Young Children Pause on Phrase Boundaries in Self-Paced Music Listening: The Role of Harmonic Cues

August 29, 2017

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, Repp, 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_{age} = 3.56 years, SD_{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 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 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

[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.

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 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.

[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*.

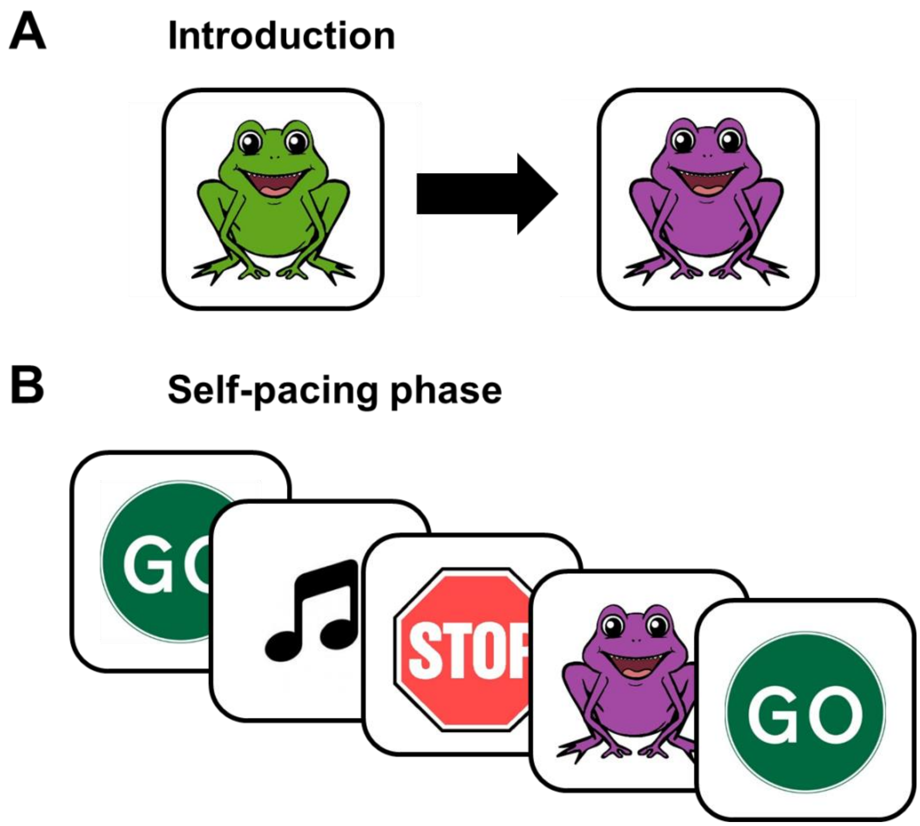

Figure 1

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.

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).

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 (Fig. 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

[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.

position 4 were dwelled on significantly longer than other non-boundary chords, but the contrast was not significant ($t(27) = 0.50$, $p = .620$).

