking U.S. Population (e.g., parent native language status and country of birth). For example, individuals whose parents were non-native English-speakers or who learned English as a second language may, therefore, be at a disadvantage on such measures due to
reduced vocabulary size, as compared to individuals who were born in the United States and/or raised by native English-speakers (e.g., Kieffer & Lesaux, 2012).

Limitations

It should be noted that there are several limitations to the current study. First, and most importantly, our study was correlational, not longitudinal, in design. While the study design allows for comparison of individuals with substantial years of musical training, the directionality of the relationship between musical training and auditory emotion perception cannot be established. There is a significant body of literature demonstrating behavioral changes and apparent neurological changes following intensive musical training, and some evidence from brief longitudinal studies that emotion perception is enhanced following a musical training intervention (For discussion of this issue, please see the “Neurological differences between musicians and non-musicians” section on page 24, the “Musical Training and Non-Emotional Mental Processes” section on page 27, and the “Musical Training and Emotional Processes” section on page 32). Nevertheless, it is possible that individuals who pursue intensive music training already tend to possess a strong faculty for either acoustic and/or emotion processing, which would account for our findings, or that these innate abilities serve to enhance benefits from musical training. As effects of musical training are identified more often in individuals with fairly extensive musical training (e.g., 8 or more years), a longitudinal study may be impractical, particularly if older individuals were to also be included in the study. Nevertheless, such a study would be crucial in establishing directionality and clarifying study findings.

In addition, the current study was also limited by the use of purely behavioral measures. Although the descriptive approach taken in this study is important in order to understand the relationship between emotion perception and musical training, it did not allow us to draw any
conclusions about underlying neural mechanisms. Although we have discussed underlying musical training-related differences in the brain in relation to the observed results, it is important to emphasize that we are not able to make any direct conclusions in this area based on the study design.

The study was also limited with regards to differentiating contributions from musical exposure alone from those of exposure combined with music practice and performance. It is possible that prolonged listening to music alone could contribute to some or all of the observed group differences. Although both groups of participants were asked to estimate exposure to both recorded and live music, many participants found this quite challenging and typically provided wide ranges as estimates. A more systematic measure to evaluate exposure could be developed, or, potentially, a randomized intervention study could be designed involving intensive music exposure only (as existing interventions have involved a combination of listening and performance).

Interestingly, Bigand and Poulin-Charronnat (2006) argued that musicians and non-musicians might differ on explicit tasks of emotion identification entirely on the basis of differences in their respective abilities to use linguistic terms to characterize their emotional responses to music. Similarly, Koelsch et al. (2015) argue that the subjective feeling states (termed “emotion precepts” by the authors) evoked by prosodic and musical emotions are likely to have more inter-individual correspondence than the actual words an individual may use to describe these feeling states (e.g., their actual feeling states are more similar than their choice of descriptive word would reflect). As our study utilized an emotion identification task in which individuals selected from a provided list of possible emotions, it is possible that this approach limited the extent to which non-musicians were able to characterize their understanding of and/or
emotional responses to music. The use of multidimensional scaling techniques to allow participants to characterize emotional content without being constrained by linguistic translation should be considered. For example, participants do not have a forced choice when identifying emotion, but, instead, can associate various images and words with emotional stimuli and/or are able to group stimuli that they perceive as having similar emotional content. These responses can then be analyzed in order to determine if they still fit within a basic emotion framework (e.g., whether or not participants routinely associate a particular stimulus with images and/or words that express anger or other basic emotions).

Regarding cognitive differences between musicians and non-musicians, as previously discussed, the measures given did not allow for a comprehensive assessment of individual cognitive domains. More comprehensive neuropsychological testing would be necessary to identify subtle cognitive differences between musicians and non-musicians. Furthermore, although group differences were found in estimated verbal intelligence, it is also important to stress that socioeconomic status (SES) was not controlled for in the present study. It is possible that the obtained difference is not a function of musical training, but, rather, that individuals who have pursued intensive musical training, especially from a young age, had the opportunity to do so because they are of a higher SES, and that this, rather than musical training, accounts for the obtained difference. Although participants were asked to provide their parents’ levels of educational and occupational attainment and estimated family income to use as proxy variables to ascertain socioeconomic status, many participants did not know this information and/or elected not to provide it. Future studies should ensure that these and related variables that may influence intellectual performances are well-controlled.

**Clinical Implications**
Declines in emotion perception and other aspects of emotion processing are present in multiple clinical populations (e.g., Schizophrenia, Parkinson’s disease, and Major Depressive Disorder), which can exacerbate other symptoms as well as interfere with treatment adherence and efficacy. Prior research suggests that musical training can affect multiple emotional functions (e.g., enhance emotion regulation, increase social engagement, and modify mood), and there is considerable interest in the possibility of using musical training as an intervention. Our findings further support the merits of developing and researching musical interventions for use with a wide range of clinical populations.

Furthermore, emotion perception declines with aging (Finley, Borod, Schmidt, et al., 2008; Lima & Castro, 2011b; Savage et al., 2013), and this is thought to potentially have adverse effects on social functioning and mood (e.g., Cacioppo & Hawkley, 2009). This is a particularly relevant area for further study and intervention, given the rapidly increasing older adult population. In line with Kraus and Anderson’s (2013a) argument that musical training may enhance cognitive reserve, our findings raise the possibility that musical training may have a preventative effect, bolstering an individual’s emotional capacities to reduce the impact of normal age related declines. Similarly, our findings provide further support for efforts to design music-based interventions for older adults.

