EXPRESSIVE TIMING IN NON-EXPERT MUSICAL PRODUCTION

EXPRESSIVE TIMING IN NON-EXPERT MUSICAL PRODUCTION

By HALEY E. KRAGNESS, B.S. (HONORS)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2019), Hamilton, Ontario (Psychology, Neuroscience & Behaviour)

TITLE: Expressive timing in non-expert musical production AUTHOR: Haley E. Kragness, B.S. (Honors, Schreyer Honors College, The Pennsylvania State University) SUPERVISOR: Professor Laurel J. Trainor NUMBER OF PAGES: xvi, 245

Lay Abstract

Musicians often deviate from the durations notated in their musical score, slowing and speeding over the course of a performance. In the absence of deviations, music sounds mechanical or computerized rather than expressive. Studies of performances by highly-trained musicians have identified patterns in the way these duration changes are implemented, but no previous research has investigated whether formal musical experiences and development contribute to these patterns. I developed a simple music production apparatus that enables musically untrained adults and children to perform music. I asked participants to "perform" chord sequences under different conditions and measured the amount of time they spent on each chord. I uncovered how young children and untrained adults use timing deviations to delineate musical phrase groups and to communicate musical emotions. Overall, my work offers a new way to examine expressive timing patterns and suggest that formal training alone does not fully account for these patterns.

Abstract

It is well established that musicians deviate substantially from regular timing in music performance, and numerous studies have sought to characterize the origin of different expressive deviations. However, this work has thus far been limited by the necessity of analyzing renditions produced by highly-trained adult musicians, which precludes the opportunity to ask questions about how development and formal experience might affect expressive timing.

In the present dissertation, I introduce a new paradigm for examining musical production in non-expert participants, the *musical dwell time paradigm*. In Chapters 2 and 3, I show that musically untrained adults and children as young as three years pause on phrase-final chords when self-pacing through chord sequences, mirroring the phenomenon of *phrase-final lengthening* that has been reported in expert music performance. I additionally demonstrate that by four years of age, this lengthening can be elicited by harmonic cues when other cues to phrase boundaries (metrical regularity and melodic contour) are controlled for. In Chapter 4, I show that when communicating different emotions through music, nonmusicians use expressive cues in a way that is highly consistent with expert musicians, and that there is striking similarity across participants despite a wide range of musical training. Finally, in Chapter 5, I demonstrate that children as young as 5 years olds' performances mirror adults' in their use of timing and loudness cues, and that their renditions become more adult-like by 7 years. Altogether, these findings corroborate previous claims that musically untrained adults are “listening experts” with substantial musical knowledge, extend these results to show that in performance musically untrained adults use timing and loudness similarly as expert musicians to delineate phrases and express emotions, and show that some elements are in place by early childhood. Overall, the *musical dwell time paradigm* offers a new, highly flexible method for examining musical production in participants with a wide variety of musical training.

Acknowledgements

I feel extremely lucky to have had a strong network of supporters throughout graduate school. Laurel, thank you for welcoming me into your lab (or rather, your labs!). Your mentorship has been invaluable. You created a lab environment in which we all build each other up. You always encouraged me to pursue research questions that spark my own curiosity, and to nurture my passions outside of the lab. Graduate school has given me opportunities beyond what I could have dreamed, and that is all thanks to you.

Matthew and Louis, your support and contributions throughout this process were invaluable. Thank you for your insightful comments, your generosity with your time, and for supporting me in my desire to gain teaching experience the past two years. I especially owe a debt of gratitude to Matthew for sharing your compositional expertise in stimulus creation, which no doubt greatly improved each chapter in this work.

To all my lab-mates – Rayna, Kate, Chris, Laura, Andrew, Sima, Sarah, Hector, David, Chantal, and Erica. Thank you all for being great sounding boards and friends. Science wouldn’t be nearly as much fun without all of you! And to my LIVELab-mates – Dan, Dave, Susan, and Dobri, I am inspired every day by your hard work and commitment to excellence. I have had such a wonderful time working with each of you.

To Elaine, thank you for sharing your boundless optimism and energy, and for inviting me to be a part of your life outside of McMaster. Playing music and sharing warm drinks with you over the years was a welcome reprieve from the stresses of daily life. Susan, I am so grateful for the opportunity to have gotten to know you better. Planning events and sharing *cold* drinks with you has been an absolute joy, and I treasure the laughs (and tears) we shared. You are an inspiration to me, and I can't wait to celebrate *your* academic milestones in the future! And Dave, I am eternally grateful for your assistance programming my experiments. I am especially thankful for your patience with my questions and clarifications, which helped me to learn and grow as a programmer.

I am grateful for the new friends I have gained through McMaster. To my cohort-mates – Lorraine, Blair, Jessica, and Andrew – and my friends from other cohorts and departments – especially Tammy, Laura, Fred, Jay, and Kendall. You each supported me, encouraged me, challenged me, and reminded me to relax once in a while. I am lucky to have forged lasting friendships with each of you, and to share softball and soccer games, drinks, and laughs through our failures and our triumphs.

I was fortunate to have wonderful mentors in both music and psychology. I thank Dave Cooper and Tom Cody, who taught me music theory and indulged me in my very first conversations about music and the mind. I especially thank Dr. David Rosenbaum, who introduced me to research psychology and taught me how to think deeply and empirically, and Dr. Sandra Trehub, who introduced me to music cognition research and encouraged me to pursue my research journey in Canada, for which I will be eternally grateful.

To my parents and grandparents, I am so thankful for all of your support throughout the years. You always encourage me to pursue my passions in life, even when they take me across international borders, and have provided me with countless opportunities to foster my natural curiosity. To Dane, my first jam partner – I am so lucky to have a brother who shares my love of sound and music, and who was willing to join me in so many wild and crazy projects over the years. And to Matt – thank you for being my programming consultant, my proofreader, my sounding board, and my camping buddy, and my best friend. I couldn’t ask for a more loving and supportive partner.

Declaration of Academic Achievement

This thesis consists of six chapters in total and is presented in the format of a sandwich thesis. The thesis consists of a general introduction, four empirical chapters, and a general discussion. Two of the four empirical chapters are published in peer-reviewed scientific journals. The remaining two empirical chapters are either currently under review or in preparation for submission to a scientific journal.

The author of this thesis is the primary author of all six chapters. I developed and implemented the *musical dwell time paradigm*. I conceptualized and designed each experiment in consultation with Laurel J. Trainor, who co-authored the four data chapters. For each study, I was the primary individual responsible for creating stimuli, collecting data, supervising data collection by undergraduate students, analyzing the data, and preparing the manuscripts. The *adapted musical dwell time paradigm* presented in Chapters 4 and 5 was implemented with technological assistance of research staff Dave Thompson. Undergraduate thesis student Ammaarah M. Baksh assisted with development, pilot testing, and data collection with the musical emotions games described in Chapter 5, and is thus a co-author on that chapter.

This thesis includes two published research articles with permission from the American Psychological Association:

Kragness, H. E. & Trainor, L. J. (2016) Listeners lengthen phrase boundaries in self-paced music. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1676-1686. http://dx.doi.org/10.1037/xhp0000245

Kragness, H. E. & Trainor, L. J. (2018). Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues. *Developmental Psychology, 54*(5), 842-456. http://dx.doi.org/10.1037/dev0000405

Table of Contents

List of Figures

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

List of Tables

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

List of all Abbreviations and Symbols

η_p^2: partial eta squared

η_G^2: generalized eta squared

AB: atonal boundary

AN: atonal nonboundary

ANOVA: analysis of variance

BWV: Bach-Werke-Verzeichnis catalog

CI: confidence interval

DSB: Digit Span Backward

DSF: Digit Span Forward

d_z: Cohen's d_z

EEG: electroencephalography

ERAN: early right anterior negativity

ERP: event-related potential

F: F-test statistic

Gold-MSI: Goldsmiths Musical Sophistication Index

IOI: inter-onset interval

M: mean

MIDI: Musical Instrument Digital Interface

N/n: sample size

nPVI: normalized Pairwise Variability Index

ms: millisecond

p: p-value

PPVT: Peabody Picture Vocabulary Test

r: Pearson's r

rho: Spearman's rho

SD: standard deviation

SEM: standard error of the mean

t: *t*-test statistic

TAAS: Test of Auditory Analysis Skills

tau: Kendall’s tau

TB: tonal boundary

TEC: Test of Emotion Comprehension

TN: tonal nonboundary

W: Shapiro-Wilk *W*

WISC-IV: Weschler Intelligence Scale for Children IV

Z: *z*-score

Chapter 1: General Introduction

1.0 Introduction

Temporal processing is vital to nearly every action and interaction in our daily lives. Knowing when past events have occurred, when future events will occur, and initiating responses at the correct time are critical for interacting with a dynamic environment (e.g., Huron, 2006). Numerous common behaviors – such as holding a conversation, hitting a thrown ball with a baseball bat, or tapping your toe to the beat of a song – require precise temporal predictions (De Ruiter, Mitterer, & Enfield, 2006; Ranganathan & Carlton, 2007; Large & Jones, 1999). The importance of temporal processing can especially be observed when it has been disrupted, such as in Parkinson's patients' difficulty completing motor tasks that otherwise seem trivial (Hausdorff, 2009). Over the past several decades, a wealth of research has sought to characterize the mechanisms of temporal processing in the brain and in behavior (e.g., Allan, 1979; Fujioka, Trainor, Large, & Ross, 2012; Grahn & Brett, 2007; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015; Merchant, Harrington, & Meck, 2013; Povel & Essens, 1985; Rosenbaum & Collyer, 1998).

Temporal processing is critical in the domain of music. Music can be conceived of as hierarchically structured sequences of sound events that unfold across time. Though the nature of these sound events varies widely across cultures (for example, differing pitch patterns and instrumentation), the presence of an underlying, temporally regular pulse is thought to be a feature of every musical system worldwide (Nettl, 2000; Savage, Brown, Sakai, & Currie, 2015). This regular pulse is referred to as the *beat.* Music is often structured into repeating two- or three-beat patterns of strong (emphasized) and weak (de-emphasized) beats, which define the *meter*. Perceptions of beat and meter are thought to arise through a combination of acoustic input and neural dynamics (Large & Jones, 1999; Tal, et al., 2017) and are additionally affected by listeners' expectations and experiences (Manning & Schutz, 2013; Nozaradan, Peretz, Missal, & Mouraux, 2011; Phillips-Silver & Trainor, 2007). Music's temporally regular structure allows for the generation of precise timing expectations, and thus has important implications for musical performance and engagement. This regularity, along with the well-established link between auditory and motor systems (e.g., Fujioka, Trainor, Large, & Ross, 2012; Grahn & Brett, 2007; Phillips-Silver & Trainor; 2005, 2007; Patel & Iverson, 2014; Zatorre, Chen, & Penhune, 2007) makes it possible to clap or dance "in time" with music, and for multiple musicians to play together in synchrony. It has been proposed that the social cohesion that is experienced by playing and moving together in time to the beat may have facilitated the evolution of music in human history (Brown, 2000; Huron, 2003; Trainor, 2015).

In the Western musical tradition, note durations are indicated in the musical score as sub-divisions or multiples of the beat level. Given a specific meter and speed (*tempo*), it is possible to determine the exact duration that is indicated by each note in a musical score. However, musicians often deviate from these durations, shortening or lengthening notes at various points in performance. Despite this, listeners are adept at tracking the underlying beat (Large & Palmer, 2002), and perceive music that is produced with exact timing rather than naturally produced timing to be relatively unpleasant (Istók, Tervaniemi, Friberg, & Seifert, 2008; Kendall & Carterette, 1990; Palmer, 1996a) and unexpressive (Broughton & Stevens, 2009).

What is the origin of timing deviations in music performance? It is likely that there are contributions of both implicit processes, such as perceptual biases, and explicit processes, such as communicative, expressive intentions (Juslin, 2003). Some deviations must come from random variability – for instance, noise in internal time-tracking mechanisms and representations, as well as in the motor system (Faisal, Selen, & Wolpert, 2008; Gilden, 2001; Wing & Kristofferson, 1973). Other deviations, while systematic, appear to arise as a result of motor planning rather than expressive intent, such as the phenomenon of motor "chunking." Just as breaking a long sequence into component parts facilitates perceptual organization (Miller, 1956), individuals tend to divide long motor sequences (such as patterns of tapping with different fingers) into shorter sequences with pauses between them (Povel & Collard, 1982; Rosenbaum, Kenny, & Derr, 1983; Sakai, Kitaguchi, & Hikosaka, 2003; van Vugt, Jabusch, & Altenmüller, 2012). Further, timing deviations are more difficult to detect at positions where they usually occur in performance (Drake, 1993; Repp, 1992b, 1998a, 1998b), which has led to speculation that grouping may "warp" time perception, leading to perceptual biases that affect both listener judgments and performances (Penel & Drake, 2004).

Performers also intentionally deviate from mechanical timing to achieve different expressive goals. Across multiple renditions of the same piece, pianists tend to slow and speed at consistent positions (Gabrielsson, 1999; Palmer, 1996b; Seashore, 1938), and when asked to play with exaggerated expression, such patterns are enhanced (Palmer, 1989). Performers are also subject to genre- and era-specific norms that influence performance. Analyzing recorded performances of early 19th century classical songs composed by Schubert, Timmers (2007) observed that timing deviations tended to diminish over the course of the twentieth century, reflecting a change in stylistic norms of the era, and contemporary popular music often incorporates isochronous beats controlled by a computer. In performances of classical music, the melodic line tends to anticipate the underlying pulse (Palmer, 1996b), while the melodic line slightly lags the underlying pulse in certain styles of jazz performance (Friberg & Sundström, 2002). Still other deviations may reflect a performer’s desire to express their individuality and creativity in a performance (Repp, 1997, 1998).

Previous investigations have been limited by the necessity of testing only highly-trained musicians in music production tasks. In the subsequent research chapters, I offer a new paradigm for investigating music production and its development in musically untrained individuals, including adults and children, the *musical dwell time paradigm*. In this method, participants press a computer key or electronic piano key to initiate the onset of each note or chord in a musical sequence. The timing of each key press is recorded. In this way, it is possible to examine timing patterns in music production of non-expert musicians.

In the present work, I use the musical dwell time paradigm to investigate timing deviations in two contexts: musical structure and musical emotions. In the following sections, I will discuss what is known about timing deviations in more depth. I additionally provide an overview of what is known about the role of development and formal music training in these areas.

2.0 Timing and Musical Structure

Phrase-final lengthening, the tendency for performers to lengthen the duration of inter-beat intervals as they approach the end of a musical phrase, is among the most well-documented systematic timing variations in music performance. Though the concept of a musical phrase is often referred to, it is not well-defined (Spiro, 2007). Perhaps the most common description refers to the level of the phrase in the musical hierarchy. At the lowest level, notes are joined together into *motives*, short patterns that are often repeated. Motives are joined to form sub-phrases and phrases, which are in turn joined together to form the full structure of a musical piece. Thus, the phrase represents an intermediate level in the hierarchical structure of a piece. Additionally, phrases are often described as conveying a single musical idea with a beginning and an end, much like a phrase in spoken language (Spiro, 2007).

Phrase structure is often notated in the musical score, but there is substantial evidence that phrases are also perceived and represented psychologically as units. Participants are consistent in their judgments of phrase boundary locations based on listening alone (Clarke & Krumhansl, 1990; Deliege, 1987; Palmer & Krumhansl, 1987; Peretz, 1989). Further evidence for the phrase as a psychological unit comes from Sloboda and Gregory (1980), who observed that when participants heard a musical sequence with randomly placed clicks during the sequence, they tended to misremember the clicks as having occured closer to the phrase boundary than they did (*click migration*). Furthermore, a pair of notes that were heard as occuring *within* a phrase were better remembered than when the same notes were heard as *crossing* a phrase boundary (Tan, Aiello, & Bever, 1981), suggesting that within-phrase pairs were represented as parts of the same unit in memory.

Numerous investigations spanning many different instruments have observed lengthening at the ends of phrases (Kendall & Carterette, 1990; Palmer, 1989, 1996b; Repp, 1992, 1998; Todd, 1985). Despite widespread documentation of phrase-final lengthening, there is little agreement about its source. Among the earliest and most widely-accepted explanations is that performers use elongated durations to emphasize the structure of a piece for naïve listeners (e.g., Clarke, 1985; Palmer, 1989; Seashore, 1936; Shaffer, Clarke, & Todd, 1985). Later, a second (but not mutually exclusive) explanation was proposed – that phrase-final lengthening mimics the constant braking force in human kinematics (Feldman, Epstein, & Richards, 1992; Friberg & Sundberg, 1999; Sundberg & Verillo, 1980). This interpretation has been criticized because it fails to account for effects of note density and rhythmic structure on slowing (Honing, 2005). Still others have observed that pianists slow down on phrase boundaries even when they intend to play mechanically, which has been taken to suggest that there is a bias to perceive group boundaries as shorter than they are, leading to lengthening in production (Penel & Drake, 1998, 2004), although the origin of the perceptual bias is not specified. The extent to which each mechanism contributes under different performance conditions has not yet been elucidated.

To date, studies of expressive musical production have been limited to testing highly-trained musicians. As a result, questions about the role of musical experience – including informal exposure to music and formal music training – have remained largely unexplored.

2.1 Development, musical structure, and timing. From birth, timing plays an important role in infants' experience of auditory stimuli. Newborns are able to differentiate between simple rhythmic patterns (Demany, McKenzie, & Vurpillot, 1977), and one study suggests newborns may be able to track patterns of strong and weak beats (Winkler, Háden, Ladinig, Sziller, & Honing, 2009). By two months, they can discriminate different tempos (Baruch & Drake, 1997), and by four months, can detect meter-disrupting melodic changes (Hannon & Trehub, 2005). By 12 months, infants show greater sensitivity to disruptions in musical meters that are common in their culture than in those that are uncommon

(Hannon, Soley, & Levine, 2011). Thus, sensitivity to temporal information in music, such as tempo and metrical hierarchies, is present very early in childhood.

With regards to timing and grouping, infants as young as 4 ½ months prefer to listen to music with pauses that are inserted *between* rather than *within* phrases, suggesting that the phrase is a psychological unit even for young infants (Krumhansl & Jusczyk, 1990). However, no studies have examined the role of timing specifically in infants' or young children's understanding of musical phrase units.

2.2 Formal training, musical structure, and timing. There is a large body of work that shows that musicians have superior temporal perception and production abilities compared to nonmusicians. For instance, musicians can detect smaller deviations from regular time (Madison & Merker, 2002; Rammsayer, Buttkus, & Altenmüller, 2012; Rammsayer & Altenmüller, 2006) and perform sensorimotor tasks, such as tapping to a beat, with more precise timing compared to nonmusicians (Matthews, Thibodeau, Gunther, & Penhune, 2016; Repp, 2010). Perhaps not surprisingly, superiority in sensorimotor timing tasks seems to be enhanced for percussionists (Butler & Trainor, 2015; Cameron & Grahn, 2014; Krause, Pollok, & Schnitzler, 2010; Manning & Schutz, 2016), particularly when using a familiar effector like a drum stick (Manning, Harris, & Schutz, 2017). Of course, it is impossible to determine the causal direction in this literature; nevertheless, in many contexts, musicians display superior timing abilities.

To what extent does musical training facilitate phrase-final lengthening? Nonmusicians appear to have similar behavioral and neurophysiological responses to phrasing as musicians. For instance, like musicians, nonmusicians find it more difficult to detect lengthened notes if they occur in a phrase-final position rather than other positions in a phrase (Repp, 1999). This suggests that listeners *expect* lengthening at locations where lengthening usually occurs, and that musical training is not the source of this expectation. The extent to which lifelong exposure to Western music accounts for this expectation has not been explored. Converging evidence comes from electrophysiological measurements of participants while listening to “phrased” or “unphrased” passages. “Phrased” passages were found to elicit a positive-going event-related potential shortly after the onset of a phrase boundary, concentrated in front and central areas and distributed along the midline (Glushko, Steinhauer, DePriest, & Koelsch, 2016). This wave (called the “music Closure Positive Shift,” or “music CPS”) manifested similarly in nonmusicians and musical experts.

Conversely, at least two studies have found that for individuals with musical training, the perceptual organization conferred by phrase boundaries is enhanced. Drake, Penel, and Bigand (2000) asked musicians and nonmusicians to tap along to mechanical and expressive versions of the same piece. Musicians tended to spontaneously tap at a higher level of the musical hierarchy than nonmusicians, which the authors interpreted as evidence for musicians’ tendency to perceptually organize the music over a longer time span than nonmusicians.

Interestingly, both groups tapped at higher hierarchical levels when tapping along with *expressive* versions of the music than *mechanical* versions, suggesting that the expressive timing plays a role in facilitating perceptual organization for those with and without music training. Chiappe and Schmuckler (1997) found that while trained musicians' memory for musical sequences was affected by the phrasing of the passage, nonmusicians' memory was not, suggesting that there was enhanced processing of the phrase structure in musicians. Overall, while nonmusicians certainly perceive phrase structure in music and are susceptible to perceptual biases related to phrase structure, there also appear to be training-related differences which emerge in certain contexts.

3.0 Timing and Emotional Expression

A second way that timing relates to musical expression is in the communication of emotions. There is no single consensus for how to define emotion as a concept. For the purposes of the present work, an emotion can be conceived of as the sum of the physiological, behavioral, and cognitive responses that tend to co-occur in response to significant events in the environment (Izard, 2009; Juslin & Sloboda, 2013). While there is disagreement about whether emotions should be modeled discretely or dimensionally (e.g., Eerola & Vuoskoski, 2013; Izard, 2009; Russell, 2003), a two-dimensional approach appears to be appropriate for music, as certain musical features tend to be associated with either *valence* (positive or negative) or *arousal* (high or low).

Though it is generally agreed that music can *induce* emotion, the present work is primarily concerned with the *expression* of emotion, which is distinct from induction (for a review, see Schubert, 2013). By adulthood, there is general agreement as to which emotion is conveyed in particular performances (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Juslin & Laukka, 2003; Mohn, Argstatter, & Wilker, 2011; Vieillard, et al., 2008), with the highest agreement for so-called "basic emotions" (including *happiness/joy, sadness, tenderness, and anger/fear*), which each belong to different quadrants of a valence-by-arousal circumplex. A number of musical cues have been implicated in listeners' determinations of intended emotions, including pitch level, timbre, mode, loudness, variations in loudness, and articulation (see Gabrielsson & Lindström, 2010; Juslin & Timmers, 2010 for reviews).

Though many expressive cues are incorporated into the musical score, performers can additionally control microvariations in loudness, timing, and timbre. Evidence that performers exercise this control is found in studies in which performers are asked to play the same score multiple times with the goal of communicating a different emotion each time. Listeners can reliably decode the emotions expressed in such performances, suggesting that performers' communicative intentions play a role in the listeners' interpretations above and beyond the content of the musical score (Behrens & Green, 1993; Juslin, 1997; Juslin & Madison, 1999; Kotylar & Morozov, 1976; Senju & Ohgushi, 1987).

Of all the cues that listeners might use to decode the emotional intent of a performance (which could include mode, instrumentation, loudness, and pitch, among others), performers have particular control over temporal features. Temporal features are most often associated with distinguishing between *low*- and *high-arousal* emotions. For example, performances typically convey high-arousal emotions, such has *happiness* and *anger* by incorporating fast tempi, low tempo variability, disconnected articulations, and high sound levels. In contrast, performances of low-arousal emotions (such as *sadness* and *peacefulness*) typically involve slow tempi, high tempo variability, connected articulations, and low sound levels (Juslin & Timmers, 2010).

Performer-controlled temporal cues to emotions also appear to be among the most salient to listeners. Sensitivity to tempo as a cue for musical emotions appears to emerge earlier than sensitivity to mode (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001), and is interpreted similarly across musical cultures (Balkwill, Thompson, & Matsunaga, 2004). Timing is related to emotional expression in domains beyond music, such as speech, perhaps explaining why it is such a powerful cue. Indeed, rate manipulations appear to affect emotion judgments similarly in music and speech (Ilie & Thompson, 2006), as well as across different languages (Thompson & Balkwill, 2006).

Though there is a large body of work examining how emotions are communicated from performers to listeners, “encoders” (performers) and “decoders” (listeners) are usually mismatched in terms of musical expertise. Most studies of musical emotions have used performances from highly trained musicians as stimuli, either obtained from performances in the lab or professional recordings. Listeners, on the other hand, can range from infants to musically untrained or trained adults. While some studies have focused on the role of cultural exposure to expressive cue use, little work has examined incidental exposure across development, and no studies have systematically examined differences between performers with different levels of musical training.

3.1 Development, musical emotions, and timing. Sensitivity to some expressive cues in music appear to be present in infancy. Infants display different behaviors and physiological responses to caregivers' lullabies versus play songs (Cirelli, Jurewicz, & Trehub, 2018; Rock, Trainor, & Addison, 1999), suggesting they are sensitive to cues that caregivers employ to differentiate between expressive goals. In addition, there is evidence that infants as young as 5 months can discriminate between musical excerpts that are judged by adults to convey *happy* and *sad* emotions (Flom & Pick, 2012; Flom, Gentile, & Pick, 2008), although it is not clear whether the infants *associate* the expressive cues in the music with each emotion, nor which expressive cues are important for such discrimination.

Studies with older children have examined emotion identification and singing production. By about 4 or 5 years, children rate fast passages as happier than slow passages (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Mote, 2011). Similarly, children at this age sing faster when they are asked to sing a song to make an experimenter happy compared to singing to make an experimenter sad, and older children manipulate tempo to a greater extent than younger children (Adachi & Trehub, 2000; Adachi, Trehub, & Abe, 2004). One major limitation of these studies is their reliance on the emotions *happy* and *sad*, which differ in both arousal (high vs. low activity) and valence (positive vs. negative). While temporal cues mainly seem to be associated with the *arousal* dimension in adults' judgments of musical emotions, thus far, there has been little exploration of the development of this association.

3.2 Formal training, musical emotions, and timing. For musicians to effectively communicate musical emotions, there must be substantial overlap between cues used by performers and their audience. Indeed, there is no evidence thus far that formal music training discernably improves emotional decoding of musical performances (Bigand et al., 2005; Juslin, 1997). However, there are many remaining questions about the role that musical training might play in musical emotion production. Production tasks may provide a window into more subtle differences than perceptual tasks. For example, at least one report noted that a novice guitarist was less effective at communicating basic emotions than his more experienced counterparts (Juslin & Laukka, 2000). Furthermore, previous work has found that student musicians with significant "informal" experience (such as performing in casual rock bands) outperformed those who had only formal classical training on an ear-playing task (Woody & Lehmann, 2010), suggesting that experience in informal music-playing settings can confer an advantage in some contexts. However, potential effects of musical training on emotional expression have not been formally investigated.

4.0 The Musical Dwell Time Paradigm

Technological advances have made it possible to simulate musical production without the barrier of years of training on a musical instrument. For example, the web application TouchPianist allows users to "perform" renditions of popular piano pieces, such as Beethoven's "Moonlight Sonata," simply by tapping the screen of a tablet to elicit each vertical onset in the piece (Bozkurt, 2015a). The application has been hugely popular, recording 2 million "plays" in the first two weeks after release (Bozkurt, 2015b). Music video games such as *Guitar Hero*, which simulate music playing, were among the most popular and lucrative games released in the 2000's. Notably, a single music simulation game, *Guitar Hero: Aerosmith*, reportly generated more revenue for Aerosmith than any single Aerosmith album (Remo, 2008).

Despite the apparent enthusiasm for simulating music production amongst the general public, music simulation as an avenue for experimental work has barely begun to be explored. Bresin and Friberg (2011) developed a slider apparatus that allowed participants to increase or decrease the values of seven different musical parameters to communicate different emotional expressions. They found that participants' use of each parameter was broadly consistent with patterns of expressive cues that have been previously observed in naturalistic performances, suggesting that even in this impoverished laboratory task, emotional expression appeared to be possible for the participants. However, even though the slider apparatus did not require musical expertise to use, they recruited only musical experts to participate.

In the current dissertation, I present a series of experiments that investigate non-musicians' musical production using a new paradigm, the *musical dwell time paradigm*. This paradigm was adapted from Hard, Recchia, and Tversky's (2011) dwell time paradigm for probing action segmentation. In their paradigm, participants self-paced through still frames lifted from a video of a person performing a series of daily tasks, such as cleaning their room. Participants' "dwell times" (which represented looking times in this case) were longer for frames that represented an event boundary (e.g., finishing making a bed) than frames that occurred within an event (e.g., tucking in the corners of the bedsheet), and that there was a hierarchical pattern such that dwell times on coarse-grained boundaries were longer than dwell times on fine-grained boundaries. Some preliminary work has suggested that this paradigm can reveal sensitivity to segmental structure in action sequences in preschool children and infants (Meyer, Baldwin, & Sage, 2011; Sage, Ross, & Baldwin, 2012).

In the *musical dwell time paradigm* developed here, participants self-pace through notes or chords in a musical sequence. Each key press elicits the onset of the next sound event, and the duration of each event is tracked. The paradigm is simple and requires no prior formal experience with music. It is also highly flexible, and can be easily adapted for different goals. For example, participants can be asked to play a familiar piece in different ways, or they can be asked to self-pace through a completely novel sequence. In addition, it is possible to alter the affordances of the apparatus such that participants can control additional aspects of each sound event, such as sound offsets and loudness. The musical dwell time paradigm presents the opportunity to examine numerous aspects of musical timing with minimal task demands.

5.0 Summary of Thesis Contributions

In the present work, I use the musical dwell time paradigm to explore musical timing as it relates to musical phrase segmentation and to emotional expression, and investigate the potential role of experience, including formal training and development.

In Chapters 2 and 3, I use the musical dwell time paradigm to investigate the relation between expressive timing and musical phrase structure, specifically with regards to harmonic cues to musical phrase structure. In Chapter 2, Experiment 1, I show that adult participants spontaneously employ phrase-final lengthening when self-pacing through a chord sequence. In Experiments 2 and 3, I show that this lengthening can be elicited by harmonic cues after controlling for other phrase boundary cues, including meter and melodic contour, and that this effect is present for those who have never taken any formal music lessons.

In Chapter 3, Experiment 1, I show that 3-year-old participants spontaneously demonstrate phrase-final lengthening in the context of metrically regular musical stimuli (Bach chorales). In Experiment 2, I show that for harmonic sequences that control for metrical and melodic contour cues, 3-year-olds do not demonstrate phrase-final lengthening, but 4-year-olds do. Thus, it appears that while 3-year-olds are sensitive to phrase boundaries in naturalistic music and intuitively increase event durations at phrase boundaries, specific sensitivity to harmonic cues develops significantly between 3 and 4 years. Together, these experiments show that phrase-final lengthening is elicited in an impoverished laboratory music production task, in a non-communicative (e.g., non-performance) setting, in nonmusicians as young as preschool age.

In Chapters 4 and 5, I adapted the musical dwell time paradigm to investigate musically untrained adults’ and young childrens’ performances of different musical emotions. In Chapter 4, I show that adults primarily use tempo and loudness cues to differentiate between emotions in music. Furthermore, I observed remarkable similarity in the means and ranges of the chosen values across individuals with a wide range of experience with formal music training. To my knowledge, this is the first experiment to systematically examine the role of formal training on expressive music production.

Finally, in Chapter 5, I asked 3-, 5-, and 7-year-old children to communicate emotions using the same paradigm described in Chapter 4. By five years, children used tempo and loudness to distinguish between high-arousal and low-arousal emotions, as do adults, though they employed a smaller range of values. Seven-year-old children's performances were qualitatively similar to those of adults', and the range of values used was intermediate to 5-year-olds' and adults' performances. Three-year-olds did not seem to be able to manipulate any of the expressive cues that were investigated. This work reinforces the importance of timing in children's understanding of musical emotions, and additionally suggests that loudness plays a significant and underexplored role.

Together, the data chapters of this thesis show that adults and children associate timing with phrase structure and use timing patterns to communicate emotions in music. Further, they are the first studies to demonstrate these patterns in a timing production task with individuals who have not had formal music training.

References

Adachi, M., & Trehub, S. E. (2000). Decoding the expressive intentions in children's songs. *Music Perception: An Interdisciplinary Journal, 18*(2), 213-224. http://dx.doi.org/10.2307/40285909

Adachi, M., Trehub, S. E., & Abe, J. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46*(4), 322-336. http://dx.doi.org/10.1111/j.1468-5584.2004.00264.x

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics, 26*(5), 340-435. http://dx.doi.org/10.3758/BF03204158

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Pscyhological Research, 46*(4), 337-349. http://dx.doi.org/10.1111/j.1468-5584.2004.00265.x

Baruch, C., & Drake, C. (1997). Tempo discrimination in infants. *Infant Behavior and Development, 20*(4), 573-577. http://dx.doi.org/10.1016/S0163-6383(97)90049-7

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20-33. http://dx.doi.org/10.1177/030753569302100102

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional response to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*(8), 1113-1139. http://dx.doi.org/10.1080/02699930500204250

Bozkurt, B. (2015). TouchPianist (1.3.0). *[Mobile application software]*, Retrieved from https://play.google.com/store/apps/details?id=com.earslap.touchpianistapp.

Bozkurt, B. (2015, May 24). Touch Pianist - Magical Piano In Your Web Browser. *[Blog post]*, Retrieved from http://earslap.com/page/touch-pianist.html.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex, 27*(9), 1068-1081. http://dx.doi.org/10.1016/j.cortex.2011.05.009

Broughton, M. & Stevens, C. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music, 37*(2), 137 – 153. http://dx.doi.org/10.1177/0305735608094511

Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. In S. Tonneau, & N. S. Thompson, *Perspectives in Ethology, Volume 13* (pp. 231-281). Boston, MA: Springer. http://dx.doi.org/10.1007/978-1-4615-1221-9_9

Butler, B. E., & Trainor, L. J. (2015). The musician redefined: A behavioural assessment of rhythm perception in professional club DJs. *Timing and Time Perception, 3(1-2)*, 88-101 http://dx.doi.org/10.1163/22134468-03002041

Cameron, D. J., & Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience, 8.* http://dx.doi.org/10.3389/fnhuman.2014.01003

Chiappe, P., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review, 4*(2), 254-259. http://dx.doi.org/10.3758/BF03209402

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2018, March). Musical rhythms in infancy: Social and emotional effects. *Paper presented at the Cognitive Neuroscience Society Annual Meeting.* Boston, MA.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. & West, *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception: An Interdisciplinary Journal, 7*(3), 213-251. http://dx.doi.org/10.2307/40285462

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1-B10. http://dx.doi.org/10.1016/s0010-0277(00)00136-0

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515-535. http://dx.doi.org/10.1353/lang.2006.0130

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl adn Jackendoff's grouping preference rules. *Music Perception, 4*(4), 325-360. http://dx.doi.org/10.2307/40285378

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature, 266*(5604), 718-719. http://dx.doi.org/10.1038/266718a0

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society, 31*(2), 107-110. http://dx.doi.org/10.3758/BF03334153

Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception: An Interdisciplinary Journal, 18*(1), 1-23. http://dx.doi.org/10.2307/40285899

Eerola, T. & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal, 30*(3), 307-340. http://dx.doi.org/10.1525/mp.2012.30.3.307

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*(4), 292-303. http://dx.doi.org/10.1038/nrn2258

Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception, 10*(2), 185-204. http://dx.doi.org/10.2307/40285606

Flom, R., & Pick, A. D. (2012). Dynamics of infant habituation: Infants' discrimination of musical excerpts. *Infant Behavior and Development, 35*(4), 697-704. http://dx.doi.org/10.1016/j.infbeh.2012.07.022

Flom, R., Gentile, D. A., & Pick, A. D. (2008). Infants' discrimination of happy and sad music. *Infant Behavior and Development, 31*(4), 716-728. http://dx.doi.org/10.1016/j.infbeh.2008.04.004

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America, 105*(3), 1469-1484. http://dx.doi.org/10.1121/1.426687

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception, 19*(3), 333-349. http://dx.doi.org/1 0.1525/mp.2002.19.3.333

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds in represented in neuromagnetic beta oscillations. *Journal of Neuroscience, 32*(5), 1791-1802. http://dx.doi.org/10.1523/JNEUROSCI.4107-11.2012

Gabrielsson, A. (1999). The performance of music. In D. Deutsch, *The Psychology of Music (2nd edition)* (pp. 501-602). Academic Press.

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In K. R. R. J. Davidson, *Series in affective science. Handbook of affective sciences* (pp. 503-534). New York, NY, US: Oxford University Press.

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

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108*(1), 33-56. http://dx.doi.org/10.1037//0033-295X.108.1.33

Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *PLoS ONE, 11*(5), e0155300. http://dx.doi.org/10.1371/journal.pone.0155300

Grahn, J. A. & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience, 19*(5), 893-906. http://dx.doi.org/10.1162/jocn.2007.19.5.893

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science, 16*(1), 48-55. http://dx.doi.org/10.1111/j.0956-7976.2005.00779.x

Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science*, *14*(4), 865-872. http://dx.doi.org/10.1111/j.1467-7686.2011.01036.x

Hard, B., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586-604. http://dx.doi.org/10.1037/a0024310

Hausdorff, J. M. (2009). Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos*, *19*(2), 026113. http://dx.doi.org/10.1063/1.3147408

Honing, H. (2005). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception: An Interdisciplinary Journal*, *23*(1), 79-85. http://dx.doi.org/10.1525/mp.2005.23.1.79

Huron, D. (2003). Is music an evolutionary adaptation? Chapter 5. In I. Peretz, & R. J. Zatorre, *The Cognitive Neuroscience of Music* (pp. 57-75). Oxford: Oxford University Press.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, Massachusetts: MIT Press.

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

Istók, E., Friberg, A., Huotilainen, M., & Tervaniemi, M. (2013). Expressive timing facilitates the neural processing of boundaries in music: Evidence from event-related potentials. *PLoS ONE, 8*(1), e55150. http://dx.doi.org/10.1371/journal.pone.0055150

Istók, E., Tervaniemi, M., Friberg, A., & Seifert, U. (2008, January). Effects of timing cues in music performances on auditory grouping and pleasantness judgments. *The 10th International Conference on Music Perception and Cognition.* Sapporo, Japan.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60*(1), 1-25. http://dx.doi.org/annurev.psych.60.110707.163539

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception: An Interdiscliplinary Journal, 14*(4), 383-418. http://dx.doi.org/10.2307/40285731

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1797-1812. http://dx.doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music, 31*(3), 273-302. http://dx.doi.org/10.1177/03057356030313003

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientiae, 4*(2), 151-183. http://dx.doi.org/10.1177/102986490000400202

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814. http://dx.doi.org/10.1037/0033-2090.129.5.770

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception, 17*(2), 197-221. http://dx.doi.org/10.2307/40285891

Juslin, P. N. & Sloboda, J. A. (2013). Music and Emotion. In D. Deutsch, *The Psychology of Music, Third Edition* (pp. 583-645). Academic Press. http://dx.doi.org/10.1016/B978-0-12-381460-9.00015-8

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin and J. A. Sloboda, *Handbook of music and emotion: Theory, research, applications* (pp. 453-489). Oxford, UK: Oxford University Press.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164. http://dx.doi.org/10.2307/40285493

Knösche, T. R., Neuhaus, C., Hauisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*(4), 259-273. http://dx.doi.org/10.1002/hbm.20088

Koponen, E., & Lacerda, F. (2003). Final lengthening in infant directed speech may function as a cue to phrase constituents. *PHONUM 9,* 9-12.

Kotylar, G., & Morozov, V. (1976). Acoustical correlates of the emotional content of vocalised speech. *Soviet Physics Acoustics, 22*, 208-211.

Krause, V., Pollok, B., & Schnitzler, A. (2010). Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychologia, 133*(1), 28-37. http://dx.doi.org/10.1016/j.actpsy.2009.08.003

Krumhansl, C. L., & Jusczyk, P. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*(1), 70-73. http://dx.doi.org/10.1111/j.1467-9280-tb000070.x

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119-159. http://dx.doi.org/10.1037/0033-295X.106.1.119

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26*(1), 1-37. http://dx.doi.org/10.1016/S0364-0213(01)00057-X

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and cultural-specific factors in the recognition and performance of musical affect expressions. *Emotion, 13*(3), 432-449. http://dx.doi.org/101.1037.a0031388

Loehr, J. D., & Palmer, C. (2009). Sequential and biomechanical factors constrain timing and motion in tapping. *Journal of Motion Behavior, 41*(2), 128-136. http://dx.doi.org/10.3200/JMBR.41.2.128-136

Ludusan, B., Cristia, A., Martin, A., Mazuka, R., & Dupoux, E. (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *Journal of the Acoustical Society of America, 140*(2), 1239-1250. http://dx.doi.org/10.1121/1.4960576

Madison, G., & Merker, B. (2002). On the limits of anisochrony in pulse attribution. *Psychological Research, 66*(3), 201-207. http://dx.doi.org/10.1007/s00426-001-0085-y

Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: Movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research, 80*(4), 532-542. http://dx.doi.org/10.1007/s00426-015-0678-5

Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythic expertise. *Experimental Brain Research, 235*(3), 861-871. http://dx.doi.org/10.1007/s00221-016-4845-8

Manning, F., & Schutz, M. (2013). "Moving to the beat" improves time perception. *Psychonomic Bulletin & Review, 20*(6), 1133-1139. http://dx.doi.org/10.3758/s13423-013-0439-7

Matthews, T. E., Thibodeau, J. N., Gunther, B. P., & Penhune, V. B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology, 7*. http://dx.doi.org/10.3389/fpsyg.2016.00069

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience, 36*(1), 313-336. http://dx.doi.org/10.1146/annurev-neuro-062012-170349

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., Fitch, W. T. (2015). Finding the beat: A neural perspective across humans and non-primates. *Philosophical Transactions of the Royal Society, 370*(1664), 20140093. http://dx.doi.lorg/10.1098/rstb.2014.0093

Meyer, M., Baldwin, D., & Sage, K. (2011). Assessing young children's hierarchical action segmentation. In H. H. L. Carlson, *Proceedings of the 33rd annual conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. http://dx.doi.org/10.1037/h0043158

Mohn, C., Argstatter, H., & Wilker, F. (2011). Perception of six basic emotions in music. *Psychology of Music, 39*(4), 503-517. http://dx.doi.org/10.1177/0305735610378183

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion, 11*(3), 818-622. http://dx.doi.org/10.1037/a0022573

Nettl, B. (2000). An enthnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown, *The Origins of Music* (pp. 463-472). Cambridge, MA: The MIT Press.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience, 31*(28), 10234-10240. http://dx.doi.org/10..1523/JNEUROSCI.0411-11.2011

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 331-346. http://dx.doi.org/10.1037/0096-1523.15.2.331

Palmer, C. (1996a). Anatomy of a performance: Sources of musical expression. *Music Perception, 13*(3), 433-453. http://dx.doi.org/10.2307/40286178

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56. http://dx.doi.org/10.2307/40285708

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 113-126. http://dx.doi.org/10.1037/0096-1523.13.1.116

Patel, A. D. & Iverson, I. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience, 8*. http://dx.doi.org/10.3389/fnsys.2014.00057

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*(10), 1367-1391. http://dx.doi.org/10.1068/p6507

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research, 61*(1), 12-32. http://dx.doi.org/10.1007/PL00008161

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562. http://dx.doi.org/10.2307/40285708

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology, 24*(1-5), 157-178. http://dx.doi.org/10.1080/00207594.1989/106000040

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science, 308*(5727), 1430. http://dx.doi.org/10.1126/science.1110922

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition, 105*(3), 533-546. http://dx.doi.org/10.1016/j.cognition.2006.11.006

Povel, D. J., & Collard, R. (1982). Structural factors in patterned finger tapping. *Acta Psychologia, 52*(1-2), 107-123. http://dx.doi.org/10.1016/0001-6918(82)90029-4

Povel, D. J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal, 2*(4), 411-440. http://dx.doi.org/10.2307/40285311

Rammsayer, A., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal, 24*(1), 37-48. http://dx.doi.org/10.1525/mp.2006.24.1.37

Rammsayer, T. H., Buttkus, F., & Altenmüller, E. (2012). Musicians do better than nonmusicans in both auditory and visual timing tasks. *Music Perception, 30*(1), 85-96. http://dx.doi.org/10.1525/mp.2012.30.1.85

Ranganathan, R., & Carlton, L. G. (2007). Perception-action coupling and anticipatory performance in baseball batting. *Journal of Motor Behavior, 39*(5), 369-380. http://dx.doi.org/10.3200/JMBR.39.5.369-380

Remo, C. (2008, September 16). *Activision Boasts Guitar Hero's Music-Selling Power, Seven Studios On Franchise.* Retrieved from Gamasutra: The Art & Business of Making Games: http://www.gamasutra.com

Repp, B. H. (1992a). Diversity and commonality in performance: An analysis of timing microstructure in Shumann's "Traumerei". *Journal of the Acoustical Society of America, 92*(5), 2546-2568. http://dx.doi.org/10.1121/1.404425

Repp, B. H. (1992b). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44*(3), 241-281. http://dx.doi.org/10.1016/0010-0277(92)90003-z

Repp, B. H. (1997). Expressive timing in a Debussy prelude: A comparison of student and expert pianists. *Musicae Scientiae, 1*(2), 257-268. http://dx.doi.org/10.1177/102986499700100206

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America, 104*(2), 1085-1100. http://dx.doi.org/10.1121/1.423325

Repp, B. H. (1998a). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research, 61*(1), 33-43. http://dx.doi.org/10.1007/s004260050011

Repp, B. H. (1998b). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 791-811. http://dx.doi.org/10.1037/0096-1523.24.3.791

Repp, B. H. (1999). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics, 61*(3), 529-548. http://dx.doi.org/10.3758/BF03211971

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science, 29*(2), 200-213. http://dx.doi.org/10.1016/j.humov.2009.08.002

Rock, A. M., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental Psychology, 35*(2), 527-534. http://dx.doi.org/10.1037/0012-1649.35.2.527

Rosenbaum, D. A., & Collyer, C. E. (1998). *Timing of Behavior: Neural, Psychological, and Computational Perspectives.* Cambridge, Massachusetts: The MIT Press.

Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance, 9*(1), 86-102. http://dx.doi.org/10.1037/0096-1523.9.1.86

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Sage, K. D., Ross, R. A., & Baldwin, D. (2012). Assessing infants' hierarchical action segmentation. *Poster presented at the International Conference on Infant Studies.* Minneapolis, Minnesota.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research, 152*(2), 229-242. http://dx.doi.org/10.1007/s00221-003-1548-8

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of the Sciences, 112*(29), 8987-8992. http://dx.doi.org/10.1073/pnas.1414495112

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in Psychology, 4.* http://dx.doi.org/10.3389/fpsyg.2013.00837

Seashore, C. E. (1936). *Objective analysis of musical performance.* Iowa City: The University Press.

Seashore, C. E. (1938). *The psychology of music.* New York: McGraw-Hill.

Senju, M., & Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception, 4*(4), 311-323. http://dx.doi.org/10.2307/40285377

Shaffer, L. H., Clarke, E. F., & Todd, N. M. (1985). Metre and rhythm in piano playing. *Cognition, 20*(1), 61-77. http://dx.doi.org/10.1016/0010-0277(85)90005-8

Silva, S., Branco, P., Barbosa, F., Marques-Teixeira, J., Petersson, K. M., & Castro, S. L. (2014). Musical phrase boundares, wrap-up and the closure positive shift. *Brain Research, 1585*, 99-107. http://dx.doi.org/10.1016/j.brainres.2014.08.025

Sloboda, J. A., & Gregory, A. H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology, 34*(3), 274-280. http://dx.doi.org/10.1037/h0081052

Spiro, N. (2007). *What contributes to the perception of musical phrases in western classical music? (Unpublished dissertation).* Cambridge, UK: University of Cambridge.