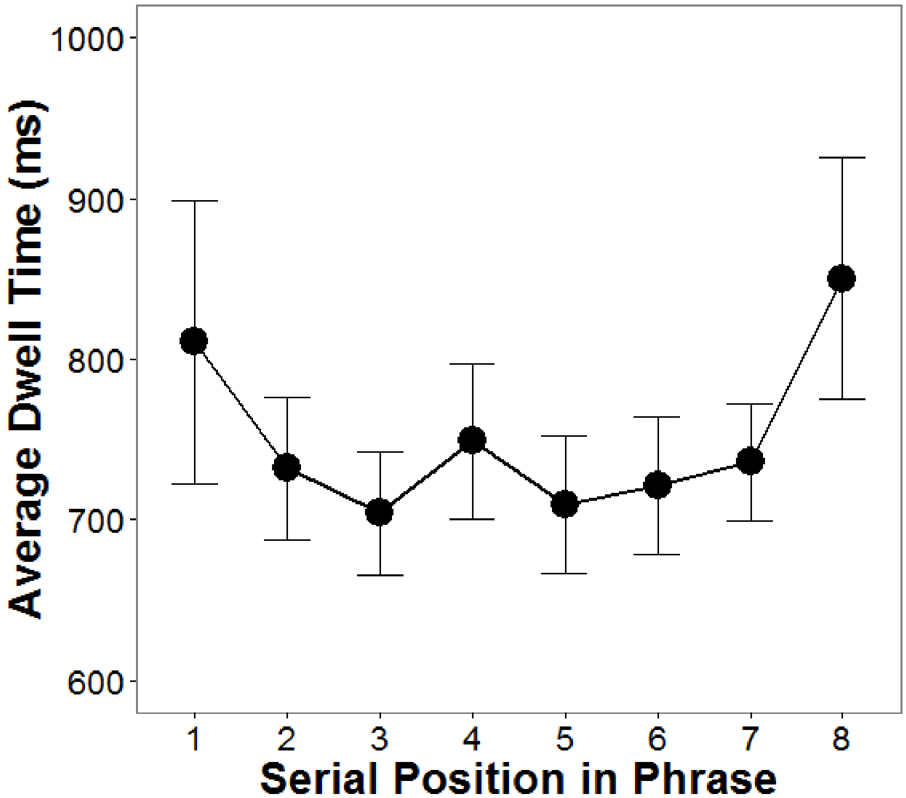

Figure 2

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

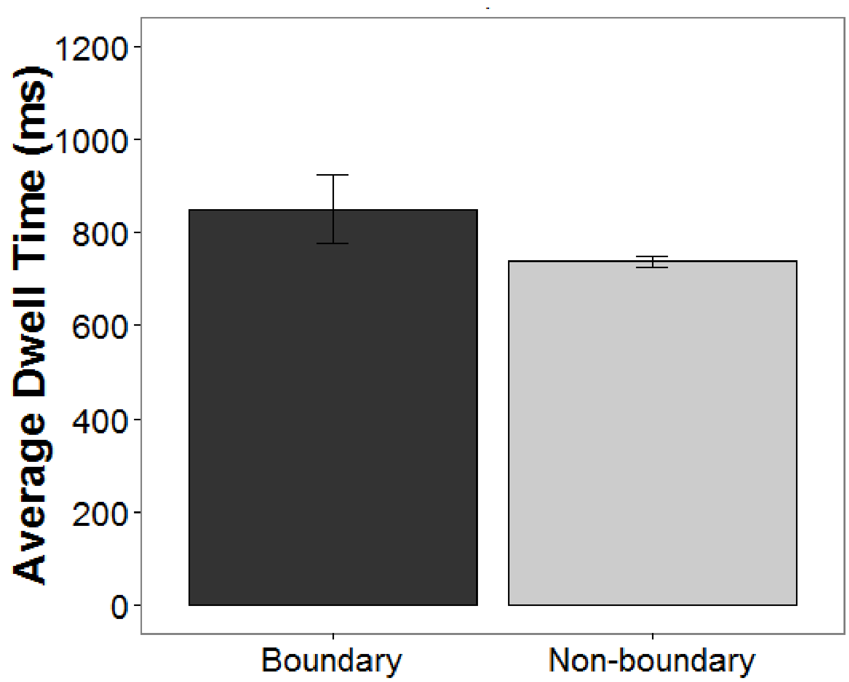

Figure 3

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

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

---

[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").

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, Haden, 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

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

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

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)

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 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 (3-year-olds, 4-year-olds, and 7-year-olds. 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 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).

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.