**Directions for Future Research**

Our findings further support musical training as a promising area for behavioral and neuroscientific research, and there are many aspects of the current research that still need to be explored more fully. First and foremost, it is important to address some of the limitations discussed above. Longitudinal studies are of particular importance in order to evaluate directionality in the relationship between musical training and emotion perception. Prior
longitudinal studies involving musical training, including those for non-emotional processes, have primarily consisted of brief musical interventions. A few have followed children over two years of training, but none of these have then examined emotional functioning. Studies that follow individuals from childhood who go on to become musicians would be ideal, although they are of course less feasible given the disadvantages of longitudinal research (e.g., cost and attrition).

In line with potential clinical applications, it would be of interest to extend this line of research to other age groups. The current study was restricted to individuals under age 40 and, in practice, primarily included individuals in early adulthood. Examining older adults would allow for explorations of interactions between musical training and emotion perception across the lifespan. Furthermore, it allows for investigation into the possibility that musical training may delay or offset cognitive or emotional declines from aging. Similarly, examining musically trained and untrained individuals within clinical populations known to have deficits in emotional processing is also of interest, as this would shed light onto the possibility that musical training may, to some degree, ameliorate these impairments.

Significant findings were not obtained for the lexical and facial channels of communication. As discussed previously, it is possible that differences were not detected as the stimuli were not sufficiently sensitive. Future studies may want to consider investigating these channels using stimuli that are more challenging, such as stimuli considered to express a mild to moderate amount of a target emotion, or ambiguous, such as blended faces. Similarly, more naturalistic stimuli, rather than posed/acted stimuli, could be used to investigate if, in daily life, musicians do have an advantage over controls in emotion perception.
Finally, our study did not include the postural/gestural channel of communication, which is a relatively more recent, but quite promising, area of emotion processing research. Emotion is conveyed via body posture, gestures, and movement during musical performance, and this visual information has been shown to modify and enhance the perception of the emotions expressed in the accompanying music (for example, participants report stronger subjective emotional experience in response to music if it is paired with video clips of a performer demonstrating emotionally congruent body movements; e.g., Krahé et al., 2013). It would be interesting to examine the perception and expression of emotion in musicians for the postural/gestural channel.

Conclusions

In summary, results of the current study provide evidence that individuals with musical training are better at identifying basic emotions in the prosodic channel of communication and in music, but not in the facial or lexical channels. This work expands on prior research by examining this relationship in a culturally diverse participant population and helps to clarify prior findings through carefully controlling for relevant demographic, perceptual, and psychological (e.g., personality) factors. Furthermore, to the best of our knowledge, this is the first study to examine the relationship between musical training and lexical emotion perception and the first to systematically examine facial emotion perception in adult musicians. Together, this work has implications for understanding relationships between music and speech and between music and emotion perception processing, it has clinical implications (e.g., in the development of cognitive remediation and psychotherapeutic interventions targeting emotion processes), and it provides additional support for the merit of music as a promising area of neuroscientific research.
Table 1: Inclusion Criteria

<table>
<thead>
<tr>
<th></th>
<th>All participants</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>18-40</td>
<td></td>
</tr>
<tr>
<td>Age learned English</td>
<td>( \leq 7 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Musician Group</strong></td>
<td><strong>Controls</strong></td>
</tr>
<tr>
<td>Age of training onset</td>
<td>( \leq 12 )</td>
<td>n/a</td>
</tr>
<tr>
<td>Years of training</td>
<td>( \geq 8 )</td>
<td>( \leq 3 ) years over lifespan</td>
</tr>
<tr>
<td>Current practice</td>
<td>( \geq 2 ) hours/week</td>
<td>n/a</td>
</tr>
<tr>
<td>Performance type</td>
<td>Instrumental musician (with or without additional voice training)</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 2: Participant Demographic Characteristics by Group

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>Musicians (n=58)</th>
<th>Controls (n=61)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>22.00 (5.15)</td>
<td>21.08 (4.09)</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13.52 (1.79)</td>
<td>13.24 (1.30)</td>
</tr>
<tr>
<td>Music Training (years)</td>
<td>13.26 (4.37)</td>
<td>n/a</td>
</tr>
<tr>
<td>Practice (Hours per week)</td>
<td>6.73 (4.01)</td>
<td>n/a</td>
</tr>
<tr>
<td>Sex (% Female)</td>
<td>55%</td>
<td>48%</td>
</tr>
<tr>
<td>Race/Ethnicity (% White)</td>
<td>53%</td>
<td>38%</td>
</tr>
<tr>
<td>Handedness (% Right)</td>
<td>95%</td>
<td>92%</td>
</tr>
</tbody>
</table>

*Note:* Values are listed as mean (standard deviation) or %
### Table 3: Descriptives for Group Scores on Questionnaires