Sundberg, J., & Verillo, V. (1980). On the anatomy of the ritard: A study of timing in music. *Journal of the Acoustical Society of America, 68*(3), 772-779. http://dx.doi.org/10.1121/1.384816

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Golumbic, E. Z. (2017). Neural entrainment to the beat: The "Missing Pulse" phenomenon. *The Journal of Neuroscience, 37*(26), 6331-6341. http://dx.doi.org/10.1523/JNEUROSCI.2500-16.2017

Tan, N., Aiello, R., & Bever, T. G. (1981). Harmonic structure as a determinant of melodic organization. *Memory & Cognition, 9*(5), 533-539. http://dx.doi.org/10.3785/BF03202347

Thompson, W. F., & Balkwill, L. (2006). Decoding speech prosody in five languages. *Semiotica, 158,* 407-424. http://dx.doi.org/10.1515/sem.2006.017

Timmers, R. (2007). Vocal expression in recorded performances of Schubert songs. *Musicae Scientiae, 11*(2), 237-268. http://dx.doi.org/10.1177/102986490701100205

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception: An Interdisciplinary Journal, 3*(1), 33-57. http://dx.doi.org/10.2307/40285321

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B, 370*(1664), 20140089. http://dx.doi.org/10.1098/rstb.2014.0089

Turino, T. (2008). *Music as social life: the politics of participation.* London: University of Chicago Press.

van Vugt, F. T., Jabusch, H., & Altenmüller, E. (2012). Fingers phrase the music differently: Trial-to-trial variability in piano scale playing and auditory perception reveal motor chunking. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsych.2012.00495

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary, and peaceful musical excerpts for research on emotions. *Cognition & Emotion, 22*(4), 720-752. http://dx.doi.org/10.1080/02699930701503567

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14*(1), 5-12. http://dx.doi.org/10.3758/BF03198607

Winkler, I., Háden, G., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences, 106*(7), 2468-2471. http://dx.doi.org/10.1073/pnas.0809035106

Woody, R. H. & Lehmann, A. C. (2010). Student musicians' ear-playing ability as a function of vernacular music experiences. *Journal of Research in Music Education, 58*(2), 101-115. http://dx.doi.org/10.1177/0022429410370785

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80-84, http://dx.doi.org/10.1111/j.1467-8721.2207.00480.x

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547-558. http://dx.doi.org/10.1038/nrn2152

Chapter 2: Listeners lengthen phrase boundaries in self-paced music

Kragness, H. E. & Trainor, L. J. (2016). Listeners lengthen phrase boundaries in self-paced music. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1676-1686. http://dx.doi.org/10.1037/xhp0000245

Preface

Expert musicians lengthen phrase boundaries in expressive music performance, but the origin of this behavior (*phrase-final lengthening*) is debated. In Chapter 2, university undergraduates were asked to self-pace through tonal and atonal chord sequences adapted from Bach chorales. Phrase-final lengthening was elicited in both conditions. In follow-up experiments, participants were asked to self-pace through chord sequences that controlled for metrical regularity and melodic contour phrase cues to examine whether phrase-final lengthening could be elicited by harmonic cues found in the tonal condition, but not the atonal condition. In this case, boundary dwelling was enhanced on phrase boundaries in the tonal condition compared to the atonal condition, and further, this effect was observed in participants who reported no previous formal music training. Together, these experiments show that phrase-final lengthening is not limited to musical experts, and suggest a need to consider explanations for phrase-final lengthening that do not rely on musical expertise.

Abstract

Previous work has shown that musicians tend to slow down as they approach phrase boundaries (*phrase-final lengthening*). In the present experiments, we used a paradigm from the action perception literature, the dwell time paradigm (Hard, Recchia, & Tversky, 2011), to investigate whether participants engage in phrase boundary lengthening when self-pacing through musical sequences. When participants used a key press to produce each successive chord of Bach chorales, they dwelled longer on boundary chords than non-boundary chords in both the original chorales *and* atonal manipulations of the chorales. When a novel musical sequence was composed that controlled for metrical and melodic contour cues to boundaries, the dwell time difference between boundaries and non-boundaries was greater in the tonal condition than in the atonal condition. Furthermore, similar results were found for a group of non-musicians, suggesting that phrase-final lengthening in musical production is not dependent on musical training and can be evoked by harmonic cues.

Introduction

Across perceptual domains, parsing events into groups as they unfold across time helps to consolidate low-level information and to focus attention on structurally important features (Chiappe & Schmuckler, 1997; Deutsch, 1980; Dowling, 1973; Large & Jones, 1999; Miller, 1956; Zacks & Swallow, 2007). Accurate parsing of real world auditory streams requires separating two or more co-occurring streams (*stream segregation*) as well as grouping elements in a stream across time (*stream integration*) (Bregman, 1990). A sequence of musical events can be grouped into *phrases*. A musical phrase is a subset of contiguous notes that culminates in a musical boundary. Students of music theory commonly learn about the features that Western composers use to indicate a boundary, and phrasing is often indicated in musical notation. Thus, musicians have explicit knowledge of phrase structures. Previous studies have shown that musicians tend to lengthen notes at the ends of phrases (*phrase-final lengthening*, Palmer, 1989; Repp, 1992a; Seashore, 1938; Todd, 1985). The current study employs a paradigm from the field of action segmentation, the dwell time paradigm, to examine whether participants, including non-musicians, engage in phrase-final lengthening when they control the timing of chord sequences. We additionally investigate whether listeners use harmonic cues (cadences) to determine phrase boundary locations, by examining whether phrase-final lengthening is larger for tonal than atonal chord sequences when other cues such as metrical (rhythmic) structure and melodic contour cues are reduced. Finally, we examine whether non-musicians with minimal musical training also exhibit phrase-final lengthening and use harmonic cues to locate phrase endings. In this way, we offer a novel method for probing listeners' implicit phrase perception defined by tonality cues as a musical sequence unfolds over time.

Of particular relevance to the present study is the idea of musical phrase boundaries as perceptual breakpoints. Perceptual grouping in music has been widely studied (e.g., Chiappe & Schmuckler, 1997; Dowling, 1973; Krumhansl & Jusczyk, 1990; Sloboda & Gregory, 1980; Tan, Aiello, & Bever, 1981; Trainor & Adams, 2000). Listeners' judgments of the locations of phrase boundaries are quite consistent, and consensus is generally even greater amongst musicians (Deliège, 1987; Palmer & Krumhansl, 1987; Peretz, 1989). Several experimental findings suggest that phrase boundaries act as anchors for attention for both musicians and non-musicians. When asked to report the location of clicks randomly inserted into musical passages, listeners reported having heard the clicks as being closer to phrase boundaries than they actually were, an effect dubbed *click migration* (Sloboda & Gregory, 1980). Several experiments have demonstrated that short passages previously heard within a musical phrase are better identified than passages previously heard that crossed phrase boundaries (Dowling, 1973; Peretz, 1989) and that this effect is enhanced for musicians compared to non-musicians when boundaries are defined by harmonic progressions (Tan et al., 1981). Chiappe and Schmuckler (1997) found better memory for musical information directly following a phrase boundary compared to that directly preceding a boundary, but only in musically trained participants. Infants also engage in basic perceptual grouping of auditory sequences. Both 4 ½- and 6-month-old infants prefer to listen to music with pauses *between* rather than *within* phrases (Krumhansl & Jusczyk, 1990), 8-month-old infants use grouping for selective attention (Smith & Trainor, 2011). English-learning infants preferentially hear long tones as phrase-ending (Yoshida et al., 2010) and their detection of pauses is worse after tones of long duration than short duration (Trainor & Adams, 2000). In sum, previous research suggests that musical training augments, but is not necessary for, grouping in music.

Additional evidence for the perception of phrase boundaries comes from neurophysiological experiments. Using event-related potentials (ERPs), language researchers have identified a characteristic waveform associated with linguistic boundary perception. The *closure positive shift*, or CPS, is a positive wave seen at the scalp in centroparietal regions that begins at the phrase boundary and lasts for several hundred milliseconds (Steinhauer, Alter, & Friederici, 1999; Steinhauer & Friederici, 2001). Knösche et al. (2005) found activity resembling the language CPS after musical phrase boundaries. This has been dubbed the “music CPS” and has since been replicated in several studies (Nan, Knösche, & Friederici, 2006; Neuhaus, Knösche, & Friederici, 2006; Silva et al., 2014). Furthermore, one study suggests that it is not affected by musical training (Nan, Knösche, & Friederici, 2009). This body of work supports the behavioral studies in establishing the musical phrase boundary as a psychological percept.

Although there are many cues for phrase boundaries (such as meter and melodic contour), here we focus particularly on *harmony*, the relationship between chords in a musical key (for additional general information on musical harmony and keys, see Supplementary Materials). A central tenet of Western music concerns the progressions from one chord to the next in a musical piece. Not all chords are equally likely in a given context (Huron, 2006). For example, there is a very high statistical dependency between the *dominant* chord (based on the fifth scale degree) and the *tonic* chord (based on the first scale degree), but a low dependency between the dominant chord and the *mediant* chord (based on the third scale degree). This hierarchy of stability between chords contributes greatly to the structure of Western music.

There is evidence that harmonic relationships need not be learned explicitly to influence perception. When asked to make speeded judgments about an unrelated feature of a target chord, such as its timbre, participants respond more quickly when the target is harmonically expected rather than unexpected, reflecting facilitated processing for expected chords (Bharucha & Stoeckig, 1986; Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001; Tillmann & Bharucha, 2002). Musical training does not seem to confer substantial advantages in this task (see Bigand & Poulin-Charronnat, 2006, for a full review of this literature) suggesting that everyday exposure to music is powerful enough to establish a high degree of sensitivity to harmonic structure. Furthermore, it has been shown that irregular, unexpected chords reliably elicit an early right anterior negativity

(ERAN) event-related potential (ERP) from both musicians and non-musicians (Koelsch, Gunter, Friederici, & Schröger, 2000; Koelsch, Jentschke, Sammler, & Mietchen, 2007; Koelsch & Jentschke, 2008, 2010; Leino, Brattico, Tervaniemi, & Vuust, 2007). Thus, both musicians and non-musicians demonstrate implicit sensitivity to harmonic structure.

It has long been observed that musicians show *phrase-final lengthening* (Palmer, 1989; Repp, 1992a; Seashore, 1938; Todd, 1985). Phrase structure in music is often hierarchical, with two or more "subphrases" occurring within a phrase (Palmer & Krumhansl, 1990), and greater lengthening tends to be produced for phrase boundaries at higher hierarchical levels in musical performances (Repp, 1992a; Todd, 1985). It has been proposed that boundary slowing is a technique used by musicians to communicate the structure of a piece to a naïve listener (*musical expression hypothesis*, Clarke, 1985; Palmer, 1989; Repp, 1992a). However, lengthening at phrase boundaries seems to be maintained even when performers are attempting to play mechanically (Penel & Drake, 1998) and listeners are less likely to detect note lengthening at phrase boundaries than within phrases, revealing an implicit expectation for boundary slowing (Repp, 1992b; Repp, 1999). Other work suggests that at least some lengthening can be accounted for by psychoacoustic phenomena that result in biases for time judgments. Thus far, both intensity differences (Tekman, 2001) and rhythmic groupings (Drake, 1993; Drake & Palmer, 1993) have been shown to affect timing judgments, but the relative contributions of different cues to boundary lengthening remain unclear.

Research in other domains has also probed the relationship between timing and boundaries. The “dwell time” paradigm was introduced in 2011 as a new methodology for investigating how observers segment actions occurring over time (Hard et al., 2011). Participants were asked to self-pace through a slideshow of an actor performing a series of action sequences, such as cleaning a room or eating breakfast. Participants controlled the onset of each slide by pressing the spacebar. It was found that participants spent more time on “breakpoint” slides perceived as boundaries between one action and the next (for example, a slide that separates the action “making a bed” from the next action “picking up clothing”) compared to “within-action” slides (for example, slides within the action sequence “making the bed”). Furthermore, there was a hierarchical pattern to participants’ dwell times, with dwell times longest at boundaries participants later identified as coarse-grained, and least on boundaries they identified as fine-grained.

There are several advantages to this experimental approach. First, it is an implicit task, with the true purpose of the task hidden from participants, thus avoiding the possibility of demand characteristics. Participants are told that they will be asked to recall the actions they saw after the slideshow, but are unaware that dwell time is the true measure of interest. Second, the task requires no specialized knowledge to complete, can even be used effectively with children as young as three years (Meyer, Baldwin, & Sage, 2011), and potentially even with children as young as 10 months (Baldwin & Sage, 2013). By adapting the dwell time paradigm to present musical sequences, we do not need to restrict our measures to perceptual judgments, but can examine listeners' timing *production* dynamically across a musical passage without the need for musical training. As far as we are aware, all previous production experiments on lengthening at musical boundaries have been done with musically trained individuals. In the experiments presented here, participants self-paced through two versions of musical excerpts chord by chord. One version (which we call the *tonal*[1] sequence) conformed to the harmonic norms of Western music, with harmonic boundaries occurring every eight chords. The second kind of sequence was atonal, wherein every other chord in a tonal sequence was shifted in pitch by a semitone (1/12 octave), obscuring the harmonic boundary cues. We predicted that listeners would dwell longer on boundary (phrase-final) chords than non-boundary chords, and also find it more difficult to detect boundaries in the atonal than tonal versions. In a second experiment, we investigated whether harmonic boundary cues contribute to phrase-final dwell times when other cues, such as metrical and melodic contour

---

[1] A musical system is called "tonal" if notes function in a hierarchical fashion relative to a central reference pitch (Huron, 2006). In music theory, "atonal" sometimes refers to musical frameworks in which notes do not function relative to a reference pitch, but rather function systematically in other ways. Here, what we refer to as "atonal" sequences were altered versions of the tonal sequences, and not atonal music in their own right.

boundary cues, are reduced. Finally, in a third experiment, we replicated the second experiment in a group of non-musicians with no formal musical training.

Experiment 1

Method

Participants. Eighteen McMaster University undergraduates participated in Experiment 1 (M_{age} = 19.4, SD_{age} = 2.79, 12 females), all of whom reported normal hearing. Four participants were excluded due to experimenter error or failure to follow experimenter instructions, leaving a total of fourteen participants (M_{age} = 19.6, SD_{age} = 3.11, 10 females). All reported fluency in English, and nine reported fluency in at least one other language (French, Urdu, Korean, Cantonese, Mandarin Chinese, Tamil, Polish, and Spanish). Seven of the fourteen reported engaging in current musical endeavors, and all but one participant reported having played an instrument at some point in their lives. Years of formal music training spanned 0 to 12 years (M_{years} = 3.92, SD_{years} = 4.05), with one participant declining to report musical experience. All but one participant were right handed. Participants received introductory psychology course credit as compensation.

Stimuli. Four 4-voice major mode chorales by J.S. Bach were selected as stimuli (see Supplementary Material; Figure 3A shows an example phrase). The first three phrases of each chorale were used. In order to be selected, the excerpt had to end with an authentic cadence (i.e., the final two chords needed to be the dominant chord [built on the 5th scale degree] and the tonic chord [built on the 1st scale degree]; Aldwell & Schachter, 2002) and be comprised of 8-chord phrases (including an anacrusis, or “pick-up” chord; Randel, 2003). Therefore, each of the four sequences (T1, T2, T3, and T4) consisted of 24 chords and three phrases. Author HK made some minor alterations to the chorales, such as removing “grace” notes and passing tones (Aldwell & Schachter, 2002) that fell as eighth or sixteenth notes between the chords. If the chorale was written in another key, it was transposed to F Major. This ensured that a key change did not alert participants to the beginning of a new chorale.

Atonal versions of the chord sequences were created by shifting every other chord down a semitone (1/12 octave). This procedure obscured the tonal center (disrupting the harmonic hierarchy) without affecting the sensory consonance of each chord or the melodic contour (Gerry, Unrau, & Trainor, 2012). The odd-numbered chords were shifted down a half-step for two of the sequences, while the even-numbered chords were shifted down for the other two, resulting in four atonal sequences (A1, A2, A3, and A4). Each chord was generated in GarageBand software with the default piano timbre and the sound level kept constant. Stimuli were presented with Presentation 16.1 06.11.12 (Neurobehavioral Systems) through Denon Stereo Headphones (AH-D501) at 57 to 60 dB, which was judged by author HK to be a naturalistic and comfortable level.

Participants experienced six blocks, three consisting of the tonal and three of the atonal sequences. In each block, all four tonal sequences or atonal sequences were played by the participant, one chord at a time, with the order of the sequences randomized and no break between sequences. Dwell times were defined as the length of time between the onset of one chord and the key-press that cued the presentation of the next chord. The atonal and tonal blocks alternated. Thus, participants played each sequence three times over the course of the experiment. This resulted in a total of 576 dwell times (24 chords per sequence for eight sequences, each played three times) for each participant. Whether the first block was tonal or atonal was counterbalanced across participants.

Upon completion of the six self-paced blocks, memory test trials were presented. Two excerpts from the tonal sequences and two excerpts from the atonal sequences were selected for the memory block. Two tonal foils and two atonal foils were created from additional comparable Bach chorales. Each excerpt and foil excerpt was seven chords long (see Supplementary Materials).

Procedure. After a brief explanation of the task and acquiring consent, participants were asked to fill out a questionnaire about their past and current musical experiences.

Participants were informed that they would hear piano chords over headphones. They were told that their task was to move the piano through the piece being played by pressing the space bar. They were informed that they could only move in the forward direction and could not replay chords. The experimenter told them that in the final part of the experiment, they would hear musical excerpts and be asked to identify whether they had heard each excerpt in the listening phase or not. Participants were not given any explicit instructions regarding pacing, timing, expression, or rhythm. If participants asked for instructions in this regard, the experimenter told them that they should move through the piece in whichever way would help them best recall it during the memory phase. A short training block with a familiar melody (15 notes from *Frère Jacques*) preceded the experiment so that participants could become familiar with the self-pacing task.

After the self-paced blocks, participants heard eight excerpts and foils presented isochronously with an inter-onset interval (IOI) of 750 milliseconds. They were asked to give their best response as to whether they had heard each excerpt in the listening phase by pressing the “1” on the keyboard number line for “yes” and the “0” for “no.”

Results

Each excerpt consisted of three 8-chord phrases, and each participant experienced a total of 36 tonal and 36 atonal phrases. The first tonal block and the first atonal block were discarded to reduce positive skew due to very long early dwell times at the beginning of the task, leaving a total of 24 tonal and 24 atonal phrases for each participant. Because participants sped up over the course of the task, the data were subjected to a linear de-trend. After this, data were Z-normalized (across all tonal and atonal trials) within each participant. We will refer to the detrended normalized data as the *transformed* data. We first examined participants' dwell times dynamically across phrases. Average transformed dwell times for each position (1-8) were calculated, generating an average timing profile for both tonal and atonal phrases (Figure 1). From the timing profiles, it is clear that the dwell times were longest for phrase boundaries (position 8) in both the tonal and atonal sequences. There also appears to be a local maximum for dwell times in positions halfway through the phrase (position 4).

To formally test whether dwell times were greater for boundaries than non-boundaries, trials were binned as Atonal Boundary (AB, 24 chords), Atonal Non-boundary, (AN, 168 chords), Tonal Boundary (TB, 24 chords), or Tonal Non-boundary (TN, 168 chords). For this analysis, boundaries were considered to be only the final chord of a phrase (the eighth, sixteenth, and twenty-fourth chords of each excerpt). The mean transformed dwell time was calculated for each bin for each participant (Figure 2). Because the data were normalized, some scores were negative. The average transformed dwell times were submitted to a two-factor repeated measures ANOVA with factors Tonality (Atonal, Tonal) and Boundary Status (Boundary, Non-boundary). The interaction and main effect for Tonality were not significant (p = .196 and p = .454, respectively), but there was a

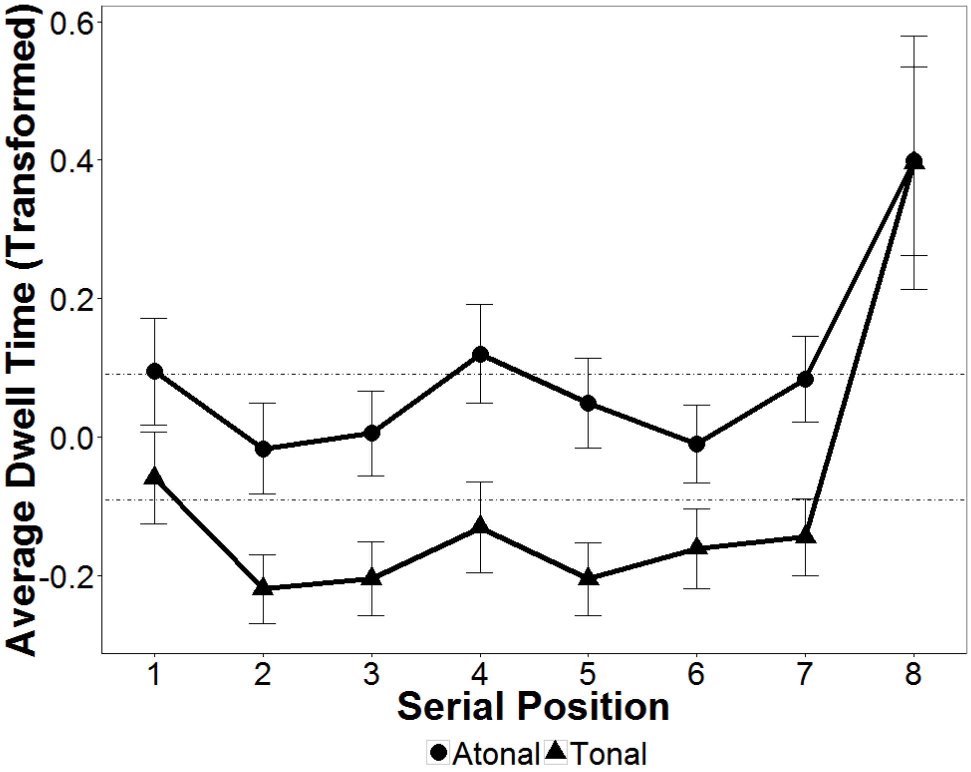


Figure 1

Average transformed dwell times for chords at each position in the 8-note phrases. Similar patterns were observed for both the Tonal condition and the Atonal condition. The dashed lines represent the grand average dwell time in each condition. For all figures, error bars represent standard error of the mean (SEM) across subjects.

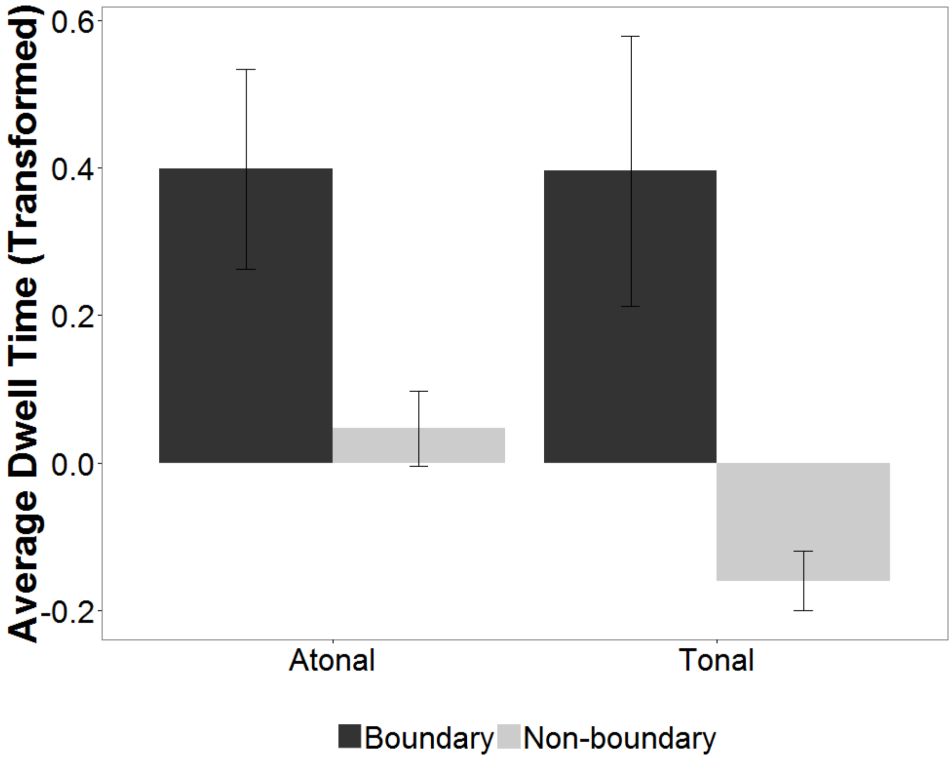


Figure 2

Mean dwell times binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard errors of the mean.

main effect of Boundary Status, $F(1,13) = 9.999$, $p < .01$, $\eta_p^2 = 0.435$[2]. Post-hoc one-way paired t-tests in each condition (Bonferroni-corrected for two comparisons, for a significance cut-off of .025) revealed that participants dwelled longer on boundaries than non-boundaries in both conditions ($t_T(13) = 3.160$, $p_T < .01$; $t_A(13) = 2.398$, $p_A < .025$).

We next investigated whether dwell times were related to hierarchical phrase structure, with shortest to longest dwell times at non-, fine-, and coarse-boundaries. For each 24-chord excerpt, only the final chord (position 24) was considered to be a coarse boundary. The final chords of the other eight-chord phrases (position 8, position 16) were considered to be fine boundaries. All other chords were considered to be non-boundaries.

We calculated the average transformed dwell time for each trial type (Figure 3). An ANOVA with factors Tonality and Level (Non, Fine, Coarse) revealed a main effect of Level, $F(2, 26) = 12.211$, $p < .001$, but no significant main effect of Tonality ($p = .765$) or interaction ($p = .357$). The main effect of Level remained significant after a Greenhouse-Geisser correction, $F(1.356, 17.628) = 12.211$, $p < .01$, $\eta_p^2 = .484$, applied due to violations of sphericity. To determine which levels differed significantly, a series of one-way paired t-tests

---

[2] An ANOVA using raw, untransformed dwell times (including the first tonal and atonal blocks) as the dependent measure resulted in the same significant effects as the ANOVA using transformed dwell times. See Table 1 for raw dwell times.

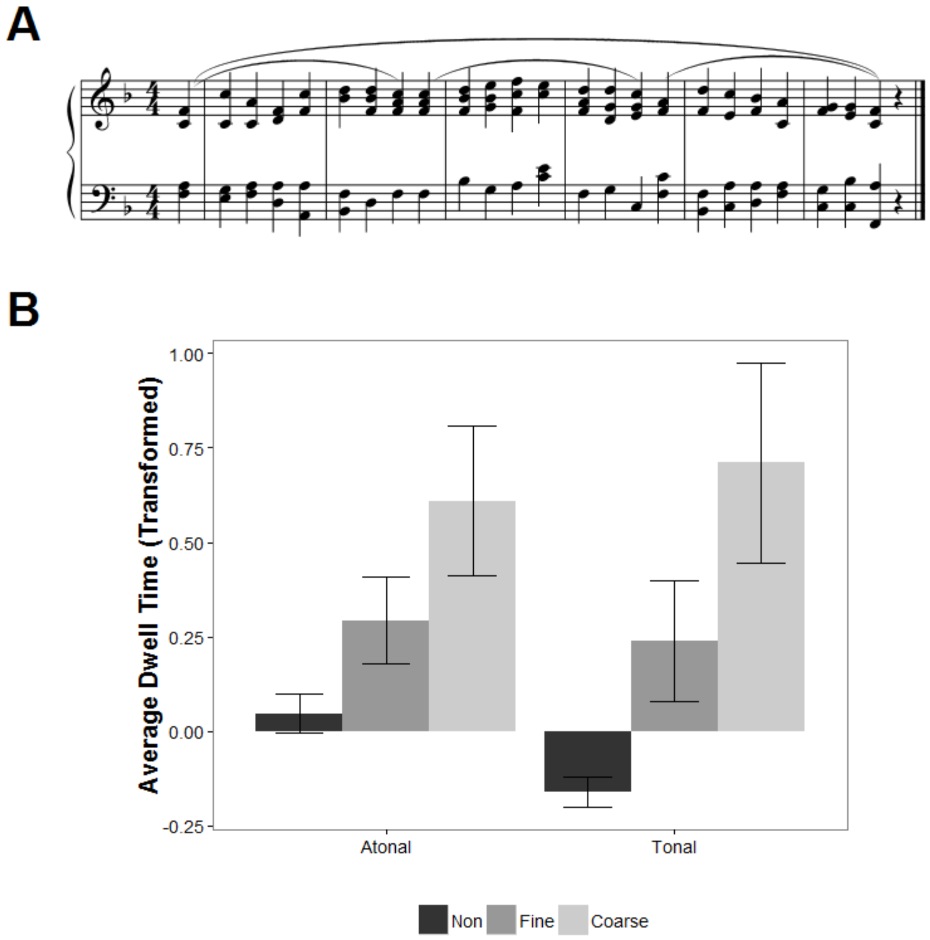


Figure 3

(A) A depiction of the two phrase levels in one excerpt (BWV 1.6) as shown by the phrase markings above the musical notation. Error bars represent SEM. The positions of coarse and fine boundaries were identical across all excerpts, which all consisted of six bars with an anacrusis. (B) Average transformed dwell times for each boundary level for atonal and tonal versions separately.

was conducted. Since three separate t-tests were performed to achieve a family-wise alpha of 95%, the significance cut-off for p was considered to be .0167 (.05/3). All pairwise comparisons were significant: dwelling was greater for fine than non-boundaries ($t(13) = -2.475$, $p < .0167$), greater for coarse than fine boundaries ($t(13) = -3.749$, $p < .01$), and greater for coarse than non-boundaries ($t(13) = -3.819$, $p < .01$).

A question of interest was whether dwelling on musical boundaries was enhanced by musical training. One participant opted not to report musical experience. Because the distribution of reported formal musical training among the remaining 13 participants approached a violation of normality ($W = .877$, $p = .066$) and the sample size was relatively small with a large number of tied ranks, Kendall’s tau was used to evaluate correlations. Directional tests predicting positive correlations between formal training and difference scores (boundary minus non-boundary) were not significant in either the Tonal condition ($r_\tau(11) = .084$, $p = .352$) or the Atonal condition ($r_\tau(11)= -.139$, $p = .737$).

In the original dwell time study for action segmentation (Hard et al., 2011), it was found that participants who looked longer at boundaries recalled more actions from the slideshow. The average score for the memory task in the present experiment was 5.14 out of 8 possible correct responses ($SD = 1.23$), and approached non-normality ($W = .882$, $p = .061$) with a large number of tied ranks. A test of Kendall’s tau predicting positive correlations between memory scores and difference scores was not significant in the Tonal condition ($r_\tau(12) = .137$, p = .263), but was significant in the Atonal condition ($r_\tau(12) = .361$, $p < .05$).

Experiment 2

Although we observed a robust effect of phrase boundaries on dwell time in Experiment 1, we expected that participants would show less sensitivity (i.e., less difference in dwell times) to boundaries in the atonal compared to tonal condition. The results revealed, however, that sensitivity to boundaries was not significantly different across conditions, suggesting that listeners used cues such as melodic contour (the up and down movement of the notes across time) and meter (beat grouping) (Lerdahl & Jackendoff, 1983) to detect boundaries. Specifically, in the Bach chorales used in Experiment 1, the highest voice tended to follow a contour of rising and then falling pitch across phrases. Furthermore, a phrase boundary occurred every 8 beats, providing a very strong metrical cue to boundary locations. Participants might well have used these cues in addition to (or instead of) harmonic cues to determine the locations of phrase boundaries.

Table 1

Raw Dwell Times (ms) by Trial Type and Experiment

Experiment	Trial Type			
	TB	TN	AB	AN
Exp 1	1699.73 *(569.22)*	1536.10 *(584.22)*	1831.96 *(819.62)*	1683.33 *(728.85)*
Exp 2	1057.09 *(464.40)*	943.63 *(390.28)*	950.05 *(366.33)*	939.17 *(358.40)*
Exp 3	937.01 *(283.63)*	825.49 *(225.70)*	790.95 *(279.35)*	803.49 *(285.41)*

Having demonstrated in Experiment 1 that the dwell time paradigm could be used successfully with musical stimuli, Experiment 2 investigated the effect of harmonic closure on musical dwell time in the absence of metrical and melodic contour cues. A novel chord sequence was composed, controlling for any grouping cues that might be elicited by meter and the contour of the highest voice. We hypothesized that participants would dwell longer on boundary chords (the last chord of *perfect authentic cadences*[3]) than non-boundary chords in the tonal condition, but that this effect would be eliminated or reduced in the atonal condition, in which the authentic cadences would be altered.

Methods

Participants. Twenty McMaster University undergraduate and graduate students participated in this study (M_{age} = 20.1, SD_{age} = 1.67, 17 females). Participants from Experiment 1 were ineligible. All reported normal hearing except one participant, who reported chronic tinnitus. Analyses were performed

---

[3] There is a strong tendency for the chord built on the fifth scale degree, V (called the *dominant*), to lead to the chord built on the first scale degree, I (called the *tonic*), particularly at points of musical closure such as phrase boundaries. When this chord sequence occurs at a phrase boundary, it is called an *authentic cadence*. Analyses of information content between two-chord successions in Bach chorales demonstrate that cadences are highly predictable compared to non-cadential musical sequences (Huron, 2006).

both with and without this participant. Because the omnibus ANOVA revealed the same effects in both cases, this participant was included in the analyses reported here. One participant reported left-handedness; all others were right handed. All participants reported English fluency, and ten reported current fluency in another language (French, Arabic, Polish, Tamil, Persian, Urdu, German, and Vietnamese). Eight participants reported currently playing an instrument, and eighteen reported either current or previous experience playing an instrument. Participants reported an average of 6.4 years of formal music lessons (*SD* = 5.25), ranging from 0 to 14 years. Participants either received credit towards an introductory psychology course or a candy bar as compensation.

Stimuli. A single long musical sequence consisting of 112 chords was composed by an assistant professor of music theory with extensive experience in harmony and improvisation in the Baroque style (see Supplementary Material; Figure 4A and 4B provide example excerpts). The sequence was composed specifically such that melodic contour and metrical cues to phrase boundaries did not align with harmonically defined boundaries (authentic cadences). To control for metrical boundary information, a series of 14 numbers was generated, such that each number was chosen pseudo-randomly from the numbers between 5 and 11 (inclusive), and each number appeared exactly twice in the series. This series was used to dictate the lengths (i.e., number of chords) of each successive phrase in the novel composition, with each phrase ending in an authentic harmonic cadence. For example, the first phrase contained 8 chords, the second phrase

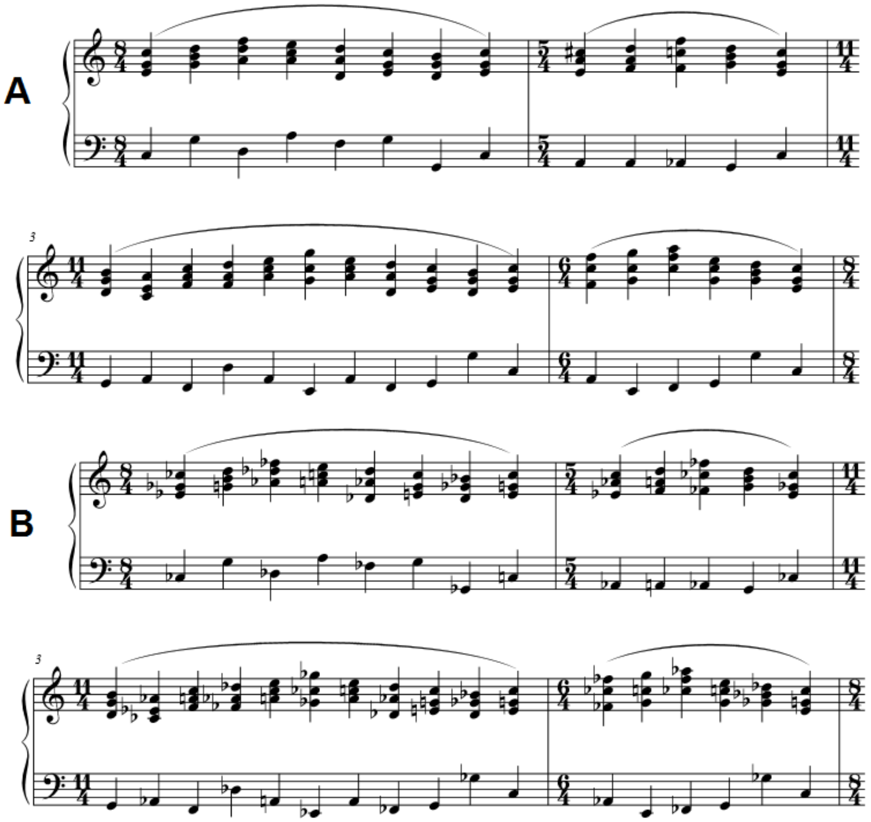



Figure 4

(A) The first four phrases of the tonal sequence in Experiment 2. (B) The first four phrases of the atonal sequence in Experiment 2. It is the same as the tonal sequence, but every second chord (starting with the first) was shifted down by a half-step.

contained 5 chords, the third phrase contained 11 chords, and so on. This was done to eliminate the possibility of participants using a consistent metrical structure (e.g., a boundary every 8 chords, as in Experiment 1) as a cue to boundary locations. Melodic contour can also offer information about boundaries. The sequence for Experiment 2 was composed such that the contour in the highest voice changed direction every five chords, so that melodic contour was uncorrelated with phrase boundaries. Thus, the piece was composed specifically such that the contour and the harmonic boundary cues did not align in a systematic way.

Overall, the composition contained 112 chords in 14 phrases. Due to the length of this sequence, participants experienced only four listening blocks in the self-pacing phase (rather than the six in Experiment 1). In atonal blocks, the odd-numbered chords in the tonal stimuli were shifted down a half step, as in Experiment 1. The chords were generated in GarageBand software with the default piano timbre and the sound level kept constant. Stimuli were presented in the same program and manner as in Experiment 1 (Presentation 16.1 06.11.12 (Neurobehavioral Systems), Denon Stereo Headphones (AH-D501) at 57 to 60 dB).

As in Experiment 1, participants were given a memory test after the self-pacing phase. Two 8-chord excerpts were lifted from the tonal version of the sequence and two 8-chord excerpts from the atonal version. Four 8-chord foil sequences (two tonal, two atonal) were composed by author HK in the same style as the original memory probes.

Procedure. The procedure was identical to that in Experiment 1 with the following exceptions. Instead of hearing a total of 576 chords, each participant heard a total of 448 chords, with each of the tonal and atonal sequences heard twice. The tonal and atonal sequences alternated, and whether the first sequence was tonal or atonal was counterbalanced across participants. The apparatus was identical to Experiment 1.

Results

As in Experiment 1, the first tonal and atonal blocks were discarded to reduce positive skew from long early looking times. Again, dwell times were subjected to a linear detrend and normalization and analyses were conducted on this transformed data. Time profiles were generated for tonal and atonal versions (Figure 5A and 5B). Because the meter was random, each solid line represents dwell times for phrases of a specific length (5 to 11 chords), resulting in seven lines in each figure. It can be seen that there is a clear jump in dwell time between the boundary chord (0) and chord directly preceding the boundary chord (-1) in the tonal condition, but not in the atonal condition.

To test whether boundary dwell times were different from non-boundary dwell times, transformed dwell time scores were binned as either Tonal Non-

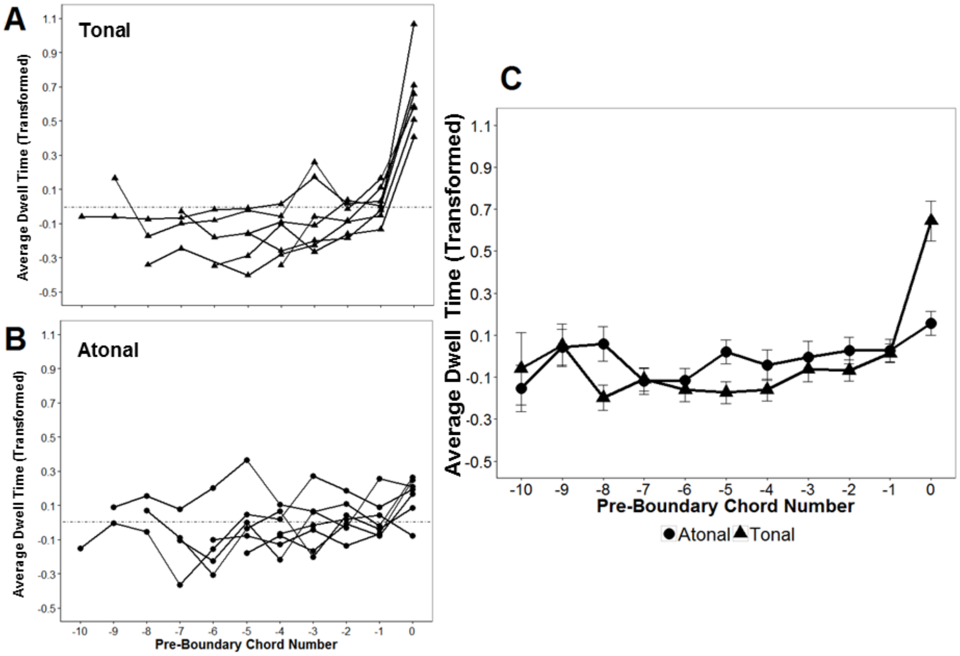


Figure 5

(A) and (B) represent Tonal and Atonal dwell time profiles, respectively, for Experiment 2. Because the meter was random, the x-axis represents the chord number before the boundary chord (0). Each of the 6 lines in each figure represents the average of phrases of a particular length (5 to 11 chords). (C) Average dwell time profiles for Experiment 2, disregarding phrase length. Because phrases were different lengths, averages for positions -4 through 0 are made up of many more data points than positions -10 through -5.

boundaries (98 chords), Tonal Boundaries (14 chords), Atonal Non-boundaries (98 chords), or Atonal Boundaries (14 chords), and the means for each bin were calculated for each participant (Figure 6). The data were submitted to a two-way repeated measures ANOVA with factors Tonality (Atonal, Tonal) and Boundary Status (Boundary, Non-boundary). The ANOVA revealed main effects of both Tonality ($F(1,19) = 9.37$, $p < .01$, $\eta_p^2 = 0.330$) and Boundary Status ($F(1,19) = 18.05$, $p < .01$, $\eta_p^2 = 0.487$), as well as a significant interaction ($F(1,19) = 13.35$, $p < .01$, $\eta_p^2 = 0.413$) such that dwelling on Boundaries compared to Non-boundaries was enhanced in the Tonal sequence compared to the Atonal sequence[4].

Paired t-tests were conducted post-hoc to investigate whether there was a significant boundary dwelling effect in each Tonality condition. After Bonferroni correction, the difference between boundaries and non-boundaries was significant in both the Atonal condition ($M_{AB} = 0.156$, $SD_{AB} = 0.228$, $M_{AN} = -0.016$, $SD_{AN} = 0.093$), $t(19) = 2.446$, $p < .025$, as well as the Tonal condition ($M_{TB} = 0.644$, $SD_{TB} = 0.685$, $M_{TN} = -0.098$, $SD_{TN} = 0.127$), $t(19) = 4.259$, $p < .001$.

---

[4] An ANOVA using raw, untransformed dwell times (including the first tonal and atonal blocks) as the dependent measure also found a significant interaction between Tonality and Boundary Status, reflected by a larger boundary dwelling effect in the tonal condition than the atonal condition.

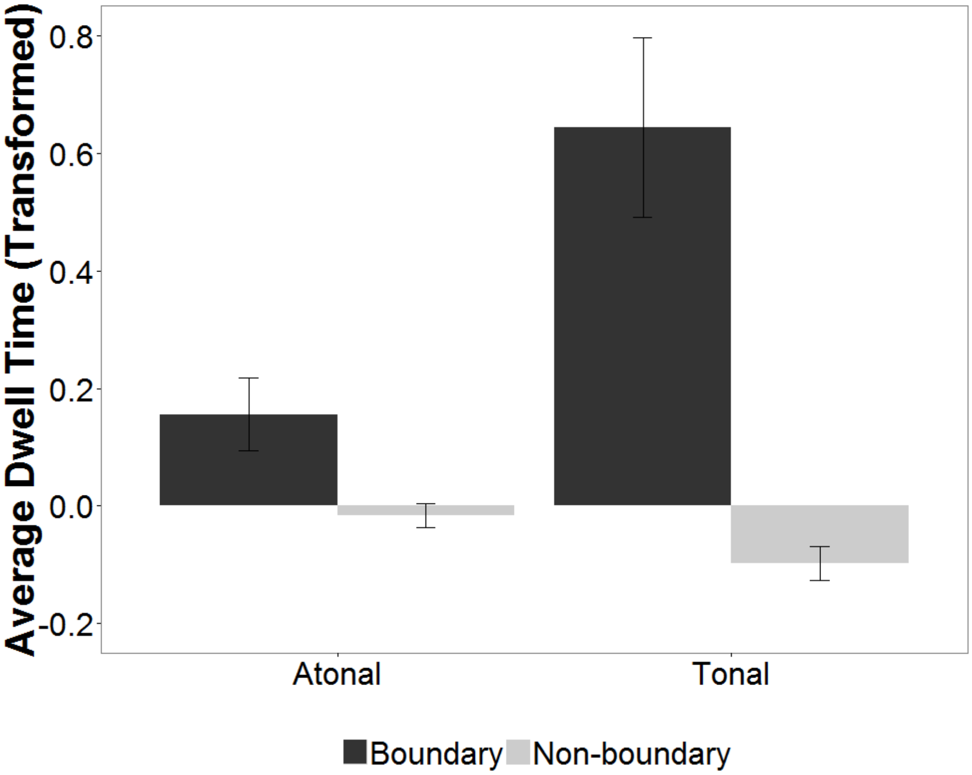


Figure 6

Mean dwell times binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard errors of the mean. Participants dwelled longer on boundaries than non-boundaries in both tonality conditions, but the difference was significantly larger in the Tonal condition than in the Atonal condition.

Reported years of musical experience and scores on the memory test violated normality ($W = .860$, $p < .01$; $W = .892$, $p < .05$) and contained several tied ranks. As such, Kendall's tau was again employed to test all correlations. Length of formal training did not have a significant positive correlation with longer boundary dwell times for either the Tonal condition ($r_\tau(18) = .120$, $p = .226$) or the Atonal condition ($r_\tau(18) = .098$, $p = .278$). We again correlated the dwell time difference scores with scores from the memory test. After the conclusion of the experiment, we discovered an error in one of the excerpts for the memory task. Thus, responses for the errant excerpt were discarded, as well as the foil trial that was matched with this trial, leaving a total of six memory probes (three excerpts, three foils). The average score for the remaining six probes was 3.8 out of 6 (*SD* = 1.8). Correlations between difference scores and memory scores approached significance in both the Tonal condition ($r_\tau(18) = .274$, $p = .059$) and the Atonal condition ($r_\tau(18) = .250$, $p = .076$).

Experiment 3

Overall, Experiment 2 replicated the main finding in Experiment 1 that participants dwell on musical boundaries in a self-paced musical production task. Experiment 2 extended this finding by demonstrating that boundary dwelling could be elicited even when stable metrical cues were eliminated and contour cues minimized. The lack of significant correlations between boundary dwelling and formal musical training in both experiments suggest that formal musical training might not be critical for boundary dwelling to be elicited. However, we sampled from university undergraduates who tend to be from mid to high socio-economic backgrounds and to have taken music lessons. Indeed, on average participants in Experiment 2 had more than 6 years of formal musical lessons. Furthermore, it is possible that some participants who did not report formal musical training were casual musicians who play on a regular basis. It is possible, then, that the majority of participants in Experiment 2 were more musically trained than the general population, which could underlie the effect we observed.

