doi: 10.1037/xhp0000245

Listeners Lengthen Phrase Boundaries in Self-Paced Music

Haley E. Kragness[1] & Laurel J. Trainor[1, 2, 3]

1. Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada

2. McMaster Institute for Music and the Mind, McMaster University, Canada

3. Rotman Research Institute, Baycrest Hospital, Toronto, Canada

Acknowledgements. 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).

Correspondence should be addressed to Laurel J. Trainor, Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON L8S 4K1, Canada; email: LJT@mcmaster.ca

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.

Listeners Lengthen Phrase Boundaries in Self-Paced Music

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 infants (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 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

[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.

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 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 ($t_T(13) = 3.160$, $p_T < .01$; $t_A(13) = 2.398$, $p_A < .025$).

[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.

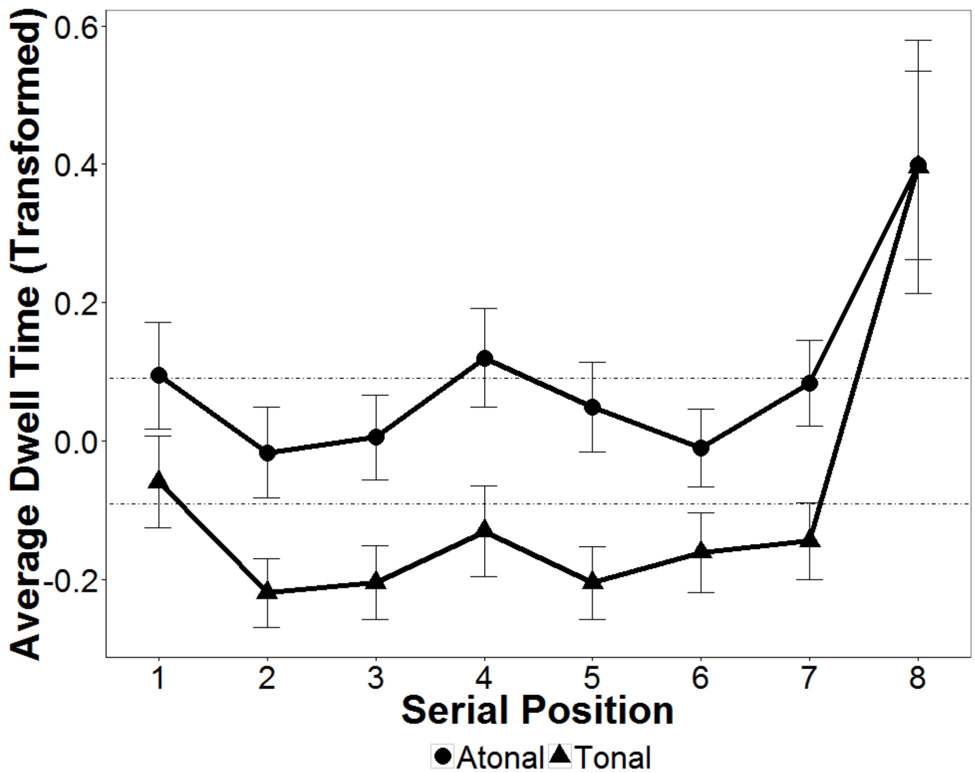

Figure 1. Average transformed dwell times for chords at each position in the 8-note phrases. Similar patterns were observed for both the Tonal condition and the Atonal condition. The dashed lines represent the grand average dwell time in each condition. For all figures, error bars represent standard error of the mean (SEM) across subjects.