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>GROUPS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians</td>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=58)</td>
<td>(n=61)</td>
<td></td>
</tr>
<tr>
<td>Toronto Alexithymia Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAS-1</td>
<td>13.05 (4.31)</td>
<td>13.72 (5.58)</td>
<td></td>
</tr>
<tr>
<td>TAS-2</td>
<td>14.77 (5.49)</td>
<td>16.08 (5.76)</td>
<td></td>
</tr>
<tr>
<td>TAS-3</td>
<td>16.53 (4.09)</td>
<td>18.44 (4.04)</td>
<td></td>
</tr>
<tr>
<td>TAS-Overall</td>
<td>44.34 (10.88)</td>
<td>48.25 (12.18)</td>
<td></td>
</tr>
<tr>
<td>Ten-Item Personality Inventory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIPI-Extraversion</td>
<td>4.50 (1.39)</td>
<td>4.41 (1.46)</td>
<td></td>
</tr>
<tr>
<td>TIPI-Agreeableness</td>
<td>4.87 (1.23)</td>
<td>5.56 (1.11)</td>
<td></td>
</tr>
<tr>
<td>TIPI-Conscientiousness</td>
<td>5.54 (1.11)</td>
<td>5.36 (1.33)</td>
<td></td>
</tr>
<tr>
<td>TIPI-Emotional stability</td>
<td>4.61 (1.37)</td>
<td>4.77 (1.20)</td>
<td></td>
</tr>
<tr>
<td>TIPI-Openness</td>
<td>6.08 (0.98)</td>
<td>5.34 (1.25)</td>
<td></td>
</tr>
<tr>
<td>Short Test of Musical Preference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOMP-Reflective &amp; Complex</td>
<td>4.73 (0.87)</td>
<td>3.91 (0.95)</td>
<td></td>
</tr>
<tr>
<td>STOMP-Intense &amp; Rebellious</td>
<td>4.54 (1.07)</td>
<td>4.14 (1.30)</td>
<td></td>
</tr>
<tr>
<td>STOMP-Upbeat &amp; Conventional</td>
<td>4.48 (9.92)</td>
<td>4.57 (1.04)</td>
<td></td>
</tr>
<tr>
<td>STOMP-Energetic &amp; Rhythmic</td>
<td>4.71 (0.98)</td>
<td>4.96 (1.05)</td>
<td></td>
</tr>
<tr>
<td>STOMP-Soundtracks</td>
<td>5.43 (1.31)</td>
<td>5.58 (1.4)</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Values are listed as mean (standard deviation)*

### Table 4: Descriptives for Non-emotional control measures

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>GROUPS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians</td>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=58)</td>
<td>(n=61)</td>
<td></td>
</tr>
<tr>
<td>NSID¹ % correct</td>
<td>74.14 (14.30)</td>
<td>73.36 (12.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61.15</td>
<td>58.91</td>
<td></td>
</tr>
<tr>
<td>BPID² % correct</td>
<td>81.15 (9.50)</td>
<td>83.22 (9.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55.38</td>
<td>64.39</td>
<td></td>
</tr>
<tr>
<td>ICID³ % correct</td>
<td>89.58 (9.28)</td>
<td>87.77 (10.09)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63.18</td>
<td>56.98</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Values are listed as mean (standard deviation) with mean rank underneath*

¹ Non-Emotional Sentence Identification
² Benton Phoneme Discrimination
³ Intonation Contours Perception
Table 5: Descriptives for Emotion Identification Tasks

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians (n=58)</td>
</tr>
<tr>
<td>PID (^1) % correct</td>
<td>59.41 (10.91) 67.53</td>
</tr>
<tr>
<td>LID (^2) % correct</td>
<td>75.07 (12.15) 62.61</td>
</tr>
<tr>
<td>FID (^3) % correct</td>
<td>71.55 (10.43) 62.65</td>
</tr>
<tr>
<td>MID (^4) % correct</td>
<td>68.55 (7.91) 70.00</td>
</tr>
</tbody>
</table>

Note: Values are listed as mean (standard deviation) with mean rank underneath

\(^1\) Prosodic Identification
\(^2\) Lexical Identification
\(^3\) Facial Identification
\(^4\) Music Identification

Table 6: Descriptives for Individual Musical Emotions

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians (n=58)</td>
</tr>
<tr>
<td>Happiness</td>
<td>66.72 (10.326) 63.77 (11.120)</td>
</tr>
<tr>
<td>Tenderness</td>
<td>62.07 (16.087) 56.89 (16.180)</td>
</tr>
<tr>
<td>Sadness</td>
<td>69.83 (15.159) 65.74 (18.299)</td>
</tr>
<tr>
<td>Anger</td>
<td>63.97 (17.566) 61.97 (16.615)</td>
</tr>
<tr>
<td>Fear</td>
<td>80.17 (13.827) 73.77 (13.437)</td>
</tr>
</tbody>
</table>

Note: Values are listed as mean (standard deviation)
Table 7: MID\(^1\) Distribution of Total Responses for Each Emotion in %

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happiness</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>66.7</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>63.8</td>
</tr>
<tr>
<td>Musician/Tender</td>
<td>27.8</td>
</tr>
<tr>
<td>Control/Tender</td>
<td>25.7</td>
</tr>
<tr>
<td>Musician/Sad</td>
<td>4.5</td>
</tr>
<tr>
<td>Control/Sad</td>
<td>9.0</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>0.3</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>0.7</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>0.7</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>0.8</td>
</tr>
</tbody>
</table>