To investigate whether non-musicians show similar results, we recruited a group of participants who considered themselves to be non-musicians and who had minimal experience singing or playing musical instruments. If musical training is necessary for harmonically-induced phrase lengthening, then there should be no difference in boundary versus non-boundary dwell times for this non-musically trained group.

Methods

Participants. Twenty-two McMaster University undergraduate and graduate students participated in this study, and one young adult from the area. Participants responded to advertisements for people who were not currently regularly playing a musical instrument, did not consider themselves to be musicians, and had never taken formal lessons. Despite this, five participants revealed in their questionnaire responses that they had taken instrumental lessons in the past. These participants' data were discarded for the current analyses. Of the 17 remaining participants, all reported normal hearing. Two reported left-handedness; all others were right handed. All reported English fluency, and 11 reported exposure to other languages (including French, Japanese, Urdu, Slovakian, Portuguese, Hindi, Punjabi, and Arabic). Though they had not taken any formal music lessons, seven reported having played an instrument at some point (including violin, clarinet, guitar, flute, piano, and organ). Participants received credit towards an introductory psychology course or a candy bar as compensation.

Stimuli and Procedure. The stimuli and procedure were identical to Experiment 2, with the exception that the erroneous stimuli in the memory phase were corrected for Experiment 3.

Results

The data were subjected to the same processing as described in Experiment 2, in which only the second half of the trials were used for each participant and the dwell times were detrended and Z-normalized. Dwell time scores were binned as either Tonal Non-boundaries (98 chords), Tonal Boundaries (14 chords), Atonal Non-boundaries (98 chords), or Atonal

Boundaries (14 chords), and the means for each bin were calculated for each participant (Figure 7). The data were submitted to a two-way repeated measures ANOVA with factors Tonality (Atonal, Tonal) and Boundary Status (Boundary, Non-boundary). The ANOVA revealed no significant main effects for Tonality or Boundary (p = .257 and p = .074, respectively). There was a significant Tonality x Boundary Status interaction, $F(1, 16) = 6.25$, p < .05, $\eta_p^2 = 0.281$[5]. One-way post-hoc paired t-tests with a family-wise confidence level of 95% were conducted to test whether Boundary dwell times were larger than Non-Boundary dwell times in each Tonality condition separately. The difference between boundaries and non-boundaries was significant in the Tonal condition ($M_{TB} = 0.230$, $SD_{TB} = 0.455$, $M_{TN} = -0.070$, $SD_{TN} = 0.079$), $t(16) = 2.452$, $p < .025$, but not the Atonal condition ($M_{AB} = -0.035$, $SD_{AB} = 0.206$, $M_{AN} = 0.041$, $SD_{AN} = 0.070$), $t(16) = -1.360$, p = .904.

Scores on the memory test ranged from 2 to 7 correct responses out of a possible eight. The average score was around chance level (M = 4.41, SD = 1.33). Because of the large number of tied ranks in memory scores, Kendall's tau was

---

[5] An ANOVA using raw, untranformed dwell times (including the first tonal and atonal blocks) also found a significant interaction between Tonality and Boundary Status, reflected by a larger boundary dwelling effect in the tonal condition than the atonal condition.

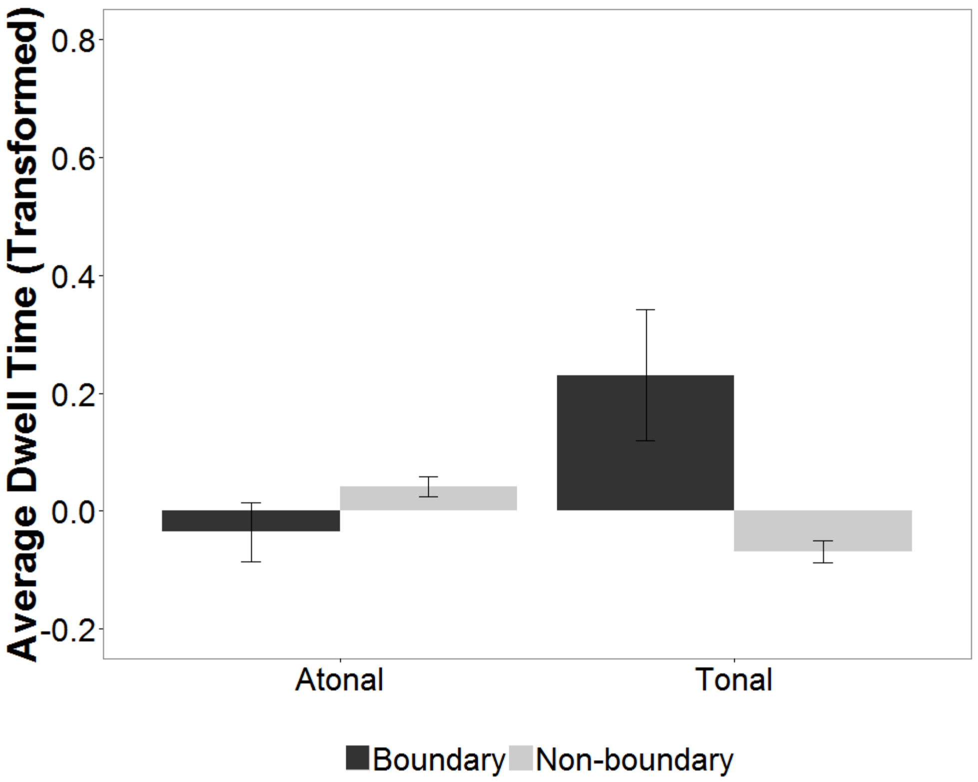


Figure 7

Mean dwell times for non-musicians binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard error of the mean. Participants dwelled longer on boundaries than non-boundaries in the tonal condition, but not the atonal condition.

used to test correlations between memory scores and difference scores (boundary minus non-boundary). The correlation was not significant in either the Tonal condition ($r_\tau(15) = -.041$, $p = .831$) or Atonal condition ($r_\tau(15) = -.122$, $p = .522$).

General Discussion

Taken together, the results of all three experiments indicate that participants engage in phrase-final lengthening when self-pacing through musical sequences. The results of Experiment 2 further demonstrate that listeners use harmonic cues to phrase boundaries when metrical predictability (a strong temporal cue for phrase boundaries) and melodic contour (a strong pitch cue for boundaries) are minimal. Finally, Experiment 3 shows that even non-musicians dwell on harmonic boundaries in a self-pacing task.

The correlations between memory for the sequences and the relative lengthening (boundary minus non-boundary) were significant or approached significance in three of the four sequences in Experiments 1 and 2, but were not significant for either condition in Experiment 3. Given that performance was quite low on the memory task, it was probably not a very sensitive index of memory. Therefore, a question to address in the future is whether participants who exaggerated phrase-final lengthening in the self-pacing task later had better memory for the musical sequences. Such a finding would suggest that the boundary dwelling effect could be some form of “chunking,” as has been found for verbal working memory (Miller, 1956).

First and foremost, these results extend the long-held observation that musicians systematically deviate from mechanical timing in musical performance. In contrast to previous studies involving musical production, none of the present participants were professional musicians and those in Experiment 3 had no formal training at all and did not play an instrument. Yet participants in all three experiments showed phrase-final lengthening in a production task. Furthermore, in Experiment 1, boundary lengthening was systematically related to the hierarchical level of boundary, such that coarse boundaries were dwelled on longer than fine boundaries. This parallels Hard et al.’s study of action segmentation (2011), and reinforces the link between hierarchical structure and lengthening described in previous studies of music (Repp, 1992a; Todd, 1985). Interestingly, participants’ raw dwell times were higher in Experiment 1 than in Experiments 2 or 3 across all conditions. Only Experiment 1 utilized real musical excerpts as stimuli, which participants may have found more pleasant overall.

The *musical expression hypothesis* (Clarke, 1985; Palmer, 1989; Repp 1992a) predicts that musical training enhances phrase-final lengthening. Interestingly, the non-musicians in Experiment 3 appear to have a reduced dwell time effect compared to the random sample in Experiment 2, consistent with the prediction of the musical expression hypothesis. However, it should be noted that each experiment drew from different populations (graduate and undergraduate students for Experiment 2; largely IntroPsych undergraduate students for Experiment 3) so it is difficult to directly compare across the two experiments. It would therefore be useful to test this hypothesis directly with the methodology of the present study, by comparing a group of participants with no musical training to a group with musical training, matched in other ways. However, it is exceedingly difficult to examine causal effects of musical training, as musicians and non-musicians are not randomly assigned and may have pre-existing population differences in neural structure and activity (e.g. Gaser & Schlaug, 2003; Schneider et al., 2002; Zatorre, 2013). While effects of musical training cannot be closely examined here, Experiment 3 clearly demonstrates the novel finding that musical training is not necessary for phrase-final lengthening in a musical production task.

A possible explanation for our finding is that participants were mimicking phrase-final lengthening of musical performances they have heard. This is possible, but seems unlikely in the present experiment. Participants experienced the music as it unfolded, and would have had little opportunity to plan their timing in a way that would closely emulate practiced musical performances, especially in Experiments 2 and 3, where metrical groupings varied with each musical bar. Further, though participants may have learned some regularities of the musical sequences over the course of the experiment, they did not have a musical score for reference, did not prepare their performances, and may have not even conceived of the task as a performance. These results corroborate past claims (e.g. Drake & Palmer, 1993) that expressive intent or imitation cannot fully explain variations in timing.

Because participants were not asked to play either expressively or mechanically, the extent to which timing variations were intentional is not known. Previous work has shown that musicians use systematic timing variations even when asked to play mechanically (Drake & Palmer, 1993), so it is likely that participants in the current experiments were unaware of their boundary dwelling. This would be consistent with studies showing that in perceptual tasks listeners are least likely to notice lengthening at points of structural importance (Repp, 1992b; Repp, 1999), revealing an implicit expectation for boundary slowing. These past results have been taken as evidence for a *perceptual compensation* explanation for some timing variations, such that some musical events are lengthened because they are perceived to be shorter than they actually are (e.g. Penel & Drake, 2004). Studies of short musical sequences have demonstrated effects of different rhythmic groupings (Drake, 1993) and intensity (Tekman, 2001) on duration judgments. Although there were no systematic rhythmic or intensity differences in experimental stimuli of the present study, it is possible that there were other stimulus factors systematically aligned with harmonic boundaries that biased time perception. For example, it has also been shown that the time between two pitches is perceived as longer when there is a larger pitch distance between the two pitches (Crowder & Neath, 1995). The perfect authentic cadences in our stimuli contain a pitch leap in the bass line from the penultimate to final note of the phrase. This may have caused participants to perceive the final bass note onset as delayed, leading them to dwell longer on the final tonic chord. This pitch leap in the bass line was present in both the tonal and atonal versions, which might explain why participants were above chance levels at dwelling longer on atonal boundaries than non-boundaries in Experiment 1. However, the bass line leap cannot explain the much greater boundary dwelling in tonal compared to atonal versions in Experiments 2 and 3, which indicates that even if pitch leaps were playing a role, participants were using tonality cues to determine phrase boundaries when these cues were available.

It would be particularly interesting in the future to investigate the origin of boundary dwelling in the atonal sequences in Experiments 1 and 2. In Experiment 1, we saw no difference in the magnitude of the effect in the tonal versus atonal conditions. We have already proposed the idea that the metrical predictability of the sequence was a main driver of boundary dwelling in Experiment 1, perhaps overriding the lack of harmonic cadences in the atonal versions. However, metrical predictability was not available in Experiment 2, where a greatly reduced but still significant effect of boundary lengthening was found in the atonal version. In Experiment 2, it is possible that participants still perceived some harmonic information in the atonal version. For example, even in the atonal version, the tonic chord was one of the most frequently occurring chords, although it was never preceded by the dominant chord. Nonetheless, chord frequency effects might have contributed to a perception of phrase boundaries.

Alternatively, participants may have perceived the overall key to be F Major to some extent, and felt that that the non-diatonic chords were ultimately resolved by the diatonic chords that followed them (as in Bharucha, 1984). Interestingly, whichever cue was driving dwelling on atonal boundaries in Experiment 2, the non-musician participants in Experiment 3 appear to have been unable to detect it, although a more sensitive measure might reveal some ability in this regard. Future work should investigate the influences of many different musical parameters on participants' boundary dwelling and the effects of individual differences and musical training.

Another idea consistent with a perceptual compensation account of boundary dwelling in the present experiment is that compensation is based on predictability. It has been posited that the perception of boundaries in melodies arises from local maxima of predictive uncertainty (Pearce & Wiggins, 2006; Pearce, Müllensiefen, & Wiggins, 2010). Specifically, the first note of a phrase (i.e., the note following the last note of the previous phrase) is less predictable than notes within a phrase. It is possible that longer dwell times may reflect greater uncertainty for the next event (the first note of the next phrase), resulting in longer processing times. In a recent study (described in Baldwin & Sage, 2013), experimenters generated nonsensical action sequences composed of three unrelated actions. Statistical dependencies between grouped actions could only be learned by passively viewing a corpus prior to the self-pacing task, and not by top-down experiences with daily actions. In the self-pacing phase of the task, dwell times were systematically related to position across an action group, but only if participants had previously viewed the exposure corpus (thereby learning the statistical dependencies). Thus, in the action perception domain, it seems that differences in dwell times may be accounted for partly by the structure imposed by transitional probabilities. Predictability was not explicitly manipulated in the present experiments, but further studies are underway to test the hypothesis that dwell times are directly related to predictive uncertainty.

In sum, we have demonstrated that the dwell time paradigm can be used to probe the relationship between timing and phrase grouping in a non-performance setting with individuals without high levels of musical training. The results offer support for the idea that there is an implicit mechanism contributing to the phenomenon of phrase-final lengthening, and offers the first evidence that musical boundary lengthening does not rely solely on training. In addition, we demonstrate the use of a new dwell time method for investigating the musical timing production of musically untrained individuals. The simplicity and flexibility of this method make it appropriate for the investigation of diverse questions, and it could be easily adapted to a variety of sequential stimuli, such as melodies. In ongoing projects we are using this method to probe the developmental trajectory of harmonic knowledge, the relation between phrase boundaries and stimulus uncertainty, and expressive timing in musical performances of non-musicians.

Author Note

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to LJT and an Ontario Trillium Scholarship to HEK. We thank Dr. Matt Woolhouse for composing the chord sequences used in Experiment 2. We additionally thank Laura Cirelli for comments on an earlier draft. Parts of this paper have been presented previously at Neuroscience & Music V (May 2014, Dijon, France) and Rhythm & Timing Symposium (February 2015, London, Canada).

References

Aldwell, E., & Schachter, C. (2002). *Harmony and voice leading*. (3rd ed.) New York, NY: Schirmer Books.

Baldwin, D., & Sage, K. (2013). Dwelling in Action. In M. Rutherford & V. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 309-330). Cambridge, MA: MIT Press.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130. http://dx.doi.org/10.1016/j.cognition.2005.11.007

Bigand, E., Tillmann, B., Poulin, B., D’Adamo, D.A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition, 81*(1), B11-B20. http://dx.doi.org/10.1016/S0010-0277(01)00117-2

Bharucha, J.J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*(4), 485-518. http://dx.doi.org/10.1016/0010-0285(84)90018-5

Bharucha, J.J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 403-410. http://dx.doi.org/10.1037/0096-1523.12.4.403

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

Chiappe, P., & Schmuckler, M.A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review, 4*(2), 254-259. http://dx.doi.org/10.3758/BF03209402

Clarke, E. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross and R. West (Eds): Musical Structure and Cognition (pp. 209-236). London, UK: Academic Press.

Crowder, R.G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception, 12*(4), 379-386. http://dx.doi.org/10.2307/40285672

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4*(4), 325-360. http://dx.doi.org/10.2307/40285378

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics, 28*(5), 381-389. http://dx.doi.org/10.3758/BF03204881

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics, 14*(1), 37-40. http://dx.doi.org/10.3758/BF03198614

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society, 31*(2), 107-110. http://dx.doi.org/10.1121/1.402158

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*(3). http://dx.doi.org/10.2307/40285574

Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240-9245. http://dx.doi.org/10.1016/s1053-8119(01)92488-7

Gerry, D., Unrau, A., & Trainor, L.J. (2012). Active music classes in infancy enhance musical, communicative, and social development. *Developmental Science, 15*(3), 398-407. http://dx.doi.org/10.1111/j.1467-7687.2012.01142.x

Hard, B.M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology, 140*(4), 586-604. http://dx.doi.org/10.1037/a0024310

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge: MIT Press.

Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*(4), 259-273. http://dx.doi.org/10.1002/hbm.20088

Koelsch, S., Gunter, T., Friederici, A.D., & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*, 520–541. http://dx.doi.org/10.1162/089892900562183

Koelsch, S., & Jentschke, S. (2008). Short-term effects of processing musical syntax: An ERP study. *Brain Research, 1212,* 55-62. http://dx.doi.org/10.1016/j.brainres.2007.10.078

Koelsch, S., & Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *Journal of Cognitive Neuroscience, 22*(1), 2251-2262. http://dx.doi.org/10.1162/jocn.2009.21338

Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology, 44*(3), 476-490. http://dx.doi.org/10.1111/j.1469-8986.2007.00517.x

Krumhansl, C.L., & Jusczyk, P. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*(1), 70-73. http://dx.doi.org/10.1111/j.1467-9280.1990.tb000070.x

Large, E.W., & Jones, M.R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review, 106*(1), 119-159. http://dx.doi.org/10.1037/0033-295X.106.1.119

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research, 1142,* 169-177. http://dx.doi.org/10.1016/j.brainres.2007.01.049

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Meyer, M., Baldwin, D., & Sage, K. (2011). Assessing young children's hierarchical action segmentation. In L. Carlson, H. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. http://dx.doi.org/10.1037/h0043158

Nan, Y., Knösche, T.R., & Friederici, A.D. (2006). The perception of musical phrase structure: A cross-cultural ERP study. *Brain Research, 1094*(1), 179-191. http://dx.doi.org/10.1016/j.brainres.2006.03.115

Nan, Y., Knösche, T.R., & Friederici, A.D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biological Psychology, 82*(1), 70-81. http://dx.doi.org/10.1016/j.biopsycho.2009.06.002

Neuhaus, C., Knösche, T.R., & Friederici, A.D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience, 18*(3), 472-493. http://dx.doi.org/10.1162/089892906775990642

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 331-346. http://dx.doi.org/10.1037/0096-1523.15.2.331

Palmer, C., & Krumhansl, C.L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 116-126. http://dx.doi.org/10.1037/0096-1523.13.1.116

Palmer, C., & Krumhansl, C.L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 728-741. http://dx.doi.org/10.1037/0096-1523.16.4.728

Pearce, M.T., Müllensiefen, D., & Wiggins, G.A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*(10), 1367-1391. http://dx.doi.org/10.1068/p6507

Pearce, M.T., & Wiggins, G.A. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23*(5), 377-405. http://dx.doi.org/10.1525/mp.2006.23.5.377

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research, 61*(1), 12-32. http://dx.doi.org/10.1007/PL00008161

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562. http://dx.doi.org/10.3758/bf03194900

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology, 24,* 157-178. http://dx.doi.org/10.1080/00207594.1989/106000040

Randel, D.M. (Ed.). (2003). *The Harvard dictionary of music.* (4th ed.). Cambridge: Belknap Press.

Repp, B.H. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92*(5), 2546-2568. http://dx.doi.org/10.1121/1.404425

Repp, B.H. (1992b). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44*(3), 241-281. http://dx.doi.org/10.1016/0010-0277(92)90003-Z

Repp, B.H. (1999). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics, 61*(3), 529-548. http://dx.doi.org/10.3758/BF03211971

Seashore, C. (1938). *Psychology of music.* New York: McGraw-Hill.

Schneider, P., Scherg, M., Dosch, H.G., Specht, H.J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*, 688-694. http://dx.doi.org/10.1038/nn871

Silva, S., Branco, P., Barbosa, F., Marques-Teixeira, J. Petersson, K.M., & Castro, S.L. (2014). Musical phrase boundaries, wrap-up and the closure positive shift. *Brain Research, 1585*, 99-107. http://dx.doi.org/10.1016/j.brainres.2014.08.025

Sloboda, J.A., & Gregory, A.H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology*, *34*(3), 274-280. http://dx.doi.org/10.1037/h0081052

Smith, N.A., & Trainor, L.J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distractors. *Infancy, 16*(6), 655-668. http://dx.doi.org/10.1111/j.1532-7078.2011.00067.x

Steinhauer, K., Alter, K., & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191-196. http://dx.doi.org/10.1038/5757

Steinhauer, K., & Friederici, A.D. (2001). Prosodic boundaries, comma rules, and brain responses: The Closure Positive Shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267-295. http://dx.doi.org/10.1023/A:1010443001646

Tan, N., Aiello, R., & Bever, T.G. (1981). Harmonic structure as a determinant of melodic organization. *Memory & Cognition*, *9*(5), 533-539. http://dx.doi.org/10.3785/BF03202347

Tekman, H.G. (2001). Accenting and detection of timing variations in tone sequences: different kinds of accents have different effects. *Perception & Psychophysics, 63*(3), 514-523. http://dx.doi.org/10.3758/bf03194417

Tillmann, B., & Bharucha, J.J. (2002). Effect of harmonic relatedness on the detection of temporal asynchronies. *Perception & Psychophysics, 64*(4), 640-649. http://dx.doi.org/10.3758/BF03194732

Todd, N.P. McA. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58. http://dx.doi.org/10.2307/40285321

Trainor, L.J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception & Psychophysics, 62*(2), 333-340. http://dx.doi.org/10.3758/BF03205553

Yoshida, K., Iverson, J.R., Patel, A.D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*(2), 356-361. http://dx.doi.org/10.1016/j.cognition.2010.01.005

Zacks, J.M., & Swallow, K.M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80-84. http://dx.doi.org/10.1111/j.1467-8721.2007.00480.x

Zatorre, R.J. (2013) Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science, 342*(6158), 585-589. http://dx.doi.org/10.1126/science.1238414

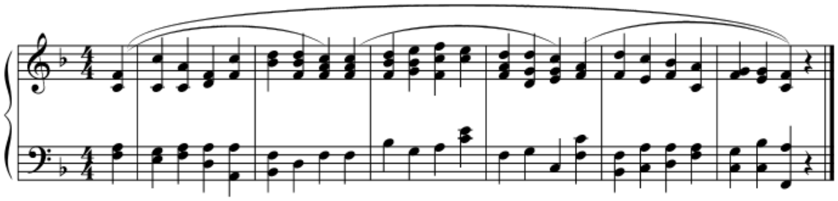
Supplementary Material

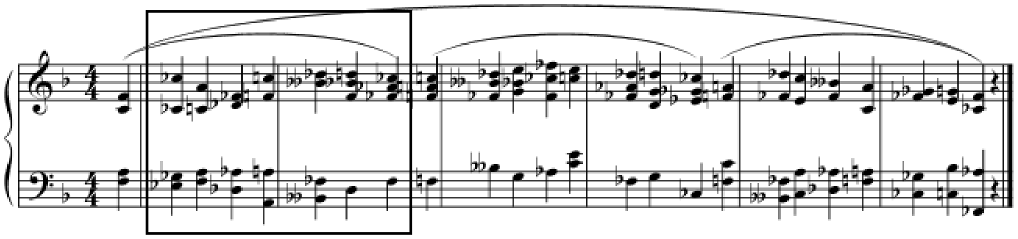
Experiment 1

The musical sequences used in Experiment 1. T refers to Tonal, A to Atonal. The chord sequences contained in the boxes were used in the memory test at the end. The Foil sequences used in the memory test are also shown.

Sequence T1

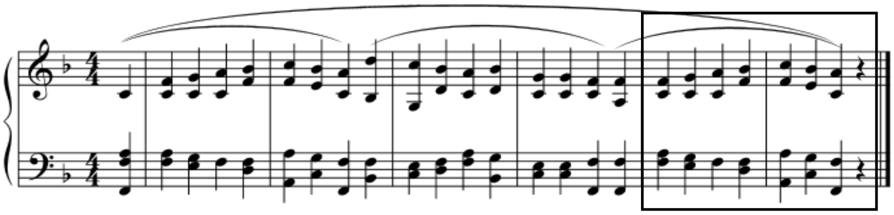
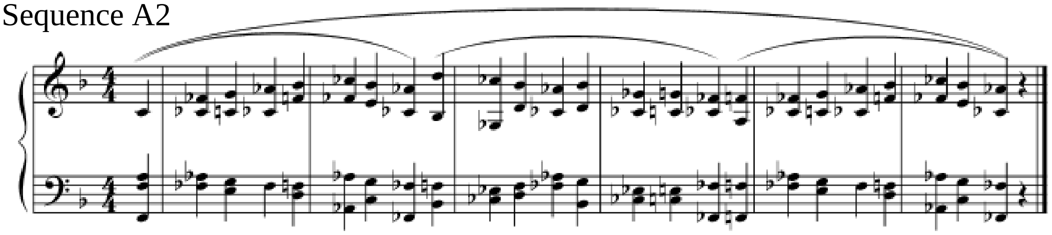
Based on BWV 1.6 harmonized by J. S. Bach

"Wie schön leuchtet der Morgenstern"

Sequence A1

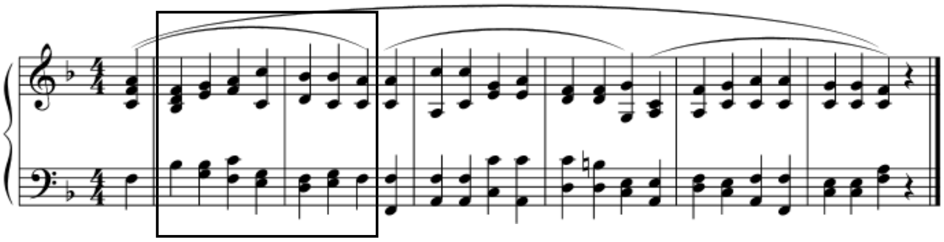


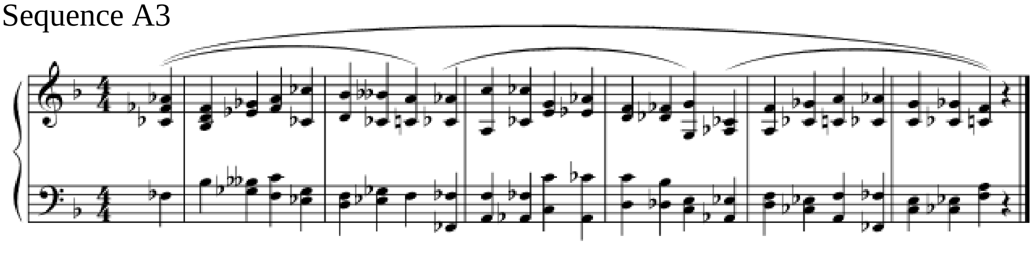
Sequence T2

Based on BWV 12.7 harmonized by J. S. Bach

“Was Gott tut, das ist wohlgetan”

Sequence A2

Sequence T3

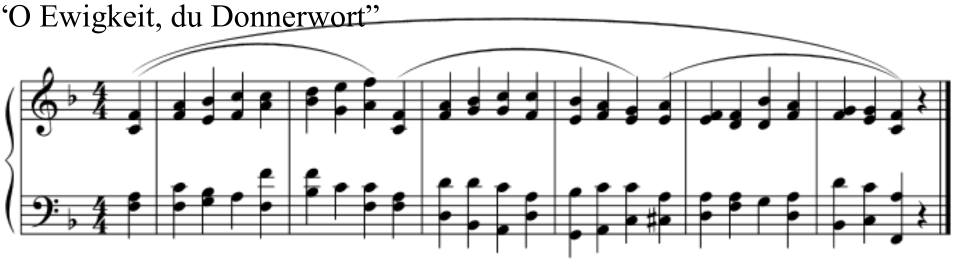
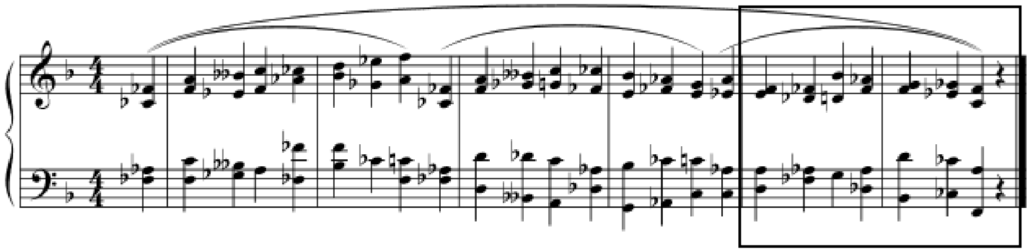
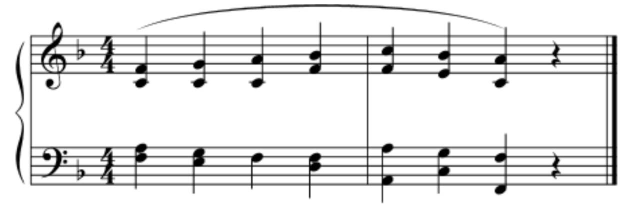
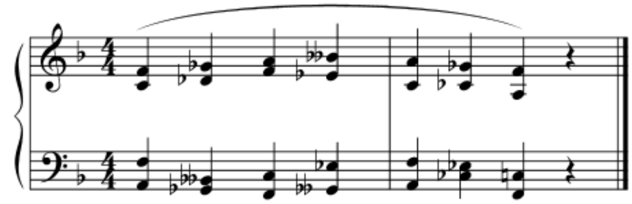
Based on BWV 13.6 harmonized by J. S. Bach

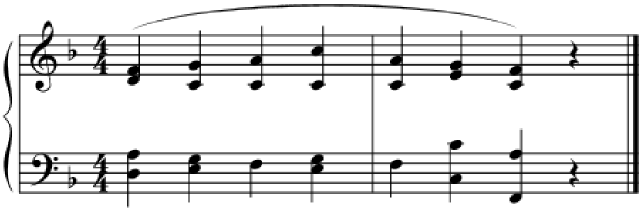
“O Welt, ich muβ dich lassen”

Sequence A3

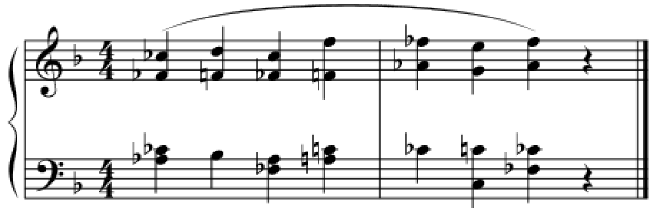


Sequence T4

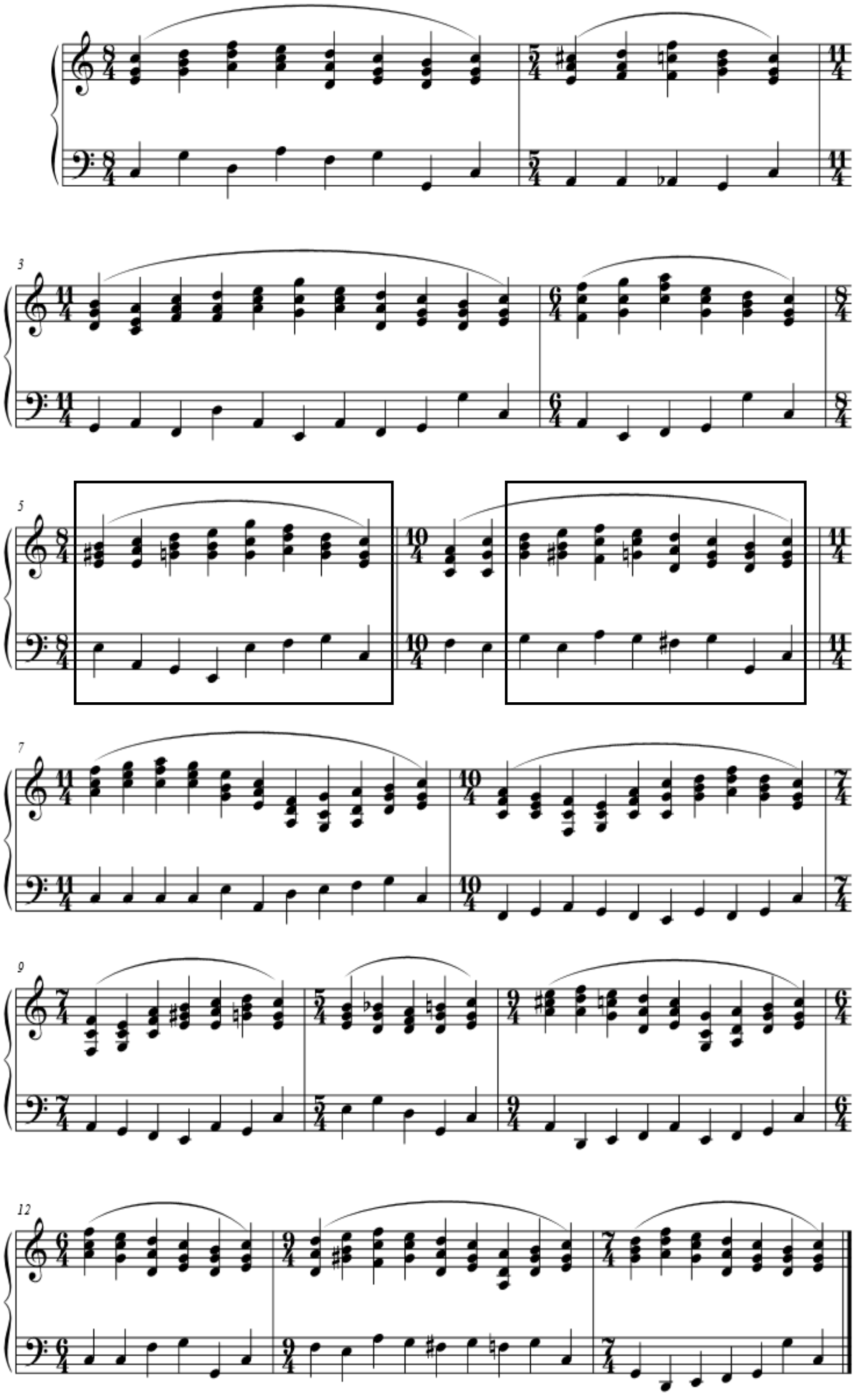
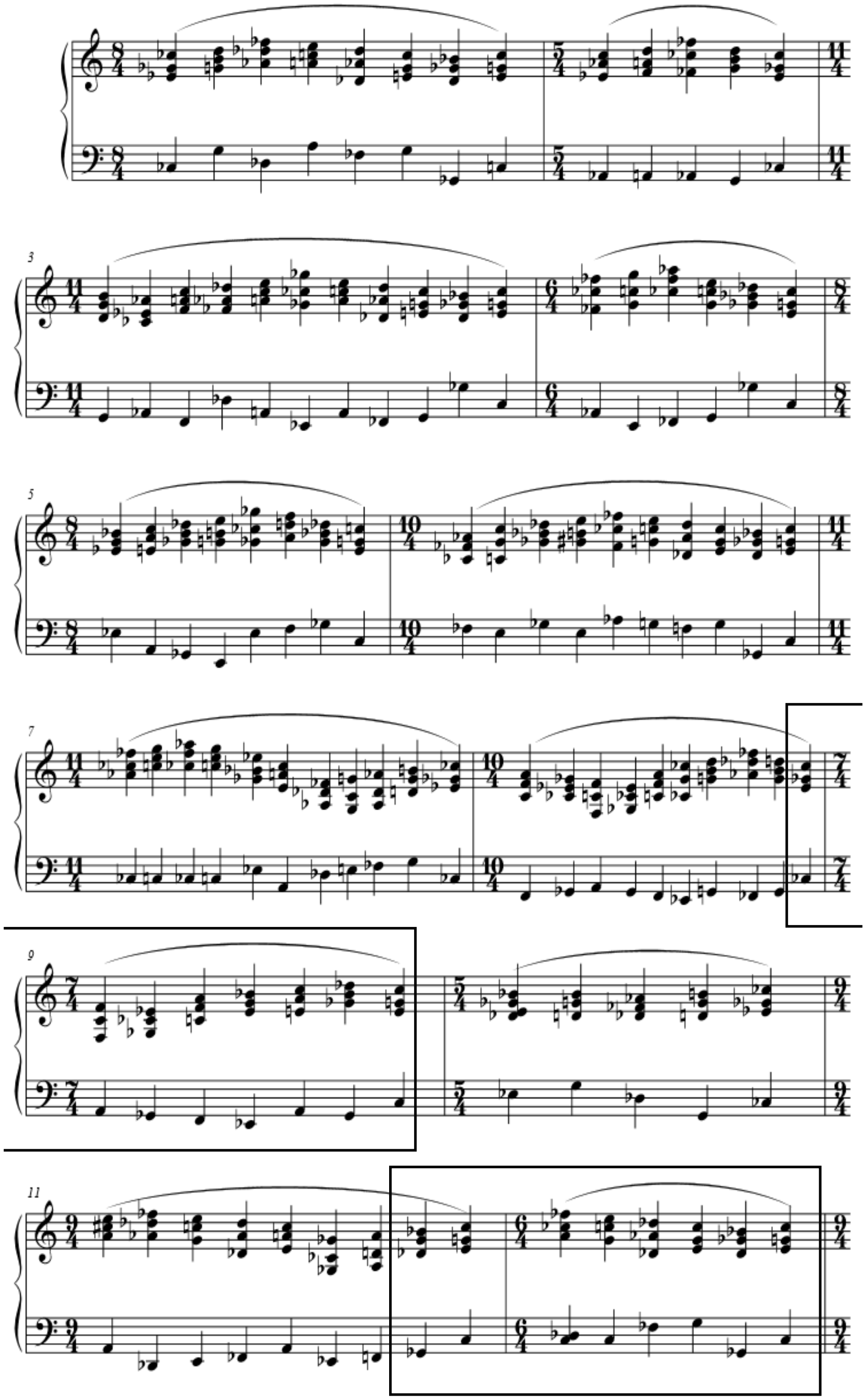
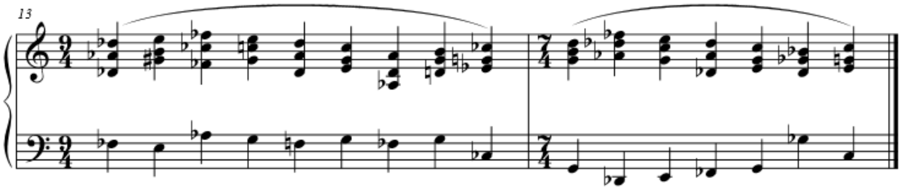
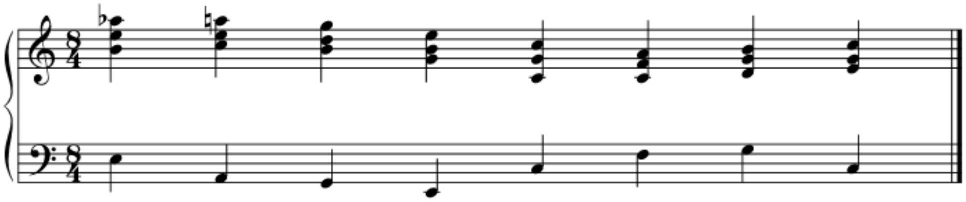
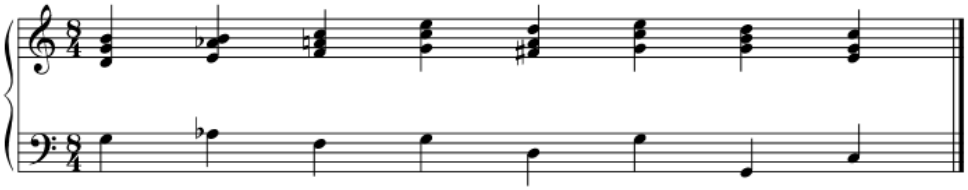
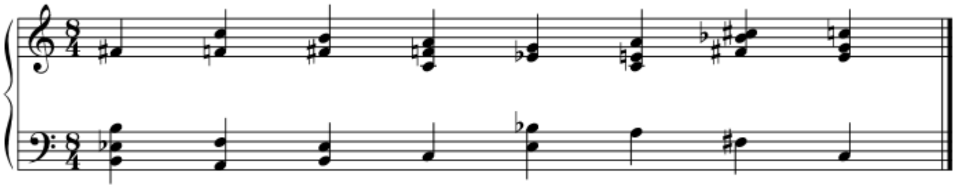
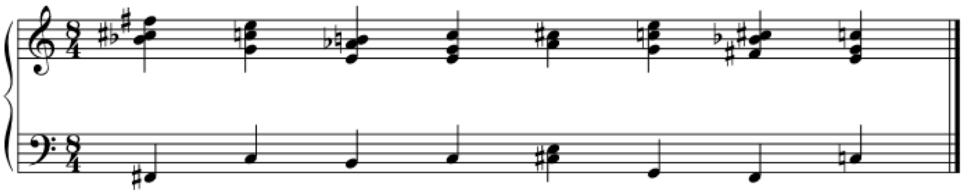
Based on BWV 20.7 harmonized by J. S. Bach

"O Ewigkeit, du Donnerwort"

Sequence A4

Foil 1
Lifted from BWV 37.6 harmonized by J. S. Bach
“Ich dank dir, lieber Herre”

Foil 2
Lifted from BWV 13.6 harmonized by J. S. Bach
“O Welt, ich muβ dich lassen”

Foil 3
Lifted from BWV 66.6 harmonized by J. S. Bach
“Christ ist erstanden”



Foil 4
Lifted from BWV 72.6 harmonized by J. S. Bach
“Was mein Gott will, das g’scheh allzeit”

Experiment 2

The tonal and atonal sequences used in Experiment 2. The length of the phrase changed randomly between 5 and 11 chords, as indicated by the time signature.

Tonal Sequence

Atonal Sequence

Foil 1

Foil 2

Foil 3

Foil 4

Notes on Key and Harmony

Most Western music is composed from a seven-note subset of the *chromatic* notes that divide the octave into 12 equally spaced intervals (on a log frequency scale), with the pitch distance between adjacent tones of the chromatic scale separated by 1/12th octave (semitone). Though different scales use somewhat different subsets of notes from the chromatic scale, for the purposes of this paper, we focus on the Major scale, for which the pitch distances between successive scale degrees (notes) follow the pattern: tone (i.e., two semitones), tone, semitone, tone, tone, tone, semitone. The *key signature* (e.g., C Major) refers to the first note of that subset, which is the tonal "center" (and scale degree 1) of the key. For example, the white notes of the piano, beginning on C, form a Major scale (C, D, E, F, G, A, B, C). The scale degrees form a hierarchy of stability (the tonal hierarchy), enabling notes to be used for different structural purposes throughout a musical piece (e.g., highly stable notes are more likely to occur at boundaries than unstable notes; Aarden, 2003; Krumhansl, 1990). Notes function according to their relationship to the tonal center rather than their absolute pitch.

In music, it is often the case that more than one note is sounded simultaneously. Each scale degree is associated with a triad, a three-note chord with that scale degree as its root and in which one note is skipped between each of the triad notes. For example the Major triad that has C as its root contains the notes C E G. Chords, like notes, function in relation to the tonal center.

Individuals with formal musical training have explicit knowledge about these relationships, but even untrained listeners provide stability ratings consistent with knowledge of the formal tonal hierarchy (Cuddy & Badertscher, 1987; Krumhansl & Kessler, 1982; Steinke et al., 1997), are better at detecting changes in melodies that violate these relationships than changes that do not (Trainor & Trehub, 1992, 1994), and show an early anterior negativity ERP component to irregular, unexpected notes or chord but not expected notes (Koelsch et al., 2007; Koelsch & Jentschke, 2008, 2010).

Chapter 3: Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues

Kragness, H. E. & Trainor, L. J. (2018). Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues. *Developmental Psychology, 54*(5), 842-456. http://dx.doi.org/10.1037/dev0000405

Preface

In Chapter 2, I found that nonmusicians lengthen phrase boundaries when self-pacing through chord sequences, and that phrase-final lengthening can be elicited by harmonic cues in adult nonmusicians. In the present chapter, I examine the developmental trajectory of this behavior. In Experiment 1, I show that that 3-year-old children demonstrate boundary lengthening when self-pacing through chord sequences adapted from Bach chorales. In Experiment 2, I show that 4-year-old and 7-year-old children, but not 3-year-old children, lengthen boundaries at harmonic cadences when controlling for metrical regularity and melodic contour. These experiments show that children as young as 3 years old associate expressive timing with phrase structure, and further show that there are significant gains in understanding how harmony relates to phrase structure between 3 and 4 years old.

Abstract

Proper segmentation of auditory streams is essential for understanding music. Many cues, including meter, melodic contour, and harmony, influence adults' perception of musical phrase boundaries. To date, no studies have examined young children's musical grouping in a production task. We used a musical self-pacing method to investigate (1) whether dwell times index young children's musical phrase grouping and, if so, (2) whether children dwell longer on phrase boundaries defined by harmonic cues specifically. In Experiment 1, we asked 3-year-old children to self-pace through chord progressions from Bach chorales (sequences in which metrical, harmonic, and melodic contour grouping cues aligned) by pressing a computer key to present each chord in the sequence. Participants dwelled longer on chords in the eighth position, which corresponded to phrase endings. In Experiment 2, we tested 3-, 4-, and 7-year-old children's sensitivity to harmonic cues to phrase grouping when metrical regularity cues and melodic contour cues were misaligned with the harmonic phrase boundaries. In this case, 7- and 4-year-olds but not 3-year-olds dwelled longer on harmonic phrase boundaries, suggesting that the influence of harmonic cues on phrase boundary perception develops substantially between 3 and 4 years of age in Western children. Overall, we show that the musical dwell time method is child-friendly and can be used to investigate various aspects of young children's musical understanding, including phrase grouping and harmonic knowledge.

Keywords: grouping, phrase, music, timing, harmony, phrase-final lengthening

Introduction

It is commonly accepted that music is present in daily life across virtually all human cultures (e.g. Brown, 1991; Huron, 2001). Exposure to music likely begins before birth and music is prominent in infants' early environments through parents singing lullabies and play songs (e.g., Ilari, 2003, 2009; Trehub & Schellenberg, 1995; Trehub & Trainor, 1998; Trehub, Unyk, & Trainor, 1993; Young, 2008). Even for non-musicians, music continues to have a daily influence across the lifespan, being a part of activities such as shopping at the mall, parties, sporting events, and religious ceremonies (Dissanayake, 2006). By adulthood, most people have experienced thousands of hours of music listening or participation in informal musical activities. Through passive exposure alone, listeners develop expectations for the pitch and timing patterns common to music in their culture (Bigand & Poulin-Charronnat, 2006; Tillmann, Bharucha, & Bigand, 2000).

Music typically consists of hierarchically nested groups of sound events. In melodies, notes are organized into short rhythmic groups, which are joined together to build phrases. The ability to parse a stream of contiguous musical events into its component phrases is essential for understanding music (Lerdahl & Jackendoff, 1983). When asked to identify the location of phrase boundaries when listening to music, there is considerable consistency among listeners, even amongst those without musical training (Clarke & Krumhansl, 1990; Deliège, 1987; Palmer & Krumhansl, 1987; Peretz, 1989; but see Pearce, Müllensiefen, &

Wiggins, 2010). A number of findings suggest that phrase boundaries additionally act as anchors for attention in music listening. For example, participants report clicks heard during music listening as having been closer to phrase boundaries than they actually were (Gregory, 1978; Sloboda & Gregory, 1980). Thus, phrase boundaries constitute an important component of musical understanding. However, little is known about how understanding of musical phrases is instantiated across development.

Although the mechanisms by which perceptual grouping of music occurs are still debated (e.g. Cambouropoulos, 2001; Narmour, 1990; Pearce & Wiggins, 2006; Schellenberg, 1997), it is generally agreed that at its simplest level, musical grouping can arise from Gestalt-like principles, such as proximity and similarity. This kind of grouping seems to be present quite early in development. Infants' grouping has been probed by presenting a sequence of tones that contains a change in some feature, such as pitch height or timbre (for example, in AAABBB formation), and introducing a pause either just before the change (between groups) or at some other point (within a group). As young as six months, infants more readily detect within-group than between-group pauses when groups differ in loudness, pitch height, and timbre, which suggests that infants perceptually group tones based on these features (Thorpe, Trehub, Morrongiello, & Bull, 1988; Thorpe & Trehub, 1989).