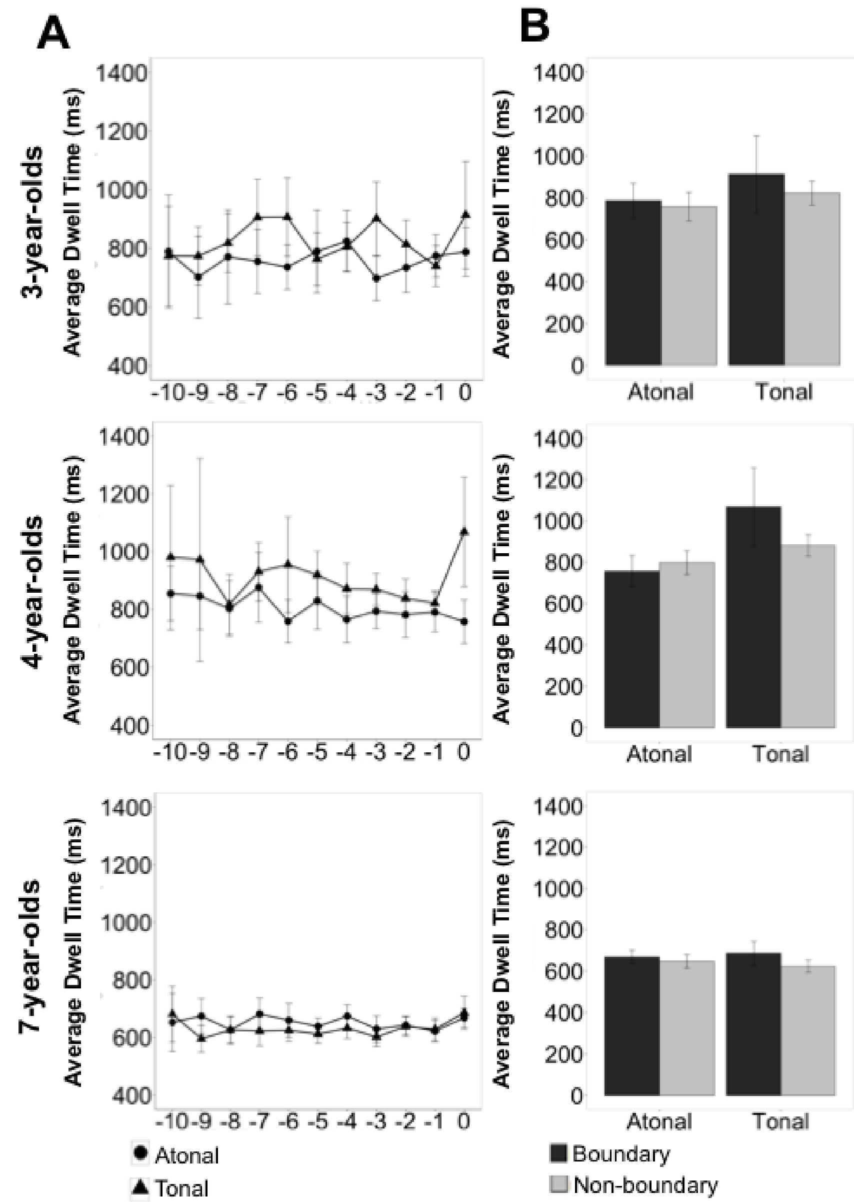

Figure 4

(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).

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.

Table 2. Correlations with boundary minus nonboundary difference scores.

Experiment	Age group	Tonality condition	Correlation coefficients: Age (rho)	Hours listening (tau)	DSF (tau)	DSB (tau)	TAAS (rho)	Blending (tau)	PPVT4 (r)
Experiment 1									
	3-year-olds	Tonal	-.215	.077	-.193	N/A	N/A	-.119	.185
Experiment 2									
	3-year-olds	Tonal	.201	-.060	.117	N/A	N/A	.221	-.053
		Atonal	.149	-.146	.069	N/A	N/A	.009	-.241
	4-year-olds	Tonal	.088	-.104	-.090	-.007	.096	.007	.376
		Atonal	-.013	.222	-.120	.007	-.127	-.066	.202
	7-year-olds	Tonal	-.428	.067	-.284	.039	-.459	-.035	.303
		Atonal	-.247	-.097	-.168	-.078	-.065	.078	.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.

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.

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. doi:10.7551/mitpress/9780262019279.003.0013

Bergeson, T. R., & Trehub, S. E. (2006). Infants' perception of rhythmic patterns. *Music Perception, 23*(4), 345–360. doi: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. doi: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. doi: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. doi: 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. doi:10.3758/bf03208069

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

Cambouropoulos, E. (2001). Melodic cue abstraction, similarity, and category formation: A formal model. *Music Perception*, *18*(3), 347–370. doi: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. doi: 10.2307/1128532

Chiappe, P., & Schmuckler, M.A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review*, *4*(2), 254-259. doi: 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. doi: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. doi: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. doi: 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. doi: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. doi: 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. doi: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. doi: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. doi: 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. doi: 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. doi: 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. doi: 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. doi: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. doi:10.1111/j.1467-7687.2012.01142.x

Gregory, A. H. (1978). Perception of clicks in music. *Perception & Psychophysics, 24*(2), 171–174. doi: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. doi: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. doi: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. doi: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. doi:10.1073/pnas.0504254102

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

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. doi: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. doi: 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 sessi on 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*. doi:10.1037/xhp0000245

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*(3), 346–374. doi: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. doi: 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. doi: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. doi: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. doi: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). doi: 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. doi: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. doi: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. doi: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. doi: 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. doi: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. doi: 10.1121/1.1914393

Olsho, L.W. (1984). Infant frequency discrimination. *Infant Behavior and Development, 7*(1), 27-35. doi: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,* 721-726. doi: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. doi: 10.1037/0096-1523.13.1.116

Patel, A. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. doi: 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. doi: 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. doi: 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. doi: 10.1007/PL00008161

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

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

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition, 98*(1), 1–11. doi: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. doi: 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. doi: 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. doi: 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. doi: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. doi: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. doi: 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. doi: 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. doi: 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*. doi: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). doi: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), 2 36–257. doi: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. doi: 10.3785/BF03202347

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25*(1), 122–127. doi: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. doi: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. doi: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. doi:10.1037/0033-295x.107.4.885

Todd, N.P. McA. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33-58. doi: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. doi: 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. doi: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. doi: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. doi: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. doi: 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. doi: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. doi: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. doi: 10.2307/1130133

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*(1), 73–99. doi: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. doi: 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. doi: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. doi: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). doi: 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. doi: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. doi:10.1121/1.402450

Winkler, I., Haden, 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. doi: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. doi: 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. doi:10.1177/0255761407085648

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