\(^1\)Music Identification
Table 8: PID\textsuperscript{1} Distribution of Total Responses for Each Emotion in %

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happiness</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>53.4</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>43.7</td>
</tr>
<tr>
<td>Musician/Pleasant</td>
<td>22.4</td>
</tr>
<tr>
<td>Control/Pleasant</td>
<td>27.9</td>
</tr>
<tr>
<td>Musician/Interest</td>
<td>10.3</td>
</tr>
<tr>
<td>Control/Interest</td>
<td>11.5</td>
</tr>
<tr>
<td>Musician/U. Surprise\textsuperscript{2}</td>
<td>2.3</td>
</tr>
<tr>
<td>Control/U. Surprise</td>
<td>2.7</td>
</tr>
<tr>
<td>Musician/Sadness</td>
<td>1.7</td>
</tr>
<tr>
<td>Control/Sadness</td>
<td>3.3</td>
</tr>
<tr>
<td>Musician/Disgust</td>
<td>8.6</td>
</tr>
<tr>
<td>Control/Disgust</td>
<td>5.5</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>1.1</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>4.4</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>0.0</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>1.1</td>
</tr>
</tbody>
</table>

\textsuperscript{1}Prosodic Identification, \textsuperscript{2}Unpleasant Surprise
Table 9: LID\(^1\) Distribution of Total Responses for Each Emotion in %

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happiness</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>72.4</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>71.6</td>
</tr>
<tr>
<td>Musician/Pleasant Surprise</td>
<td>10.3</td>
</tr>
<tr>
<td>Control/Pleasant Surprise</td>
<td>8.7</td>
</tr>
<tr>
<td>Musician/Interest</td>
<td>15.5</td>
</tr>
<tr>
<td>Control/Interest</td>
<td>15.8</td>
</tr>
<tr>
<td>Musician/U. Surprise(^2)</td>
<td>0.0</td>
</tr>
<tr>
<td>Control/U. Surprise</td>
<td>1.6</td>
</tr>
<tr>
<td>Musician/Sadness</td>
<td>0.0</td>
</tr>
<tr>
<td>Control/Sadness</td>
<td>0.5</td>
</tr>
<tr>
<td>Musician/Disgust</td>
<td>0.6</td>
</tr>
<tr>
<td>Control/Disgust</td>
<td>1.1</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>0.6</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>0.5</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>0.6</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>0.0</td>
</tr>
</tbody>
</table>

\(^1\)Lexical Identification, \(^2\)Unpleasant Surprise
### Table 10: FID\(^1\) Distribution of Total Responses for Each Emotion in %

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happiness</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>72.0</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>74.6</td>
</tr>
<tr>
<td>Musician/Pleasant Surprise</td>
<td>9.5</td>
</tr>
<tr>
<td>Control/Pleasant Surprise</td>
<td>13.1</td>
</tr>
<tr>
<td>Musician/Interest</td>
<td>11.2</td>
</tr>
<tr>
<td>Control/Interest</td>
<td>7.8</td>
</tr>
<tr>
<td>Musician/U. Surprise(^2)</td>
<td>1.3</td>
</tr>
<tr>
<td>Control/U. Surprise</td>
<td>1.6</td>
</tr>
<tr>
<td>Musician/Sadness</td>
<td>2.6</td>
</tr>
<tr>
<td>Control/Sadness</td>
<td>2.0</td>
</tr>
<tr>
<td>Musician/Disgust</td>
<td>3.0</td>
</tr>
<tr>
<td>Control/Disgust</td>
<td>0.8</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>0.0</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>0.0</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>0.4</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>0.0</td>
</tr>
</tbody>
</table>

\(^1\)Facial Identification, \(^2\)Unpleasant Surprise
Table 11: MID\(^1\) Distribution of Inaccurate Responses for Each Emotion

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happy</td>
<td>Tender</td>
<td>Sad</td>
<td>Anger</td>
<td>Fear</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>-</td>
<td>38.2</td>
<td>2.9</td>
<td>7.2</td>
<td>15.7</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>-</td>
<td>30.4</td>
<td>5.7</td>
<td>12.1</td>
<td>16.9</td>
</tr>
<tr>
<td>Musician/Tender</td>
<td>83.4</td>
<td>-</td>
<td>64.4</td>
<td>1.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Control/Tender</td>
<td>71.0</td>
<td>-</td>
<td>58.4</td>
<td>3.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Musician/Sad</td>
<td>13.5</td>
<td>55.0</td>
<td>-</td>
<td>3.4</td>
<td>18.3</td>
</tr>
<tr>
<td>Control/Sad</td>
<td>29.4</td>
<td>66.5</td>
<td>-</td>
<td>11.6</td>
<td>11.2</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>1.0</td>
<td>0.0</td>
<td>9.7</td>
<td>-</td>
<td>57.4</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>1.8</td>
<td>0.0</td>
<td>12.0</td>
<td>-</td>
<td>61.9</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>2.1</td>
<td>6.8</td>
<td>22.9</td>
<td>78.0</td>
<td>-</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>2.3</td>
<td>3.0</td>
<td>23.9</td>
<td>73.3</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^1\) Music Identification
Table 12: PID\(^1\) % Distribution of Inaccurate Responses for Each Emotion