There are many other cues associated with musical groups, especially in naturalistic music. For example, in the Western musical tradition, it is common for the *melodic contour* of a phrase (the up and down movement of the lead voice, usually the highest in pitch) to follow an "inverted U" pattern, with the melodic line rising and then falling over the course of the phrase (Huron, 1996). Metrical structure represents another contributor to hierarchical groups in music. In music, *meter* refers to the regular pattern of stressed and unstressed beats (for example, stresses every second or every third beat create groups of two or three beats, respectively, and correspond to "march" or "waltz" meters, respectively). Though meter and grouping are functionally distinct, the length of musical phrases is related to meter – for example, pieces in 4/4 time often have phrases or sub-phrases that contain 4, 8, or 16 beats, while pieces in 3/4 time usually have phrases that contain 6 or 12 beats.

It is possible for different grouping cues to conflict, and the relative contribution of different cues to the perception of musical phrases, especially for young children, is not well understood. One study probing infants' phrase perception in naturalistic music found that they preferred to listen to Mozart minuets that had pauses inserted *between* phrases over versions with pauses inserted *within* phrases (Krumhansl & Jusczyk, 1990), suggesting that they expect lengthening at phrase endings. The musical cues to phrase boundaries in these studies included falling pitch and longer durations, which tend to occur at the ends of phrases (Jusczyk & Krumhansl, 1993). Overall, the capacity to group tones is present relatively early in development, though few studies have attempted to investigate the development of children's use of different cues in phrase perception.

One cue that is tightly associated with group-endings in music performance is lengthened duration. *Phrase-final lengthening,* the tendency for music performers to slow on phrase endings, has been well-documented (Repp, 1992a; Seashore, 1938; Todd, 1985). One study reported that performers lengthen boundaries even when they are attempting to play mechanically, although it is difficult to say whether this is due to practice or a perceptual bias (Penel & Drake, 1998). Listeners expect phrase-final lengthening, as well. When participants are asked to detect lengthened notes in a musical sequence with notes of even durations, they are less able to detect lengthened phrase-final notes than lengthened within-phrase notes (Repp, 1992b, 1999), suggesting that listeners expect boundary notes to be longer. Both infants and adults appear to experience relatively long tones as group-ending in simple tone sequences (Trainor & Adams, 2000; although this effect may be modulated by language experience (see Iversen, Patel, & Ohgushi, 2008 and Yoshida et al., 2010). Phrase-final lengthening is found in speech production, as well, and has been reported in a number of different languages (e.g. Oller, 1973; Turk & Shattuck-Hufnagel, 2007; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992), in infant-directed as well as adult-directed speech (Koponen & Lacerda, 2003), and even in the babbled phrases of deaf infants (Nathani, Oller, & Cobo-Lewis, 2003). Though the origins of linguistic group-final lengthening are still debated, it is evident that lengthened duration in phrase-final positions is widespread in speech as well as music, even at young ages.

One cue for closure at the phrase level relies on functional relationships between musical notes and chords. Western music is typically based on a scale, or collection of notes (e.g., the Major Scale) and a musical piece is said to be in a particular *key* depending on the particular collection of notes on which it is based. The notes of a scale form a *tonal hierarchy* wherein notes of the scale function in relation to a reference note. The reference note (the *tonic*) is repeated often, particularly in structurally important positions such as phrase boundaries, and is judged to be the most stable (Krumhansl, 1990). Notes not in the scale of a piece are highly unstable. Within the notes of the scale, some are more stable than others. Notes that are more stable in the tonal hierarchy are preferred at phrase-final positions (Boltz, 1989; Aarden, 2003).

Western music additionally employs *harmony*, often represented by *chords* (clusters of notes that are sounded simultaneously with the melodic line; see Aldwell & Schacter, 2002). As with notes, some chords in a sequence (*harmonic progression)* are more stable than others. Both musically trained and untrained Western adults have preferences and expectations about the order of chords in harmonic progressions that are consistent with the descriptions formalized in Western tonal music theory (e.g. Bigand & Poulin-Charronnat, 2006; Bharucha, 1984; Bigand, 1997; Krumhansl, 1979; Krumhansl & Kessler, 1982; Tillmann, 2005). The influence of harmonic cues is also evident in phrase perception. Listeners expect that phases will end with relatively stable chords. Participants better recognize musical excerpts that that do not cross harmonic boundaries than excerpts that cross harmonic boundaries (Tan, Aiello, & Bever, 1981), suggesting that harmonic boundaries influence the encoding of musical sequences.

Less is known about how young children relate pitch information to phrase boundaries. By 6 months, infants can discriminate pitches differing by much less than a semi-tone, which is the smallest functional unit difference between notes in Western music (e.g. He, Hotson, & Trainor, 2009; Olsho, 1984; Olsho, Schoon, Sakai, Turpin, & Sperduto, 1982). Infants of this age can also remember short melodies for days (Plantinga & Trainor, 2005), and discriminate between consonant and dissonant note pairs (pairs of frequencies that can be represented as simple and complex ratios, respectively; Masataka, 2006; Trainor & Heinmiller, 1998; Trainor, Tsang, & Cheung, 2002; Zentner & Kagan, 1998).

Although infants' pitch perception and melodic memory is already quite sophisticated by 6 months, understanding of the tonal hierarchy, including key and harmony, appears to have a more protracted development. While six-month-old infants can detect mistuned notes in native-scale and non-native-scale sequences equally well, adults have an advantage for detecting mistuned notes in music that uses a native musical scale, suggesting that a process of enculturation to native pitch structure occurs sometime after six months (Lynch, Eilers, Oller, & Urbano, 1990). Similarly, Western adults' detections of melodic changes are superior for changes outside the scale on which the piece is based (violating key membership) than changes consistent with the key, but 8-month-old infants detect both types of changes equally well (Trainor & Trehub, 1992). In North American infants, enculturation to Western key structure ("tonality") seems to take place after 12 months for the most part, although some general familiarity with the major scale appears to be present around 12 months (Lynch & Eilers, 1992), and sensitivity to tonality at this age can be enhanced by active musical experience between 6 and 12 months (Gerry, Unrau, & Trainor, 2012). Such findings parallel results from research in language development, which suggest that infants' perceptual abilities are "universal" at birth, but narrow to reflect the stimuli in their environments around 12 months (e.g. Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; Werker & Tees, 2005).

Studies asking children to rate the "goodness" of probe tones found evidence for knowledge of key membership as young as six years, the youngest age tested (Cuddy & Badertscher, 1987; Krumhansl & Keil, 1982). Trainor and Trehub (1994) found that both 5-year-olds and 7-year-olds demonstrated knowledge of key membership by easily detecting changes in a melody that went outside the key. However, they found that harmonic knowledge appears to develop later; 7-year-olds better detected changes that remained within the key but violated implied harmonic structure compared to 5-year-olds. In another study, children aged 6 years and older also demonstrated facilitated processing for stable chords in a harmonic priming study, showing faster reaction times to stimuli paired with expected chords than unexpected chords (Schellenberg, Bigand, Poulin-Charronnat, Garnier, & Stevens, 2005), suggesting that 6-year-olds have some harmonic knowledge.

Recent studies utilizing more child-friendly approaches suggest that some key membership and harmonic knowledge is likely present earlier than previously thought. One such study employed an interactive forced-choice paradigm in which children viewed videos of puppets that appeared to be playing a song on the piano. Children were tasked with awarding a prize to the puppet that played better music (Corrigall & Trainor, 2014). Five-year-old children awarded fewer prizes to puppets whose chord sequences ended with an out-of-key chord than to puppets whose chord sequences ended with an in-key chord, but did not prefer sequences ending with in- vs. out-of-harmony chords. Four-year-olds, in contrast, did not award more prizes to in-key *or* in-harmony versions. However, EEG recordings showed a bilateral positive event-related potential (ERP) component in response to out-of-key chords compared to in-key chords in 4-year-old children, suggesting that they are sensitive when chords are out-of-key, even though it was not exhibited in the behavioral task of awarding prizes. Other ERP studies have shown that the signature waveform associated with harmonic syntactic violations, ERAN (early right anterior negativity), is present by 5 years (Jentschke, Koelsch, Sallat, & Friederici, 2008; Koelsch et al., 2003), and one study has reported an ERAN-like response to music-syntactic violations in children as young as 30 months (Jentschke, Friederici, & Koelsch, 2014). Finally, Gerry et al. (2012)

reported that 12-month-old infants involved in early Suzuki music classes preferred to listen to a tonal compared to atonal version of a classical piano piece, although infants without musical training did not show any preference. Thus, it seems likely that young children have some implicit tonal knowledge, though they may be unable to meet some behavioral demands of some of the tasks used to date. Overall, we can be quite confident that some key membership and harmonic knowledge is established by the age of 5, but very little is known about the development of this knowledge between 1 and 5 years, largely due to methodological difficulties in testing children at these ages.

Another way to probe children's knowledge is to use a task that is behavioral in nature but does not require any explicit judgments. In one such procedure, a *dwell time* method, participants repeatedly press a computer key to self-pace through a sequence (e.g., a chord sequence). They are not given any instructions for *how* to time their key presses and they experience the sequence as it unfolds. The dwell time method has previously been used to investigate action segmentation, showing that participants "dwell" (spend relatively more time on) slides depicting event boundaries (Hard, Recchia, & Tversky, 2011). Preliminary work has found converging results with preschoolers and infants (Meyer, Baldwin, & Sage, 2011; Sage, Ross, & Baldwin, 2012). A musical adaptation of the dwell time method recently used with adults demonstrated that even non-musicians dwelled on phrase-final chords in chord sequences (Kragness & Trainor, 2016). Participants do not need special knowledge or instructions in order to perform the dwell time task, and they do not need to make any verbal responses, which makes it useful for exploring the implicit knowledge of young children.

In the following experiments, we adapted the musical dwell time method to explore the developmental trajectory of children's understanding of musical phrase boundaries and use of harmonic cues to phrase closure. Because previous studies of children's phrase understanding have used perceptual judgments, we were particularly interested in whether young children would show phrase-final lengthening in the *production* of music using the dwell time method. Second, we investigated whether young children would lengthen harmonic phrase boundaries when they were misaligned with other phrase boundary cues, as adults have been shown to do (Kragness & Trainor, 2016). We were particularly interested in 3- and 4-year-olds as previous studies have suggested some sensitivity to harmonic knowledge at age 4 but not at age 3. In Experiment 1, we tested only 3-year-olds and used a chord sequence from J. S. Bach with phrase boundaries that were regularly spaced (every 8 chords) to demonstrate that dwell times can index 3-year-old children's segmentation of chord sequences in naturalistic music. Specifically, we expected that preschoolers would spend more time on phrase-final chords (hereafter referred to as "boundary" chords) than non-final chords (hereafter referred to as "nonboundary" chords), reflecting segmentation of the sequences into phrases. In Experiment 2, we tested children's sensitivity to harmonic phrase boundaries by using chord sequences in which metrical and melodic cues to boundaries were not aligned with the harmonic boundaries. We tested 3- and 4-year-olds in Experiment 2 as previous studies suggest that this might be a critical time for the development of sensitivity to harmonic information, and tested 7-year-olds because we expected harmonic sensitivity to be well in place at this age. We additionally included tests of phonemic blending, segmentation, and digit span to examine whether individual differences in the music task were related to mechanisms that underlie language or working memory development. To our knowledge, no previous studies have examined potential relationships between children's language abilities and music segmentation, so these tests were exploratory. Thus, this research addresses the important questions of how early children produce phrase-final lengthening in music and when they develop sensitivity to harmonic cues to phrase boundaries.

Experiment 1

Method

Participants. 37 3-year-olds were recruited to participate. Nine children were excluded due to parent interference (3), shyness or fussiness (4), or equipment failure (2), resulting in a final sample of 28 participants ($M_{\text{age}} = 3.56$ years, $SD_{\text{age}} = 0.09$, 14 boys, 14 girls). A power analysis conducted in the software G*Power (3.1.9.2) estimated a sample size of 22 would be sufficient to detect a significant result for an effect size Cohen's $d_Z = .62$ (alpha = .05, 80%

power). This effect size estimate was based on the contrast of interest (difference between boundaries and nonboundaries in the Bach chord sequence) from a previous study with adults (Kragness & Trainor, 2016). Since children's performance is expected to be more variable than adults', we recruited a number of additional participants for a final sample of 28 after exclusion.

All children spoke English. Five parents reported that their child had significant exposure to another language, with other parents reporting very little or no exposure to languages other than English (French, Cantonese, Cantonese/Mandarin, Italian, Filipino/Arabic; see Table 1 for more information). One parent reported that their child attended early music classes, but no child had taken any formal instrumental classes. All but four parents indicated that their child predominately used their right hand to complete tasks: three reported that their child used both hands equally and one reported left hand dominance. Children received a small prize as compensation.

Stimuli. The stimuli consisted of four different 4-voice major mode chorale excerpts, harmonized by J. S. Bach (see Supplemental Material). These stimuli were previously validated with adult participants in a similar experimental design (Kragness & Trainor, 2016). Each of the four sequences contained 24 chords (one every quarter note). Every eighth chord (chords in position 8, 16, and 24) was a "boundary" (phrase-final) chord. Eight-chord phrases are very common in Western music. Minor alterations were made to remove eighth and sixteenth notes that fell between quarter note chords. Chorales were all transposed to F

Table 1
Demographic breakdown of participants in all experiments

Experiment	Experiment 1	Experiment 2		
	3-year-olds	3-year-olds	4-year-olds	7-year-olds
N	28	26	18	19
Gender distribution	14 boys, 14 girls	13 boys, 13 girls	7 boys, 11 girls	10 boys, 9 girls
Age (years), M(*SD)*	3.56 (*0.09*)	3.55 (*0.11*)	4.77 (*0.25*)	7.70 (*0.25*)
Average hours music listening/week	9.22 (*6.35*)	9.19 (*7.35*)	9.94 (*9.71*)	7.00 (*9.44*)
Number reporting > 10% exposure to non-English	5	5	4	5
Languages	French, Cantonese, Cantonese/Mandarin, Italian, Filipino/Arabic	French (3), Arabic, Korean	Italian/Spanish, French, French/Polish Filipino	French (3), Polish/Italian, Tamil
Distribution of family income	60k or less (6) 60k-120k (8) 120k or more (12)	60k or less (6) 60k-120k (8) 120k or more (10)	60k or less (5) 60k-120k (8) 120k or more (4)	60k or less (1) 60k-120k (8) 120k or more (10)
Number in early music classes (past or present)	1	7	4	3
Formal music training	0	0	0	Piano (3), violin (1), guitar (1), drums (1)
Formal dance training (past or present)	Ballet (2), creative movement (2), dancing toddlers (2), intro to dance (1)	Ballet (4), creative movement (1), acro (1)	Jazz (2), creative movement (1), ballet/creative movement (1), ballet/hip-hop (1)	Ballet (1), creative movement (1), hip-hop (1), mixed (1), jazz/tap (1), ballet/jazz/tap (1)

Major, ensuring that a key change did not alert participants to the beginning of a new chorale. Each chord was generated in GarageBand software with the default piano timbre set at a constant level. Stimuli were presented with Presentation 16.1 06.11.12 (Neurobehavioral Systems) through speakers (Altec Lansing Amplified Speaker System, ACS2213W) at a comfortable level. Participants experienced all four sequences played in the same order twice for a total of 192 chords (24 phrase boundaries, 168 non-boundaries).

Survey. After acquiring consent, parents were asked to fill out a survey about their child's home life, language experiences, and musical experiences while the child completed the tasks of the experiment.

Self-Pacing Procedure. The child was seated at the computer, which first displayed a picture of a green cartoon frog, who was introduced as "Freddie" (Figure 1). The next slide showed the same cartoon of Freddie as a *purple* frog. The experimenter explained that Freddie was under a spell that turned him purple, and that the only way to turn him green again was to play beautiful music, "not too fast, not too slow, but just right." After the explanation, a green "GO" sign appeared on the screen, which signaled that the participant should begin to play the music.

Participants controlled the onset of each chord by pressing the "enter" key on the number pad. This was the key furthest to the right and closest to the bottom, and thus the easiest to access without hitting other keys. The chord only

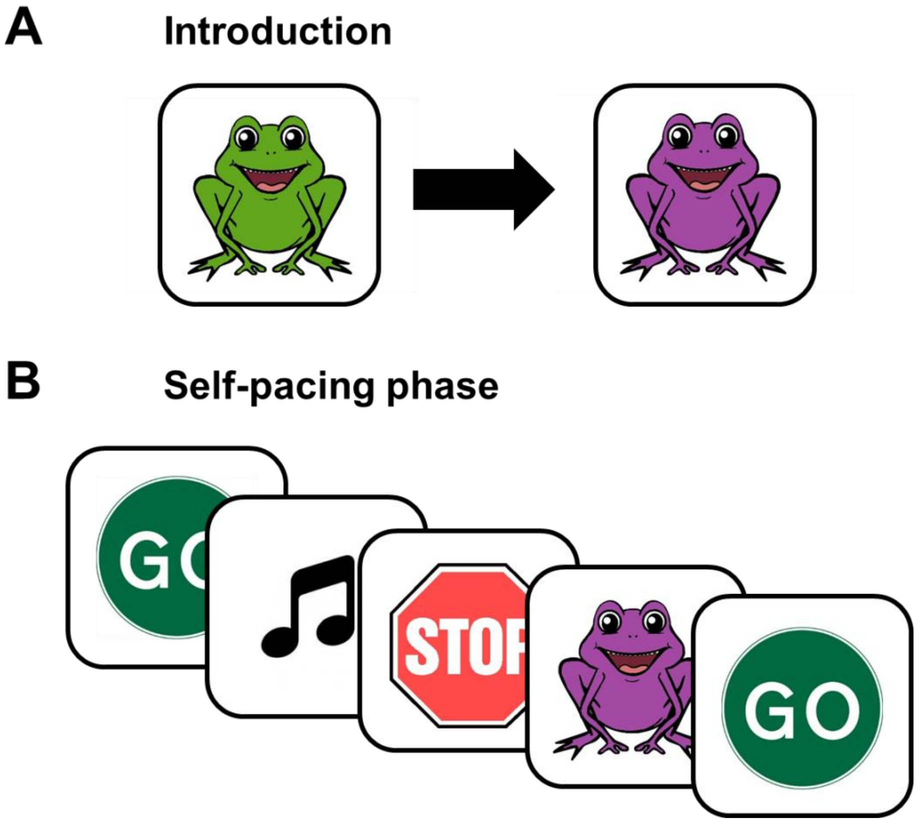


Figure 1

A representation of the procedure as experienced by the participants.

(A) First, children were introduced to Freddie, the green frog who had been turned purple. They were told beautiful music would turn him green again. They were instructed to play “not too fast and not too slow, but just right.” (B) In the self-pacing phase, children saw a “GO” sign, indicating they could begin to press the number pad “enter” key to play music. They heard 29-39 chords, and then saw a “STOP” sign, which indicated to stop pressing the “enter” key. They then saw Freddie turn partially green. During the transformation slide, the experimenter offered verbal praise and encouraged the participant to continue when the “GO” sign reappeared after the transformation.

stopped playing when participants pressed the same key to trigger the next chord. Thus, participants only controlled the onset of the next chord. They could not control the specific chord played and did not know what was coming next. The time of each key press was recorded.

A short break was given every 25 to 40 chords during which children could see Freddie turning more green. A series of pseudorandom numbers between 25 and 40 were generated, such that the same number was not used consecutively. These numbers were used to determine how many chords children played before the "STOP" sign appeared. For example, because the first two numbers generated were 33 and 34, the first "STOP" sign followed the 33rd chord, and the second "STOP" sign followed the 67th chord (34 chords after the last "STOP" sign). After each "STOP" sign, there was a brief animation of Freddie turning partially green. During this time, the experimenter offered verbal praise and asked the participant to continue playing the music when a "GO" sign re-appeared. After each "GO" sign, rather than re-starting at the next chord in the sequence, the sequence re-started with the first chord of the current phrase. This was done so that participants heard each chord in the context of a complete phrase. The maximum number of chords that were repeated directly prior to the "STOP" sign and directly after the "GO" sign was three. Stopping points were not systematically aligned with phrase boundaries, and did not differ across participants. The frog's transformation happened in the same way at the same time across participants regardless of the child's performance. This procedure was used to create a game-like atmosphere, in order to motivate children to continue in the experiment without interfering with the task. When the participant had played through the sequence fully, the final animation showed Freddie turning completely green.

Prior to testing, all children participated in a short training phase, which emulated this task but with a different cartoon character and the first fifteen beats of the melody *Frère Jacques*[1]. The only difference was that the experimenter played the first three notes to demonstrate, and encouraged the participant to play the rest of the notes, simulating a tapping motion to encourage them.

Additional Tests. After participating in the main task, children participated in several additional tests to examine potential correlations between knowledge of musical phrase boundaries and language skills that might share an underlying mechanism. Auditory working memory and manipulative memory were assessed with the Digit Spans Forward and Backward, respectively (WISC-IV; Wechsler et al., 2004). To examine children's awareness of blending and segmenting phonemes/words, we used a blending task (Helfgott, 1976) and the Test of

---

[1] The first fifteen beats of *Frère Jacques* include a half note. For the training session, this note was played twice so that the melody would sound correct if participants played each note with approximately similar lengths. This was done because participants did not know what they would be playing ahead of time, and we did not want the training session to be metrically confusing.

Auditory Analysis Skills (TAAS; Rosner, 1975)[2]. Finally, we tested participants' receptive vocabulary on the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007) to ensure that poor performance in the self-pacing task was not the result of misunderstanding task instructions.

Results

Dwell times. Durations for each chord ("dwell times") were calculated by subtracting the time of the key press terminating the chord (eliciting the onset of the next chord) from the time of the key press that elicited the chord onset. Some chords were heard more than once in close proximity, right before and then again directly after a STOP signal and frog transformation (see Procedure section). For those chords, we only included the *first* time they were heard, ignoring dwell times for these chords the second time. This was done because we wanted to assess dwell times for the first time participants heard each chord, and we wanted to measure dwell times for each chord within the context of the full sequence (see Procedure section).

---

[2] In pilot testing, many participants demonstrated discomfort with Item 7 of the TAAS (in which the test item is "Say *game*. Now say it again, but don't say /m/" and the desired response is *gay*). Because of this, in Experiments 1 and 2, the test item was changed to *name* with the desired response of *nay.*

We first wanted to investigate participants' timing across each eight-chord phrase. Dwell times for each position in the phrase (1-8) were averaged. As can be seen in Figure 2, the longest dwell times were on position 8, the phrase-ending chord. In order to test whether children dwelled longer on phrase-ending chords (position 8) than chords that were not phrase-ending (positions 1-7), each trial was binned as a boundary chord or nonboundary chord (Figure 3). Each participant received an average score for each trial type. A planned one-way paired *t*-test demonstrated that participants' boundary scores were significantly greater than their nonboundary scores (Figure 2, $M_B = 850.045$, $SD_B = 332.600$, $M_N = 737.306$, $SD_N = 198.883$, $t(27) = 2.565$, $p = .016$, Cohen's $d_Z = .485$)[3]. We additionally observed a small local peak in dwelling at chords in position 4, which would represent the sub-phrase boundary. We did a post-hoc *t*-test to test whether chords in position 4 were dwelled on significantly longer than other non-boundary chords, but the contrast was not significant ($t(27) = 0.50$, $p = .620$).

Demographic considerations. One consideration was whether the magnitude of boundary over nonboundary dwelling was greater in older than younger 3-year-olds in our sample. We used the difference score (boundary minus nonboundary) as a measure of magnitude. We then tested for a correlation

---

[3] We used Lakens' (2013) recommendations for reporting effect sizes, reporting Cohen's d_Z for correlated *t*-tests and generalized eta-squared for ANOVAs that had repeated measures.

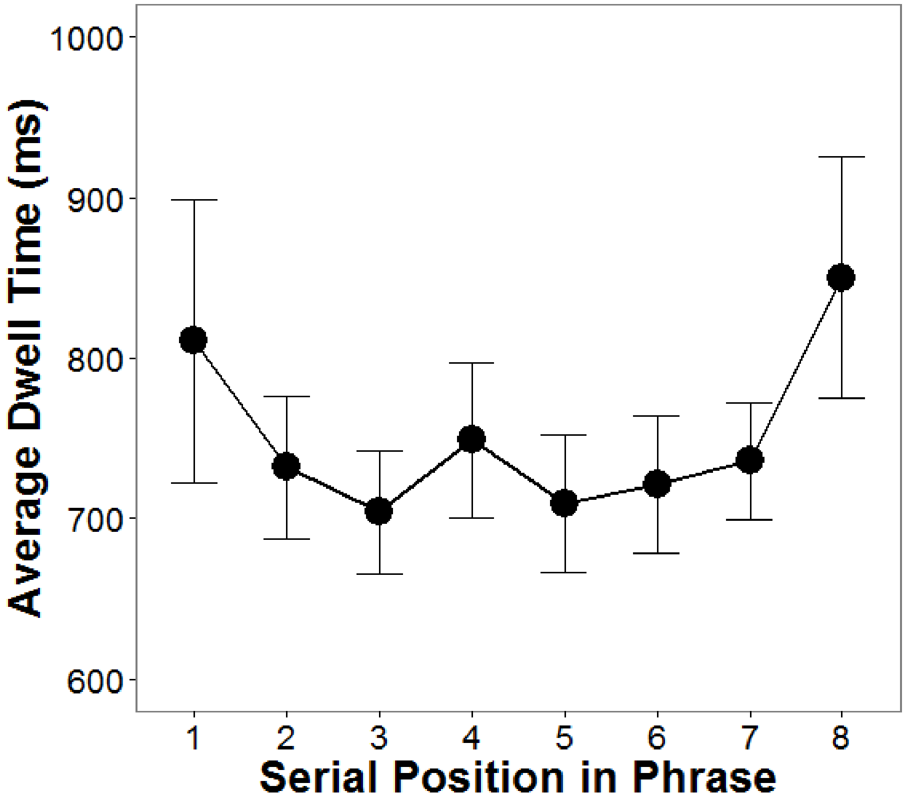


Figure 2

Average dwell times for chords in each position in the phrase. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005).

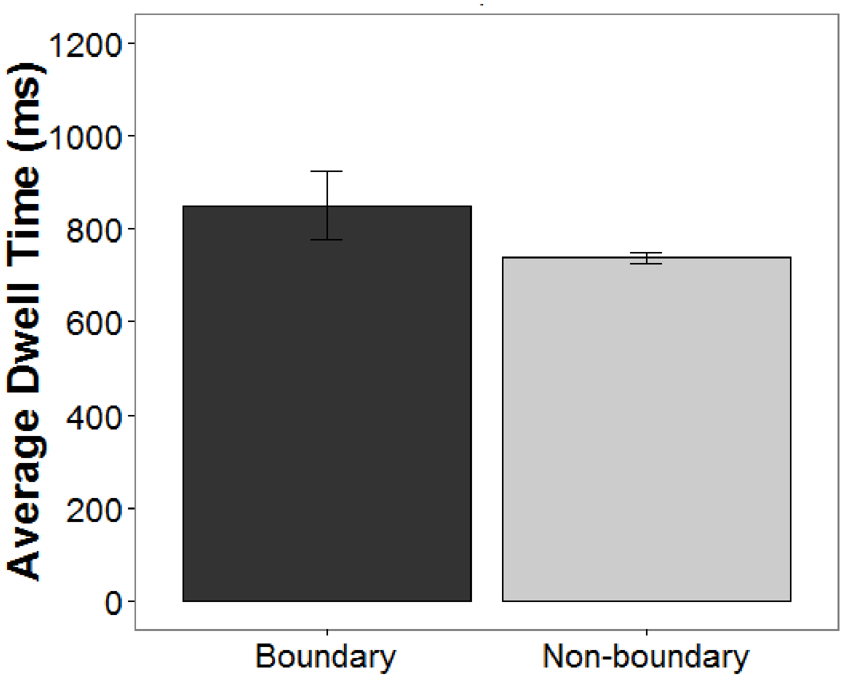


Figure 3

Average dwell times for boundary chords and nonboundary chords. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005).

between actual numerical age (test date minus date of birth) and difference score, but found no evidence for a correlation (Spearman's $r_S = -.215$, $p = .272$; see Table 2). We further tested for a correlation between difference scores and reported hours of music listening per week, to examine whether children with more music exposure would show a stronger effect[4]. Five parents declined to offer information about music listening, and the correlation for the remaining participants was not significant (Kendall's tau = .077, $z(21) = .504$, $p = .614$).

Digit Spans, Blending, TAAS, and PPVT. Because there are no normed scores available for 3-year-old participants for the Digit Spans, Blending, or TAAS, we used raw scores. Three participants did not participate in the DSF. Among the remaining participants, the average score was 3.56 out of a possible 16 points on the Digit Span Forward (*SD* = 1.78, ranging from 0 to 6), and all participants scored 0 out of a possible 16 points on the Digit Span Backward. Therefore, further analyses were done only on the Digit Span Forward (DSF), which is thought to reflect working memory (Wechsler, 2007). Because a Shapiro-Wilk test demonstrated the data were non-normally distributed ($W = .904$, $p = .022$) and included a number of tied ranks, Kendall's tau was employed to test

---

[4] If a parent reported a range of hours, the average of the minimum and maximum values given was used for this analysis (e.g., "7.5" was used instead of "7 to 8").

the correlation between tonal difference scores and DSF scores, which was negative but non-significant (tau = -0.193, $z(23) = -1.269$, $p = .204$).

All but one participant was willing to participate in the Blending task and the average score was 1.44 out of a 10 possible points ($SD = 1.48$, ranging 0 to 6). However, most scores fell within a very small range (0 to 3). Kendall's tau was employed because there were a number of tied ranks. Again, a negative but non-significant correlation was found between performance on the Blending task and difference scores (tau = -.119, $z(25) = -0.802$, $p = .422$).

Two of the participants were too fussy to participate in the TAAS. Of the remaining 26 participants, 20 received a score of 0. Correlations were therefore not run on the remaining 6 participants, whose scores ranged from 1 to 3 out of a possible 13.

Finally, normed receptive vocabulary scores (PPVT) were examined. The average normed score was 112.25, which represents approximately the 76th percentile. Difference scores were not significantly correlated with PPVT scores ($r = .185$, $t(26) = .961$, $p = .346$), suggesting that individual differences in the magnitude of boundary dwelling likely cannot be attributed to differences in comprehension of task instructions.

Experiment 2

In Experiment 1, 3-year-old participants self-paced through chord sequences written by J. S. Bach, in which every eighth chord was a phrase-final chord. It was hypothesized that if participants differentiated between boundaries and nonboundaries, the length of time that they spent on each type of chord would differ. It was found that participants spent more time on boundary chords than nonboundary chords, which is consistent with previous work using dwell time methods for studying musical boundaries in adults (Kragness & Trainor, 2016) as well as for action segmentation (Hard et al., 2011; Meyer et al., 2011).

A major question is the extent to which harmonic cues specifically contributed to the children's perception of closure. As outlined in the Introduction, Western harmony refers to the relative stability of chords in a piece of music. The chord built on the reference note is the most stable, and stable chords tend to be used in structurally important positions, such as phrase-endings. Previous studies have failed to show that children younger than 5 years old are sensitive to the relative stability of some chords over others. The phrase-ending chords in Experiment 1 were harmonically stable, but co-occurred with a number of other cues that might have influenced participants' expectations about phrase-endings and phrase lengths.

One cue that could generate expectations for particular phrase lengths is meter. Meter refers to the patterning of strong and weak beats in musical sequences that leads to temporal regularity in hierarchical levels above the beat level. Sensitivity to metrical cues for grouping emerges early in development. Infants seem to be able to extract patterns of strong and weak beats in rhythmic patterns shortly after birth (Winkler, Háden, Ladnig, Sziller, & Honing, 2009), can detect changes to melodies that disrupt meter as early as 4 months (Hannon & Trehub, 2005a; Hannon, Soley, & Levine, 2011), and associate metrical accents with movement experience by 7 months (Phillips-Silver & Trainor, 2005). By 12 months, Western infants show an advantage for processing metrical disruptions in musical meters that are common in Western music. For example, at this age infants perform better with simple meter music compared to complex meter music (Hannon & Trehub, 2005b, Hannon et al., 2011) and with simple duple meters over simple triple meters (Bergeson & Trehub, 2006; Gerry, Faux, & Trainor, 2010). The Bach chorales were written in duple meter, and eight-chord phrases are very common in Western duple-meter music. Thus, it is possible that the strong duple metrical structure generated expectations for four- or eight-chord phrase lengths in Experiment 1.

A second cue that could influence grouping is melodic contour, the up and down movement of the leading voice (the melody). Western melodies tend to have an arched contour over a phrase (Huron, 1996). Contour is thought to be the most salient of all pitch cues for infant melodic processing (Trehub & Hannon, 2006), as changes to melodies that disturb contour are readily detected (Chang & Trehub, 1977; Trehub, Bull, & Thorpe, 1984) and infants categorize melodies on the basis of contour (Ferland & Mendelson, 1989; Trehub, Thorpe, & Morrongiello, 1987). Pre-school children also demonstrate categorization of melodic sequences based on contour (Creel, 2015; Morrongiello, Trehub, Thorpe, & Capodilupo, 1985) and recognize melodies of familiar songs when the contour is preserved but timing is disrupted (Volkova, Trehub, Schellenberg, Papsin, & Gordon, 2014). Considering young children's sensitivity to contour, it is possible that contour cues may also have contributed to children's processing of the phrase boundaries in the chord sequences in Experiment 1.

The goal of Experiment 2 was to investigate children's perception of harmonic boundary cues. We used a chord sequence that (1) was not metrically regular and (2) had a melodic contour that did not systematically align with harmonic phrase boundaries (previously validated in Kragness & Trainor, 2016). Because previous work has found evidence for some harmonic understanding in 4-year-old children using EEG measures (Corrigall & Trainor, 2014), but not behavioral measures, and little work has been done with 3-year-old children in this domain, we tested both age groups in Experiment 2. We additionally included a sample of 7-year-old children, who should demonstrate robust harmonic understanding based on previous work (e.g. Costa-Giomi, 2003; Krumhansl & Keil, 1982, Schellenberg et al., 2009). Based on this work, we hypothesized that 7-year-olds would dwell on harmonic boundaries. We also predicted that 4-year-olds would dwell on harmonic boundaries, but that 3-year-olds would not.

Method

Participants. Thirty-two three-year-olds, 18 four-year-olds, and 20 seven-year-olds were recruited to participate. Seven children were excluded from the analyses due to experimenter error (3), failing to complete the task due to shyness or fussiness (2, both three-year-olds), apparatus failure (1), or parent-reported hearing problems (1). The final sample included a total of 26 three-year-olds (M_{age} = 3.55 years, SD_{age} = 0.11, 13 boys, 13 girls), 18 four-year-olds (M_{age} = 4.77 years, SD_{age} = 0.25, 7 boys, 11 girls), and 19 seven-year-olds (M_{age} = 7.70 years, SD_{age} = .25, 11 boys, 9 girls). A power analysis conducted in the software G*Power (3.1.9.2) estimated a sample size of 16 would be sufficient to detect a significant result for an effect size Cohen's d_Z = .77 (alpha = .05, 80% power). This effect size estimate was based on the contrast of interest (difference between harmonic boundaries and nonboundaries in the tonal sequence) from a previous study with adults (Kragness & Trainor, 2016, Experiments 2 and 3). As in Experiment 1, since children's performance is expected to be more variable than adults', we recruited a number of additional participants for a final sample of 18-26 in each age group.

As in Experiment 1, parents provided information about their children's language and musical experiences, as well as some demographic information (see Table 1 for more information). All participants were English speakers, with 14 of the 63 participants' parents reporting 10% or greater exposure to one or more other language. Parents reported that 14 had experienced some kind of early music class as an infant, such as Kindermusik (3-year-olds, *N* = 7; 4-year-olds, *N* = 4; 7-year-olds, *N* = 3). None of the 3- or 4-year-olds reported any formal music training, and six 7-year-olds had formal music training (including violin, piano, guitar, and drums). Of the 63 participants, 50 were reported to be right-handed, 5 were reported to be left-handed, and 9 parents reported uncertainty about their child's handedness (eight 3-year-olds and one 4-year-old).

Stimuli. In Experiment 2, rather than using Bach chorales, we used a composition in which phrase length expectations elicited by metrical predictability and melodic contour were misaligned with harmonic boundaries (see Supplemental Material). This provided a test of whether children could use harmonic cues to phrase boundaries. This chord progression was composed by an assistant professor of music theory at McMaster University in the Baroque style for a previous dwell time experiment with adults (Kragness & Trainor, 2016). The number of chords between each harmonic phrase boundary varied pseudo-randomly between 4 and 10 chords such that each harmonic phrase length appeared twice in the sequence, and no adjacent harmonic phrases had the same number of chords. Thus, there was no temporal regularity above the level of the beat that could lead to expectations for groups of particular lengths. The melodic contour was controlled such that the upper voice changed pitch direction every fifth chord. This ensured that the melodic contour did not correlate with harmonic phrase endings. Finally, every harmonic phrase ended with a robust harmonic cue for phrase closure[5] with the most stable chord in the phrase-final position. In total, the sequence was 112 chords long and contained 14 harmonic phrases (and, therefore, 14 boundary chords and 98 nonboundary chords).

An atonal version of this sequence was generated by shifting every other chord (odd numbered chords) down by a semitone (1/12 octave; see Supplemental Material). This technique has been used in previous experiments to disrupt the tonal hierarchy while maintaining the sensory consonance of each chord, as well as preserving the melodic and bass line contours of the tonal version (Gerry et al., 2012; Kragness & Trainor, 2016). Thus, the atonal version of the sequence retained most of the same features of the tonal sequence, but did not contain harmonic closure cues. If participants were sensitive to harmonic cues for closure, they should dwell on chords in the phrase-final position (harmonic boundaries) in the *tonal* version, but not chords in the same position in the *atonal* version.

Children self-paced through the tonal sequence and the atonal sequence once each (with the order counter-balanced within each age group and gender) for a total of 224 chords. There was no break between sequences, so participants were not made explicitly aware that a new sequence had begun. As in Experiment 1, chords were generated in GarageBand software, using the default piano timbre

---

[5] Each phrase concluded with an *authentic cadence*. For more information, see Aldwell & Schachter (2002), pg. 85.

with the sound level was kept constant. Stimuli were presented with Presentation 16.1 06.11.12 (Neurobehavioral Systems) through speakers (Altec Lansing Amplified Speaker System, ACS2213W).

Self-Pacing Procedure. The training (beginning of *Frère Jacques*) and self-pacing procedures were identical to those of Experiment 1, except that because there were more chords (224 compared to 192), the participants experienced one more episode of frog transformation in Experiment 2 than in Experiment 1. As in Experiment 1, after the frog transformation, self-pacing re-started at the beginning of the current harmonic phrase (i.e., the chord that followed the most recent harmonic boundary) rather than the next chord in sequence. The maximum number of chords that were heard both before and after the stopping point was four. Participants were all tested in the same room with the same procedure. Dwell times were derived in the same way as described in Experiment 1.

Additional Tests. As in Experiment 1, we were interested in learning whether working memory, language blending and parsing skills, and receptive vocabulary contributed to participants' performance. Thus, children also participated in the Digit Spans Forward and Backward (Wechsler et al., 2004, to assess verbal working memory and manipulative working memory), a blending task (Helfgott, 1976, to assess children's ability to blend phonemes), the Test of Auditory Analysis Skills (TAAS; Rosner, 1975, to assess children's ability to parse words into segments), and the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007, to test receptive vocabulary).

Results

Dwell times. Dwell times were calculated in the same manner as described in Experiment 1. The only difference was that, because of the tonality manipulation, there were four instead of two trial types: tonal boundary (TB), tonal nonboundary (TN), atonal boundary (AB), and atonal nonboundary (AN). For one participant in the 3-year-old group, one additional dwell time (AN) was eliminated due to a technical error that resulted in a dwell time of 0 ms. Finally, each trial was binned into the appropriate trial type. Each participant thus had an average score for each of the four trial types.

Three 2x2 within-subjects ANOVAs were performed to test the effects of tonality (tonal or atonal) and boundary status (boundary or nonboundary) on dwell times in each age group (see Figure 4). For 7-year-olds, the boundary status x tonality interaction did not reach the threshold for significance (p = .079) and there was no significant main effect of tonality (p = .928), but there was a significant main effect of boundary status (F(1, 18) = 6.560, p = .019, $\hat{\eta}^2_G$ = .009), with longer dwelling at boundaries. For 4-year-olds, there was a significant boundary status x tonality interaction (F(1, 17) = 5.355, p =.033, $\hat{\eta}^2_G$ = .016). Paired directional *t*-tests corrected for multiple comparisons revealed that scores

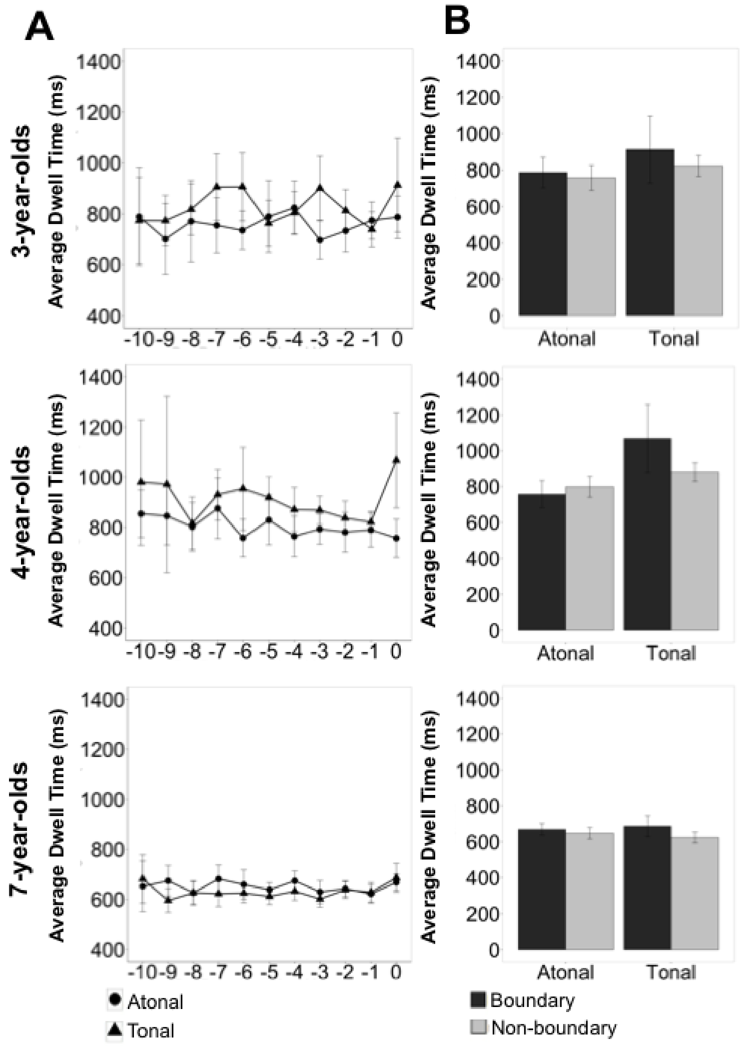


Figure 4

(A) Time profiles showing the average dwell time for chords in each position. Note that because the length of phrases varied from 5 to 11 chords, there are many more data points in the average times for chords in positions 0 to -4 than chords in other positions. (B) Average dwell times for TB, TN, AB, and AN for each age group. Three-year olds did not demonstrate slowing at boundaries, while 4-year-olds slowed at boundaries in the tonal condition only, and 7-year-olds slowed at boundaries in both conditions. Error bars represent within-subjects' 95% confidence intervals (Cousineau, 2005).

were greater for boundary trials than nonboundary trials in the tonal condition for 4-year-olds (M_{TB} = 1067.942, SD_{TB} = 652.247, M_{TN} = 880.680, SD_{TN} = 414.754, $t(17) = 2.078$, $p = .027$, Cohen's $d_Z = .490$), but not in the atonal condition (M_{AB} = 757.163, SD_{AB} = 329.193, M_{AN} = 798.493, SD_{AN} = 348.157, $t(17) = -2.190$, $p = .979$). For 3-year-olds, there were no significant interactions ($p = .547$) or main effects of tonality or boundary status ($p = .256$, $p = .214$, respectively). Because the residuals for 3-year-olds and 4-year-olds were not normal ($W_{3yo} = .754$, $p < .001$; $W_{4yo} = .956$, $p = .013$), the analyses were also performed on log10-transformed data, which did not violate the assumption for normality. All results were consistent with those reported here (see Supplemental Material for ANOVA tables).

Demographic considerations. We first examined whether boundary dwelling was greater for older children within each age group. As in Experiment 1, we operationalized boundary dwell time magnitude by taking the difference scores for each tonality condition for each subject (boundary minus nonboundary). For 7-year-olds, difference scores of raw times were used; for 4- and 3-year-olds, difference scores of log10-transformed dwell times were used. Thus, each participant was assigned a score for both tonal and atonal boundary dwelling magnitude. Because age was distributed non-normally within each of the three age groups ($W_3 = .850$, $p_3 = .001$; $W_4 = .799$, $p_4 =.001$; $W_7 = .885$, $p_7 = .026$), Spearman's rho was used to examine correlations. There were no significant correlations between difference scores and age in any age group (Table 2).

Table 2
Correlations with boundary minus nonboundary difference scores.

			Correlation coefficients						
Experiment	Age group	Tonality condition	Age (rho)	Hours listening (tau)	DSF (tau)	DSB (tau)	TAAS (rho)	Blending (tau)	PPVT4 (r)
Experiment 1									
	3-year-olds	Tonal	-0.215	0.077	-0.193	N/A	N/A	-0.119	0.185
Experiment 2									
	3-year-olds	Tonal	0.201	-0.06	0.117	N/A	N/A	0.221	-0.053
		Atonal	0.149	-0.146	0.069	N/A	N/A	0.009	-0.241
	4-year-olds	Tonal	0.088	-0.104	-0.09	-0.007	0.096	0.007	0.376
		Atonal	-0.013	0.222	-0.12	0.007	-0.127	-0.066	0.202
	7-year-olds	Tonal	-0.428	0.067	-0.284	0.039	-0.459	-0.035	0.303
		Atonal	-0.247	-0.097	-0.168	-0.078	-0.065	0.078	0.088

Note: If either variable failed Shapiro-Wilk test for normality, Spearman's rho was used. For large numbers of tied ranks, Kendall's tau was used. In other cases, Pearson's *r* was used. The correlation between tonal difference scores and TAAS scores was negative (p = .048), but was not significant after correcting for multiple comparisons. No other correlation was significant, with all $p \geq .067$ before correcting for multiple comparisons.

All but 5 parents (7-year-olds, n = 2; 4-year-old, n = 1; 3-year-olds, n = 2) reported an estimated hours of music exposure per week. Because there were many tied ranks, Kendall's tau was used to evaluate correlations with music exposure. Boundary dwelling magnitude did not significantly correlate with reported hours of music exposure in either tonality condition for any age group.

Digit Spans, Blending, TAAS, and PPVT. There were no significant correlations found between the magnitude of the boundary effect in either tonality condition with scores on the DSF, DSB, Blending, TAAS, or PPVT in any age group. The correlation between tonal difference scores and TAAS scores was negative (p = .048), but was not significant after correcting for multiple comparisons. Of the 26 3-year-olds, 21 achieved a score of 0 on the TAAS, and correlations were not tested for the remaining 5. All 3-year-olds scored 0 on the DSB, and 18 of the 26 3-year-olds were willing to participate in the DSF and Blending tasks. For all other correlations, none were significant, with $p \geq .067$ before correcting for multiple comparisons.

General Discussion

The results indicate that the musical dwell time method can be used successfully in children as young as 3 years of age. Specifically, when self-pacing through the chords of Bach chorales in Experiment 1, 3-year-olds dwelled longer on phrase-final chords than chords within phrases. This indicates that 3-year-olds are sensitive to musical phrase structure and, furthermore, that they spontaneously use phrase-final lengthening when producing musical timing. In Experiment 2, however, we found that 4- and 7-year-old children, but not 3-year-old children, dwelled longer on harmonic boundaries than nonboundaries. In this instance, the number of chords between each harmonic phrase boundary varied unpredictably and melodic contour cues were uncorrelated with the occurrence of harmonic boundaries. This study provides the first behavioral evidence that 4-year-old children are sensitive to harmony as a cue for phrase-endings, even in the context of a relatively artificial chord sequence. It also suggests that although 3-year-olds have the ability to parse chord progressions into phrases, harmonic cues are not salient enough to elicit segmentation when other cues, like meter and contour, are not consistent with the harmonic cues.

An abundance of evidence demonstrates that even individuals with no formal music training are highly sensitive to harmonic structure, indicating that mere exposure to music leads to internalization of musical pitch structure (for a review, see Bigand & Poulin-Charronnat, 2006). Different musical systems employ different pitch structures and, through everyday experience with a particular pitch system, children acquire knowledge of that system, just as they learn language without formal instruction (e.g., Hannon & Trainor, 2007; Trainor & Hannon, 2013). There appears to be an orderly progression in musical acquisition, with early sensitivity to key membership (knowing what notes belong in the key a piece is composed in) and later sensitivity to harmony. Previous behavioral studies using infant preference methods have demonstrated some knowledge of key membership as early as 12 months in infants who attended 6 months of Suzuki music classes (Gerry et al., 2012; Trainor et al., 2012). On the other hand, behavioral evidence for harmonic knowledge in children is typically not reported until age 5 or 6 years (Corrigall & Trainor, 2014; Schellenberg et al., 2005), although very basic harmonic knowledge has been reported as early as 4 years when simple, familiar music is used (Corrigall & Trainor, 2009, 2010). The current experiments offer the first behavioral evidence that 4-year-old children are sensitive to a harmonic cue to phrase boundaries, even when listening to unfamiliar chord progressions. Furthermore, the results show that there is a significant increase in this sensitivity between 3 and 4 years of age.

Previous studies examining pre-conscious neurophysiological responses to violations in harmonic syntax have reported such responses to be present in 4-year-old children (Corrigall & Trainor, 2014; Jentschke et al., 2008; Koelsch et al., 2003). The behavioral data reported here is consistent with these studies and shows, for the first time, that 4-year-old children can exhibit this knowledge behaviorally with the dwell-time paradigm. There is one report of an immature ERAN-like response in 2-year-old children in response to harmonic syntactic violations (Jentschke, Friederici, & Koelsch, 2014), although to our knowledge it has not been replicated. It is possible, then, that initial learning of harmonic knowledge might occur in brain networks responsible for automatic preconscious auditory processing, and that the ability to use such knowledge in a conscious behavioral manner emerges later. It is also possible the current study imposed greater working memory demands than that of Jentschke et al. (2014). The trials used in Jentschke et al. (2014) were always five chords in length (four priming chords followed by a target chord) with pauses between trials, so sequences were relatively short and the position of possible target chords in the sequence was predictable. In contrast, our stimuli unfolded continuously, the average phrase length was 8 chords, and the location of boundary chords was not metrically predictable. It is possible that children at this age do have some harmonic knowledge, but that they cannot use it to effectively parse long sequences or sequences with variable phrase lengths. What is clear from the present study is that the ability to use harmonic cues to phrase boundaries increases greatly between 3 and 4 years of age.

One possible explanation for this change in sensitivity is the increase in cumulative exposure to music between 3 and 4 years old. If this were the case, we would expect to see that older 3-year-olds would dwell on harmonic boundaries more than younger 3-year-olds, although our sample may not have been large enough to detect such a difference. We additionally investigated whether changes in working memory, speech parsing ability, or receptive vocabulary were related to harmonic boundary dwelling, as these abilities all improve significantly over the preschool years. However, we found no significant correlations at any of the three ages between boundary-dwelling magnitude and any of these measures. This suggests that sensitivity to harmonic structure may be unrelated to these variables, but care must be taken before accepting the null hypothesis, particularly because these correlational tests may have been underpowered. In recent years, a number of researchers have proposed that processing of musical syntax and language syntax draw on similar or identical cognitive resources (e.g. Fiveash & Pammer, 2012; Koelsch, Gunter, Wittforth, & Sammler, 2005; Patel, 2003; Slevc, Rosenberg, & Patel, 2009). One possibility that we did not test in the current study is that understanding of syntax in language could be related to understanding of harmonic syntax in music. Future work could examine the trajectories of syntactic understanding in music and language across childhood, to determine the extent to which their developmental trajectories are related.

We hypothesized that 7-year-olds would dwell on boundaries in the tonal version but not in the atonal version, demonstrating a contribution of harmonic cues above and beyond any other cues that might be present in the chord sequence that were not controlled. Contrary to our hypothesis, we found that seven-year-olds dwelled on boundaries in *both* tonal and atonal sequences, in contrast to 4-year-olds, who dwelled on boundaries *only* in the tonal version, and 3-year-olds, who did not significantly dwell on boundaries in either version. Interestingly, previous work has found that adults also dwelled on boundaries in both tonal and atonal versions of these sequences (Kragness & Trainor, 2016). This finding suggests that there are additional cues to boundaries in the atonal sequences to which older children and adults are sensitive, even though metrical, harmonic, and melodic contour cues would not support the perception of boundaries in those positions. For instance, C Major was one of the most frequent chords in the atonal sequences overall, which could have led to a sense of relative stability when this chord was played. This stability could, in turn, have led to the perception of a musical boundary, especially if the chord was perceived to have resolved the preceding chord (as in Bharucha, 1984). Neither 3- nor 4-year-olds dwelled significantly on boundaries in the atonal sequences, showing no evidence of sensitivity to such cues in the atonal case. Interestingly, adults in Kragness and Trainor (2016) showed a *stronger* boundary dwelling effect in the tonal than atonal condition, whereas this contrast was not significant in the 7-year-olds. It is possible that the present study simply did not have enough power to observe a difference between tonal and atonal conditions in 7-year-olds (the means in Figure 4 demonstrate the expected pattern). Future work should investigate how different cues contribute to the perception of musical phrase boundaries in participants who differ in age and musical experience.

Our results additionally extend previous research on phrase-final lengthening, the well-documented tendency for music performers to slow on phrase endings (Repp, 1992a; Seashore, 1938; Todd, 1985). A leading explanation holds that phrase-final lengthening is a cue that performers use with the explicit intent of emphasizing the structure of a piece to listeners (*music expression hypothesis;* Clarke, 1985; Todd, 1985). Results from several studies, however, suggest that there may be an implicit component to boundary lengthening – for example, performers lengthen boundaries even in "mechanical" performances (Penel & Drake, 1998) and listeners expect phrase-ending notes to be lengthened (Repp, 1992b; Repp, 1999). Previous work on phrase-final lengthening has found that the amount of lengthening seems to be related to the hierarchical level of the phrase boundary, with higher-level boundaries being lengthened more than lower-level boundaries. Interestingly, in Experiment 1, 3-year-old participants appear to demonstrate this pattern, as a small local peak can be seen at the sub-phrase boundary (Figure 1, position 4). This is strikingly similar to the pattern seen in adults' timing profiles while self-pacing through the same chord sequences (Kragness & Trainor, 2016; Figure 1). That study additionally showed that even adults with no formal music training lengthened musical boundaries when self-pacing through chord sequences, so it is clear that musical training is not necessary for phrase-final lengthening. The present results extend the results of that study, showing that the relationship between lengthened duration and musical phrase boundaries is present early in childhood. It is highly unlikely that 3-year-olds who have never taken music lessons would employ an intentionally communicative strategy in a lab-based self-pacing task, so these results further implicate a contribution of implicit mechanisms to the phenomenon of phrase-final lengthening. The potential role of implicit mechanisms in phrase-final lengthening is a question for future work to address in more detail. An additional question of interest for future work is whether there are individual differences in the age at which phrase-final lengthening emerges, particularly for harmonic phrase boundaries, and whether this could be used as a marker of musical precocity.

Self-pacing studies have now been shown to index segmentation in the domains of both action and music perception (Baldwin & Sage, 2013; Hard et al., 2011; Kragness & Trainor, 2016). Future research should investigate what psychological mechanism the dwell time measure captures. Previous work has shown that in watching film clips, viewing event boundaries is associated with increased cognitive load (as measured by pupillary dilation, Smith, Whitwell, & Lee, 2006; Swallow & Zacks, 2004) and surges in attention (Swallow, Zacks, & Abrams, 2009), suggesting that boundaries are more cognitively challenging than nonboundaries. We propose that music, which has structure that can be statistically modeled (e.g. Pearce, 2005), is uniquely well-suited to investigating this question further. Currently, we are examining whether longer boundary dwell times are associated with predictive uncertainty in adults self-pacing through melodies (Kragness, Hansen, Vuust, Trainor, & Pearce, 2016).

There is some evidence that group-final lengthening may enhance processing of sequences. Hard et al. (2011) found that participants who dwelled on phrase boundaries when self-pacing through action sequences later had better memory for what the actions they had seen. However, it was not clear in this study whether group-final lengthening directly caused enhanced memory. Nevertheless, whether group-final lengthening could play a role in structural learning in music is a question worth pursuing. For example, some have proposed that phrase-final lengthening in infant-directed speech could facilitate infants' learning of phrases (Koponen & Lacerda, 2003). Understanding what drives dwelling on boundaries, whether this is shared across modalities, and whether durational cues to boundaries enhance learning and memory could be important for future work on understanding the mechanisms underlying structural learning in many domains, both for infants learning about structure in their native environment, and potentially also for older children and adults learning novel structures, such as foreign languages.

Overall, the current experiments extend the finding that self-pacing tasks can be used to index young children's segmentation of action sequences (Meyer et al., 2011) to the musical domain. The implicit nature of the measure makes the dwell time method particularly well-suited for testing young children. The paradigm was effective down to three years, the youngest age tested. Several experiments have used the paradigm to probe infants' understanding of action sequences using self-paced slideshows on a touchscreen (reviewed in Baldwin & Sage, 2013). Whether the self-pacing method will work with infants for auditory sequences could be investigated in future studies. Although there is an abundance of studies examining infants' perceptual grouping of short syllable and tone sequences using looking time methods (e.g. Saffran, Johnson, Aslin, & Newport, 1999; Trainor and Adams, 2000; Yoshida et al., 2010), very few studies have investigated infants' perception of larger units, like phrases (but see Krumhansl & Juscyzk, 1990; Jusczyk and Krumhansl, 1993), and isolating the specific cues that lead to infants' phrase grouping has proved especially difficult. The dwell time method could potentially provide a new way to learn about infants' discovery of structure in auditory sequences, and could perhaps even be used to investigate infants' and young children's emerging understanding of grouping structure in other domains, such as language.

Author Note

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Laurel J. Trainor and an Ontario Trillium Scholarship to Haley E. Kragness. We thank Matthew Woolhouse for composing the chord sequences used in Experiment 2 and Laura Cirelli for her comments on an earlier version of this manuscript. We additionally thank Ammaarah Baksh, Abirami Kirubarajan, and Pushp Narula for their help in recruitment and/or data collection. Finally, we thank our participants and their families for their time.

References

Aarden B. (2003). Dynamic melodic expectancy. Unpublished doctoral dissertation, School of Music, Ohio State University.

Aldwell, E., & Schachter, C. (2002). Harmony and voice leading. (3rd ed.) New York, NY: Schirmer Book.

Baldwin, D. & Sage, K. (2013). Dwelling on action. In M. Rutherford & V. Kuhlmeier (Eds.) *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 309-330). Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/9780262019279.003.0013

Bergeson, T. R., & Trehub, S. E. (2006). Infants' perception of rhythmic patterns. *Music Perception, 23*(4), 345–360. http://dx.doi.org/10.1525/mp.2006.23.4.34

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*(4), 485–518. http://dx.doi.org/10.1016/0010-0285(84)90018-5

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 808–822. http://dx.doi.org/10.1037/0096-1523.23.3.808

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*(1), 100-130. http://dx.doi.org/10.1016/j.cognition.2005.11.007

Boltz, M. (1989). Rhythm and "good endings": Effects of temporal structure on tonality judgments. *Perception & Psychophysics, 46*(1), 9–17. http://dx.doi.org/10.3758/bf03208069

Brown, D. E. (1991). *Human universals.* New York: McGraw-Hill. http://dx.doi.org/10.4135/9781452276311.n135

Cambouropoulos, E. (2001). Melodic cue abstraction, similarity, and category formation: A formal model. *Music Perception, 18*(3), 347–370. http://dx.doi.org/10.1525/mp.2001.18.3.347

Chang, H. & Trehub, S.E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child Development, 48*(4), 1666-1670. http://dx.doi.org/10.2307/1128532

Chiappe, P., & Schmuckler, M.A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review, 4*(2), 254-259. http://dx.doi.org/10.3758/BF03209402

Clarke, E. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross and R. West (Eds): *Musical Structure and Cognition* (pp. 209-236). London, UK: Academic Press

Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception: An Interdisciplinary Journal, 7*(3), 213–251. http://dx.doi.org/10.2307/40285462

Corrigall, K. A., & Trainor, L. J. (2009). Effects of musical training on key and harmony perception. *Annals of the New York Academy of Sciences, 1169*(1), 164–168. http://dx.doi.org/10.1111/j.1749-6632.2009.04769.x

Corrigall, K. A. & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception, 28*(2), 195-200. http://dx.doi.org/10.1525/mp.2010.28.2.195

Corrigall, K. A. & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142-158. http://dx.doi.org/10.1111/desc.12100

Costa-Giomi, E. (2003). Young children's harmonic perception. *Annals of the New York Academy of Sciences, 999*(1), 477-484. http://dx.doi.org/10.1196/annals.1284.058

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42-45. http://dx.doi.org/10.20982/tqmp.01.1.p042

Creel, S. C. (2015). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science, 40*(2), 373–403. http://dx.doi.org/10.1111/cogs.12237

Cuddy, L.L. & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception & Psychophysics, 41*(6), 609–620. http://dx.doi.org/10.3758/bf03210493

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4*(4), 325-360. http://dx.doi.org/10.2307/40285378

Dissanayake, E. (2006). Ritual and ritualization: Musical means of conveying and shaping emotion in humans and other animals. In S. Brown, & U. Voglsten (Eds.), *Music and manipulation: On the social uses and social control of music* (pp. 31–56). Oxford and New York: Berghahn Books.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test fourth edition, Manual*. Minneapolis, MN, USA: Pearson Assessments.

Ferland, M. B. & Mendelson, M. J. (1989). Infants' categorization of melodic contour. *Infant Behavior and Development, 12*(3), 341-355. http://dx.doi.org/10.2307/1129143

Fiveash, A. & Pammer, K. (2012). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. http://dx.doi.org/10.1177/0305735612463949

Gerry, D. W., Faux, A. L., & Trainor, L. J. (2010). Effects of Kindermusik training on infants' rhythmic enculturation. *Developmental Science, 13*(3), 545–551. http://dx.doi.org/10.1111/j.1467-7687.2009.00912.x

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science, 15*(3), 398–407. http://dx.doi.org/10.1111/j.1467-7687.2012.01142.x

Gregory, A. H. (1978). Perception of clicks in music. *Perception & Psychophysics, 24*(2), 171–174. http://dx.doi.org/10.3758/bf03199545

Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: effects of interval ratio complexity and enculturation. *Developmental Science, 14*(4), 865–872. http://dx.doi.org/10.1111/j.1467-7687.2011.01036.x

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*(11), 466–472. http://dx.doi.org/10.1016/j.tics.2007.08.008

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science, 16*(1), 48–55. http://dx.doi.org/10.1111/j.0956-7976.2005.00779.x

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences, 102*(35), 12639–12643. htp://dx.doi.org/10.1073/pnas.0504254102

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology, 140*(4), 586-604. http://dx.doi.org/10.1037/a0024310

He, C., Hotson, L., & Trainor, L. J. (2009). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience, 29*(4), 861–867. http://dx.doi.org/10.1111/j.1460-9568.2009.06625.x

Helfgott, J. A. (1976). Phonemic segmentation and blending skills of kindergarten children. Implications for beginning reading acquisition. *Contemporary Educational Psychology, 1,* 157-169. http://dx.doi.org/10.1016/0361-476x(76)90020-5

Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology, 10,* 3-23.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences, 930*, 43-61. http://dx.doi.org/10.1111/j.1749-6632.2001.tb05724.x

Ilari, B. (2003). Singing and listening to music with infants: An interview study with Canadian mothers. *Abstracts of the 5th Triennial ESCOM Conference,* 364.

Ilari, B. (2009). Songs of belonging: Musical interactions in early life. In J. L. Kerchner & C.R. Abril (Eds.), *Musical experiences in our lives: Things we learn and meanings we make.* (pp. 21-38). Lanham, MD: National Association for Music Education/Rowman & Littlefield Education.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*(4), 2263. http://dx.doi.org/10.1121/1.2973189

Jentschke, S., Friederici, A. D., & Koelsch, S. (2014). Neural correlates of music-syntactic processing in two-year old children. *Developmental Cognitive Neuroscience, 9*, 200–208. http://dx.doi.org/10.1016/j.dcn.2014.04.005

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*(11), 1940–1951. http://dx.doi.org/10.1162/jocn.2008.20135

Jusczyk, P. W. & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19*(3), 627-640. http://dx.doi.org/10.1037/0096-1523.19.3.627

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schröger, E., & Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience, 15*(5), 683–693. http://dx.doi.org/10.1162/jocn.2003.15.5.683

Koelsch, S., Gunter, T. C., Wittforth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577. http://dx.doi.org/10.1162/089892905774597290

Koponen, E., & Lacerda, F. (2003). Final lengthening in infant directed speech may function as a cue to phrase constituents. *PHONUM, 9*, 9–12.

Kragness, H. E., Hansen, N. C., Vuust, P., Trainor, L. J., & Pearce, M. (2016, July). *Information dynamics of boundary perception: Evidence from self-paced music listening.* Poster session presented at the 14th International Conference on Music Perception and Cognition, San Francisco, CA, USA.

Kragness, H. E. & Trainor, L. J. (2016). Listeners lengthen boundaries in self-paced music listening. *Journal of Experimental Psychology: Human Perception and Performance.* http://dx.doi.org/10.1037/xhp0000245

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346–374. http://dx.doi.org/10.1016/0010-0285(79)90016-1

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* Oxford: Oxford University Press.

Krumhansl, C.L., & Jusczyk, P. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*(1), 70-73. http://dx.doi.org/10.1111/j.1467-9280.1990.tb000070.x

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition*, *10*(3), 243–251. http://dx.doi.org/10.3758/bf03197636

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. http://dx.doi.org/10.1037/0033-295x.89.4.334

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. http://dx.doi.org/10.1111/j.1467-7687.2006.00468.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*(863). http://dx.doi.org/10.3389/fpsyg.2013.00863

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics, 52*(6), 599–608. http://dx.doi.org/10.3758/bf03211696

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*(4), 272–276. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00213.x

Masataka, N. (2006). Preference for consonance over dissonance by hearing newborns of deaf parents and of hearing parents. *Developmental Science, 9*(1), 46–50. http://dx.doi.org/10.1111/j.1467-7687.2005.00462.x

Meyer, M., Baldwin, D., & Sage, K. (2011). Assessing young children's hierarchical action segmentation. In L. Carlson, H. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Morrongiello, B. A., Trehub, S. E., Thorpe, L. A., & Capodilupo, S. (1985). Children's perception of melodies: The role of contour, frequency, and rate of presentation. *Journal of Experimental Child Psychology, 40*(2), 279-292. http://dx.doi.org/10.1016/0022-0965(85)90090-6

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Nathani, S., Oller, D.K., & Cobo-Lewis, A.B. (2003). Final Syllable Lengthening (FSL) in infant vocalizations. *Journal of Child Language, 30*(1), 3–25. http://dx.doi.org/10.1017/s0305000902005433

Oller, D.K. (1973). The effect of position-in-utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54*(5), 1235-1247. http://dx.doi.org/10.1121/1.1914393

Olsho, L.W. (1984). Infant frequency discrimination. *Infant Behavior and Development*, *7*(1), 27-35. http://dx.doi.org/10.1016/s0163-6383(84)80020-x

Olsho, L.W., Schoon, C., Sakai, R., Turpin, R., and Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology, 18*(5), 721-726. http://dx.doi.org/10.1037/0012-1649.18.5.721

Palmer, C., & Krumhansl, C.L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 116-126. http://dx.doi.org/10.1037/0096-1523.13.1.116

Patel, A. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. http://dx.doi.org/10.1038/nn1082

Pearce, M.T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis, London, UK: Department of Computing, City University

Pearce, M.T., Müllensiefen, D., & Wiggins, G.A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*(10), 1367-1391. http://dx.doi.org/10.1068/p6507

Pearce, M.T., & Wiggins, G.A. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23*(5), 377-405. http://dx.doi.org/10.1525/mp.2006.23.5.377

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research, 61*(1), 12-32. http://dx.doi.org/10.1007/PL00008161

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology, 24*(1-5), 157-178. http://dx.doi.org/10.1080/00207594.1989/106000040

Phillips-Silver, J. & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science, 308*(5727), 1430. http://dx.doi.org/10.1126/science.1110922

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition, 98*(1), 1–11. http://dx.doi.org/10.1016/j.cognition.2004.09.008

Repp, B.H. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92*(5), 2546-2568. http://dx.doi.org/10.1121/1.404425 7

Repp, B.H. (1992b). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44*(3), 241-281. http://dx.doi.org/10.1016/0010-9

Repp, B.H. (1999). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics, 61*(3), 529-548. http://dx.doi.org/10.3758/BF03211971

Rosner, J. (1975). Helping children overcome learning difficulties. New York: Walker & Co.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. http://dx.doi.org/10.1016/s0010-0277(98)00075-4

Sage, K. D., Ross, R. A., & Baldwin, D. (2012). Assessing infants' hierarchical action segmentation. Poster presented at the International Conference on Infant Studies, Minneapolis, Minnesota.

Schellenberg, E. G. (1997). Simplifying the implication-realization model. *Music Perception: An Interdisciplinary Journal, 14*(3), 295-318. http://dx.doi.org/10.2307/40285723

Schellenberg, E. G., Bigand, E., Poulin-Charronnat, B., Garnier, C., & Stevens, C. (2005). Children's implicit knowledge of harmony in Western music. *Developmental Science, 8*(6), 551-566. http://dx.doi.org/10.1111/j.1467-7687.2005.00447.x

Seashore, C. (1938). *Psychology of music*. New York: McGraw-Hill.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin and Review, 16*(2), 374-381. http://dx.doi.org/10.3758/162.374

Sloboda, J.A., & Gregory, A.H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology, 34*(3), 274-280. http://dx.doi.org/10.1037/h0081052

Smith, T. J., Whitwell, M., & Lee, J. (2006). Eye movements and pupil dilation during event perception. *Proceedings of the Eye Tracking Research and Applications conference 2006*. http://dx.doi.org/10.1145/1117309.1117333

Swallow, K.M., Zacks, J.M. (2004). Hierarchical grouping of events revealed by eye movements. *Abstracts of the Psychonomics Society, 9*(81). http://dx.doi.org/10.1037/e527352012-530

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*(2), 236–257. http://dx.doi.org/10.1037/a0015631

Tan, N., Aiello, R., & Bever, T.G. (1981). Harmonic structure as a determinant of melodic organization. *Memory & Cognition, 9*(5), 533-539. http://dx.doi.org/10.3785/BF03202347

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25*(1), 122–127. http://dx.doi.org/10.1037/0012-1649.25.1.122

Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology, 24*(4), 484–491. http://dx.doi.org/10.1037/0012-1649.24.4.484

Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Annals of the New York Academy of Sciences, 1060*(1), 100–110. http://dx.doi.org/10.1196/annals.1360.007

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*(4), 885–913. http://dx.doi.org/10.1037/0033-295x.107.4.885

Todd, N.P. McA. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33-57. http://dx.doi.org/10.2307/40285321

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception & Psychophysics, 62*(2), 333-340. http://dx.doi.org/10.3758/BF03205553

Trainor, L. J., & Hannon, E. E. (2013). Musical development. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 423–98). London, UK: Elsevier Inc. http://dx.doi.org/10.1016/b978-0-12-381460-9.00011-0

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music. *Infant Behavior and Development, 21*(1), 77–88. http://dx.doi.org/10.1016/s0163-6383(98)90055-8

Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences, 1252*(1), 129–138. http://dx.doi.org/10.1111/j.1749-6632.2012.06462.x

Trainor, L.J. & Trehub, S.E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2). 394-402. http://dx.doi.org/10.1037/0096-1523.18.2.394

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics, 56*(2), 125–132. http://dx.doi.org/10.3758/bf03213891

Trainor, L. J., Tsang, C. D., & Cheung, V. H. W. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception, 20*(2), 187–194. http://dx.doi.org/10.1525/mp.2002.20.2.187

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants’ perception of melodies: The role of melodic contour. *Child Development, 55*(3), 821-830. http://dx.doi.org/10.2307/1130133

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*(1), 73–99. http://dx.doi.org/10.1016/j.cognition.2005.11.006

Trehub, S. E., & Schellenberg, E. G. (1995). Music: Its relevance to infants. In R.Vasta (Ed.), Annals of child development (Vol. 11, pp. 1-24). New York: Jessica Kingsley Publishers.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants’ perception of auditory patterns. *Child Development, 58*(3), 741-749. http://dx.doi.org/10.2307/1130211

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

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior and Development, 16*(3), 285–295. http://dx.doi.org/10.1016/0163-6383(93)80036-8

Turk, A.E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*(4), 445–472. http://dx.doi.org/10.1016/j.wocn.2006.12.001

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2014). Children's identification of familiar songs from pitch and timing cues. *Frontiers in Psychology, 5*(863). http://dx.doi.org/10.3389/fpsyg.2014.00863

Wechsler, D.t, Kaplan, E., Fein, D., Kramer, J., Morris, R. (2004). *Wechsler intelligence scale for children – Fourth edition integrated (WISC-IV Integrated): administration and scoring manual.* San Antonio, TX: Pearson.

Werker J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233–251. http://dx.doi.org/10.1002/dev.20060

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*(3), 1707–1717. http://dx.doi.org/10.1121/1.402450

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences, 106*(7), 2468–2471. http://dx.doi.org/10.1073/pnas.0809035106

Yoshida, K., Iverson, J.R., Patel, A.D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*(2), 356-361. http://dx.doi.org/10.1016/j.cognition.2010.01.005

Young, S. (2008). Lullaby light shows: everyday musical experience among under-two-year-olds. *International Journal of Music Education, 26*(1), 33–46. http://dx.doi.org/10.1177/0255761407085648

Zentner, M. R. & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Child Development, 21*(3), 483-492. http://dx.doi.org/10.1016/S0163-6383(98)90021-2

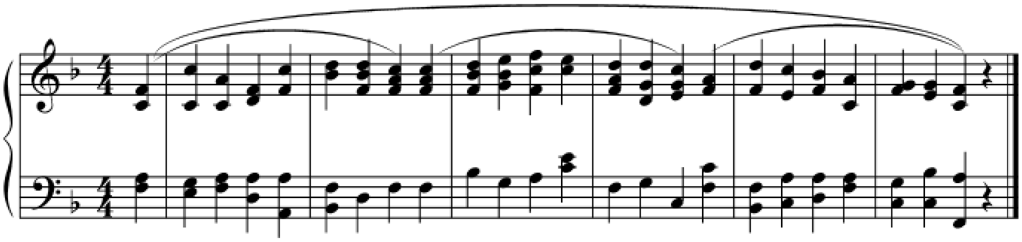
Supplementary Material

Experiment 1 Stimuli

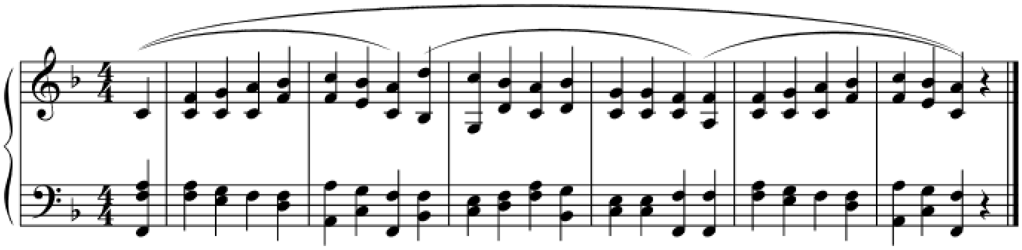
Reprinted from “Listeners lengthen phrase boundaries in self-paced music,” by H. E. Kragness and L. J. Trainor, 2016, *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), p. 1676-1686. Copyright 2016 by the American Psychological Association. Adapted with permission.

Sequence 1

Based on BWV 1.6 harmonized by J. S. Bach

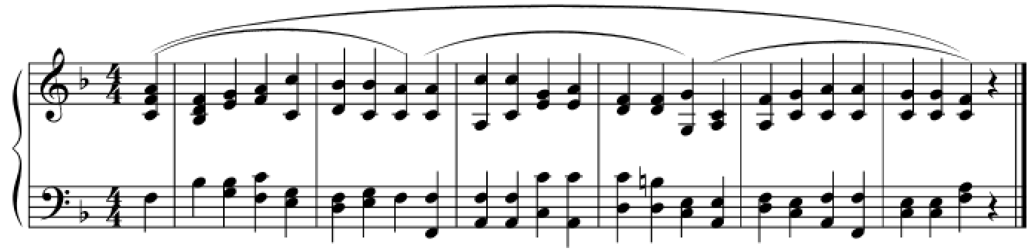
“Wie schön leuchtet der Morgenstern”

Sequence 2

Based on BWV 12.7 harmonized by J. S. Bach

“Was Gott tut, das ist wohlgetan”

Sequence 3

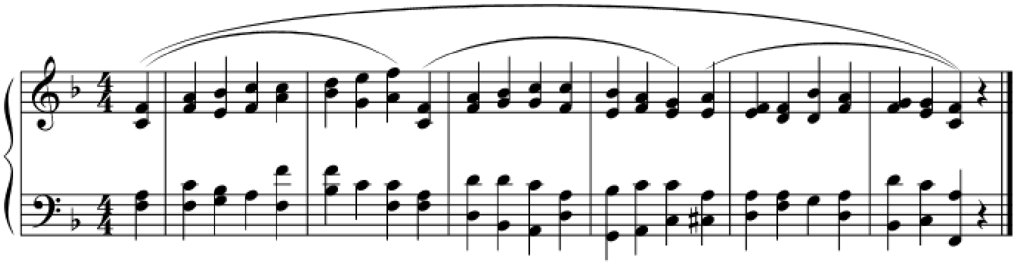
Based on BWV 13.6 harmonized by J. S. Bach

“O Welt, ich muβ dich lassen”



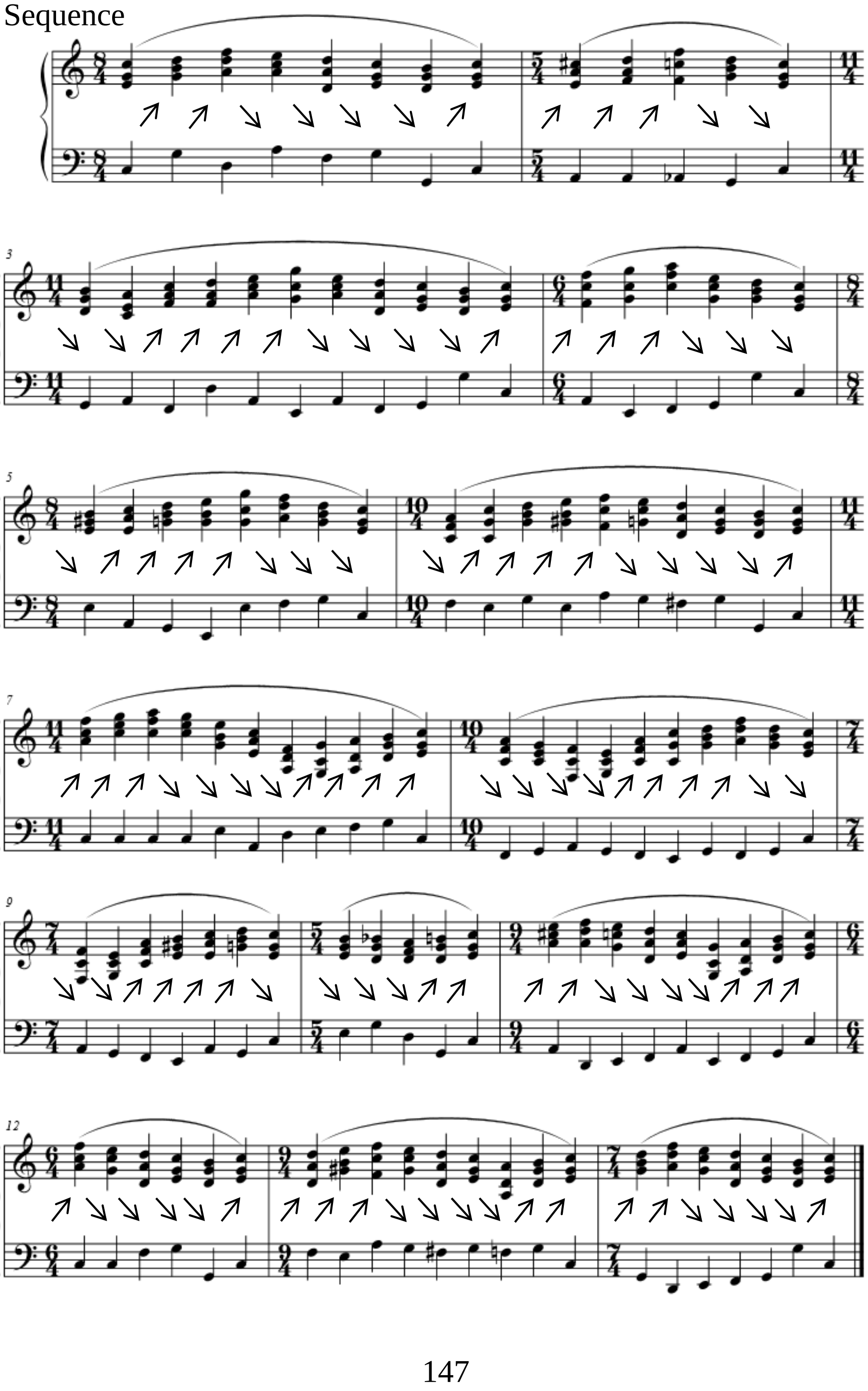
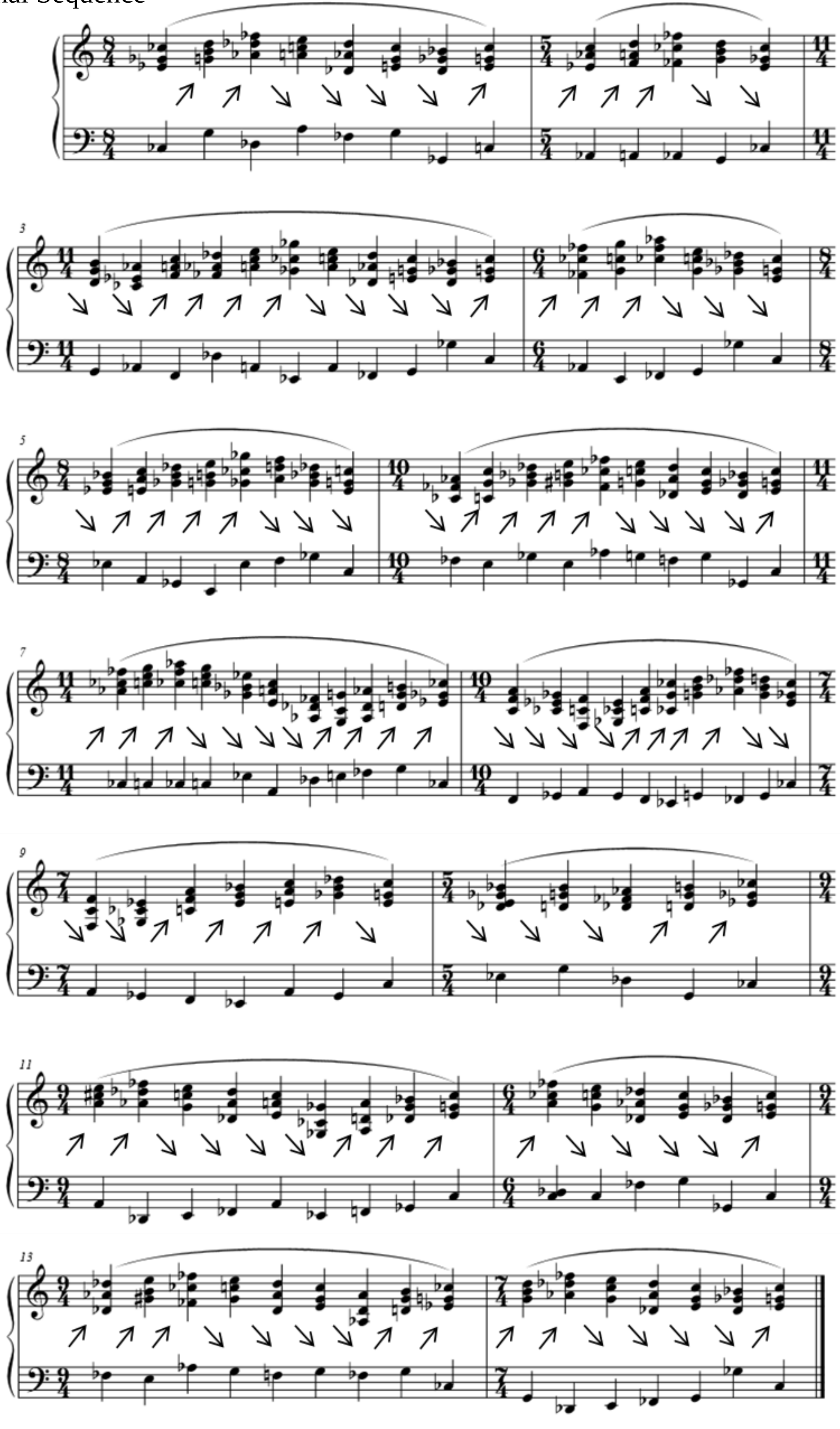
Sequence T4

Based on BWV 20.7 harmonized by J. S. Bach

“O Ewigkeit, du Donnerwort”

Experiment 2 Stimuli

Adapted from “Listeners lengthen phrase boundaries in self-paced music,” by H. E. Kragness and L. J. Trainor, 2016, *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), p. 1676-1686. Copyright 2016 by the American Psychological Association. Adapted with permission.

Arrows indicate the melodic contour (the pitch direction of the highest voice).

Tonal Sequence

Atonal Sequence

Supplemental Table 1. ANOVA table for log10-transformed data (Experiment 2, 4-year-olds)

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Boundary Status	1	17	1.981	.177	0.002
Tonality	1	17	8.651	.009*	0.038
Boundary Status x Tonality	1	17	10.159	.005*	0.012

Note: Significance at the level of p < .05 is designated with a *.

Supplemental Table 2. t-test results describing boundary minus non-boundary difference scores (Experiment 2, 4-year-olds)

Tonality condition	df	t	p	Cohen's d_Z
Tonal	17	2.450	.012*	0.584
Atonal	17	-2.658	.992	-0.626

Note: Significance at the level of p < .05 is designated with a *.

Supplemental Table 3. ANOVA table for log10-transformed data (Experiment 2, 3-year-olds)

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Boundary Status	1	25	1.070	.311	0.002
Tonality	1	25	0.955	.338	0.006
Boundary Status x Tonality	1	25	0.421	.522	0.000

Note: Significance at the level of p < .05 is designated with a *.

Chapter 4: Nonmusicians express emotions in musical productions using conventional cues

Kragness H. E. & Trainor, L. J. (under review). Nonmusicians express emotions in musical productions using conventional cues.

Preface

In the previous two chapters, I examined the association between expressive timing and phrase structure. In Chapter 4, I examined how individuals with different levels of formal music training use timing and loudness cues to express emotions in a music production task. I adapted the musical dwell time task to enable participants to control the onset, offset, and loudness of each successive chord, thereby controlling the tempo, articulation, and dynamics. Participants were asked to express joy, anger, sadness, and peacefulness through music production. I observed that, consistent with previous observations of expert performance, timing and loudness cues were mainly associated with conveying the *arousal* associated with the intended emotion, rather than the valence of the intended emotion. Further, I found that participants with a wide range of formal training used each expressive cue in a similar fashion. To my knowledge, this is the first investigation of nonmusicians' knowledge of expressive cues in a production task.

Abstract

Expert musicians use a number of expressive cues to communicate specific emotions in musical performance. In turn, listeners readily identify the intended emotions. Previous studies of cue utilization have exclusively studied the performances of expert or highly-trained musicians, limiting the generalizability of the results. Here we use a musical self-pacing paradigm to investigate expressive cue use, by non-expert individuals with varying levels of formal music training. Participants controlled the onset and offset of each chord in a musical sequence by repeatedly pressing and lifting a single key on a MIDI piano, controlling *tempo* and *articulation*. In addition, the velocity with which they pressed the key controlled the sound level (*dynamics*). Participants were asked to "perform" the music to express basic emotions that were (1) positively or negatively valenced and (2) high- or low-arousal (*joy, sadness, peacefulness,* and *anger*). Nonmusicians' expressive cue use was consistent with patterns of cue use by professional musicians described in the literature. In a secondary analysis, we explored whether those with differing levels of formal training expressed the target emotions differently through tempo, articulation, dynamics, and rhythm. We observed that the patterns of cue use were strikingly consistent across groups with differing levels of formal musical training. Future work could investigate whether expertise is implicated in the expression of more complex emotions and/or in the expression of more complex musical structures, as well as explore the role of emotional intelligence and informal musical experiences in expressive performance.

Keywords: emotion, music, expression, performance, training

Introduction

Music is pervasive in our everyday lives. One particularly compelling aspect of music is that it can be used as a nonverbal medium for emotional communication. Caregivers use affective songs and song-like speech in interactions with infants (e.g. Fernald et al., 1989; Ilari, 2003; Trehub & Trainor, 1998; Young, 2008) and adults commonly report using music to regulate their own emotions (Lonsdale & North, 2011). Surveys indicate that affective value is a primary motivation for listening to music (Juslin & Laukka, 2004; Sloboda & O'Neill, 2001). Likewise, musicians attest that communicating emotions is a central goal in their performances (Lindström, Juslin, Bresin, & Williamon, 2003).

It is widely accepted that there is a distinction between emotions that are *induced* by music ("felt emotions") and emotions that are *communicated* by music ("perceived emotions"; Gabrielsson, 2001). A number of studies have found that listeners' felt and perceived emotional responses to music can be different (for a review, see Schubert, 2013). For example, many report experiencing pleasant feelings when listening to sad music, despite being able to identify the music as sad-sounding (e.g. Garrido & Schubert, 2011; Huron, 2011; Kawakami, Furukawa, Katahira, & Okanoya, 2013). Even when the felt and perceived emotions are the same, listeners tend to rate perceived emotions as more intense than felt emotions (Schubert, 2013). Such contrasts suggest that musical emotions are not identified by simply reflecting on felt emotions; rather, listeners make an appraisal about performers' intended emotions based on acoustic cues (Juslin, 1997; Juslin & Laukka, 2003).

In general, adults agree on the emotion communicated by musical excerpts (Bigand et al., 2005; Juslin & Laukka, 2003; Mohn, Argstatter, & Wilker, 2011; Vieillard et al., 2008). In childhood, *happiness* and *sadness* tend to be the most readily identified and earliest recognized emotions (Cunningham & Sterling, 1988; Dalla Bella, Peretz, Rousseau, & Gosselin, 2000; Dolgin & Adelson, 1990; Kastner & Crowder, 1990; Mote, 2011; Nawrot, 2003; Terwogt & Van Grinsven, 1988). *Happiness* and *sadness* are followed in identification accuracy by *anger* and *fear*, all of which are also recognized cross-culturally to some extent (Balkwill & Thompson, 1999; Fritz et al., 2009; Laukka et al., 2013). In some circumstances, more complex emotions, such as *longing*, *humor*, and *awe* have also been identified by listeners (Huron, 2006; Senju & Ohgushi, 1987; Laukka et al., 2013). Such findings have led to questions about the emotional "code" used by composers, performers, and listeners in expressing and interpreting musical emotions. Amplitude envelope, rhythm, pitch contour, melodic range, mode, loudness, variations in loudness, articulation, pitch level, tempo, and timbre are among the musical cues that have been implicated in emotional communication (see Gabrielsson & Lindström, 2010 and Juslin & Timmers, 2010 for reviews).

Cues for emotional expression are embedded to some extent in the score by the composer. For example, complex rhythms are often perceived to express *joy* or *anger*, and simple, regular rhythms to convey *sadness* or *boredom* (e.g. Scherer & Oshinsky, 1977; Thompson & Robitaille, 1992). However, studies of performance using the *standard content paradigm* (Seashore, 1947; Gabrielsson & Juslin, 1996) have demonstrated that performers also have control over communicating emotions. In such experiments, performers are asked to perform a prescribed score with the goal of communicating different emotions[1] and listeners are asked to identify the intended emotion. The extent to which the intended and judged emotions match is considered a measure of the communicative success. Studies using this paradigm with opera singers, electric guitarists, violinists, trumpeters, percussionists, and pianists (amongst others) have found that both musically trained and untrained listeners can indeed decode expressive intent in such performances (Behrens & Green, 1993; Juslin, 1997; Juslin, 2000; Juslin & Madison, 1999; Kotlyar & Morozov, 1976, Laukka & Gabrielsson, 2000).

Researchers have considered how various cues relate to different dimensions of emotions. Russell's (1980) circumplex model of emotions, which

---

[1] As noted by Juslin (2013), "expression" in this context does not suggest that performers are necessarily experiencing the emotion they intend to convey, but rather that they are performing a symbolic representation of that emotion (as would an actor in a play).

characterizes emotions on the dimensions of valence (negative to positive) and arousal (low to high), has been widely used in this regard. Performers seem to have particular control over cues that distinguish between high- and low-arousal emotions. For example, performances conveying high-arousal emotions, such has *happiness* and *anger*, typically incorporate fast tempi, low tempo variability, disconnected articulations, and high sound levels. In contrast, performances of low-arousal emotions (such as *sadness* and *peacefulness*) involve slow tempi with final ritardandi, high tempo variability, connected articulations, and low sound levels (Juslin & Timmers, 2010). These cues are redundant to some extent – it is not necessary to utilize *all* cues to communicate an emotion, but using multiple cues increases the chances of correct identification (Juslin, 2000).

Though these studies have been very informative, it is interesting to note that "encoders" and "decoders" are usually mismatched – encoders must be musically experienced in order to perform the music, but decoders may have had no experience with formally producing music at all. It is not clear that formal music training discernably improves emotional decoding of professional performances (Bigand et al., 2005; Juslin, 1997), and these results are consistent with research demonstrating that day to day exposure to music confers a degree of sensitivity to many aspects of music, resulting in "musically experienced listeners" (for a review, see Bigand & Poulin-Charronnat, 2006). Nevertheless, there is reason to believe that the situation is different for musical production. While many societies have a *participatory* music culture in which non-experts are regularly active in music production and performance, Western culture is largely *presentational*, with production activities limited to highly trained individuals (Turino, 2008). In one study, Juslin and Laukka (2000) noted that a novice guitarist (one year of experience) communicated *happiness*, *sadness*, *anger*, and *fear* less accurately than more experienced guitarists (up to 15 years of experience). Experience was not explicitly investigated, however, and it is impossible to determine whether this difference in accuracy was due to differences in expressive intentions, differences in technical proficiency, or some other factor.