<table>
<thead>
<tr>
<th>Group/Emotion Selected</th>
<th>Correct Emotion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happy</td>
<td>P.Surprise</td>
</tr>
<tr>
<td>Musician/Happy</td>
<td>-</td>
<td>25.0</td>
</tr>
<tr>
<td>Control/Happy</td>
<td>-</td>
<td>22.9</td>
</tr>
<tr>
<td>Musician/Pleasant Surprise</td>
<td>48.1</td>
<td>-</td>
</tr>
<tr>
<td>Control/Pleasant Surprise</td>
<td>49.5</td>
<td>-</td>
</tr>
<tr>
<td>Musician/Interest</td>
<td>22.2</td>
<td>31.2</td>
</tr>
<tr>
<td>Control/Interest</td>
<td>20.4</td>
<td>36.5</td>
</tr>
<tr>
<td>Musician/U. Surprise(^2)</td>
<td>4.9</td>
<td>38.5</td>
</tr>
<tr>
<td>Control/U. Surprise</td>
<td>4.9</td>
<td>28.1</td>
</tr>
<tr>
<td>Musician/Sadness</td>
<td>3.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Control/Sadness</td>
<td>5.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Musician/Disgust</td>
<td>18.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Control/Disgust</td>
<td>9.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Musician/Anger</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Control/Anger</td>
<td>7.8</td>
<td>2.1</td>
</tr>
<tr>
<td>Musician/Fear</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Control/Fear</td>
<td>1.9</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\(^1\)Prosodic Identification, \(^2\)Unpleasant Surprise
Table 13: Patterns of Misidentified Emotions for MID for Musicians & Controls

<table>
<thead>
<tr>
<th>Correct Emotion</th>
<th>Misidentified Emotions</th>
<th>MID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>Tender</td>
<td></td>
</tr>
<tr>
<td>Tender</td>
<td>Happiness/Sadness</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>Tender</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>Fear</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>Anger</td>
<td></td>
</tr>
</tbody>
</table>

Table 14: Patterns of Misidentified Emotions for PID, FID, and LID for Musicians & Controls

<table>
<thead>
<tr>
<th>Correct Emotion</th>
<th>Misidentified Emotions by Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PID</td>
</tr>
<tr>
<td>Happiness</td>
<td>P.Surprise</td>
</tr>
<tr>
<td>P. Surprise¹</td>
<td>Interest/U.Surprise</td>
</tr>
<tr>
<td>Interest</td>
<td>U.Surprise</td>
</tr>
<tr>
<td>U. Surprise²</td>
<td>P.Surprise/Interest/Disgust</td>
</tr>
<tr>
<td>Sadness</td>
<td>Disgust/U.Surprise</td>
</tr>
<tr>
<td>Disgust</td>
<td>Anger/Interest/U.Surprise</td>
</tr>
<tr>
<td>Anger</td>
<td>Disgust</td>
</tr>
<tr>
<td>Fear</td>
<td>Sadness</td>
</tr>
</tbody>
</table>

¹Pleasant Surprise, ²Unpleasant Surprise
### Table 15: Descriptives for Cognitive measures

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>GROUPS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians (n=58)</td>
<td>Controls (n=61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montreal Cognitive Assessment (MoCA)</td>
<td>28.31 (1.52)</td>
<td>27.92 (1.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wechsler Test of Adult Reading (WTAR)</td>
<td>112.35 (10.00)</td>
<td>103.39 (11.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven’s Advanced Progressive Matrices (RAPM)</td>
<td>19.26 (5.10)</td>
<td>18.92 (4.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Values are listed as mean (standard deviation)*

### Table 16: Group by Ethnicity by Emotion ANCOVA Statistics

<table>
<thead>
<tr>
<th>Group</th>
<th>Ethnicity</th>
<th>Emotion</th>
<th>Group x Ethnicity</th>
<th>Group x Emotion</th>
<th>Ethnicity x Emotion</th>
<th>Group x Ethnicity x Emotion</th>
</tr>
</thead>
</table>

*Note: * = significant at the p≤0.05 level
Figure 1. Group Accuracy for Emotion Identification Tasks
Figure 2. Group accuracy for Music Identification by Individual Emotion
Figure 3. MID – Musicians - Distribution of Responses for Each Emotion
Figure 4. MID – Controls – Distribution of Responses for Each Emotion
Figure 5. PID – Musicians – Distribution of Responses for Each Emotion
Figure 6. PID– Controls – Distribution of Responses for Each Emotion
Figure 7. MID – Musicians – Distributions of Inaccurate Responses
Figure 8. MID – Controls – Distributions of Inaccurate Responses
Figure 9. PID – Musicians – Distributions of Inaccurate Responses
Figure 10. PID – Controls - Distributions of Inaccurate Responses
Appendix