Technological advances have made it possible to examine musical production without using musical instruments. In a recent study, Bresin and Friberg (2011) used a novel “slider” apparatus to allow participants to systematically vary different musical features (including tempo, sound level, articulation, phrasing, register, timbre, and attack speed) to communicate *happy*, *sad*, *peaceful*, or *scary* expressions. They found that participants’ use of expressive cues to communicate each emotion were broadly consistent with those that have been found in previous studies, confirming that using a non-instrumental apparatus is sufficient for expression. Although the apparatus did not require musical expertise to use, only musical experts were recruited. Thus, no existing research to our knowledge has examined how nonmusicians use expressive cues to convey musical emotions.

In the present study, we used a simple self-pacing apparatus to examine the use of expressive cues in musical production across non-expert undergraduate performers with different levels of formal music training. A similar apparatus has been used in previous self-pacing studies to examine expressive timing and musical phrase structure in nonmusicians and children (Kragness & Trainor, 2016; Kragness & Trainor, 2018). In previous studies, participants controlled the onset of each successive chord in a prescribed musical sequence. In the present study, this setup was adapted such that participants additionally controlled the offset and sound level of each chord. Thus, they could control the timing and loudness of each chord in the sequence. We predicted that nonmusician participants would use similar patterns of expressive cues as expert participants in previous studies, based on shared representations of musical expression from years of music listening, and that expressive cues differentiating emotions would be enhanced in participants in our sample who had relatively high levels of training.

Method

Participants

This research was approved by the university's research ethics board. Twenty-four undergraduates (M_{age} = 18.65 years, *SD* = 1.22 years, ranging from 18 to 22 years) were recruited through the participant pool. Participants were granted course credit as compensation. Because a pilot study (*N* = 30) conducted with a slightly different methodology and apparatus (using a computer keyboard instead of a MIDI keyboard) demonstrated large differences in participants' tempo and articulation use across the target emotions used here, a sample size of 24 was deemed likely to be sufficient in the present study. First, they filled out a questionnaire that asked about handedness, as well as experiences with languages, music, and dance. They additionally filled out the self-report inventory portion of the Goldsmith Music Sophistication Index, v1.0 ("Gold-MSI," Müllensiefen, Gingras, Stewart, & Musil, 2014).

Stimuli

We selected four excerpts from the chorales of J. S. Bach. Each excerpt contained three sub-phrases that were each eight chords in length, for a total of 24 chords, and began with an anacrusis (or "pick up" chord). Two of the excerpts ("Maj1" and "Maj2") were originally composed in the major mode and two of the excerpts ("Min3" and "Min4") were originally composed in the minor mode. Each excerpt was transposed to the key of F and small alterations were made to eliminate passing tones and ornamentations between each quarter-note-length chord. For each excerpt, a second version was created in the parallel major or minor mode (“Min1,” “Min2,” “Maj3,” and “Maj4”)[2]. Thus, there were eight excerpts in total.

Each participant “performed” the emotions using four of the eight excerpts: two major and two minor (except one participant, who played only one major excerpt due to a technical error). Thus each participant produced 16 “performances”. Of the two major and two minor excerpts, one was original and the other was modally altered. In a single experimental session, at no time did a participant perform both an excerpt *and* its modal alteration. Major and minor excerpts always alternated. The possible orders of excerpts can be seen in Table 1.

Apparatus

During the experiment, each chord was generated online in the default piano timbre in Max MSP (version 5). The interface used by participants was an M-AUDIO Oxygen-49 MIDI keyboard, and the sounds were presented through a pair of external speakers located to the left and right of the participant in the sound booth (WestSun Jackson Sound, model JSI P63 SN 0005). Participants

---

[2] Alterations and transpositions were done with the support and advice of an Assistant Professor of Music Theory with significant experience in Baroque-era harmony.

Table 1

Order Number

	Excerpt 1	Excerpt 2	Excerpt 3	Excerpt 4
Order 1	Maj1	Min2	Maj3	Min4
Order 2	Maj2	Min1	Maj4	Min3
Order 3	Min3	Maj4	Min1	Maj2
Order 4	Min4	Maj3	Min2	Maj1

were seated in a chair in front of a monitor at a distance of approximately 3.5 feet. In front of the participant was a desk on which the MIDI keyboard sat.

Using the MIDI keyboard, participants could control the onset and the offset of each chord in succession by pressing and releasing the key. Each chord was sustained until the key was released, which terminated the chord. Only the middle C key elicited a chord; the other keys did not respond if pressed. The middle C key was marked with an orange sticker to remind participants which key to use. Additionally, participants could control the sound level of each chord by pressing with more or less velocity (greater velocity produced louder chords).

Procedure

Training phase. After completing the questionnaires, participants were trained to use the MIDI keyboard apparatus. They were informed that only the key marked with a sticker (middle C) would elicit a note or chord, and instructed to use the index finger of their dominant hand to “perform” music. They were told to continue to press the key for each chord until pressing the key elicited no sound, which was an indication that the excerpt was complete. They were told that they did not control *what* chord was being played, but *when, how loudly,* and *for how long* each chord was played. First, participants were asked to practice “performing” an excerpt, which was a C Major scale (C4 to C5), without any instructions about how to play it. Next, they were asked to perform the same excerpt in six different ways: with short notes, with long notes, with loud notes, with soft notes, with fast notes, and with slow notes. This training session was done to give participants familiarity with the breadth of expressive options available to them during the testing phase.

Testing phase. Prior to the testing phase, we asked participants to jot down words or pictures to remind themselves of a time they felt each of the target emotions: *joy*, *sadness, peacefulness*, and *anger*. These emotions were selected because they each fall into a different quadrant of a two-dimensional valence-by-arousal circumplex (Russell, 1980). We asked participants to jot down some words or pictures that “reminded them of a time they felt each emotion.” The purpose of this activity was not to act as an emotion “induction,” since, as previously noted (Footnote 1), there is no reason to believe that “feeling” an emotion is necessary for communicating it. Rather, we wanted the participants to have a consistent reference point to consider for their performances throughout the experiment.

Next, there were four separate performance blocks. The experimenter informed the participant that they would be asked to perform each excerpt with the four different emotions, and that they should do their best to communicate each emotion, because a new set of participants would later be asked to guess their intended emotion. Then, the experimenter left the sound booth and they were guided for the rest of the experiment by instructions on the computer monitor.

In each block, participants first *heard* the excerpt that they would be performing. The excerpt was played in the default piano timbre at an inter-onset interval of 450 ms (approximately 133 onsets per minute). This rate was chosen because it is well within the range of adults' spontaneous motor tempo and was therefore likely to be around the chosen tempi (Drake, Jones, & Baruch, 2000). After listening, participants were instructed to play the excerpt "mechanically, without emotion." Next, one of the four target emotions was displayed and they were instructed to practice performing the excerpt with that emotion. Finally, they were instructed to undertake the "performance" for that emotion. The practice and performance sessions were repeated for each target emotion. The listening, mechanical, and practice sessions were all included to give participants an opportunity to familiarize themselves with the chord sequences and plan their performances to the best of their ability.

After the testing phase was complete, participants were given a questionnaire. They were asked to indicate how easy they found it to play each of the emotions (1 "Very easy" – 5 "Very difficult"), as well as the extent to which they *felt* the emotion while they were playing it (1 "Not at all" – 5 "Very much"). Finally, they were asked to report in a free response format whether they had used any strategies to try to convey different emotions.

Results

We investigated three expressive cues: tempo (time between consecutive onsets), key velocity (which was experienced as loudness), and articulation (for each interval between onsets, the proportion of that interval in which the chord was played). We additionally analyzed durational variability using the normalized pairwise variability index (*nPVI*; Grabe & Low, 2002). This measure of variability describes the degree of contrast between pairs of consecutive durations (e.g. Patel & Daniele, 2003; Hannon, Lévêque, Nave, & Trehub, 2016, Huron & Ollen, 2003; Quinto, Thompson, & Keating, 2013). Though rhythmic patterning is usually considered to be a compositional cue rather than expressive cue, in the current study, participants could use rhythmic patterns to communicate emotions if they desired.

Because planned emotional performances were of particular interest, the “mechanical” trial and the practice trials were excluded in the present analyses. For each cue, a separate 2x2 within-subjects ANOVA was performed with factors arousal (high and low) and valence (positive and negative). All participants experienced all four performance blocks, except for one participant, who completed only three performance blocks due to technical malfunction. Archived data are available at the link provided in the Supplemental Material.

Tempo

Tempo was defined as the average number of onsets per minute. The ANOVA indicated a significant main effect of arousal, $F(1,23) = 81.620$, $p < .0001$, $\eta_G^2 = .459$, demonstrating that across valence conditions, high-arousal emotions were played at a higher rate of onsets per minute than low-arousal emotions (Figure 1A). No significant main effect of valence ($F(1,23) = 4.064$, $p = .056$, $\eta_G^2 = .015$) or interaction effect ($F(1,23) = 1.287$, $p = .268$, $\eta_G^2 = .009$) were observed.

Normalized Pairwise Variability Index (nPVI)

Normalized pairwise variability (nPVI) was calculated using the formula developed for speech analysis by Grabe and Low (2002) and subsequently used in the context of music by Patel and Daniele (2006):

$$nPVI = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right| / (m-1) \right]$$

where m is the number of intervals in the excerpt and d is the duration of the kth interval. nPVI values range from 0 to 200, such that 0 indicates perfect isochrony (no durational contrast between pairs at all) and 200 represents maximal durational contrast (Figure 1B). We observed no significant main effects for arousal ($F(1,23) = 2.940$, $p = .10$, $\eta_G^2 = .015$) or valence ($F(1,23) = 0.099$, $p =$

.757, η_G^2 = .0005), but there was a significant interaction, *F*(1,23) = 6.238, *p* = .020, η_G^2 = .020. Further analysis revealed that the significant interaction was driven by greater nPVI for low-arousal than high-arousal emotion in the negative valence condition (*t*(23) = -2.964, *p* = .007, 95% CI = -17.352 to -3.088), but not in the positive valence condition (*t*(23) = 0.170, *p* = .867, 95% CI = -6.91 to 8.137).

Sound Level (Key Velocity)

Sound level was measured by the velocity with which the chord was pressed (recorded as MIDI velocity), on a scale of 0 (minimum velocity) to 125 (maximum velocity). The ANOVA revealed main effects of both valence (F(1,23) = 62.970, *p* < .0001, η_G^2 = .141) and arousal (*F*(1,23) = 137.68, *p* < .0001, η_G^2 = .597), as well as a significant interaction (*F*(1,23) = 45.621, *p* < .0001, η_G^2 = .142). Further analyses demonstrated that the significant interaction was driven by the fact that high-arousal emotions were played with greater velocity than low-arousal emotions, while at the same time, anger was played with significantly greater velocity (higher sound level) than joy (*t*(23) = 8.456, *p* < .0001, 95% CI = 18.803 to 30.981), but there was no significant difference in the velocity participants used for sadness and peace (*t*(23) = -0.033, *p* = .974, 95% CI = -3.674 to 3.558).

Articulation

Articulation was considered to be the proportion of the onset-to-onset interval in which the chord was played (for example, if a chord was played for 300 ms and the next onset was initiated 300 ms later, the articulation value would be .5; if the next onset was initiated 900 ms later, the articulation value would be .25). The ANOVA revealed main effects of both valence ($F(1,23) = 19.457$, $p < .001$, $\eta_G^2 = .042$) and arousal ($F(1,23) = 85.021$, $p < .0001$, $\eta_G^2 = .358$), such that low-arousal emotions were played with more connected chords than high-arousal emotions, and negatively valenced emotions were played with more connected chords than positively valenced emotions (Figure 1D). No significant interaction was observed ($F(1,23) = 2.584$, $p = .121$, $\eta_G^2 = .013$).

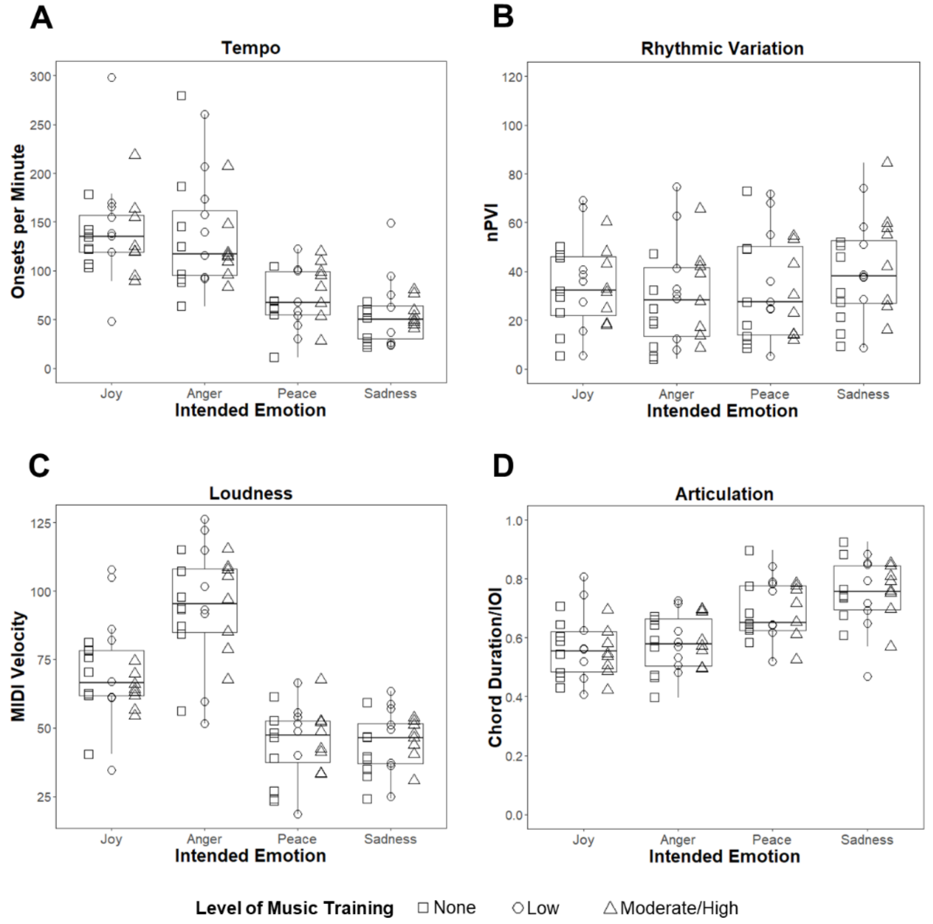


Figure 1

Use of expressive cues for each emotion. (A) Average tempo used to convey each emotion. Higher values indicate longer inter-onset intervals and slower tempi. (B) Average nPVI used to convey each emotion. Higher values indicate more pairwise durational contrast. (C) Average velocity used to convey each emotion. Higher values indicate greater velocity (and higher sound level). (D) Average articulation used to convey each emotion. Higher values indicate more connected chords.

Correlations Between Expressive Cues

To examine relationships between expressive cues, Spearman's rho was calculated for each pair of expressive cues (Table 2). Because six correlations were examined, the Bonferroni-corrected significance cut-off of $p = .008$ was used (.05/6). Tempo was correlated with sound level and articulation, such that faster tempi were correlated with greater sound level ($r_S = .697$, $p < .0001$) and less connected notes ($r_S = -.791$, $p < .0001$). Articulation was also significantly correlated with loudness ($r_S = -.526$, $p < .0001$), such that more connected notes (higher articulation values) were played more softly (lower sound levels). No correlations with nPVI were observed at the Bonferroni-corrected significance threshold (tempo, $r_S = -.225$, $p = .028$; articulation, $r_S = .239$, $p = .019$; sound level, $r_S = -.090$, $p = .385$).

The Role of Musical Training

Although all participants were undergraduates and non-experts musically, they represented a wide range of musical experiences as revealed by their responses to the self-report inventory of the Gold-MSI. Specifically, eight participants reported no formal music training at all (0 years), while others reported upwards of 10 years of training. In order to examine whether those with formal music training used cues differently from those with no training, we separated participants into three equally-sized groups based on their scores on the

Table 2

Correlation Coefficients Between Expressive Cues

	Tempo	nPVI	Sound Level	Articulation
Tempo	-			
nPVI	-.225*	-		
Sound Level	.697***	-.090	-	
Articulation	-.791***	.239*	-.526***	-

Note. Correlation coefficients represent Spearman's rho. Asterisks indicate statistical significance: *$p < .05$, **$p < .01$, ***$p < .001$

Formal Musical Training sub-scale of the Gold-MSI (see Table 3), which combines information about years of formal music lessons, music practice, music theory training, etc. Participants in the "no training" group reported 0 years of formal lessons (*n* = 8), participants in the "low training" group reported 0.5-9 years of formal lessons (mean = 2.25 years, *n* = 8) and participants in the "moderate/high training" group reported 3-10 or more years of formal lessons (mean = 7.56 years or higher[3], *n* = 8). One-way ANOVAs performed on participants' scores on the Active Musical Engagement and Sophisticated Emotional Engagement subscales of the Gold-MSI suggested that the groups did not significantly differ on these measures, $F(2,21) = 1.456$, $p = .256$ and $F(2,21) = 0.783$, $p = .470$, respectively.

Next, all of the previous ANOVAs were rerun with the additional between-subjects variable of Musical Training (no, low, moderate/high). However, there were no significant main effects or interactions with Musical Training in any condition (see Supplemental Material for ANOVA tables).

---

[3] Years of training were determined based on participants' responses to question 35 of the self-report questionnaire of the Gold-MSI. In instances where a response represented a range of values (e.g. "4-6 years"), individuals' response was coded as the middle value of the range to compute the average years of training for each group (i.e., "4-6 years" was coded as 5 years). The highest possible response, "10 or more" years, was coded as 10 years. Because of this, the actual average years of training for the "moderate/high training" group may be higher than 7.88.

Table 3

Participant Demographics (Gold-MSI Self-Report Questionnaire)

	No Training	Low Training	Moderate/High Training
Age	18.83 years (1.60) 2 not reporting	18.40 years (0.89) 3 not reporting	18.67 years (1.21) 2 not reporting
Gender	5 F, 1 M 2 not reporting	5 F, 0 M 3 not reporting	5 F, 1 M 2 not reporting
Active Musical Engagement	32.88 (6.56)	35.63 (3.74)	38.50 (8.57)
Perceptual Abilities	41.75 (5.37)	43.75 (7.05)	46.25 (9.29)
Musical Training	11.53 (3.38)	24.38 (4.21)	33.50 (2.51)
Singing Abilities	25.88 (3.72)	25.38 (5.88)	33.00 (9.27)
Sophisticated Emotional Engagement	31.75 (3.58)	33.88 (4.73)	34.25 (4.53)
General Musical Sophistication	59.75 (8.64)	69.75 (12.84)	83.63 (11.40)

Ratings

After participants completed all four performance blocks of the experiment, they were asked to rate the difficulty of conveying each emotion (Figure 2A). Because a Shapiro-Wilks test indicated that the data were non-normally distributed ($W = .852$, $p < .0001$), a Friedman rank sum test was used, revealing significant differences in difficulty ratings among emotions (Friedman chi-squared = 8.521, $N = 24$, $p = .036$). However, post-hoc pairwise comparisons using a Nemenyi multiple comparison test revealed no significant differences between emotions. A series of paired t-tests revealed significant differences between *joy* and *peace* ($t(23) = 2.450$, $p = .022$) and *sadness* and *peace* ($t(23) = 2.557$, $p = .018$), but no other comparisons were significant (all p's > .245). As the significant comparisons would not survive corrections for post-hoc testing and multiple comparisons, they should be interpreted with caution.

Participants were also asked to rate the extent to which they felt the intended emotion while playing it (Figure 2B). Because the data were non-normally distributed ($W = .874$, $p < .0001$), a Friedman rank sum test was used to evaluate whether there were differences in difficulty ratings. The Friedman ranked sum test did not reveal any significant differences (Friedman chi-squared = 7.723, $p = .065$).

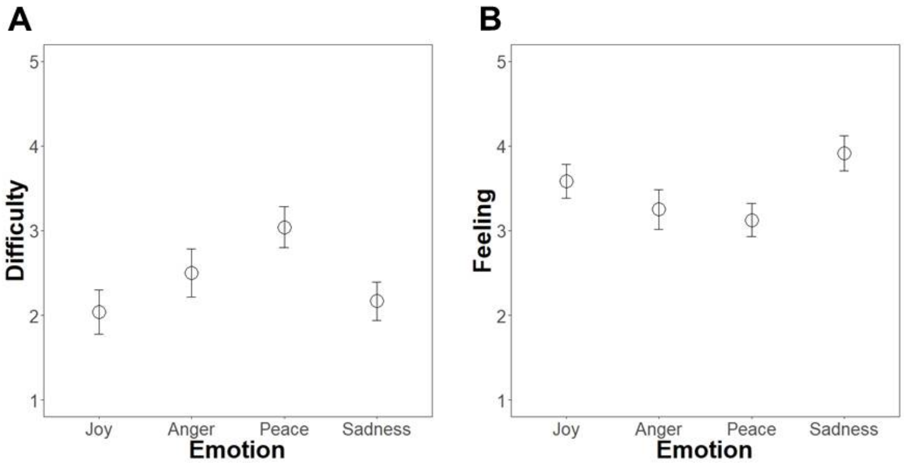


Figure 2

Participants' ratings after the experiment. (A) Participants' average rating for difficulty to convey each emotion (1 = not difficult, 5 = very difficult). Error bars represent within-subject SEM (Cousineau, 2005). (B) Participants' ratings for how strongly they felt the emotions that they performed (Cousineau, 2005).

Discussion

Although many studies have investigated how performers use expressive cues to communicate emotions in music, no previous experiments have examined nonmusicians' expressive productions. This is the first study to demonstrate that those with little to no musical training use timing and loudness cues to differentiate musical emotions in a production setting. Moreover, musically untrained participants used the available cues in ways that were nearly identical to those used by musically trained individuals in our sample. Consistent with previous studies using highly trained musicians, differences in expression were most strongly associated with differences in the arousal of the intended emotion, such that performances of joy and anger were played with a faster tempo, more loudly, and with more disconnected chords than peacefulness and sadness (e.g. Gabrielsson & Juslin, 1996; Juslin, 1997). Valence was also represented to a lesser extent – anger was played more loudly than joy, and within each arousal level (anger vs. joy; sadness vs. peacefulness) negatively-valenced emotions were played with more connected articulation than positively-valenced emotions. Finally, durational contrast as measured by the nPVI was used to portray sadness more than other emotions. This is broadly consistent with Quinto et al.'s (2013) previous finding that high-level musicians used greater nPVI values in brief compositions intended to portray sadness than other emotions, though this contrast was not significant in their analysis. Our finding is somewhat inconsistent, however, with previous reports that rhythms with durational contrast are often perceived to convey positive emotions (Thompson & Robitaille, 1992; Keller & Schubert, 2011). One possible explanation for this apparent inconsistency is that the chord sequences were drawn from chorales by J. S. Bach in which the chords are primarily equally spaced. As well, the task itself tended to encourage an isochronous interpretation, so rhythmic opportunities for the production of different categories of note length (e.g., eighth notes and quarter notes, the latter being twice as long as the former) were constrained. If this is the case, larger nPVI values may have been observed in *sadness* and *peace* than in *anger* and *joy* simply because participants intended to maintain isochrony and found it more difficult to do so when the overall tempo was slow than when it was fast. Overall, the role of rhythm in emotional communication has received little attention, and warrants future investigation.

Although this is among the first studies to examine non-expert participants, multiple levels of music training were represented. Thus, as a secondary analysis, we explored whether participants with formal training used cues in a different way from musically untrained participants. The profile of expressive cues used by those with *no* formal training was strikingly similar to those with training, even in magnitude. This result could have been obtained if participants simply used the maximum and minimum values of each cue (e.g., playing “anger” and “joy” as loudly as possible and “sadness” and “peace” as quietly as possible). However, the consistent use of intermediate values (especially for loudness and articulation) suggests that similarities across training levels were not simply due to use of extreme values.

Interestingly, a previous observational study of music lessons found that instructors spend surprisingly little time on expressive instruction, tending to focus more on technique (Karlsson & Juslin, 2008), despite the widespread belief that expressivity is central to music performance (Juslin & Laukka, 2004; Lindström et al., 2003). It has been proposed that this is because teachers often conceptualize expressivity as instinctual and difficult to verbally communicate (Hoffren, 1964; Lindström, Juslin, Bresin, & Williamon, 2003). This lack of expressive instruction is especially interesting considering that feedback from music teachers *can* improve emotional communication accuracy (Juslin, Karlsson, Lindström, Friberg, & Schoonderwaldt, 2006). In combination with the present findings, these studies suggest that formal training can enhance expression of musical emotions, but that formal music curricula likely do not utilize expression instruction to its fullest potential.

The present results are consistent with previous perceptual studies that have found that nonmusicians are equally adept as musicians at identifying emotions in music (Bigand et al., 2005; Juslin, 1997). In the present study, participants mainly controlled *performer cues* (timing and loudness), rather than *composer cues*, such as mode. Indeed, performer cues are typically found to be the most consistently decoded across ages and cultures (Dalla Bella et al, 2001; Thompson & Balkwill, 2010; Quinto, Thompson, & Taylor, 2013). Whether awareness of these cues rests on a "universal affect code" shared across domains including vocal expression (e.g. Juslin, 1997, 2001, 2013; Juslin & Laukka, 2003) or other mechanisms such as cultural transmission is unknown, but the present study shows that nonmusicians use performance cues in production tasks as well as in perceptual tasks.

We asked participants to report any strategies that they implemented while completing the task using a free response format (see Supplemental Material). Though no statistical analyses were conducted on these responses, they offer several interesting insights. First, it is clear that regardless of level of training, participants were often explicitly aware of the cues that they intended to use. For example, S22 (moderate/high training) reported, "I sustained the notes more to match the peaceful and sad emotion. I cut the notes short for angry" and S09 (no training) wrote, "sad = slower/louder, angry = faster/louder, peaceful = slow/quiet, happy = loud/fast." A number of strategies were reported, including "imagining movie scenes (as well as their soundtracks) to fit with each emotion," and "thinking about what music provokes these emotions." Though participants' free responses were not explicitly analyzed, they suggest that participants have some capacity to introspect about expressive cues and emotional musical production. Exploring the relationship between participants' introspections and their expressive productions would be an interesting future direction.

Many participants expressed that one primary strategy was to reflect on memories that incorporated the target emotions, as they were instructed. This is consistent with past studies with musicians, many of whom reported that they believed that feeling the intended emotion is important for expressive performance (Lindström, Juslin, Bresin, & Williamon, 2003) and that they used recalling emotional memories as one strategy for inducing the intended mood (Persson, 2001). Interestingly, although differences in "difficulty" and "feeling" ratings were not significant, participants indicated greater mood induction for *joy* and *sadness*, which were also reported to be the easiest to communicate. *Peace* was the least-felt of the four target emotions, and also rated as the most difficult to communicate. This pattern suggests that there may indeed have been an association between *feeling* and *communicating* emotions in this context, although whether *feeling* the emotion resulted in more accurate communication is unknown.

The present results should not be taken as evidence that music training has no effect on expressivity. It is important to note that the target emotions each belonged to different quadrants of the two-factor model (Russell, 1980) and had been previously observed to elicit relatively high agreement across listeners in emotion recognition tasks (e.g., Gabrielsson & Juslin, 2003; Juslin & Laukka, 2003; Vieillard et al., 2008). Mixed emotions and aesthetic emotions (such as longing, love, awe, and humor), which generally have lower levels of agreement, are likely to be more difficult to express. Future studies could investigate whether music training alters expressive cues in the context of complex emotions. Additionally, though our No Training participants had no formal lessons at all, a post-hoc one-way ANOVA did not offer any evidence for differences between the three formal training groups in Active Musical Engagement subscale of the Gold-MSI. Given that expressivity is not typically a focus in music lessons, it is possible that active engagement with music is more important for expressive production than formal training. The participants also did not differ significantly on their self-reported emotional engagement in music. However, one previous study has reported that people who score higher on a measure of emotional intelligence are better at recognizing emotions in music (Rescinow, Salovey, & Repp, 2004). Recent work has shifted toward more sophisticated and multidimensional conceptions of musical experience using subscales and composite measures of overall engagement (e.g. Chin & Rickard, 2012; Müllensiefen, Gingras, Stewart, & Musil, 2014). Informal musicianship and emotional engagement with music may be important contributors to expressive production, although they were not investigated in the present study.

Overall, the present study demonstrated that participants with a variety of musical backgrounds can use the self-pacing apparatus to express emotions in music. Future work could use similar paradigms to investigate questions of expressive timing in a variety of other populations. For example, previous studies investigating cross-cultural expressions of emotion have been limited by the need to use different musical scores and instruments (Balkwill, Thompson, & Matsunaga, 2004; Laukka et al., 2013). Using our self-pacing apparatus, it would be possible to examine amateur and expert performers’ expressive tendencies while performing either music congruent with their own musical system or music of a foreign musical system, which could illuminate questions about the universality of expressive cues for basic emotions. Similarly, previous studies examining children's expressive productions have been limited to analyzing singing (Adachi & Trehub, 2000; Adachi, Trehub, & Abe, 2004), and we are presently using this apparatus to investigate children's expressive musical production.

References

Adachi, M., & Trehub, S. E. (2000). Decoding the expressive intentions in children's songs. *Music Perception: An Interdiscliplinary Journal, 18*(2), 213-224. http://dx.doi.org/10.2307/40285909

Adachi, M., Trehub, S. E., & Abe, J. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46*(4), 322-336. http://dx.doi.org/10.1111/j.1468-5584.2004.00264.x

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hinustani music by Japanese listeners. *Japanese Psychological Research, 46*(4), 337-349. http://dx.doi.org/10.1111/j.1468-5584.2004.00265.x

Balkwill, L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*(1), 43-64. http://dx.doi.org/10.2307/40285811

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20-33. http://dx.doi.org/10.1177/030573569302100102

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130. http://dx.doi.org/10.1016/j.cognition.2005.11.007

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional response to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*(8), 1113-1139. http://dx.doi.org/10.1080/02699930500204250

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex, 47*(9), 1068-1081. http://dx.doi.org/10.1016/j.cortext.2011.05.009

Chin, T., & Rickard, N. S. (2012). The music use (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception: An Interdisciplinary Journal,* 29*(4)*, 429-446. http://dx.doi.org/10.1525/mp.2012.29.4.429

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Massion's method. *Tutorial in Quantitative Methods for Psychology,* 1*(1)*, 4-45. http://dx.doi.org/10.20982/tqmp.01.1.p042

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion*, *12*(4), 399-413. http://dx.doi.org/10.1007/bf00992362

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), 1584-1595. http://dx.doi.org/10.1016/s0010-0277(00)00136-0

Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music*, *18*(1), 87-98. http://dx.doi.org/10.1177/0305735690181007

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251-288. http://dx.doi.org/10.1016/s0010-0277(00)00106-2

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477-501. http://dx.doi.org/10.1017/S0305000900010679

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573-576. http://dx.doi.org/10.1016/j.cub.2009.02.058

Gabrielsson, A. (2001). Perceived emotion and felt emotion: same or different? *Musicae Scientiae*, *5*(1), 123-148. http://dx.doi.org/10.1177/10298649020050s105

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the perfomer's intention and the listeners' experience. *Psychology of Music*, *24*(1), 68-91. http://dx.doi.org/10.1177/0305735696241007

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In K. R. R. J. Davidson, *Series in affective science. Handbook of affective sciences* (pp. 503-534). New York, NY, US: Oxford University Press.

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

Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception: An Interdiscliplinary Journal*, *28*(3), 279-2964. http://dx.doi.org/10.1525/mp.2011.28.3.279

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gusenhoven, & N. Warner, *Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Hannon, E. E., Lévêque, Y., Nave, K. M., & Trehub, S. E. (2016). Exaggeration of language-specific rhythms in English and French children's songs. *Frontiers in Psychology, 7,* 939. http://dx.doi.org/10.3389/fpsyg.2016.00939

Hoffren, J. (1964). A test of musical expression. *Bulletin of the Council for Research in Music Education, 2,* 32-35.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, Massachusetts: MIT Press.

Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae*, *15*(20), 146-158. http://dx.doi.org/10.1177/1029865911401171

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception: An Interdisciplinary Journal, 21*(2), 267-271. http://dx.doi.org/10.1177/102986491101500202

Ilari, B. (2003). Singing and listening to music with infants: An interview study with Canadian mothers. *Abstracts of the 5th Triennial ESCOM Conference*, 364.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception: An Interdisciplinary Journal*, *14*(4), 383-418. http://dx.doi.org/10.2307/40285731

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology, 4*. http://dx.doi.org/10.3389/fpsyg.2013.00596

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientiae*, *4*(2), 151-183. http://dx.doi.org/10.1177/102986490000400202

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770-814. http://dx.doi.org/10.1037/0033-2090.129.5.770

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217-238. http://dx.doi.org/10.1080/0929821042000317813

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, *17*(2), 197-221. http://dx.doi.org/10.2307/40285891

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin, & J. A. Sloboda, *Handbook of music and emotion: Theory, research, applications* (pp. 453-489). Oxford, UK: Oxford University Press.

Juslin, P., Karlsson, J., Lindström, E., Friberg, A., & Schoonderwaldt, E. (2006). Play it again with feeling: Computer feedback in musical communication of emotions. *Journal of Experimental Psychology: Applied*, *12*(2), 79-95. http://dx.doi.org/10.1037/1076-898x.12.2.79

Karlsson, J., & Juslin, P. N. (2008). Musical expression: An observational study of instrumental teaching. *Psychology of Music, 36*(3), 309-334. http://dx.doi.org/10.1177/0305735607086040

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189-201. http://dx.doi.org/10.2307/40285296

Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in Psychology*, *4*. http://dx.doi.org/10.3389/fpsyg.2013.00311

Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology, 7,* 142–156. http://dx.doi.org/10.2478/v10053-008-0094-0

Kotlyar, G. M., & Morozov, V. P. (1976). Acoustic correlates of the emotional content of vocalized speech. *Soviet Physics Acoustics*, *22*(3), 208-211.

Kragness, H. E., & Trainor, L. J. (2016). Listeners lengthen phrase boundaries in self-paced music. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1676-1686. http://dx.doi.org/10.1037/xhp0000245

Kragness, H. E., & Trainor, L. J. (2018). Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues. *Developmental Psychology*, *54*(5), 842-856. http://dx.doi.org/10.1037/dev0000405

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, *28*(2), 181-189. http://dx.doi.org/10.1177/0305735600282007

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434-449. http://dx.doi.org/101.1037/a0031388

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). “Expressivity comes from within your soul”: a questionnaire study of music students’ perspectives on expressivity. *Research Studies in Music Education*, *20*(1), 23-47. http://dx.doi.org/10.1177/1321103X030200010201

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, *102*(1), 108-134. http://dx.doi.org/10.1348/000712610X506831

Mohn, C., Argstatter, H., & Wilker, F. (2011). Perception of six basic emotions in music. *Psychology of Music*, *39*(4), 503-517. http://dx.doi.org/10.1177/0305735610378183

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*(3), 618-622. http://dx.doi.org/10.1037/a0022573

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

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music*, *31*(1), 75-92. http://dx.doi.org/10.1177/0305735603031001325

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-B45. http://dx.doi.org/10.1016/s0010-0277(02)00187-7

Persson, R. (2001). The subjective world of the performer. In P. N. Juslin, & J. A. Sloboda, *Music and Emotion: Theory and Research* (pp. 275-289). Oxford: Oxford University Press.

Quinto, L., Thompson, W. F., & Keating, F. L. (2103). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4*, 184. http://dx.doi.org/10.3389/fpsyg.2013.00184

Quinto, L., Thompson, W. F., & Taylor, A. (2013). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music*, *42*(4), 503-524. http://dx.doi.org/10.1177/0305735613482023

Rescinow, J. E., Salovey, P., & Repp, B. H. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception*, *22*(1), 145-158. http://dx.doi.org/10.1525/mp.2004.22.1.145

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

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotional attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346. http://dx.doi.org/10.1007/bf00992539

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in Psychology, 4.* http://dx.doi.org/10.3389/fpsyg.2013.00837

Seashore, C. E. (1947). *In search of beauty in music. A scientific appraoch to musical aesthetics.* Westport, CT: Grenwood Press.

Senju, M., & Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception, 4*(4), 311-323. http://dx.doi.org/10.2307/40285377

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everday listening to music. In P. N. Juslin, & J. A. Sloboda, *Music and Emotion: Theory and Research* (pp. 415-429). Oxford, UK: Oxford University Press.

Terwogt, M. M., & Van Grinsven, F. (1988). Recognition of emotions in music by children and adults. *Perceptual and Motor Skills*, *67*(3), 697-698. http://dx.doi.org/10.2466/pms.1988.67.3.697

Thompson, W. F., & Balkwill, L. L. (2010). Cross-cultural similarities and differences. In P. N. (eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 755-788). Oxford, UK: Oxford University Press.

Thompson, W. F., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts*, *10*(1), 79-89. http://dx.doi.org/10.2190/nbny-awdw-gw58-mtel

Timmers, R. (2007). Vocal expression in recorded performances of Schubert songs. *Musicae Scientiae*, *11*(2), 237-268. http://dx.doi.org/10.1177/102986490701100205

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

Turino, T. (2008). *Music as social life: the politics of participation.* London: University of Chicago Press.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720-752. http://dx.doi.org/10.1080/02699930701503567

Young, S. (2008). Lullaby light shows: everyday musical experience among under-two-year-olds. *International Journal of Music Education*, *26*(1), 33-46. http://dx.doi.org/10.1177/0255761407085648

Young, V., Burwell, K., & Pickup, D. (2003). Areas of study and teaching strategies in instrumental teaching: A case study research project. *Music Education Research*, *5*(2) 139-155. http://dx.doi.org/10.1080/1461380032000085522

Supplementary Material

Supplemental Table 1. ANOVA table for Tempo

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Training	2	21	1.044	.369	.039
Valence	1	21	3.063	.095	.018
Arousal	1	21	35.499	<.0001*	.310
Training x Valence	2	21	0.055	.947	.001
Training x Arousal	2	21	1.469	.253	.036
Valence x Arousal	1	21	1.559	.226	.014
Training x Valence x Arousal	2	21	0.271	.765	.005

Note: Significance at the level of p < .05 is designated with a *.

Supplemental Table 2. ANOVA table for nPVI

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Training	2	21	0.842	.445	.053
Valence	1	21	0.105	.749	.000
Arousal	1	21	2.699	.115	.017
Training x Valence	2	21	1.792	.191	.016
Training x Arousal	2	21	0.058	.944	.001
Valence x Arousal	1	21	6.182	.021*	.021
Training x Valence x Arousal	2	21	0.898	.423	.006

Note: Significance at the level of p < .05 is designated with a *.

Supplemental Table 3. ANOVA table for Sound Level (Velocity)

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Training	2	21	0.461	.647	.026
Valence	1	21	64.056	<.0001*	.147
Arousal	1	21	129.766	<.0001*	.609
Training x Valence	2	21	1.198	.321	.006
Training x Arousal	2	21	0.339	.716	.008
Valence x Arousal	1	21	46.260	<.0001*	.149
Training x Valence x Arousal	2	21	1.161	.332	.009

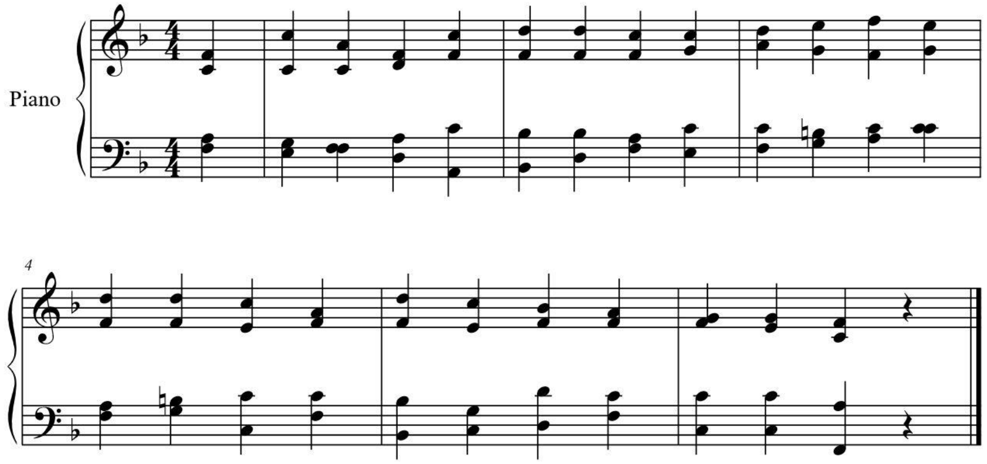
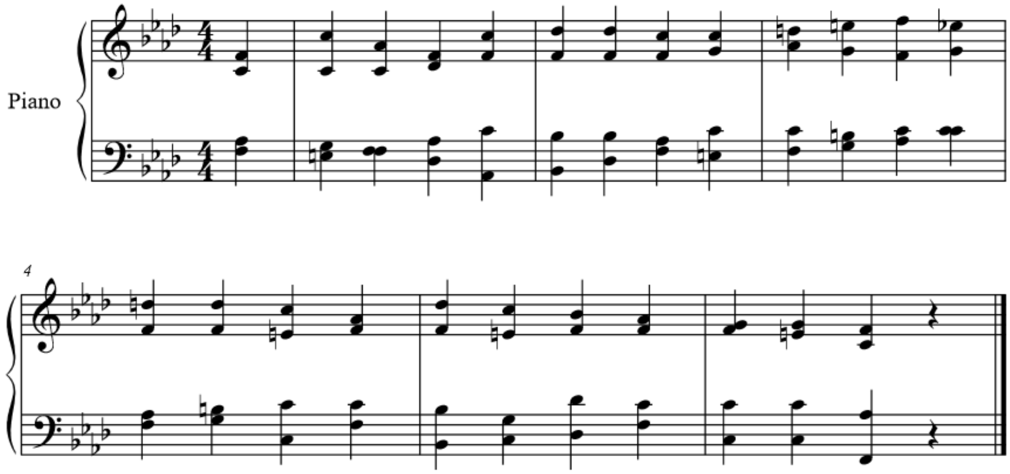
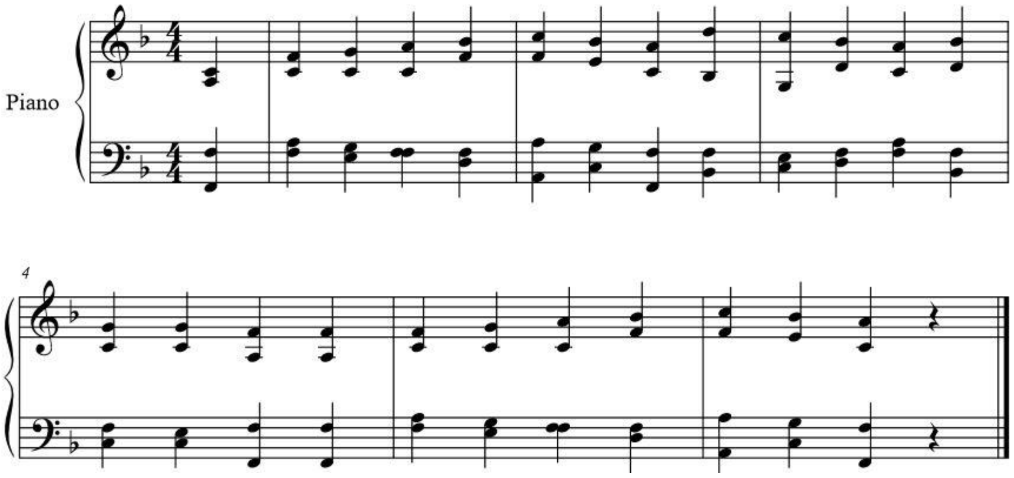
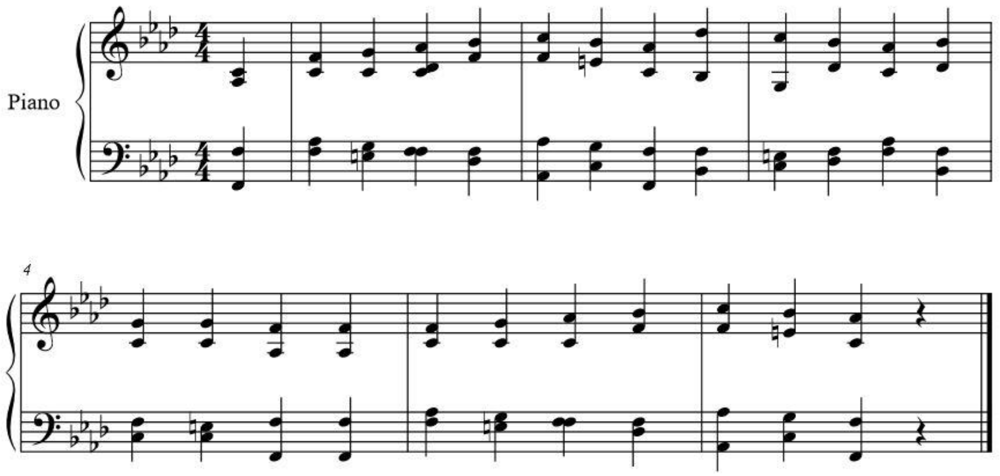
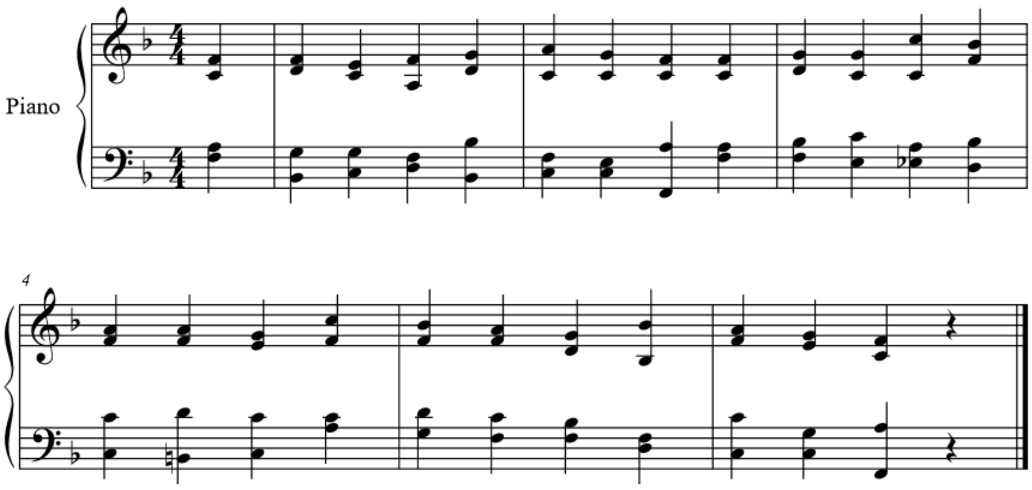
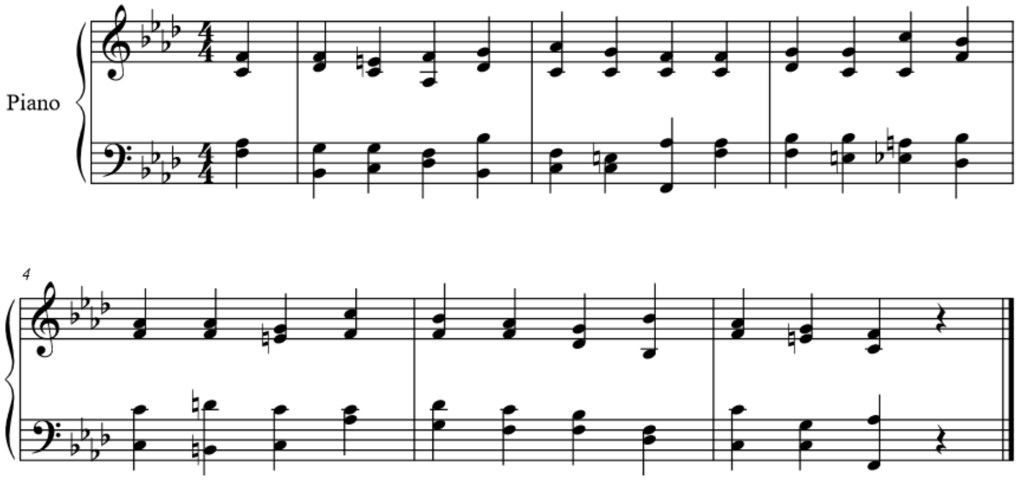
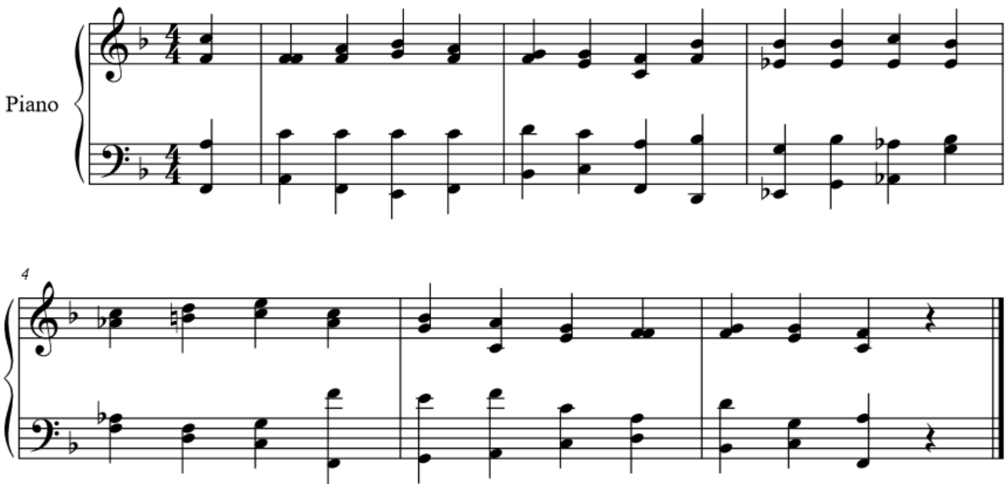
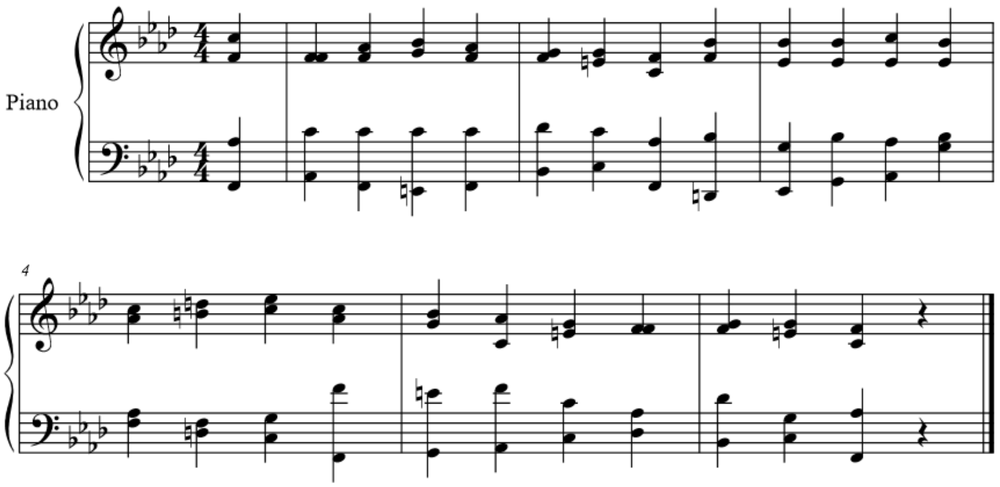
Note: Significance at the level of p < .05 is designated with a *.