Table 1: Power Analysis

<table>
<thead>
<tr>
<th>Significant Effects by Study</th>
<th>Individual Task Reported Effect Size (η²_p)</th>
<th># of Participants Needed</th>
<th>Calculated Effect Size (η²_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lima &amp; Castro, n=80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>.11</td>
<td>108</td>
<td>66</td>
</tr>
<tr>
<td>Age</td>
<td>.24</td>
<td>44</td>
<td>28</td>
</tr>
<tr>
<td>Emotion (all)</td>
<td>.33</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>Thompson et al.*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1, n = 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>.31</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>Emotion (all)</td>
<td>.33</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>Study 2, n = 56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>.12</td>
<td>98</td>
<td>60</td>
</tr>
<tr>
<td>Emotion (all)</td>
<td>.33</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>Expertise x Emotion</td>
<td>.10</td>
<td>159</td>
<td>103</td>
</tr>
<tr>
<td>Individual Emotion: Sad</td>
<td>.10</td>
<td>119</td>
<td>73</td>
</tr>
<tr>
<td>Individual Emotion: Fear</td>
<td>.10</td>
<td>119</td>
<td>73</td>
</tr>
</tbody>
</table>
*Thompson et al. did not report effect sizes, so effect sizes were calculated based on reported F values

Table 2: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Experimental Emotion Discrimination Tasks

<table>
<thead>
<tr>
<th>Group</th>
<th>Shapiro-Wilk Statistic</th>
<th>Df</th>
<th>Sig.</th>
<th>Levene Statistic Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MID</td>
<td>Musicians</td>
<td>.940</td>
<td>58</td>
<td>.007*</td>
<td>0.114</td>
<td>1,117</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>.972</td>
<td>61</td>
<td>.168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LID</td>
<td>Musicians</td>
<td>.962</td>
<td>58</td>
<td>.070</td>
<td>1.980</td>
<td>1,117</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>.932</td>
<td>61</td>
<td>.002*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FID</td>
<td>Musicians</td>
<td>.950</td>
<td>58</td>
<td>.019*</td>
<td>0.108</td>
<td>1,117</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>.933</td>
<td>61</td>
<td>.002*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PID</td>
<td>Musicians</td>
<td>.972</td>
<td>58</td>
<td>.191</td>
<td>0.001</td>
<td>1,117</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>.947</td>
<td>61</td>
<td>.010*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * = significant at the p= ≤0.05 level
Table 3: Shapiro-Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Music Identification Individual Emotions

<table>
<thead>
<tr>
<th>Group</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
<th>Statistic</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MID-IE Anger Musicians</td>
<td>.960</td>
<td>58</td>
<td>.051</td>
<td>0.167</td>
<td>1,117</td>
<td>.683</td>
</tr>
<tr>
<td>Controls</td>
<td>.940</td>
<td>61</td>
<td>.005*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID-IE Fear Musicians</td>
<td>.913</td>
<td>58</td>
<td>.001*</td>
<td>0.296</td>
<td>1,117</td>
<td>.588</td>
</tr>
<tr>
<td>Controls</td>
<td>.930</td>
<td>61</td>
<td>.002*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID-IE Happy Musicians</td>
<td>.910</td>
<td>58</td>
<td>.000*</td>
<td>0.041</td>
<td>1,117</td>
<td>.839</td>
</tr>
<tr>
<td>Controls</td>
<td>.892</td>
<td>61</td>
<td>.000*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID-IE Sad Musicians</td>
<td>.942</td>
<td>58</td>
<td>.008*</td>
<td>3.036</td>
<td>1,117</td>
<td>.084</td>
</tr>
<tr>
<td>Controls</td>
<td>.952</td>
<td>61</td>
<td>.018*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID-IE Tender Musicians</td>
<td>.956</td>
<td>58</td>
<td>.034*</td>
<td>0.133</td>
<td>1,117</td>
<td>.716</td>
</tr>
<tr>
<td>Controls</td>
<td>.963</td>
<td>61</td>
<td>.059</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * = significant at the p= ≤0.05 level

Table 4: Shapiro Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Non-emotional experimental tasks

<table>
<thead>
<tr>
<th>Group</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSID Musicians</td>
<td>.966</td>
<td>58</td>
<td>.103</td>
<td>1.595</td>
<td>1,117</td>
<td>.209</td>
</tr>
<tr>
<td>Controls</td>
<td>.977</td>
<td>61</td>
<td>.290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICID Musicians</td>
<td>.898</td>
<td>58</td>
<td>.000*</td>
<td>0.146</td>
<td>1,117</td>
<td>.703</td>
</tr>
<tr>
<td>Controls</td>
<td>.914</td>
<td>61</td>
<td>.000*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPID Musicians</td>
<td>.926</td>
<td>58</td>
<td>.002*</td>
<td>0.453</td>
<td>1,117</td>
<td>.502</td>
</tr>
<tr>
<td>Controls</td>
<td>.874</td>
<td>61</td>
<td>.000*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * = significant at the p= ≤0.05 level
Table 5: Shapiro Wilk Tests of Normality and Levene’s Test of Homogeneity of Variance for Cognitive measures

<table>
<thead>
<tr>
<th>Group</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
<th>Levene’s Test Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MoCA</td>
<td>.829</td>
<td>58</td>
<td>.000*</td>
<td>2.510</td>
<td>1,117</td>
<td>.214</td>
</tr>
<tr>
<td></td>
<td>.883</td>
<td>61</td>
<td>.000*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAPM</td>
<td>.967</td>
<td>58</td>
<td>.114</td>
<td>.342</td>
<td>1,117</td>
<td>.109</td>
</tr>
<tr>
<td></td>
<td>.979</td>
<td>61</td>
<td>.372</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTAR</td>
<td>.942</td>
<td>58</td>
<td>.008*</td>
<td>2.522</td>
<td>1,117</td>
<td>.000*</td>
</tr>
<tr>
<td></td>
<td>.980</td>
<td>61</td>
<td>.431</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * = significant at the p≤0.05 level; Musicians = 1; Controls = 2

Table 6: One-Way ANOVAs comparing White and Non-White participants’ emotion perception accuracy percentage scores

<table>
<thead>
<tr>
<th>Group</th>
<th>MID</th>
<th>PID</th>
<th>LID</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Musicians</td>
<td>F = 1.393</td>
<td>F = 1.278</td>
<td>F = 1.270</td>
<td>F = .628</td>
</tr>
<tr>
<td>(n= 58, 53% White)</td>
<td>p = .243</td>
<td>p = .263</td>
<td>P = .264</td>
<td>p = .431</td>
</tr>
<tr>
<td>Controls</td>
<td>F = .426</td>
<td>F = 2.065</td>
<td>F = .220</td>
<td>F = 2.433</td>
</tr>
<tr>
<td>(n = 61, 38% White)</td>
<td>p = .516</td>
<td>p = .156</td>
<td>p = .641</td>
<td>p = .124</td>
</tr>
</tbody>
</table>
Table 7: Spearman Correlations between WTAR score and the MID or PID score for Musicians and Controls

<table>
<thead>
<tr>
<th></th>
<th>MID</th>
<th>PID</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musicians</td>
<td>$\rho = .026$</td>
<td>$\rho = .180$</td>
</tr>
<tr>
<td>Controls</td>
<td>$p = .844$</td>
<td>$p = .175$</td>
</tr>
</tbody>
</table>

Table 8: Descriptives PID, LID, and FID Individual Emotions (% correct)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>PID</th>
<th>LID</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Musicians (n=58)</td>
<td>Controls (n=61)</td>
<td>Musicians (n=58)</td>
</tr>
<tr>
<td>Happiness</td>
<td>60.92 (30.03)</td>
<td>56.28 (26.21)</td>
<td>54.31 (19.94)</td>
</tr>
<tr>
<td>P. Surprise</td>
<td>56.28 (26.21)</td>
<td>64.37 (27.12)</td>
<td>57.76 (18.26)</td>
</tr>
<tr>
<td>Interest</td>
<td>63.22 (29.08)</td>
<td>54.64 (30.45)</td>
<td>52.16 (20.02)</td>
</tr>
<tr>
<td>U. Surprise</td>
<td>52.30 (27.30)</td>
<td>41.53 (31.42)</td>
<td>60.78 (16.96)</td>
</tr>
<tr>
<td>Sadness</td>
<td>69.54 (28.12)</td>
<td>65.57 (27.87)</td>
<td>71.98 (8.21)</td>
</tr>
<tr>
<td>Disgust</td>
<td>68.97 (24.87)</td>
<td>61.20 (28.01)</td>
<td>49.57 (21.70)</td>
</tr>
<tr>
<td>Anger</td>
<td>55.74 (23.68)</td>
<td>53.01 (23.86)</td>
<td>48.28 (22.39)</td>
</tr>
<tr>
<td>Fear</td>
<td>40.23 (27.75)</td>
<td>40.98 (27.48)</td>
<td>55.60 (19.89)</td>
</tr>
</tbody>
</table>

Table 9: PID, LID, and FID ANCOVA Statistics

<table>
<thead>
<tr>
<th>Task</th>
<th>Main Effect</th>
<th>Group*Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>$F_{(1)} = 6.91$</td>
<td>$F_{(7)} = .918$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.01^*$</td>
<td>$p = 0.49$</td>
</tr>
<tr>
<td>LID</td>
<td>$F_{(1)} = 0.49$</td>
<td>$F_{(7)} = 0.00$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.84$</td>
<td>$p = 1.00$</td>
</tr>
<tr>
<td>FID</td>
<td>$F_{(1)} = 0.02$</td>
<td>$F_{(7)} = 1.09$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.89$</td>
<td>$p = 0.37$</td>
</tr>
</tbody>
</table>

Note: $^*$ = significant at the $p \leq 0.05$ level
Table 10: Multiple Wald Test Statistics for LID and FID