Supplemental Table 4. ANOVA table for Articulation

Effect	df_n	df_d	F	p	$\hat{\eta}^2_G$
Training	2	21	0.050	.951	.003
Valence	1	21	20.930	.0001*	.043
Arousal	1	21	80.482	<.0001*	.361
Training x Valence	2	21	1.871	.179	.008
Training x Arousal	2	21	0.386	.684	.005
Valence x Arousal	1	21	2.413	.135	.013
Training x Valence x Arousal	2	21	0.240	.789	.002

Note: Significance at the level of p < .05 is designated with a *.

Scores for musical excerpts

Maj 1 (original mode)

Min 1(altered mode)

Maj 2 (original mode)

Min 2 (altered mode)

Maj 3 (altered mode)

Min 3 (original mode)

Maj 4 (altered mode)

Min 4 (original mode)

Were there any strategies that you used to communicate different emotions?

No Training

“No.”
"Use the high and low notes to the advantage of the emotion being asked to play"
"Thinking about what made me feel those emotion. Recollecting the past"
"Different tempo, loudness"
"Tried thinking about what music provokes these emotions and past memories"
"Sad = slower/louder, Angry = faster/louder, Peaceful = slow/quiet, Happy = loud/fast"
"Previous songs that made me feel the different emotions"
"Control the speed"

Low Training

"I thought about memories with different emotions"
"Loud for angry. Long and soft for sad. Quick and soft for joy."
"Recalling events from past, breathing between each emotion, mindfulness + awareness"
"Use different strength when I press the key"
"Try to remember the scenarios when I feel those emotions"
"Picture a scene in mind and imagining the music is playing in that place"
"Angry was faster shorter notes. Peaceful was longer and slow and same for sad, hence was a bit difficult to play those two. Happy was a mix of fast and slow keys."
"Happy: had played shorter, faster notes. Angry: longer, faster notes played with some force. Sad: softer, longer notes played. Peaceful: softer, shorter notes."

Moderate/High Training

"The tempo, the duration of the notes, the magnitude of it"
"Close my eyes, imagine the events that I wrote down"
"Imagining movie scenes (as well as their soundtracks) to fit with each emotion"
"Imagining scenarios where I felt those specific emotions"
"Louder/shorter for angry, shorter for happy, long, softer for sad, softer, medium speed for peaceful"
"Tried to speed up specific parts of the excerpt in comparison to others/ play at varying paces throughout. Tried to group certain notes together when playing. Staccato vs. elongated notes; loud vs. quiet/soft"
"Soft/loud tones, quick/slow tones, different rhythms & timing"
"I sustained the notes more for it to match the peaceful and sad emotion. I cut the notes short for angry. I cut and sustained the notes for happy."

Archived Data

Archived data can be accessed at the following URL:
https://osf.io/zjtm7/?view_only=e604698f0bc146d59c2282f8e1725ec0

Chapter 5: How do young children communicate different musical emotions? Changes in the use of speed and loudness between three and seven years of age

Kragness, H. E., Baksh, A. M. & Trainor, L. J. (in preparation for submission). How do young children communicate different musical emotions? Changes in the use of speed and loudness between three and seven years of age.

Preface

In Chapter 4, I adapted the music dwell time paradigm to examine how adults with different levels of music training use tempo, articulation, and loudness to communicate emotions in music. In this chapter, I developed a music emotions game to investigate musical emotion expression in 3-, 5-, and 7-year-old children. In this experiment, the children were asked to use the MIDI piano apparatus to express each of four emotions (joy, anger, sadness, and peacefulness) that were expressed by a character in a vignette and a matching photograph. I found that, like adults, children as young as 5 years old used timing and loudness cues to communicate the arousal dimension of the intended emotion. By 7 years, children additionally incorporated articulation to differentiate emotions based on arousal. Three-year-old children did not differentiate between emotions. Previous studies have examined children's communication of happy and sad emotions through song, and found that children tend to rely on song lyrics than expressive cues. To my knowledge, this is the first study to examine children's music production in a non-singing task, and to examine how children's use of expressive cues relates to arousal and valence separately.

Abstract

Music is ubiquitous in children's lives. Caregivers sing to infants to regulate their emotional state, and much of children's entertainment and interactions incorporate play songs. Adults can readily identify emotions in musical performances. How children learn to interpret emotions in music, however, is not well understood. Even less is known about how children use expressive cues to communicate musical emotions, as young children are usually not proficient musicians, and often focus on lyrics more than musical features. In the present work, 3-, 5- and 7-year-old children "performed" music to express basic emotions using a self-pacing paradigm. Participants controlled the onset and offset of each chord in a musical sequence by repeatedly pressing and lifting the same key on a MIDI piano. Key press velocity controlled the loudness of each chord. Participants performed the music to match vignettes and accompanying facial expressions conveying joy, sadness, peacefulness, or anger. We measured how children incorporated speed, loudness, and articulation (connectedness from chord to chord) to express these emotions. By five years old, children distinguished emotional arousal, using faster and louder chords for joy and anger than sadness and peacefulness, as do adults. By 7 years old, children’s performances of different emotions became more differentiated and more similar to adults, using more disconnected chords for joy than peacefulness and for anger than sadness, and incorporating different levels of loudness to distinguish between anger and joy. Three-year-olds showed little differentiation, suggesting that children’s expressions of musical emotions develop dramatically in early childhood.

Keywords: music, emotion, expression, timing, loudness

Introduction

From the newborn period, infants are exposed to music in their environment, such as songs from their caregivers (Trehub & Trainor, 1998). Infants respond differently to music that adults perceive to be calming compared to arousing (Rock, Trainor, & Addison, 1999; Tsang & Conrad, 2010), and caregiver's songs can prevent or delay infant stress, as well as modify arousal levels (Cirelli, Jurewicz, & Trehub, 2018; Corbeil, Trehub, & Peretz, 2005; Shenfield, Trehub, & Nakata, 2003). In adulthood, both music listeners and performers identify music-related emotions as among the primary motivations for engaging with music (Juslin & Laukka, 2004; Lindström, Juslin, Bresin, & Williamon, 2003). Indeed, this aspect of music has been proposed as one possible explanation for the observation that music is found across human cultures (e.g. Trainor, 2015). Thus, the association between music and emotion appears to be present as early as infancy and persists throughout the lifetime.

The ability to use and interpret emotional signals is vital for successful social interactions in general. Gestures, facial expressions, language and music can all contribute to the communication of emotions between humans. Though much of the emotional intent in spoken language is derived from the content of a sentence, emotional intent is also conveyed through the speaker's use of expressive cues, or *prosody* (Morton & Trehub, 2001; Nygaard & Queen, 2008; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006). For example, higher speech rate and intensity are often associated with communicating happiness and anger, and lower speech rates with tenderness and sadness (for a review, see Juslin & Laukka, 2003). Emotive prosody can arise from physiological changes in arousal that affect the vocal tract (Scherer, 1986) but, in addition, individuals can intentionally manipulate prosodic cues to convey different affective states (Banse & Scherer, 1996; Juslin & Laukka, 2003).

Similarly, emotional expression in music is influenced both by the content of the notated music (or “score”) and the expressive cues used by the performing musicians. Both musically trained and untrained adults can identify intended emotions in musical performances, with the most widespread agreement for so-called “basic” emotions, including *happiness, sadness, anger, fear,* and *tenderness* (Balkwill, Thompson, & Matsunaga, 2004; Fritz et al., 2009; Juslin & Laukka, 2003; Mohn, Argstatter, & Wilker, 2010). There is some evidence that these basic emotions, and even some more complex emotions, are recognized above chance levels cross-culturally, although cultural congruency between the performer and listener appears to enhance recognition (Laukka, Eerola, Thingujam, Yamasaki, & Gregory, 2013).

Though there is no consensus on the best way to model emotions in music, a simple two-dimensional approach including the dimensions valence (negative to positive) and arousal (low to high activity) has been widely used (e.g. Schubert, 1999; Vieillard, et al., 2008). This conceptualization is based on the observation that musical features often map onto either *valence* (positive/negative) or *arousal* (high/low activity level). For instance, in Western music, the major mode is usually associated with positively-valenced emotions, such as happiness, and the minor mode with negatively-valenced emotions, such as anger or sadness (Hevner, 1935; Gagnon & Peretz, 2003; Lindström, 2006; Peretz, Gagnon, & Bouchard, 1998). Instruments that play in a low pitch register can usually more effectively communicate negatively-valenced emotions than positively-valenced emotions (Hevner, 1937; Scherer & Oshinsky, 1977). With respect to arousal, fast tempi, high intensity levels, and staccato articulation (playing notes in a disconnected way) are usually associated with high-arousal emotions, such as joy, excitement, and fear, while slow tempi, low sound levels, and legato articulation (playing notes in a connected way) are associated with low-arousal emotions, such as peacefulness, tenderness, and grief (Gabrielsson & Juslin, 1996; Gagnon & Peretz, 2003; Ilie & Thompson, 2006; Juslin, 2000).

Composers notate many of these cues to some extent into musical scores to convey specific emotions. However, a number of studies have shown that musicians additionally incorporate expressive elements into their performances to communicate emotions, including variations in timing, loudness and articulation. In laboratory settings, this has been investigated by asking musicians to perform the same excerpt with different emotional intentions. In such performances, listeners are able to decode the intended emotions above chance levels (Behrens & Green, 1993; Juslin, 1997, 2000; Laukka & Gabrielsson, 2000), demonstrating that performers have a role in communicating emotions in performances beyond what is present in the musical notation.

An important question is how children associate expressive cues with emotions in music, and though a number of studies have examined this question, differences in methods and in stimuli have led to somewhat inconsistent results. Overall, it appears that by 5 to 7 years old, children tend to agree with adults about basic emotions expressed in orchestral music, including *happiness*, *sadness*, *anger*, and *fear*, although the latter two are often confused (Andrade, Vanzella, Andrade, & Schellenberg, 2017; Cunningham & Sterling, 1988; Giomo, 1993; Hunter, Schellenberg, & Stalinski, 2011; Kratus, 1993; Terwogt & van Grinsven, 1991; Vidas, Dingle, & Nelson, 2018). Other studies have reported successful identification of *happiness*, *sadness*, and *anger* by children as young as 3 or 4 years old (Gentile, 1998; Franco, Chew, & Swaine, 2016), and several studies found that that infants as young as 5 months old could discriminate between excerpts that had been selected previously by adults and pre-schoolers as representative of *happy* and *sad* (Flom & Pick, 2012; Flom, Gentile, & Pick, 2008). However, from these studies, it is impossible to know which expressive cues pre-schoolers used in their judgments, and whether infant discrimination was based on differences in perceived affect per se or simply on differences in certain salient acoustic cues (for example, tempo differences between the *happy* and *sad* excerpts).

To our knowledge, only two studies have examined the relative contribution of different expressive cues to children's identifications of emotions by manipulating features in music. Dalla Bella and colleagues (2001) systematically manipulated the tempo and mode (major or minor) of musical passages and asked children to judge whether each passage sounded happy (by selecting a cartoon happy face) or sad (by selecting a cartoon sad face). Children as young as 5 years rated fast passages as happier than slow passages, and by 6 years, children used mode (major, minor) as well as tempo to identify valence. Mote (2011) asked children to perform the same task with musical passages that varied in tempo and familiarity, and found associations between happy/sad distinctions and tempo as young as 4 years, but no associations with familiarity. However, neither of these studies examined children's understanding of musical cues from a dimensional approach – in other words, it is not possible to know whether children associated fast tempi with happiness due to its high arousal level or to its positive valence.

Though studies with adults have examined both the encoding (performer) and decoding (listener) of emotional communication in music, studies with children have been largely limited to decoding – children usually are not able to perform music at a high level. Only one study to our knowledge has investigated the development of musical communication of emotion (Adachi & Trehub, 1998). In this study, 4- to 12-year-old children were asked to sing an experimenter-selected, but familiar, song with the goal of making the experimenter feel happy or sad. In subsequent studies (Adachi & Trehub, 2000; Adachi, Trehub, & Abe, 2004), same-age children and adults were asked to guess the children's intended emotions. Children tended to sing faster, louder, and at a higher pitch when trying to invoke happiness compared to sadness. Older children manipulated tempo to a greater extent than younger children between the happy and sad conditions. At all ages, children primarily used cues found in both music and speech (e.g. tempo/speed, dynamics/loudness), and used cues that are primarily music-specific (e.g. articulation) relatively infrequently.

Although singing studies can be useful, there are notable limitations. First, it is necessary to use familiar songs, which likely have pre-existing emotional associations and/or expressive performance characteristics. A second drawback is reliance on the voice as the instrument, as singing range and accuracy increase significantly throughout childhood (Hedden, 2012). In the present study, we used a self-pacing paradigm to test young children’s expressive musical productions. Children pressed a single key on a MIDI piano (marked with a sticker) to self-pace through a musical excerpt several times, portraying a different emotion in each performance. They were able to control the onset, offset, and loudness of each chord in sequence. The children were asked to use the piano apparatus to play a music game. In this game, children played chord sequences to accompany vignettes and expressive faces that conveyed four different emotions, each of which was in a different quadrant of a valence-by-arousal circumplex. Aside from being widely used in the music emotion literature, this dimensional representation is thought to represent children’s earliest conceptualization of emotion schemes (Widen & Russell, 2008).

The simplicity of the self-pacing tasks removes the requirement of musical training and experience. In this way, we examined young children's associations between expressive cues and emotion dimensions in the context of music.

Method

Participants

Because no previous studies have examined such a question using a self-paced performance paradigm in children, we did not have any basis for estimating expected variance. In the absence of the ability to perform a power analysis, we tested 36 children in total (12 at each age). Forty-five children were recruited to participate. Of those, seven were excluded from analyses due to technical errors (n = 2), expressed desire to withdraw (n = 2), reporting greater than two years of formal musical training (n = 1), or parent-reported developmental delays including gross motor delays (n = 1) and receptive language barrier (n = 1). Two additional participants scored below the third percentile in receptive vocabulary (PPVT-4) and were excluded due to concerns about their ability to understand the task. Thus, in the final sample, 36 children served as participants, including 12 3-year-olds (5 boys, 7 girls; M_{age} = 3.60, SD = .08), 12 5-year-olds (6 boys, 6 girls; M_{age} = 5.64, SD = .10), and 12 7-year-olds (6 boys, 6 girls; M_{age} = 7.57, SD = .04) (see Table 1 for more participant details). Participants were recruited from the Developmental Studies Database at the university and received a certificate and a

Table 1
Demographic breakdown

	3-year-olds	5-year-olds	7-year-olds
N	12	12	12
Gender distribution	5 boys, 7 girls	6 boys, 6 girls	6 boys, 6 girls
Age (years), M(*SD)*	3.60*(.08)*	5.64*(.10)*	7.57*(.04)*
Hours of music listening/week	8.38*(5.72),* 4 non-responders	6.40*(6.35)* 2 non-responders	5.21*(2.78)*
Number reporting > 10% exposure to non-English	3, 5 non-responders	0, 4 non-responders	2, 1 non-responder
Languages	Spanish (1), Croatian (1), Punjabi (1)	N/A	French (2)
Distribution of family income	60k or less (4) 60k-120k (5) 120k or more (3)	60k-120k (8) 120k or more (4)	60k or less (1) 60k-120k (3) 120k or more (6) 2 non-responders
Number in early music classes (past or present)	2	2	4
Formal music training (past or present)	None	None	Piano (3, < 2 years), guitar (1, < 2 years)
Formal dance training (past or present)	Dance movement/ballet (1), dance movement (1), ballet (1), Mom's Baby Dance (1)	Ballet/tap/jazz (1), jazz (1)	Tap/jazz (1), ballet/jazz (1)
TEC scores	3.78*(0.91)*	5.48*(1.48)*	6.86*(1.44)*
PPVT percentile	74.25*(22.45)*	81.25*(18.25*)	79.83*(22.07)*

small prize of their choice as compensation. The procedures were approved by the McMaster Research Ethics Board (#2011 139), “Perception of Sound Structure in Infants, Children, and Adults.”

Stimuli

Music selection. Four excerpts from 4-voice Bach chorales were selected as stimuli: two that were originally composed in the major mode and two originally composed in the minor mode. With the guidance of an expert in Baroque music theory, each excerpt was transposed to the key of F (major or minor) and small alterations were made such that each consisted of twenty-four chords of quarter-note length. Four additional musical excerpts were created by transforming the excerpts originally in a major mode to F minor, and by transforming the minor mode excerpts to F major (see Supplemental Material for the scores). These chord sequences were used in a previous study with adults (Kragness & Trainor, submitted).

Vignette selection. Seven vignettes were selected from previous studies examining children’s understanding of emotions in stories (see Table 2). Two vignettes were selected for each emotion, except for *peaceful*, for which only a single vignette could be found in the previous literature. An additional *peaceful* vignette was composed by the authors, for a total of eight. Some of the vignettes were slightly adapted such that each story consisted of a character responding to a

Table 2
Vignettes

Emotion	Story	Reference
Joy (Positive, high-arousal)	“It was _____’s birthday. All of his/her friends came to his/her birthday party and gave her presents ._____ jumped up and down and clapped her hands. How do you think he/she felt? He/she felt happy and excited.”	Based on Widen & Russell (2010)
	“________ wanted his/her friends to come over and play. So he/she asked them, and they came to play at his/her house. He/she smiled and jumped up and down. How do you think he/she felt? He/she felt happy and excited.”	Based on Ribordy et al. (1998)
Sadness (Negative, low-arousal)	“_________ went to feed his/her pet goldfish. But it was not swimming. It was not even in the fish tank. ______’s fish had died. ________ walked over to a chair and sat down. Tears came to his/her eyes. He/she didn’t want to talk to anyone. How do you think he/she felt? He/she felt sad.”	Based on Widen & Russell (2002)
	“_________’s friend, who he/she really liked to play with, moved away. He/she couldn’t play with his/her friend anymore. How do you think he/she felt? He/she felt sad.”	Based on Ribordy et al. (1988)
Anger (Negative, high-arousal)	“______ was waiting in line. Then a boy cut in line in front of her. He didn’t even ask. _____ shoved him out of line and yelled at him. How do you think he/she felt? He/she felt angry.”	Based on Widen & Russell (2010)
	“________s little brother broke his/her favourite toy on purpose. He/she clenched his/her fists and stomped his/her feet. How do you think he/she felt? He/she felt angry.”	Based on Ribordy et al. (1988)
Peace (Positive, low-arousal)	“______ drew a picture and showed it to his/her father. His/her father really liked it and said _____ did a good job. That made him/her feel very content and happy, and he/she gave his/her dad a hug.”	Based on Ribordy et al. (1988)
	“____ was feeling very tired and likes to sleep with his/her favourite stuffed animal. He/she searched around his/her whole room for the stuffed animal. When he/she found his/her stuffed animal, how do you think he/she felt? He/she felt happy and sleepy. He/she hugged his/her stuffed animal, smiled, and went to bed.”	Original

specific situation and the explicit statement of the emotion associated with the situation. In case 3-year-old children might have difficulty understanding the words “joy” and “peaceful,” we used the phrases “happy and excited” (joy) and “happy and content”/”happy and sleepy” (peaceful), respectively.

Facial expression selection. Sixteen faces were selected from the NimStim set of facial expressions (Tottenham, et al., 2009; see Table 3) to accompany the vignettes. This was done to reinforce the verbal instructions. Each of the four emotions being tested corresponded to four different faces (two male faces and two female faces). For happy/excited, we selected NimStim pictures that were “happy, open-mouth, exuberant,” and for peaceful, we used the “happy, closed-mouth” pictures. No individual actor was represented in more than one image across the set.

Procedure

After a brief explanation of the task, informed consent was obtained from the parent and verbal consent from the children, who were informed that they could choose to stop at any time. The child was told that they would play through a musical excerpt by pressing middle C (indicated with a sticker) on a MIDI piano keyboard. Pressing the key would initiate the onset of each successive chord in the musical excerpt and releasing the key would cause the offset of the chord.

Table 3

Emotional faces used in the study

Emotion	NimStim faces – female	NimStim faces - male
Joy	01F_HA_X	30M_HA_X
(Positive, high-arousal)	09F_HA_X	36M_HA_X
Sadness	02F_SA_C	22M_SA_C
(Negative, low-arousal)	07F_SA_C	34M_SA_C
Anger	10F_AN_C	20M_AN_C
(Negative, high-arousal)	05F_AN_C	23M_AN_C
Peace	03F_HA_C	28M_HA_C
(Positive, low-arousal)	08F_HA_C	32M_HA_C

Participants could alter the loudness by changing the level of force used to press the key.

A practice phase was used to familiarize participants with the task and to allow participants to practice manipulating the musical cues. Prior to the testing blocks, the experimenter asked the child to perform a short practice excerpt (a C Major scale) with different musical features (quickly, slowly, with short notes, with long notes, loudly, and quietly) to ensure that they were aware of the expressive cues available to them.

Participants experienced two testing blocks. One block consisted of a musical excerpt in a major mode and the other block consisted of a different musical excerpt in a minor mode. The order in which participants experienced the major and minor mode excerpts was counterbalanced across participants. Each testing block began with the participant listening to the musical excerpt through loudspeakers. Participants were told that they were going to play a musical game, and that they should play the music in a manner that matched the feeling of the story. The experimenter read a short vignette to the participant, while showing the participant a picture of a face that matched the emotion portrayed in the vignette. Then, the experimenter asked the participant how the character in the vignette felt. After the participant had verbally provided an answer, the experimenter indicated whether the answer was correct or incorrect by explicitly stating the target emotion (e.g. “That’s right, the character felt angry” or “Actually, the character felt angry”). Then, the experimenter asked the participant to play the musical excerpt in a way that sounded like the target emotion (e.g. “Can you play the music in a way that sounds angry?”). This process was repeated for each of the four emotions being tested for each of the two excerpts, for a total of 8 test trials. The order in which the emotions were presented was randomized within each block. The stories and faces used in the first block were not repeated in the second testing block. The participant either viewed only male or only female faces from the face set throughout the two testing blocks.

After the two testing blocks had been completed, the experimenter administered the Test of Emotional Comprehension (TEC, Pons & Harris, 2000) to obtain a measure of each participant’s understanding of emotions in general. Finally, the Peabody Picture Vocabulary Test (PPVT-4, Dunn & Dunn, 2007) was administered to examine each participant’s receptive vocabulary skills, which could affect their ability to understand the task instructions.

Results

Tempo

We first examined tempo (speed) differences (see Figure 1A). Tempo was defined as chord onsets per minute, such that higher values indicate faster speeds. The data were submitted to a mixed ANOVA with within-subjects factors valence (positive, negative) and arousal (high, low), and the between-subjects factor age group (3-year-old, 5-year-old, 7-year-old). Because this ANOVA revealed a significant three-way interaction, $F(2, 33) = 3.363$, $p = .0469$, each age group was

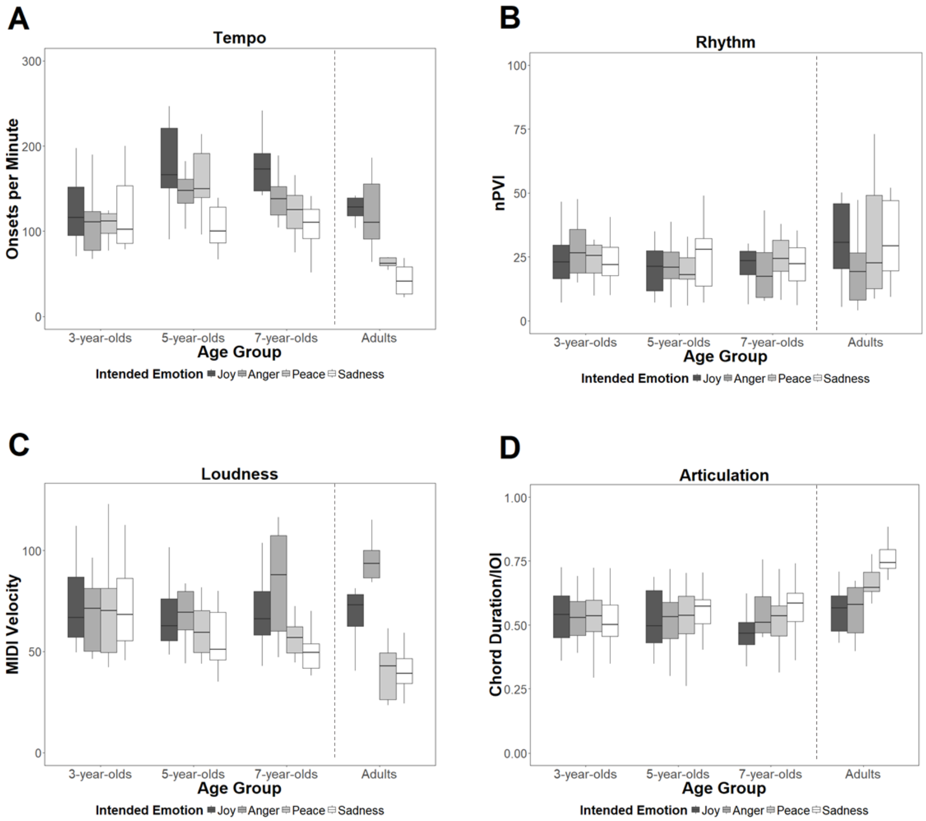


Figure 1

The patterns of cue use by 3-, 5-, and 7-year-old children as compared to adults tested in a previous study (Kragness & Trainor, submitted). (A) Portrays tempo, defined as the number of onsets per minute, such that higher values are faster and lower values are slower. (B) Portrays rhythm as defined by the normalized pairwise variability index (nPVI), such that higher values imply more pairwise variability (e.g., a dotted rhythm) while 0 represents isochrony. (C) Portrays loudness, which was manipulated via MIDI key press velocity. Higher values were experienced as louder, and lower values as softer. (D) Portrays articulation. Lower values indicate greater disconnect between chords (*staccato*), and higher values represent more connected chords (*legato*).

analyzed with separate ANOVAs to examine the effects of valence and arousal in each age group. For the three-year-old children, there were no significant main effects of valence or arousal ($p = 0.847$ and $p = 0.631$, respectively), and no significant interaction ($p = 0.212$). For five-year-old children, there were significant main effects of valence, $F(1, 11) = 11.041$, $p = 0.006$, $\eta_G^2 = .180$, and arousal, $F(1, 11) = 8.825$, $p = 0.013$, $\eta_G^2 = .121$, and a significant interaction, $F(1, 11) = 8.771$, $p = 0.013$, $\eta_G^2 = .067$. Post-hoc paired t-tests revealed that the speed difference between the high- and low-arousal conditions for the negative-valence emotions (anger and sadness) was significant ($t(11) = -3.877$, $p = 0.003$, $d_Z = 1.119$), but the tempo difference was not significant between the positive-valence emotions ($p = 0.482$). For seven-year-old children, there was a significant main effect of arousal ($F(1,11) = 13.201$, $p = 0.004$, $\eta_G^2 = .135$), but no significant effect of valence ($p = .184$).

Rhythm

To examine rhythmic patterning, we employed the normalized pairwise variability index (nPVI). This formula for calculating nPVI has been used to examine rhythmic patterns in speech (Grabe & Low, 2002) and music (Patel & Daniele, 2003):

$$nPVI = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right| /(m-1) \right]$$

where *m* is the number of intervals in the excerpt and *d* is the duration of the *k*th interval. An nPVI value of 0 indicates perfect isochrony (no durational contrast between pairs at all), while a value of 200 represents maximal durational contrast (Figure 1B). The omnibus ANOVA revealed a significant interaction between age group and valence, *F*(2,33) = 3.532, *p* = .041, motivating separate analysis of each age group. No significant effects or interactions were observed for 3-year-olds or 5-year-olds, but there was a significant main effect of arousal for 7-year-olds, *F*(1,11) = 5.714, *p* = .036, such that low-arousal emotions were played with more rhythmic variation than high-arousal emotions. There was no significant main effect of valence (*p* = .195) nor significant interaction (*p* = .972).

Loudness

Participants varied the loudness of each chord by varying the velocity of their key presses. Here, we used the MIDI velocity information recorded by the piano apparatus as a proxy for loudness. MIDI velocity measures range from 1-127, with 1 representing the minimum level (experienced as the quietest sound possible on the piano) and 127 representing the maximum level (experienced as the loudest sound possible on the piano; Figure 1C). The omnibus ANOVA revealed a three-way interaction with arousal, valence, and age group (*F*(2, 33) = 6.535, *p* = .004). Further analyses were performed within each age group separately. For three-year-olds, there were no significant main effects (valence*: p*

= .740, arousal: $p = .899$), nor significant interaction ($p = .234$). For five-year-olds, there was no main effect of valence ($p = .287$) or interaction ($p = .232$), but there was a significant main effect of arousal ($F(1,11) = 18.182$, $p = .001$, $\eta_G^2 = .099$), such that high-arousal emotions were played more loudly than low-arousal emotions. For seven-year-olds, we observed a significant valence-by-arousal interaction ($F(1, 11) = 12.945$, $p = .004$, $\eta_G^2 = .102$), such that the loudness difference between high- and low-arousal emotions was larger for negative valence than positive valence.

Articulation

In music, *articulation* refers to the connectedness between notes, ranging from smooth and connected (“legato”) to disconnected, with silent gaps between notes (“staccato”). In the present study, this was measured by calculating the proportion of the onset-to-onset interval in which the chord was sounded. For example, if a chord was sounded for 500 ms of a 1000 ms interval, the articulation value would be .5. If a chord was sounded for 200 ms of a 1000 ms interval, the value would be .2. Thus, values closer to 0 represent more disconnected chords, while values closer to 1 represent more connected chords (Figure 1D). The mixed ANOVA revealed no significant effects or interactions.

Discussion

The purpose of the present study was to examine how young children of different ages use acoustic cues to communicate emotions in music. Previous studies of children's emotional expression in music have been limited to singing and have exclusively investigated the emotions "happy" and "sad." Using a simple self-pacing apparatus, we found evidence that five- and seven-year-old children express emotional arousal using tempo and loudness, with patterns of cue use mirroring those observed in adults. With respect to tempo, both five-year-old and seven-year-old children played faster to differentiate high-arousal emotions from low-arousal emotions, although in five-year-olds this was observed primarily for differentiating between the negative emotions *anger* and *sadness* rather than the positive emotions *joy* and *peacefulness*. This is consistent with previous research using emotion identification tasks that found that tempo is among the earliest cues used by children when evaluating emotions in music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Kratus, 1993; Mote, 2011). The present results reinforce and extend these findings to the production domain. Loudness was also used to differentiate between high- and low-arousal emotions for children as young as five years. To our knowledge, no previous studies have examined the role of loudness in children's emotion judgments, although Adachi and Trehub (1998) found that children as young as four years manipulated loudness to differentiate between sadness and happiness in vocal music. The present study extends this observation by demonstrating that both five-year-old and seven-year-old children played high-arousal emotions louder than low-arousal emotions in both positive and negative valence conditions, similarly to adults. This suggests that loudness may be a particularly salient cue for children when judging musical emotions, and should be considered in future studies of children's emotional judgments.

To our knowledge, the present study is the first to examine children's emotional productions with regards to articulation and rhythm. Although statistically significant differences were not observed in any age group, the articulation data show an adult-like pattern of differentiation as early as five years, with increased differentiation by seven years. Visual inspection of the data shows that five- and seven-year-old children performed with the most disconnected chords to convey *joy*, followed by *anger*, *peace*, and *sadness*, for which they performed with the most connected chords. One possible explanation for the lack of significance is that articulation was more difficult to manipulate than tempo or loudness, requiring the children to control both the onset *and* the offset of each note, resulting in noisier measurements. An alternative possibility is that articulation is not a particularly salient cue for emotional expression for children. Results from one study (Kratus, 1993) suggest that children as young as six years do rely on articulation to make emotion judgments, but that study did not examine potential correlations among various cues, so it is difficult to know the contribution of articulation specifically. Interestingly, articulation is considered to be relatively specific to the music domain, compared to other cues such as loudness and speed that are used prevalently in speech, as well as music (Juslin & Laukka, 2003). From this perspective, we might predict that sensitivity to articulation as an emotional cue would develop later than other, more domain-general cues.

Interestingly, in contrast to all other cues, we did not observe any consistent developmental pattern for rhythm use. Even the oldest children did not show a pattern that resembled that shown by adults. Rhythm in music is much more regular and periodic than in speech, which is consistent with the idea that domain-specific cues are less salient to children. Furthermore, rhythm can be considered largely a compositional cue rather than a performer cue, as it is dictated by the score, although performers do take liberties with the timing on a micro level. As well, the chord sequences used in the present study were based on chorales by J. S. Bach in which the chords were written, for the most part, to be equally spaced from each other, which might encourage an isochronous interpretation from participants, and discourage rhythmic variation. Interestingly, a previous study using the same stimuli as in the present study examined adult performances and found that participants, regardless of amount of music training, consistently incorporated higher nPVI values in their *sadness* performances compared to their *anger, happiness,* and *peacefulness* performances (Kragness & Trainor, under review), although another study examining compositions by highly-trained violinists and vocalists did not find any differences in nPVI use amongst basic emotions (Quinto, Thompson, & Keating, 2013). Because nPVI is a measure of pairwise durational contrast, it does not represent all possible variations in rhythm that could be utilized. The use of rhythm in emotionally expressive performance could therefore be studied in more detail by examining expert listener ratings of rhythmic complexity across performances expressing different emotions. In general, much work remains to be done to understand the role of rhythm in emotion identification and expression, both in adults and in children.

In the present experiment, three-year-olds did not show evidence of differentiating between the four target emotions using any of the available cues. This is consistent with prior research using emotion identification tasks that have found generally poor ability to identify musical emotion in 3- and 4-year-old children (Dalla Bella et al., 2001; Gregory, Worrall, & Sarge, 1996; Kastner & Crowder, 1990; Mote, 2011). To our knowledge, the effects of loudness on emotion identification tasks with three-year-olds have yet to be explored, although slightly older children (4 - 12 year-olds) show evidence of manipulating loudness to convey happiness versus sadness in vocal music (Adachi & Trehub, 1998). Based on the present results alone, it is not clear whether three-year-old participants' failure to differentiate emotions is the result of failure to understand emotions in music, difficulty understanding the task, or difficulty with the motor demands involved in the task. Though *happy, sad,* and *angry* are among the earliest emotion labels acquired, typically beginning to be used between two and three years of age (Ridgeway, Waters, & Kuczaj, 1985; Widen & Russell, 2003), incorrect labeling of faces and emotion stories are common well into the preschool years, and 3-year-olds' lack of differentiation in the performance task of the present experiment is consistent with these observations. At the same time, a few studies demonstrate discrimination between happy and sad music in infants (Flom & Pick, 2012; Flom, Gentile, & Pick, 2008) and even that infants associate positive visual stimuli more strongly with "happy" music than with "sad" music (Nawrot, 2003; Xiao, et al., 2017). This apparent discrepancy across studies of infants and preschoolers merits further investigation into which acoustic cues infants are using to differentiate between excerpts and to make associations – and it highlights the challenges in selecting age-appropriate methods for toddlers and preschool children.

An interesting question to be explored in future research is the role of formal music training in childhood. Past research in adults suggests that music training does not significantly enhance identification of emotions in music (Bigand et al., 2005; Juslin, 1997; Song, Dixon, Pearce, & Halpern, 2016; but see Castro & Lima, 2014). This was corroborated by recent work using the self pacing paradigm of the present paper, which found that adults with no music training use tempo, loudness and articulation to express emotions similarly to those with training (Kragness & Trainor, submitted). This suggests that the level of informal experiences with music accumulated by adulthood is sufficient for identifying and expressing musical emotions. Following that idea, we might expect to see formal training differences earlier in development, when less exposure has been acquired. To our knowledge, only one previous study has examined the effect of formal training on children's emotional judgments of excerpts and found no differences between 12-year-olds with six years of formal training and 6-year-olds without training (Kratus, 1993). In this case, the participants were asked to make dichotomous decisions about the valence and arousal of excerpts, so the null result may have resulted from universally high performance across participants. Effects of musical training might be observed in a more difficult task, such as the performance task in the present paper, or in a rating task with a larger gradient to choose from.

An enduring question in this literature relates to the specific music stimuli that are used. While many studies have employed orchestral music (e.g., Cunningham & Sterling, 1988; Giomo, 1993; Nawrot, 2003), others have found superior performance when using culturally familiar music that is intended for children (Mote, 2011; Schellenberg, Nakata, Hunter, and Tamoto, 2007). In the present study, we used modified chord sequences from Bach chorales, which are not child-directed or likely to be overly familiar (though they are likely to be culturally familiar). We selected these sequences as stimuli for several reasons. First, the sequences have a reliable phrase structure, such that each sequence is made of three 8-chord sub-phrases. Using these stimuli also enabled direct comparison with previous work on adults using the same stimuli (Kragness & Trainor, submitted). Finally, these stimuli are unlikely to be familiar to children, thus avoiding previously-held associations with recognized pieces. However, it may be the case that selecting child-directed stimuli would lead to greater task engagement, potentially resulting in evidence of emotional differentiation in performances at younger ages.

The present results demonstrate that when using cues such as tempo and loudness, the younger children differentiated their performances primarily based on arousal, with further differentiation by valence emerging later. However, previous research on children's understanding of emotion in music and speech prosody has largely focused on "happy" and "sad" emotion categories, in which valence and arousal are confounded ("happy" is high arousal and positive valence whereas sad" is low arousal and negative valence). When other emotions have been included, "angry" and "neutral" are most commonly used. Notably, two developmental studies of musical emotion understanding which accounted for both valence and arousal dimensions found that accurate identification of arousal level preceded accurate identification of valence in development (Andrade et al., 2017; Hunter et al., 2011). Results from these studies and the current experiment suggest that it is important to include both valence and arousal dimensions in investigations of children's understanding of emotions.

In the present work, we present a new method for testing children's expressive musical productions that does not rely on the vocal apparatus or the ability to play a musical instrument. This method is easy to explain to young children, highly flexible, and can be used to examine children's productions of familiar music or unfamiliar music. In Western society, singing in a social setting can be an intimidating task. Though singing is useful for studying children's musical development, the task used in the present study can corroborate and extend those findings. Further, we demonstrated adult-like patterns of cue use for tempo and loudness that emerge as early as five years and become more adult-like by seven. Additionally, music-specific cues like articulation and rhythm were not observed to be used significantly for differentiation, suggesting a pattern in which use of domain-general cues occurs earlier than music-specific cues. Overall, in the present work, we offer a new way to examine children's expressive musical productions, and demonstrate striking age-related differences in the using of timing and loudness to communicate musical emotions in children between three and seven years old.

References

Adachi, M., & Trehub, S. E. (1998). Children's expression of emotion in song. *Psychology of Music, 26*(2), 133-153. http://dx.doi.org/10.1177/0305735698262003

Adachi, M., & Trehub, S. E. (2000). Decoding the expressive intentions in children's songs. *Music Perception: An Interdisciplinary Journal, 18*(2), 213-224. http://dx.doi.org/10.2307/40285909

Adachi, M., Trehub, S. E., & Abe, J. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46*(4), 322-336. http://dx.doi.org/10.1111/j.1468-5584.2004.00264.x

Andrade, P. E., Vanzella, P., Andrade, O. V. C. A., & Schellenberg, E. G. (2017). Associated emotions with Wagner's music: A developmental perspective. *Psychology of Music, 45*(5), 752-760. http://dx.doi.org/10.1177/0305735616678056

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46*(4), 337-349. http://dx.doi.org/10.1111/j.1468-5584.2004.00265.x

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614-636. http://dx.doi.org/10.1037/0022-3514.70.3.614

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20-33. http://dx.doi.org/10.1177/030573569302100102

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional response to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*(8), 1113-1139. http://dx.doi.org/10.1080/02699930500204250

Bullock, M., & Russell, J. A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioural Development, 7*(2), 193-214http://dx.doi.org/10.1177/016502548400700207

Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception: An Interdisciplinary Journal, 32*(2), 125-142. http://dx.doi.org/10.1525/mp.2014.32.2.125

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2018, March). Musical rhythms in infancy: Social and emotional effects. *Paper presented at the Cognitive Neuroscience Society Annual Meeting*, Boston, MA.

Corbeil, M., Trehub, S. E., & Peretz, I. (2005). Singing delays the onset of infant distress. *Infancy, 21*(3), 373-391. http://dx.doi.org/10.1111/infa.12114

Cunningham, J., & Sterling, R. (1988). Developmental changes in the understanding of affective meaning in music. *Motivation and Emotion*, *12*(4), 399-413. http://dx.doi.org/10.1007/bf00992362

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1-B10. http://dx.doi.org/10.1016/s0010-0277(00)00136-0

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test.* Minneapolis, MN: Pearson Assessments.

Flom, R., & Pick, A. D. (2012). Dynamics of infant habituation: Infants' discrimination of musical excerpts. *Infant Behavior and Development, 35*(4), 697-704. http://dx.doi.org/10.1016/j.infbeh.2012.07.022

Flom, R., Gentile, D. A., & Pick, A. D. (2008). Infants' discrimination of happy and sad music. *Infant Behavior and Development, 31*(4), 716-728. http://dx.doi.org/10.1016/j.infbeh.2008.04.004

Franco, F., Chew, M., & Swaine, J. S. (2016). Preschoolers' attribution of affect to music: A comparison between vocal and instrumental performance. *Psychology of Music, 45*(1), 131-149. http://dx.doi.org/10.1177/0305735616652954

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., . . . Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19*(7), 573-576. http://dx.doi.org/10.1016/j.cub.2009.02.058

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*(1), 68-91. http://dx.doi.org/10.1177/0305735696241007

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to happy-sad judgements in equitone melodies. *Cognition & Emotion, 17*(1), 25-40. http://dx.doi.org/10.1080/02699930302279

Gentile, D. A. (1998). An ecological approach to the development of perception of emotion in music. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 59*, 2454.

Giomo, C. J. (1993). An experimental study of children's sensitivity to mood in music. *Psychology of Music, 21*(2), 141-162. http://dx.doi.org/10.1177/030573569302100204

Grabe, E., & Low, E. L. (2002). Durational variability in speed and the rhythm class hypothesis. In C. G. (Eds.), *Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Hedden, D. (2012). An overview of existing research about children's singing and the implications for teaching children to sing. *Update: Applications of Research in Music Education, 30*(2), 52-62. http://dx.doi.org/10.1177/8755123312438516

Hevner, K. (1935). The affective character of the major and minor mode in music. *American Journal of Psychology, 47*(1), 103-118. http://dx.doi.org/10.2307/1416710

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology, 49*(4), 621-630. http://dx.doi.org/10.2307/1416385

Hunter, P. G., Schellenberg, E. G., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology*, *110*(1), 80-93. http://dx.doi.org/10.1016/j.jecp.2011.04.001

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

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception: An Interdiscliplinary Journal, 14*(4), 383-418. http://dx.doi.org/10.2307/40285731.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Pereption and Performance, 26*(6), 1797-1813. http://dx.doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814. http://dx.doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*(3), 217-238. http://dx.doi.org/10.1080/0929821042000317813

Kragness, H. E., & Trainor, L. J. (2018). Young children pause on phrase boundaries in self-paced music listening: The role of harmonic cues. *Developmental Psychology*, *54*(5), 842-856. http://dx.doi.org/10.1037/dev0000405

Kratus, J. (1993). A developmental study of children's interpretation of emotion in music. *Psychology of Music*, *21*(1), 3-19. http://dx.doi.org/10.1177/030573569302100101

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, *28*(2), 181-189. http://dx.doi.org/10.1177/0305735600282007

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and cultural-specific factors in the recognition and performance of musical affect expressions. *Emotion, 13*(3), 432-449. http://dx.doi.org/101.1037.a0031388

Lindström, E. (2006). Impact of melodic organization on perceived structure and emotional expression in music. *Musicae Scientiae*, *10*(1), 85-117. http://dx.doi.org/10.1177/102986490601000105

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": a questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education, 20*(1), 23-47. http://dx.doi.org/10.1177/1321103x030200010201

Mohn, C., Argstatter, H., & Wilker, F. (2010). Perception of six basic emotions in music. *Psychology of Music, 39*(4), 503-517. http://dx.doi.org/10.1177/0305735610378183

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72*(3), 834-843. http://dx.doi.org/10.1111/1467-8624.00318

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion, 11*(3), 818-622. http://dx.doi.org/10.1037/a0022573

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music*, *31*(1), 75-92. http://dx.doi.org/10.1177/0305735603031001325

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017-1030. http://dx.doi.org/10.1037/0096-1523.34.4.1017

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-B45. http://dx.doi.org/10.1016/S0010-0277(02)00187-7

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

Pons, F., & Harris, P. (2000). *Test of Emotion Comprehension - TEC.* Oxford: University of Oxford.