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>LID</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>$\chi^2(7) = 0.58, p = 1.00$</td>
<td>$\chi^2(7) = 15.95, p = 0.02^*$</td>
</tr>
<tr>
<td>Pleasant Surprise</td>
<td>$\chi^2(7) = 0.53, p = 0.99$</td>
<td>$\chi^2(7) = 11.00, p = 0.14$</td>
</tr>
<tr>
<td>Interest</td>
<td>$\chi^2(7) = 3.79, p = 0.80$</td>
<td>$\chi^2(7) = 9.03, p = 0.25$</td>
</tr>
<tr>
<td>Unpleasant Surprise</td>
<td>$\chi^2(7) = 0.42, p = 1.00$</td>
<td>$\chi^2(7) = 6.77, p = 0.45$</td>
</tr>
<tr>
<td>Sadness</td>
<td>$\chi^2(7) = 0.49, p = 1.00$</td>
<td>$\chi^2(7) = 4.55, p = 0.71$</td>
</tr>
<tr>
<td>Disgust</td>
<td>$\chi^2(7) = 0.33, p = 1.00$</td>
<td>$\chi^2(7) = 0.56, p = 0.99$</td>
</tr>
<tr>
<td>Anger</td>
<td>$\chi^2(7) = 1.00, p = 1.00$</td>
<td>$\chi^2(7) = 2.82, p = 0.90$</td>
</tr>
<tr>
<td>Fear</td>
<td>$\chi^2(7) = 2.78, p = 0.90$</td>
<td>$\chi^2(7) = 4.03, p = 0.73$</td>
</tr>
</tbody>
</table>

Note: * = significant at the $p \leq 0.05$ level

Table 11: Relationships Between Training Duration & Accuracy

<table>
<thead>
<tr>
<th>Task</th>
<th>Spearman Correlation</th>
</tr>
</thead>
</table>
| MID  | $q = 0.368$  
$p = 0.00^*$ |
| PID  | $q = 0.240$  
$p = 0.0^*$ |
| LID  | $q = 0.134$  
$p = 0.15$ |
| FID  | $q = 0.121$  
$p = 0.19$ |

Note: * = significant at the $p \leq 0.05$ level
Figure 1: Normal and detrended normal Q-Q plots for Music Identification (1=Musicians; 2=Controls)
Figure 2: Normal and detrended normal Q-Q plots for Facial Identification (1=Musicians; 2=Controls)
Figure 3: Normal and detrended normal Q-Q plots for Lexical Identification (1=Musicians; 2=Controls)
Figure 4: Normal and detrended normal Q-Q plots for Prosodic Identification (1=Musicians; 2=Controls)

Normal Q-Q Plot of Prosodic Total% for Group Code = 1.0

Normal Q-Q Plot of Prosodic Total% for Group Code = 2.0
Figure 5: Normal and detrended normal Q-Q plots for Music Identification Happy (1=Musicians; 2=Controls)
Figure 6: Normal and detrended normal Q-Q plots for Music Identification Sad (1=Musicians; 2=Controls)
MUSICAL TRAINING AND EMOTION PERCEPTION

Detrended Normal Q-Q Plot of MusSad Total%
for Group Code = 1.0

Detrended Normal Q-Q Plot of MusSad Total%
for Group Code = 2.0
Figure 7: Normal and detrended normal Q-Q plots for Music Identification Anger (1=Musicians; 2=Controls)
Figure 8: Normal and detrended normal Q-Q plots for Music Identification Fear (1=Musicians; 2=Controls)
Figure 9: Normal and detrended normal Q-Q plots for Music Identification Tender (1=Musicians; 2=Controls)

Normal Q-Q Plot of MusTender Total%  
for Group Code: 1.0

Normal Q-Q Plot of MusTender Total%  
for Group Code: 2.0
Detrended Normal Q-Q Plot of Mustender Total%

for GroupCode = 1.0

Detrended Normal Q-Q Plot of Mustender Total%

for GroupCode = 2.0
Figure 10: Group Accuracy for Prosodic Identification By Individual Emotion
Figure 11: Group accuracy for Lexical Identification By Individual Emotion

Group Accuracy for Lexical Experimental by Individual Emotion

- Happiness
- Pleasant surprise
- Interest
- Unpleasant surprise
- Sadness
- Disgust
- Anger
- Fear

Group
- Musician
- Control
Figure 12: Group accuracy for Facial Identification By Individual Emotion
Figure 13: LID Distribution of Responses for Each Emotion - Musicians
MUSICAL TRAINING AND EMOTION PERCEPTION

Figure 14: LID Distribution of Responses for Each Emotion - Controls
Figure 15: FID Distribution of Responses for Each Emotion - Musicians
Figure 16: FID Distribution of Responses for Each Emotion - Controls
Figure 17: LID Distributions of Inaccurate Responses - Musicians
Figure 18: LID Distributions of Inaccurate Responses - Controls
Figure 19: FID Distributions of Inaccurate Responses - Musicians
Figure 20: FID Distributions of Inaccurate Responses - Controls
References


Bidelman, G., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of music and language. *PLos One, 8*, E60676.


Erkkilä, J., Punkanen, M., Fachner, J., Ala-Ruona, E., Pöntiö, I., Tervaniemi, M., Vanhala, M., &


and correlates with aspects of peoples’ lives. Social Psychiatry and Psychiatric Epidemiology, 43, 140-150.
Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance:


Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., Schlaug, G. (2005). Are there pre-
existing neural, cognitive, or motoric markers for musical ability? *Brain & Cognition, 59*, 124-134.


Stewart, L. (2008). Do musicians have different brains? *Clinical Medicine, 8*, 304-308.


Zatorre (Eds.), The Cognitive Neuroscience of Music, (310-324), Oxford, United Kingdom: Oxford University Press.