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4,* 184. http://dx.doi.org/10.3389/fpsyg.2013.00184

Ribordy, S. C., Camras, L. A., Stefani, R., & Spaccerelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *Journal of Clinical Child Psychology, 17*(4), 322-325. http://dx.doi.org/10.1207.s15374424iccp1704_4

Ridgeway, D., Waters, E., & Kuczaj, S. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology, 21*(5), 901-908. http://dx.doi.org/10.1037.0012-1649.21.5.901

Rock, A. M., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Develpomental Psychology, 35*(2), 527-534, http://dx.doi.org/10.1037/0012-1649.35.2.527

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Pscyhology, 39*(6), 1161-1178. http://dx.doi.org/10.1037/h0077714

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165. http://dx.doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346. http://dx.doi.org/10.1107/BF00992539

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92. http://dx.doi.org/10.1177/0022022101032001009

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*(3), 154-165. http://dx.doi.org/10.1080/00495399908255353

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music, 31*(4), 365-375. http://dx.doi.org/10.1177/03057356030314002

Song, Y., Dixon, S., Pearce, M. T., & Halpern, A. R. (2016). Perceived and induced emotion responses to popular music: Categorical and dimensional models. *Music Perception: An Interdiscliplinary Journal, 33*(4), 472-492. http://dx.doi.org/10.1525/mp.2016.33.4.472

Terwogt, M. M., & van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music, 19*(2), 99-109. http://dx.doi.org/10.1177/0305724591192001

Thompson, W. F., & Balkwill, L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*, 407-424. http://dx.doi.org/10.1515/sem.2006.017

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242-249. http://dx.doi.org/10.1016/j.psychres.2008.05.006

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B, 370,* (1664), 20140089. http://dx.doi.org/10.1098/rstb.2014.0089

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

Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior and Development, 33*(1), 96-100, http://dx.doi.org/10.1016/j.infbeh.2009.11.006

Vidas, D., Dingle, G. A., & Nelson, N. L. (2018). Children's recognition of emotion in music and speech. *Music & Science,* http://doi.org/10.1177/2059204318762650

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary, and peaceful musical excerpts for research on emotions. *Cognition & Emotion, 22*(4), 720-752. http://dx.doi.org/10.1080/02699930701503567

Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merill-Palmer Quarterly, 48*(3), 248-262. http://dx.doi.org/10.1353/mpq.2002.0013

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39*(1), 114-128. http://dx.doi.org/10.1037/0012-1649.39.1.114

Widen, S. C., & Russell, J. A. (2004). The relative power of an emotion's facial expression, label, and behavioral consequences to evoke preschoolers' knowledge of its cause. *Cognitive Development, 19*(1), 111-125. http://dx.doi.org/10.1016/j.cogdev.2003.11.004

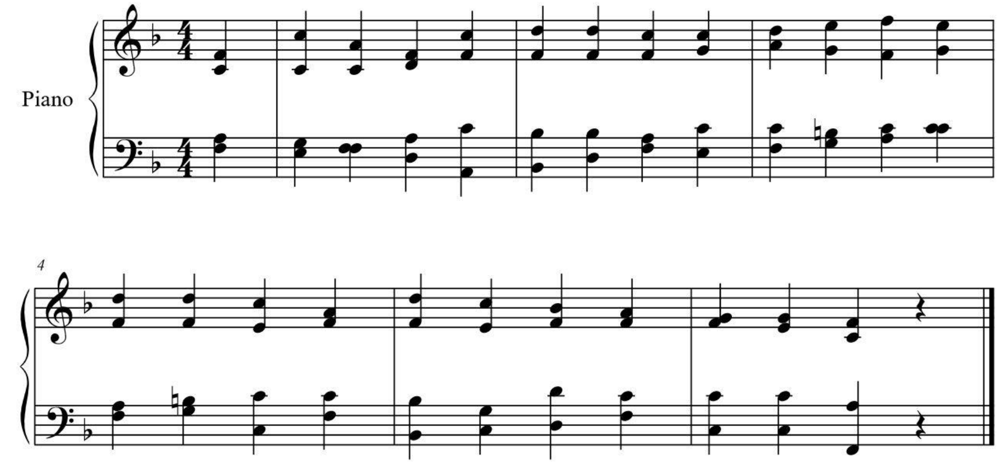
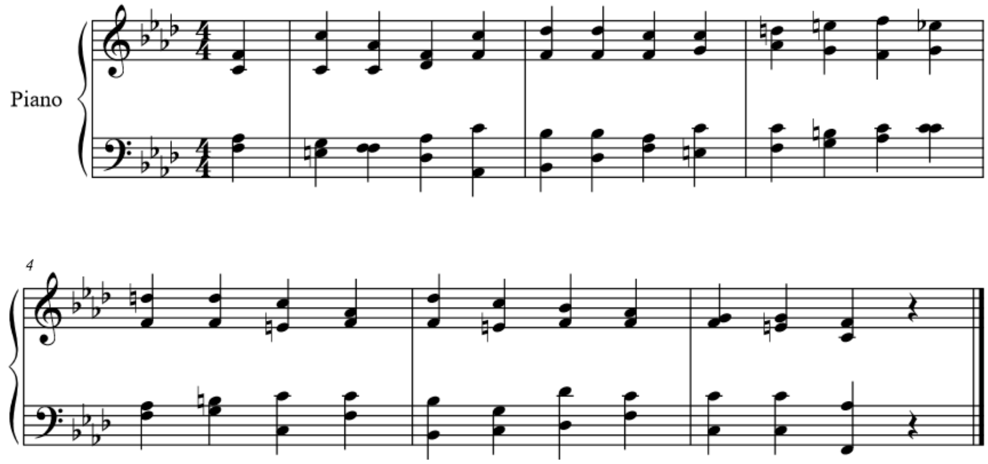
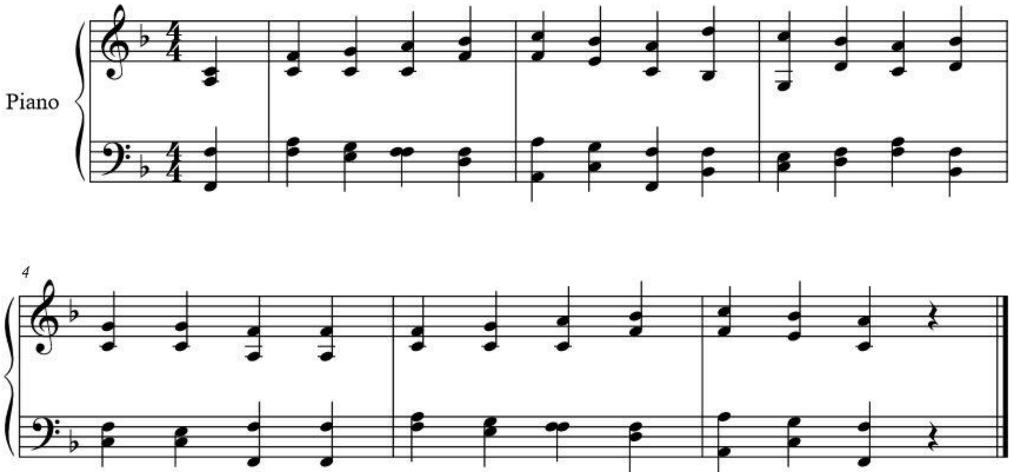
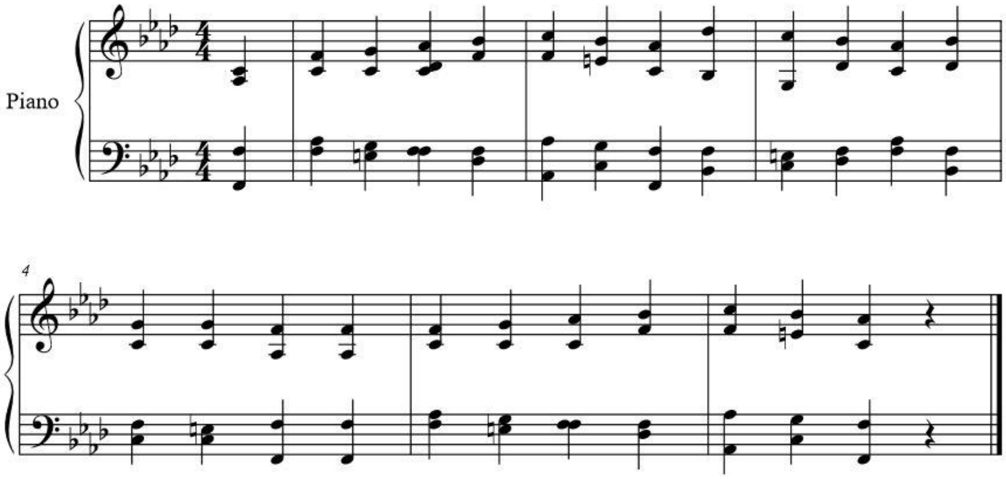
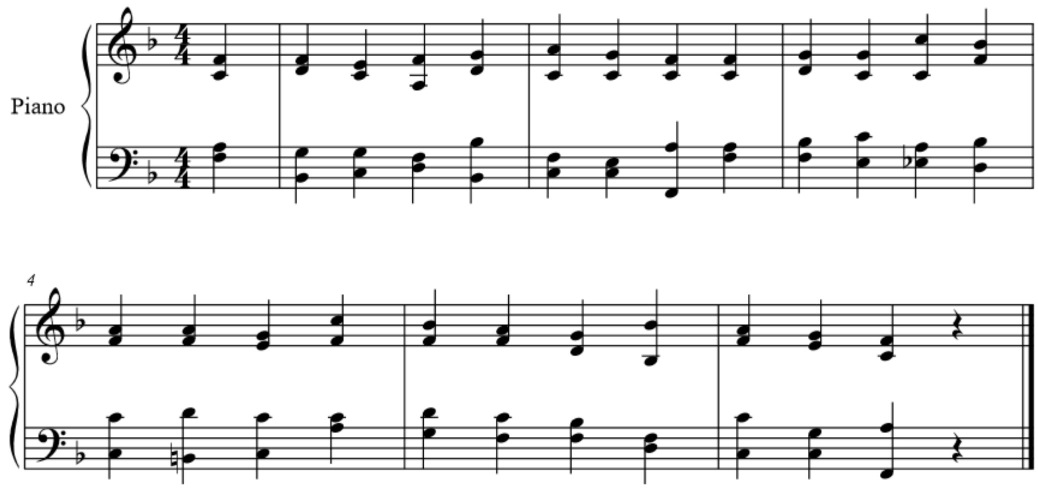
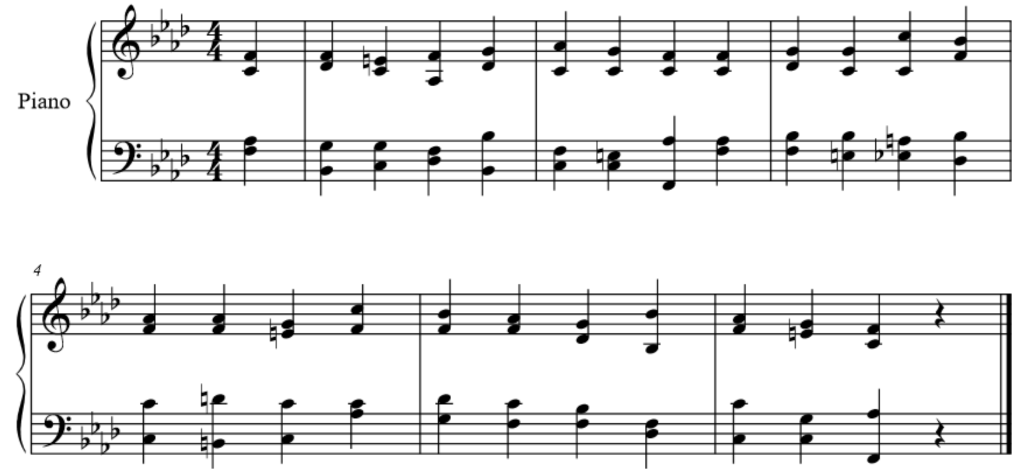
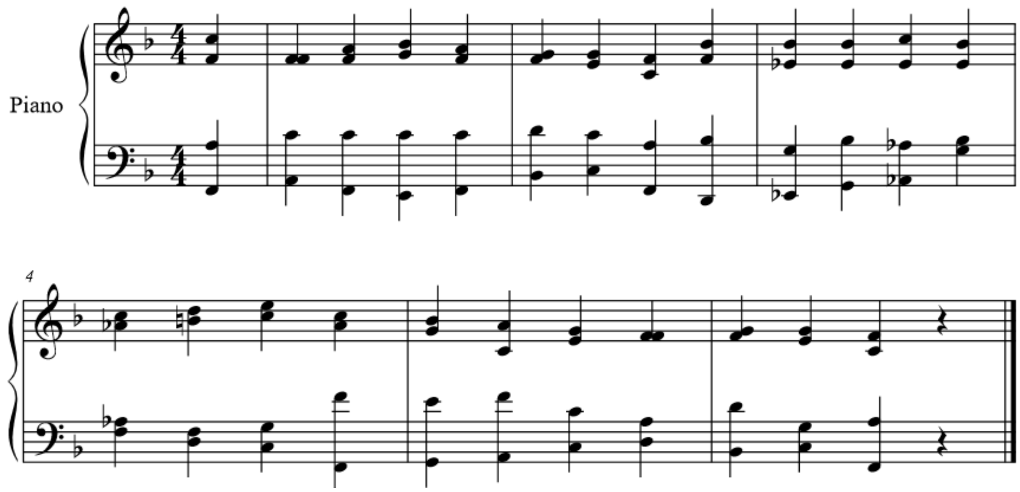
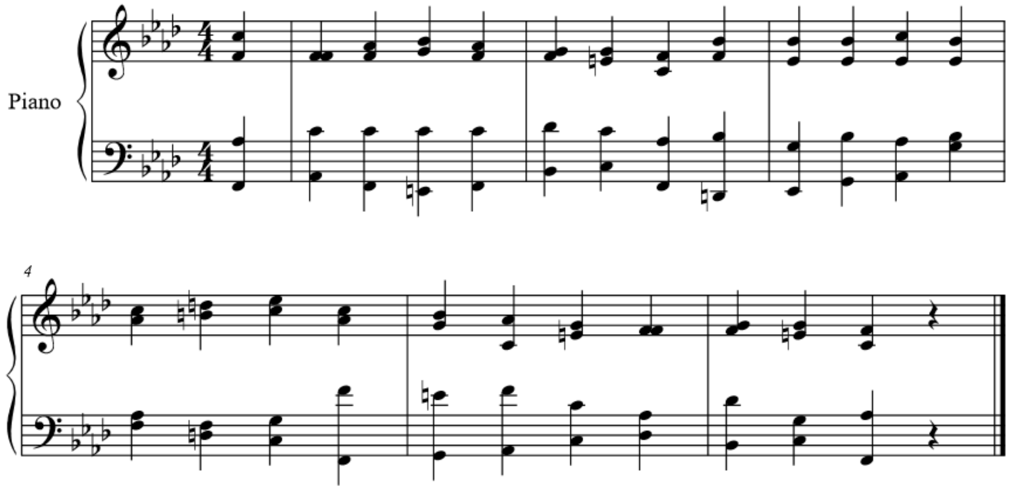
Widen, S. C., & Russell, J. A. (2008). Young children's understanding of other's emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett, *Handbook of emotions* (pp. 348-363). New York, NY, US: Guildford Press.

Widen, S. C., & Russell, J. A. (2010). Children's scripts for social emotions: Causes and consequences are more central than facial expressions. *British Journal of Developmental Psychology*, *28*(3), 565-581. http://dx.doi.org/10.1348/026151009X457550d

Xiao, N. G., Quinn, P. C., Liu, S., Ge, L., Pascalis, O., & Lee, K. (2017). Older but not younger infants associate own-race faces with happy music and other-race faces with sad music. *Developmental Science*, *21*(2), e12537. http://dx.doi.org/10.1111/desc.12537

Supplementary Material

Scores for musical excerpts

Maj 1 (original mode)

Min 1(altered mode)

Maj 2 (original mode)

Min 2 (altered mode)

Maj 3 (altered mode)

Min 3 (original mode)

Maj 4 (altered mode)

Min 4 (original mode)

Chapter 6: General discussion

As early as the 1920s and 1930s, scientists sought to quantify expressive aspects of music performance through analysis of data derived from piano rolls (Seashore, 1927, 1936, 1938). In the subsequent decades, innovations in music recording have made analysis of performance features vastly more feasible, resulting in dozens of studies quantifying expressive music performance (e.g., Gabrielsson, 1999; Gabrielsson & Lindström, 2010; Juslin & Laukka, 2003; Juslin & Timmers, 2010; Palmer, 1989, 1996; Repp, 1992, 1998). Stemming from this work, researchers have theorized that communicative intention, perceptual biases, and cognitive biases might all contribute to timing deviations in music performance (Honing, 2005; Juslin, 2003; Penel & Drake, 2004). However, previous work has been constrained by the necessity of testing highly trained musicians, precluding the ability to determine how formal training might contribute to the observed patterns. In the previous chapters, I have presented a *musical dwell time paradigm* for testing musical production across individuals with varying levels of music training, and I showed that this paradigm can be used to probe questions about the origins of expressive timing deviations in various contexts.

Chapters 2 and 3

Unique contributions. Chapters 2 and 3 focused on the phenomenon of phrase-final lengthening. In Chapter 2, I demonstrated that adults lengthen phrase boundaries when self-pacing through musical chord sequences. Slowing was greater on coarse than fine boundaries, consistent with patterns observed in expert performances. Further, I found that when metrical predictability and melodic contour were controlled for, lengthening could be elicited by phrase boundaries that were marked by harmonic cadences. This effect was observed in a group of participants who were recruited without regard to formal training, as well as a group of participants who reported no formal music training. To my knowledge, these studies are the first to show that nonmusicians spontaneously produce phrase-final lengthening in a non-communicative, non-performance setting, and that phrase-final lengthening can be evoked by harmonic cues.

In Chapter 3, I investigated the developmental trajectory of the results found in Chapter 2. In Experiment 1, I demonstrated that children as young as three years of age spontaneously lengthen phrase boundaries when self-pacing through metrically and harmonically stable chord sequences. In Experiment 2, I found that 4- and 7-year-old children, but not 3-year-old children, lengthen phrase boundaries marked by harmonic cues when metrical and melodic contour cues are controlled for. No previous work has examined how young children associate timing cues with phrase segmentation in music, and although one previous study has shown evidence for implicit harmonic knowledge in 4-year-olds using EEG techniques (Corrigall & Trainor, 2014), this is the first study to demonstrate 4-year-olds' knowledge of harmony in a behavioral task. Further, 3-year-olds' boundary lengthening in Experiment 1 suggests that the difference between 3- and 4-year-old children in Experiment 2 reflects a substantial increase in harmonic knowledge over that age range rather than improvement in the task itself. Finally, the work in Chapter 3 supported the conclusion drawn in Chapter 2, that music training is not necessary for phrase-final lengthening.

Limitations and future directions. Chapters 2 and 3 demonstrated that both children and musically untrained adults lengthen phrase boundaries when self-pacing through music. These results naturally point to a larger question: why does phrase-final lengthening occur in this context? It has been proposed that boundary lengthening is an avenue by which performers communicate musical structure to listeners (e.g., Clarke, 1985; Todd, 1985) or emulate the kinematics of braking (e.g., Friberg & Sundberg, 1999). The results presented in Chapters 2 and 3 suggest that performer intentions alone likely do not explain phrase-final lengthening, as it would be surprising for deliberate performance strategies to be present in a self-pacing task taking place in a solitary laboratory setting, particularly for young children.

In Chapter 3, General Discussion, I put forward the idea that phrase-final lengthening could result from listeners' expectations about the statistical structure of the music. It has been widely argued that expectations play an important role in our experience of music (e.g., Huron, 2006; Meyer, 1956; Narmour, 1992) and recent work has shown that musical expectations are influenced by statistical structure (Agres, Abdallah, & Pearce, 2018; Hansen & Pearce, 2014). One implementation of a statistical model for music involves examining the number of likely outcomes after a particular musical event by estimating Shannon entropy. Briefly, in this context, Shannon entropy can be thought of as characterizing the sharpness of a likelihood distribution for an upcoming outcome. If one outcome is particularly likely, and others unlikely, entropy is low. On the other hand, if there are many equally likely outcomes, entropy is high. In the context of melodies, entropy is a proxy for predictive uncertainty, such that higher entropy is associated with greater predictive uncertainty (Hansen & Pearce, 2014). Statistical modeling of melodies has demonstrated that phrase boundaries tend to be points of relatively high melodic entropy (Hansen, Vuust, Pearce, & Huron, 2017) – in other words, the likelihood distribution for melodic continuations is sharper after *within-phrase* events than after *phrase-final* events. Based on this work, we hypothesized that boundary lengthening might be partially attributable to the relative uncertainty experienced at boundaries, rather than its classification as a boundary, per se.

To test this hypothesis, we selected a number of melodic excerpts from a corpus of Western music. Each excerpt included a note that was (1) notated in the score as either a phrase boundary or not, and (2) afforded either high or low entropy continuations. Entropy was estimated by the Information Dynamics of Music Model (IDyOM), a variable-order Markov model that acquires knowledge through unsupervised statistical learning (Pearce, 2005). IDyOM's output incorporates information gleaned from *long-term* learning, which is acquired through an experimenter-selected corpus, and *short-term* learning, which is acquired through the current stimulus material. Here, the model was trained on a corpus of Western folk songs to simulate the experience of the average Western participant. Participants were asked to self-pace through each excerpt, and we measured the amount of time they spent on the target notes. In a second experiment, new participants were asked to rate the extent to which they felt the excerpt was complete after hearing the target note in context. We found that participants demonstrated more lengthening and rated higher feelings of completeness on notes that were high in entropy (i.e., generated relatively greater uncertainty), regardless of their phrase boundary status (Kragness*, Hansen*, Vuust, Trainor, & Pearce, 2016).

Overall, these experiments suggest that predictive uncertainty contributes to boundary perception. Further, they suggest that boundary lengthening may reflect psychological mechanisms related to the statistical structure of the music, specifically, the spike in uncertainty that is often experienced at phrase boundaries, in addition to intentional communicative strategies. This kind of mechanism could operate in parsing continuous information streams in any modality. Indeed, boundary-lengthening phenomena have been observed in studies in which participants self-pace through action sequences (Hard, Recchia, & Tversky, 2011; Meyer, Baldwin, & Sage, 2011), as well as in spoken languages across the globe (Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). This offers the intriguing possibility that statistical structure may relate to timing production across domains. More work is needed to elucidate the hypothesized association between lengthening and uncertainty in the context of music and beyond.

It is useful to note that the aforementioned experiments were done with melodic excerpts, and it is therefore an open question whether this can explain what I observed with *harmonic* sequences, such as those in Chapters 2 and 3. IDyOM has recently been adapted for modeling harmonic sequences (Sears, Pearce, Caplin, & McAdams, 2017). Future work could examine whether IDyOM would estimate relatively high entropy at phrase boundaries in the stimuli used in Chapters 2 and 3. Modelling entropy could prove particularly revealing in the case of atonal sequences. In Chapter 2, Experiments 1 and 2, and Chapter 3, Experiment 2, we observed lengthening at boundary positions, even in the atonal conditions in which the tonal center was obscured, thereby eliminating harmonic cadences. While global expectations would likely not assist listeners in identifying atonal “boundaries,” local expectations could be generated through statistical learning over the course of self-pacing through the sequence (Saffran, Johnson, Aslin, & Newport, 1999). Recent work has suggested that statistical learning gradually improves across early to middle childhood (Shufaniya & Arnon, 2018). Given that 7-year-old children and adults lengthened tonal *and* atonal boundaries, but that 4-year-old children *only* lengthened tonal boundaries, it is plausible that statistical learning of local probabilities might underlie this pattern of results.

Alternative explanations for “atonal boundary lengthening” should also be considered. In the absence of a familiar tonal structure, participants might have relied on other information in the stimulus to infer the presence of a phrase boundary, such as idiomatic contours of the different voices. For instance, we proposed in Chapter 2 that participants might have inferred a boundary in instances where the bass note (the lowest note) made a large “leap.” In the tonal sequences, the bass line often “leapt” the interval of a fourth between the phrase-final chord and the chord directly preceding it. In the atonal sequences, the corresponding interval was a semitone larger or smaller, which would still represent a relatively large interval, in turn perhaps leading to the perception of a boundary. Potential contour effects could be eliminated, or at least reduced, by asking participants to self-pace through sequences that use “Shepard tones” (Shepard, 1964) rather than piano tones as stimuli. A Shepard tone consists of a set of simultaneous sine waves that are each separated by an octave, with the amplitude of each wave dictated by its position in a bell-shaped spectral envelope. Because the height of each pitch is obscured, the perceived contour of two consecutive Shepard tones relies on the proximity their two pitch classes – that is, participants usually report hearing a Shepard tone with the pitch class A followed by a Shepard tone with the pitch class B as an ascension, because the ascending relationship is more proximal (a distance of 2 semitones) than the descending relationship (10 semitones). Thus, the use of such tones would control for the perception of voice leading by obscuring the contour of each voice as written in the score. Boundary detection based on voice leading and melodic contours should disappear in this case, but boundary detection relying on harmonic movement should be maintained.

It is also possible that the perception of tonality may have been maintained to some extent in the atonal conditions. Recent work shows that previously-heard tonal contexts influence listeners' judgments up to 20 seconds after modulation to a different key (Farbood, 2016; Spyra, Stodolak, & Woolhouse, 2017; Woolhouse, Cross, & Horton, 2016). In the tonal condition, authentic cadences were considered to be the harmonic cue for phrase endings. Though there were no authentic cadences in the atonal condition, participants heard either a dominant *or* tonic chord (in the key of the tonal condition), which may have signaled the presence of a cadence-like structure. Effects of residual memory for the key of the tonal condition could be investigated post-hoc by examining participants' dwell time patterns when they played through the atonal condition *first* compared to participants who played through the atonal condition *second*. However, there would likely not be sufficient number of trials per participant to obtain reliable dwell time estimates with the data collected here. In a future experiment, this question could be explored using either a between-subjects design (in which participants self-pace through *only* the tonal or atonal sequences) or a within-subjects design with two sessions separated in time (in which participants self-pace through only the tonal or atonal condition in each session, thus reducing residual memory for the key of the tonal condition). The extent to which participants perceived tonality, perceived boundaries, and lengthened boundaries in the atonal sequences are open questions to be explored in the future.

A related area in which this paradigm could prove useful is examining enculturation to pitch structure across early childhood. Studies of North American infants and children have shown that enculturation to Western key membership emerges some time between 12 months and four or five years, with sensitivity to harmony emerging slightly later, between five and six years (Corrigall & Trainor, 2014; Lynch, Eilers, Oller, & Urbano, 1990; Schellenberg, Bigand, Poulin-Charronnat, Garnier, & Stevens, 2005; Trainor & Trehub, 1992, 1994). Brain responses to key- and harmony-violating events, likely representing pre-conscious enculturation, appear to be present slightly earlier than behavioral sensitivity has been measured (Corrigall & Trainor, 2014; Jentschke, Koelsch, Sallat, & Friederici, 2008; Koelsch et al., 2003), and one study has reported brain responses to harmonic violations in children as young as two years (Jentschke, Friederici, & Koelsch, 2014). In Chapter 3, I showed that enculturation to the harmonic stability of the tonic chord emerges behaviorally between 3 and 4 years. However, there are still substantial gaps in our understanding of when and how knowledge of melodic stability is instantiated, largely because of challenges in testing and interpreting null results in children between infancy and three years old.

It would be very useful to have a paradigm that could reveal children's knowledge about tonal stability across the time span in question. Previous work has shown that around 8 months, infants are sensitive to the structure of tone sequences after a brief exposure period (Saffran, Johnson, Aslin, & Newport, 1999), but are not yet sensitive to Western key membership (Trainor & Trehub, 1992). If infants' dwell times demonstrated lengthening on group boundaries after exposure to an artificial tone sequence, but *not* group boundaries in the context of Western melodies, that would offer compelling evidence that enculturation had not yet taken place. By testing young children between 8 months and 4 years using the same paradigm, it may be possible to more closely examine when behavioral sensitivity to melodic stability emerges.

Dwell time experiments in the action segmentation domain have reported success with children spanning this age range. For instance, 3-year-old children's dwell times can reveal knowledge of structure in action sequences (Meyer, Baldwin, & Sage, 2011). The dwell time paradigm has also been used effectively with 10-month-old infants, with the simple modification of using a touch screen rather than a computer key to elicit each event (Sage, Ross, & Baldwin, 2012). To my knowledge, no studies have examined whether auditory adaptations of the dwell time paradigm could be used with children younger than 3 years old. In a pilot study not reported here, we adapted the dwell time paradigm to be used with a tablet touch screen for 2-year-old participants. In this case, participants elicited the onset of each chord by touching the screen with any part of their hand, which was thought to be much easier than a computer-based apparatus. However, the tablet registered toddler's palm taps as being multiple consecutive taps, probably because different parts of their hands touched the screen at slightly different times. Future studies could address this problem simply by introducing a 50-100 ms restriction on consecutive taps. Addressing this technical difficulty might make it possible to use the dwell time paradigm with children from infancy through to young childhood, spanning the period in which enculturation to Western melodic conventions appears to emerge.

An outstanding question concerns which factors influence enculturation to pitch structure. In Chapter 3, Experiment 2, I found that knowledge of harmonic structure changes substantially between three and four years of age, but it is not clear whether this is attributable to accumulated music exposure or gains in general cognition in this time. We did not find evidence that the strength of boundary dwelling correlated with working memory, phonemic awareness, or vocabulary, although it is difficult to interpret null correlations with small groups of participants, and future work should explore the influence of these variables more closely. We also did not find any significant correlations with reported hours of music listening. Interestingly, previous work has shown that active, but not passive, musical experiences lead to a preference for tonality over atonality by around 12 months (Gerry, Unrau, & Trainor, 2012). Follow-up work could experimentally examine whether sensitivity to harmonic structure observed in

Chapter 3, Experiment 2 could be accelerated by exposing children to active or passive musical experiences.

Chapters 4 and 5

Unique contributions. Chapters 4 and 5 focus on the communication of emotions in music. In Chapter 4, I adapted the musical dwell time paradigm to give participants additional control over their musical productions. Participants used a single key on a MIDI piano to control the onset, offset, and velocity of each key press, thereby controlling the speed, connectedness, and loudness of each chord. I found that participants used each expressive cue in ways that were consistent with previous studies of musical experts in naturalistic performance settings. In subsequent analyses, participants were divided into three groups with differing levels of music training, ranging from None (0 years), Low (mean of 2 years), and Moderate/High (mean of 8 years). I observed that despite this wide range in music training, there was significant overlap across groups in their use of each cue. To my knowledge, this is the first study to investigate non-musicians' expressions of musical emotions and to examine how formal music training might relate to expressive cue use in music production.

In Chapter 5, I developed a music emotions "game" for use with children. Three-, five-, and seven-year-old children heard a vignette depicting a particular emotion while concurrently viewing an emotionally expressive face congruent with the vignette. They were asked to identify the emotion of the character in the vignette and then to perform music to convey that emotion using the MIDI piano apparatus. I found that children as young as 5 years used tempo and loudness to distinguish arousal, and that 7-year-old children's use of tempo and loudness was more exaggerated than 5-year-olds', but less exaggerated than adults'. Additionally, by 7 years, children incorporated articulation as an expressive cue to distinguish arousal. While other studies have used singing to examine children's expressive cues, singing studies have been limited by reliance on the vocal apparatus, as well as by children's focus on words and familiar songs. The present work extends previous findings by examining how children's expressive cues relate to both valence and arousal.

Limitations and future directions. The results presented in Chapter 4 showed significant overlap in cue use across groups with different levels of music training. Because no differences were observed, no conclusions can be drawn about whether music training improves emotional expression in music production. It is important to note that we selected four emotions, each from a different quadrant of a valence-by-arousal circumplex, three of which are considered to be "basic" and among the emotions that are easiest to convey and recognize across modalities (Ekman, 1992, 1993). Complex, mixed, or aesthetic emotions (such as *awe, humor, longing*, or *nostalgia*) are likely to be more difficult to convey, and formal training differences might emerge in this context. In addition, it might be the case that formal training does enhance emotion expression early in development, but that nonmusicians “catch up” after accruing decades of exposure to music through listening experience.

At the same time, the present results are consistent with previous work suggesting that instruction on emotional expression may be under-utilized in musical instrumental training. Despite the fact that musicians consider communication of emotions to be a vital component of their performances (Laukka, 2004), observational studies have reported that surprisingly little time is spent on discussion of expression during formal instrumental instruction (Karlsson & Juslin, 2008). It could be argued that highly-trained instrumentalists do not require instruction on emotional expression and are already capable of achieving highly accurate communication. However, Juslin and colleagues (2006) showed that receiving feedback about expressive cue use improved semi-professional guitarists’ accuracy in communicating emotions in music, even when that feedback was delivered by a computer program. Overall, the results presented in Chapter 4 support the idea that expression instruction is potentially under-utilized in traditional instrumental instruction, although whether this holds true across different curricula and at different developmental timepoints remains to be seen.

An important question in this literature is the extent to which expressive cues overlap across musical cultures (e.g., Fritz et al., 2009; Higgins, 2012; Sievers, Polansky, Casey, & Wheatley, 2012). Previous work in this area has largely focused on “decoders,” or listeners, rather than performers. This is likely at least partially due to the fact that different musical systems utilize different instruments, complicating attempts to directly compare musical expression. One previous study examined bowed-string performers across four cultures, reporting that accuracy of communication was superior when the performers and listeners were matched compared to mismatched in culture (Laukka, Eerola, Thingujam, Yamasaki, & Gregory, 2013). Though the instrumentalists were matched in the sense that they all performed on a similar *type* of instrument (bowed-string instruments), the authors acknowledged that the differences in instrumentation and in musical structure of the stimuli likely contributed to their findings. Future work could implement the musical dwell time paradigm in a cross-cultural design. Controlling stimulus material and instrumental familiarity in this way would enable stronger inferences about the nature of shared and non-shared cues across cultures.

It is important to note that the present work has focused exclusively on the *expression* of musical emotions. Expression, however, represents only one side of musical communication. Measuring successful communication requires both “encoders,” or performers, and “decoders”, or listeners (Brunswik, 1956; Juslin, 2000). While we show significant overlap between the cues used by performers with different levels of music training, there may be additional features of their performance that we did not characterize in our analyses, such as the interaction of different cues, which might lead to more or less effective decoding. In the future, it would be very useful to examine both encoding and decoding to obtain a fuller picture of musical communication, rather than expression alone.

The question of communication could prove especially interesting in the case of child development. In the majority of previous studies, children were asked to identify emotions in excerpts that were previously selected by adults. Unsurprisingly, most such studies found that children's ratings more closely match adults' as they age. Few studies, however, have examined the role that stimulus material might play in this observed trend. One study that used culturally familiar, child-directed music (Mote, 2011) found discrimination between happy and sad excerpts based on tempo in younger children than in studies using Western classical music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001), suggesting that studies using adult-directed music may underestimate children's emotional identification abilities. In one of the few studies examining children's production, Adachi and colleagues (2004) found that children outperformed adults in recognizing "happy" and "sad" intent in renditions performed by same-age children. This finding offers the intriguing possibility that children's performances and interpretations are not simply more variable than adults' (although this is likely true), but are also qualitatively different – in other words, that children may "weigh" cues differently from adults when performing and recognizing musical emotions. To my knowledge, no studies have attempted to replicate this same-age superiority effect, nor explore what mechanisms might underlie such an effect. Future work using the dwell time paradigm could be very informative in this regard, as the experimenter is able to control many aspects of the stimulus material (including pitch and timbre), while allowing the “performer” to easily vary other aspects.

Previous work examining expressive cues in music performance have noted that expressive cues often map onto either the valence or arousal dimension. Interestingly, research examining children’s recognition of emotions in faces and vignettes suggest that a valence-by-arousal dimensional approach is appropriate for describing early differentiation in children’s emotion recognition in these domains, as well (Widen & Russell, 2008). Despite this, only a small number of studies have examined children’s perception of musical emotions from such a dimensional approach, likely due to methodological challenges in obtaining valence and arousal ratings from children, and Chapter 5 describes the first study of production to consider both dimensions. Even in perceptual studies that consider both arousal and valence (Andrade, Vanzella, Andrade, & Schellenberg, 2017; Hunter, Schellenberg, & Stalinski, 2011), children have been asked to make categorical judgments of emotions rather than dimensional assessments, such as on a sliding scale. Recently, Widen and Russell (2016) developed the Children’s Scales of Pleasure and Arousal, and reported that children reliably identified valence and arousal in facial expressions as young as 3 years old. Future work in children’s emotion recognition in music would benefit from incorporating these novel scales into work in the domain of music.

Although we observed increasing differentiation in 3-, 5-, and 7-year-olds' expressions of emotion, the reason for this developmental change is as yet unknown. Myriad developmental changes occur over this time period, all of which could affect the expression of musical emotions. For example, motor development, emotional development, language development, and gains in attention and working memory could all potentially contribute. Disentangling the relative contribution of each of these abilities across this age range would require a large number of participants and significant time testing in the lab. However, a group-based approach might shed light on how children's production of expressive music might relate to other developmental skills. For instance, it has been noted that children with autism often have impairments in skills related to communication, and yet can recognize musical emotions at rates comparable to children without autism (for a review, see Molnar-Szakacs & Heaton, 2012). It would be useful to examine whether autistic children's preserved musical abilities extend from perception to production and, if so, whether engaging in expressive music production could scaffold the learning of other expressive skills.

Summary

In the present thesis, I presented a new paradigm for exploring music production, the musical dwell time paradigm. I demonstrated that this methodology can be used to examine different aspects of music performance, including variations related to musical structure and emotional expression. Finally, I have shown that the dwell time paradigm can be effective for testing individuals with a variety of musical experiences and backgrounds, from children as young as three years old to university undergraduates. Together, the findings presented in Chapters 2-5 support previous work showing that nonmusicians have a relatively high degree of knowledge about musical patterns and structures (Bigand & Poulin-Charronnat, 2006). In addition, this body of work raises novel questions about developmental and cognitive origins of these patterns, and provides a new avenue for testing hypotheses about the mechanisms involved in perception and production of music, rhythm, and timing more generally.

References

Adachi, M., Trehub, S. E., & Abe, J. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46*(4), 322-336. http://dx.doi.org/10.1111/j.1468-5584.2004.00264.x

Agres, K., Abdallah, S., & Pearce, M. T. (2018). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cognitive Science, 42*(1), 43-76. http://dx.doi.org/10.1111/cogs.12477

Andrade, P. E., Vanzella, P., Andrade, O. V. C. A., & Schellenberg, E. G. (2017). Associated emotions with Wagner's music: A developmental perspective. *Psychology of Music, 45*(5), 752-760. http://dx.doi.org/10.1177/0305735616678056

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130. http://dx.doi.org/10.1016/j.cognition.2005.11.0077

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

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. & West, *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science, 17*(1), 142-158. http://dx.doi.org/10.1111/desc.12100

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1-B10. http://dx.doi.org/10.1016/s0010-0277(00)00136-0

Demos, A. P., Lisboa, T., Begosh, K. T., Logan, T., & Chaffin, R. (2018). A longitudinal study of the development of expressive timing. *Psychology of Music,* Advanced online publication. http://dx.doi.org/10.1177/0305735618783563

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*(3), 550–553. http://dx.doi.org/10.1037/0033-295X.99.3.550

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384–392. http://dx.doi.org/10.1037/0003-066X.48.4.384

Farbood, M. M. (2016). Memory of a tonal center after modulation. *Music Perception: An Interdisciplinary Journal, 34*(1). 71-93. http://dx.doi.org/mp.2016.34.1.71

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America, 105*(3), 1469-1484. http://dx.doi.org/10.1121/1.426687

Fritz, T., Jentschke, S., Gosslin, N., Sammler, D., Peretz, I., Turner, R., . . . Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19*(7), 573-576. http://dx.doi.org/10.1016/j.cub.2009.02.058

Gabrielsson, A. (1999). The performance of music. In D. Deutsch, *The Psychology of Music (2nd edition)* (pp. 501-602). Academic Press.

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

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative, and social development. *Developmental Science, 15*(3), 398-407. http:// dx.doi.org/10.1111/j.1467-7687.2012.01142.x

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology, 5*, 1052. http://dx.doi.org/10.3389/fpsyg.2014.01052

Hansen, N. C., Vuust, P., Pearce, M., & Huron, D. (2017, August). Entropic ebbs and flows: The expectancy dynamics of musical phrases. Presented at the Society for Music Perception and Cognition, San Diego, USA.

Hard, B., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586-604. http://dx.doi.org/10.1037/a0024310

Higgins, K. M. (2012). Biology and culture in music emotions. *Emotion Review, 4*(3), 273-282. http://dx.doi.org/10.1177/1754073912439762

Honing, H. (2005). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception: An Interdisciplinary Journal, 23*(1), 79-85. http://dx.doi.org/10.1525/mp.2005.23.1.79

Hunter, P. G., Schellenberg, E. G., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology*, *110*(1), 80-93. http://dx.doi.org/10.1016/j.jecp.2011.04.001

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectations.* Cambridge, MA: MIT Press.

Jentschke, S., Friederici, A. D., & Koelsch, S. (2014). Neural correlates of music-syntactic processing in two-year old children. *Developmental Cognitive Neuroscience, 9*(1), 200-209. http://dx.doi.org/10.1016/j.dcn.2014.04.005

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*(11), 1940-1951. http://dx.doi.org/10.1162/jocn.2008.20135

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1797-1813. http://dx.doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music, 31*(3), 273-302. http://dx.doi.org/10.1177/03057356030313003

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814. http://dx.doi.org/10.1037/0033-2090.129.5.770

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Sloboda, *Handbook of music and emotion: Theory, research, applications* (pp. 453-489). Oxford, UK: Oxford University Press.

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schroger, E., & Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience, 15*(5), 683-693. http://dx.doi.org/10.1162/jocn.2003.15.5.683

Kragness, H. E.*, Hansen, N. C.*, Vuust, P., Trainor, L. J., & Pearce, M. T. (2016, July). Information dynamics of boundary perception: Entropy in self-paced music listening. Presented at the International Conference for Music Perception and Cognition, San Francisco, USA. *These authors contributed equally to the work.

Laukka, P. (2004). Instrumental music teachers' views on expressivity: A report from music conservatoires. *Music Education Research, 6*(1), 45-56. http://dx.doi.org/10.1080/1461380032000182821

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Gregory, B. (2013). Universal and cultural-specific factors in the recognition and performance of musical affect expressions. *Emotion, 13*(3), 432-449. http://dx.doi.org/101.1037.a0031388

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science, 1*(4), 272-276. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00213.x

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

Meyer, M., Baldwin, D., & Sage, K. (2011). Assessing young children's hierarchical action segmentation. In H. H. L. Carlson, *Proceedings of the 33rd annual conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Molnar-Szakacs, I., & Heaton, P. (2012). Music: A unique window into the world of autism. *Annals of the New York Academy of Sciences, 1252,* 318-324. http://dx.doi.org/10.1111/j.1749-6632.2012.06465.x

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion*, *11*(3), 818-622. http://dx.doi.org/10.1037/a0022573

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The Implication-Realization model.* Chicago, IL: Chicago University Press.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America, 54*(5), 1235-1247. http://dx.doi.org/10.1121/1.1914393

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 21-54.

Pearce, M. T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. Doctoral Dissertation, Department of Computing, City University, London, UK.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences,* Online advanced publication. http://dx.doi.org/10.1111/nyas.13654

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562. http://dx.doi.org/10.3758/bf03194900

Repp, B. H. (1992). Diversity and commonality in performance: An analysis of timing microstructure in Shumann's "Traumerei". *Journal of the Acoustical Society of America, 92*(5), 2546-2568. http://dx.doi.org/10.1121/1.404425

Repp, B. H. (1997). Expressive timing in a Debussy prelude: A comparison of student and expert pianists. *Musicae Scientiae, 1*(2), 257-268. http://dx.doi.org/10.1177/102986499700100206

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America, 104*(2), 1085-1100. http://dx.doi.org/10.1121/1.423325

Repp, B. H. (1998). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research, 61*(1), 33-43. http://dx.doi.org/10.1007/s004260050011

Repp, B. H. (1998). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 791-811. http://dx.doi.org/10.1037/0096-1523.24.3.791

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tones sequences by human infants and adults. *Cognition, 70*(1), 27-52. http://dx.doi.org/10.1016/S0010-0277(98)00075-4

Sage, K. D., Ross, R. A., & Baldwin, D. (2012). Assessing infants' hierarchical action segmentation. Poster presented at the International Conference on Infant Studies, Minneapolis, Minnesota.

Schellenberg, E. G., Bigand, E., Poulin-Charronnat, B., Garnier, C., & Stevens, C. (2005). Children's implicit knowledge of harmony in Western music. *Developmental Science, 8*(6), 551-566. http://dx.doi.org/10.1111/j.1467-7687.2005.00447.x

Sears, D. R., Pearce, M. T., Caplin, W. E., & McAdams, S. (2017). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research, 47*(1), 29-52. http://dx.doi.org/10.1080/09298215.2017.1367010

Seashore, C. E. (1927). Phonophotography in the measurement of the expression of emotion in music and speech. *Scientific Monthly, 24*, 463–471.

Seashore, C. E. (1936). *Objective analysis of musical performance.* Iowa City: The University Press.

Seashore, C. E. (1938). *The psychology of music.* New York: McGraw-Hill.

Shepard, R. N. (1964). Circularity judgments of relative pitch. *Journal of the Acoustical Society of America, 36*(12), 2346-2353. http://dx.doi.org/10.1121/1.1919362

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, Online advanced publication. http://dx.doi.org/10.1111/cogs.12692

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2012). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences, 110*(1), 70-75. http://dx.doi.org/10.1073/pnas.1209023110

Spyra, J., Stodolak, M., & Woolhouse, M. (2017, August). Effects of intervening keys manipulated by events and time on nonadjacent harmonic perceptions. Presented at the Society for Music Perception and Cognition, San Diego, USA.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception: An Interdisciplinary Journal, 3*(1), 33-57. http://dx.doi.org/10.2307/40285321

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 394-402. http://dx.doi.org/10.1037/0096-1523.18.2.394

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics, 56*(2), 125-132. http://dx.doi.org/10.3758/bf03213891

Widen, S. C., & Russell, J. A. (2008). Young children's understanding of other's emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett, *Handbook of emotions* (pp. 348-363). New York, NY, US: Guildford Press.

Widen, S. C., & Russell, J. A. (2016). Children's scales of pleasure and arousal. *Journal of Nonverbal Behavior, 40*(3), 187-203. http://dx.doi.org/10.1007/s10919-016-0228-x

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*(3), 1707-1717. http://dx.doi.org/10.1121/1.402450

Woolhouse, M., Cross, I., & Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychology of Music*, *44*(4), 1-14. http://dx.doi.org/10.1177/0305735615593409