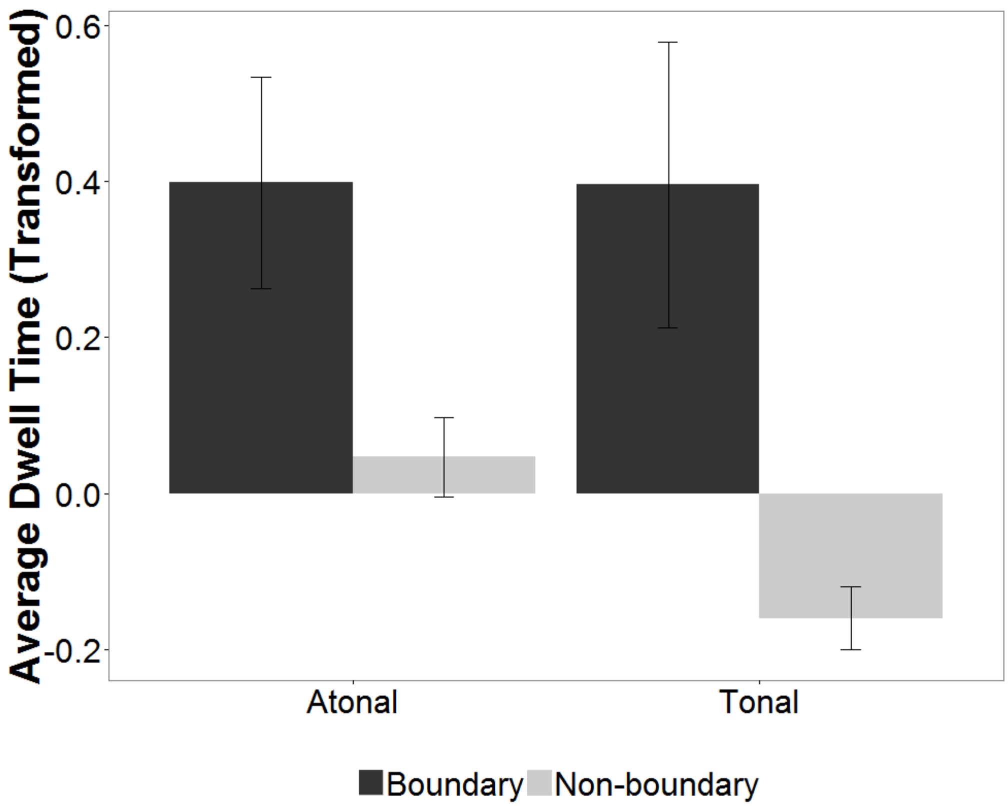

Figure 2. Mean dwell times binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard errors of the mean.

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 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 ($t(13) = -3.819$, $p < .01$).

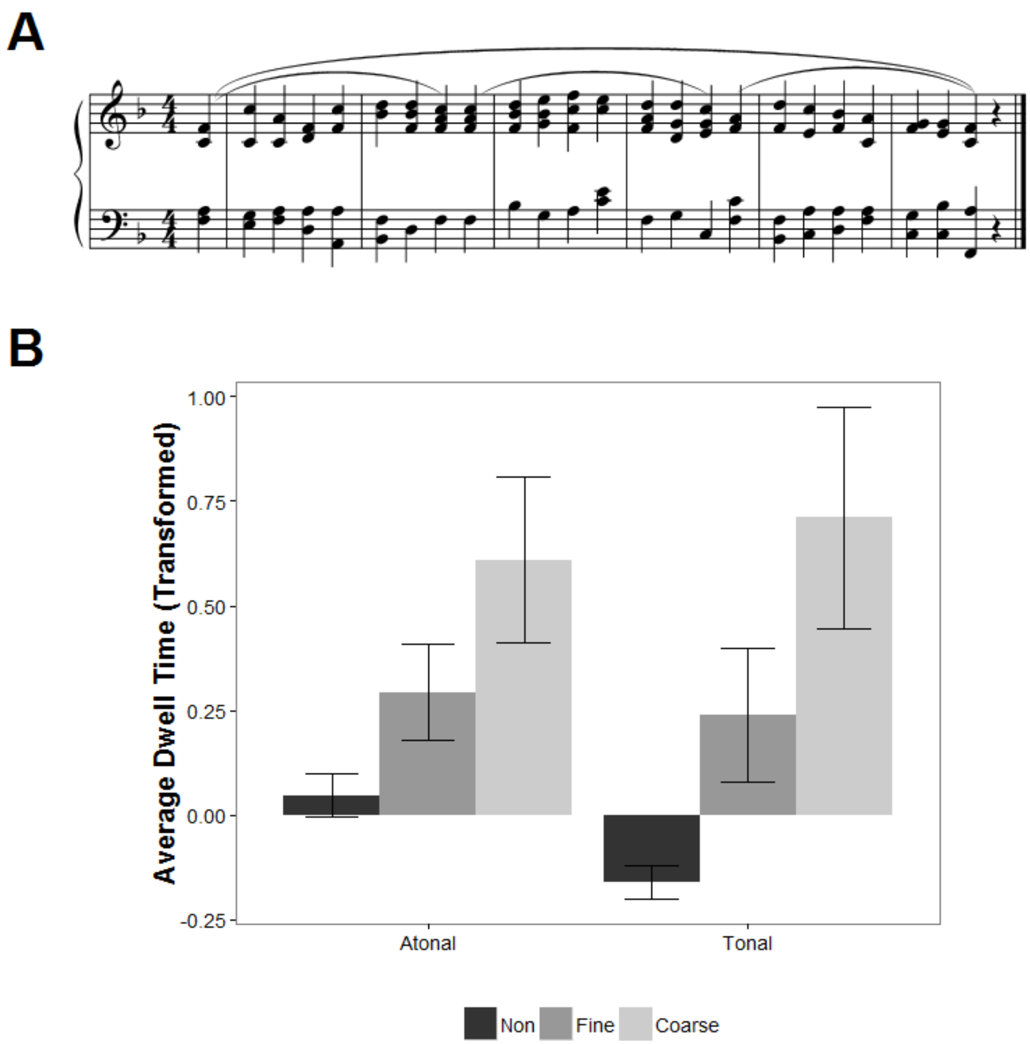

Figure 3 (A) A depiction of the two phrase levels in one excerpt (BMV 1.6) as shown by the phrase markings above the musical notation. Error bars represent SEM. The positions of coarse and fine boundaries were identical across all excerpts, which all consisted of six bars with an anacrusis. (B) Average transformed dwell times for each boundary level for atonal and tonal versions separately.

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.

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

---

[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).

were ineligible. All reported normal hearing except one participant, who reported chronic tinnitus. Analyses were performed 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. 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 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.

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.

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.

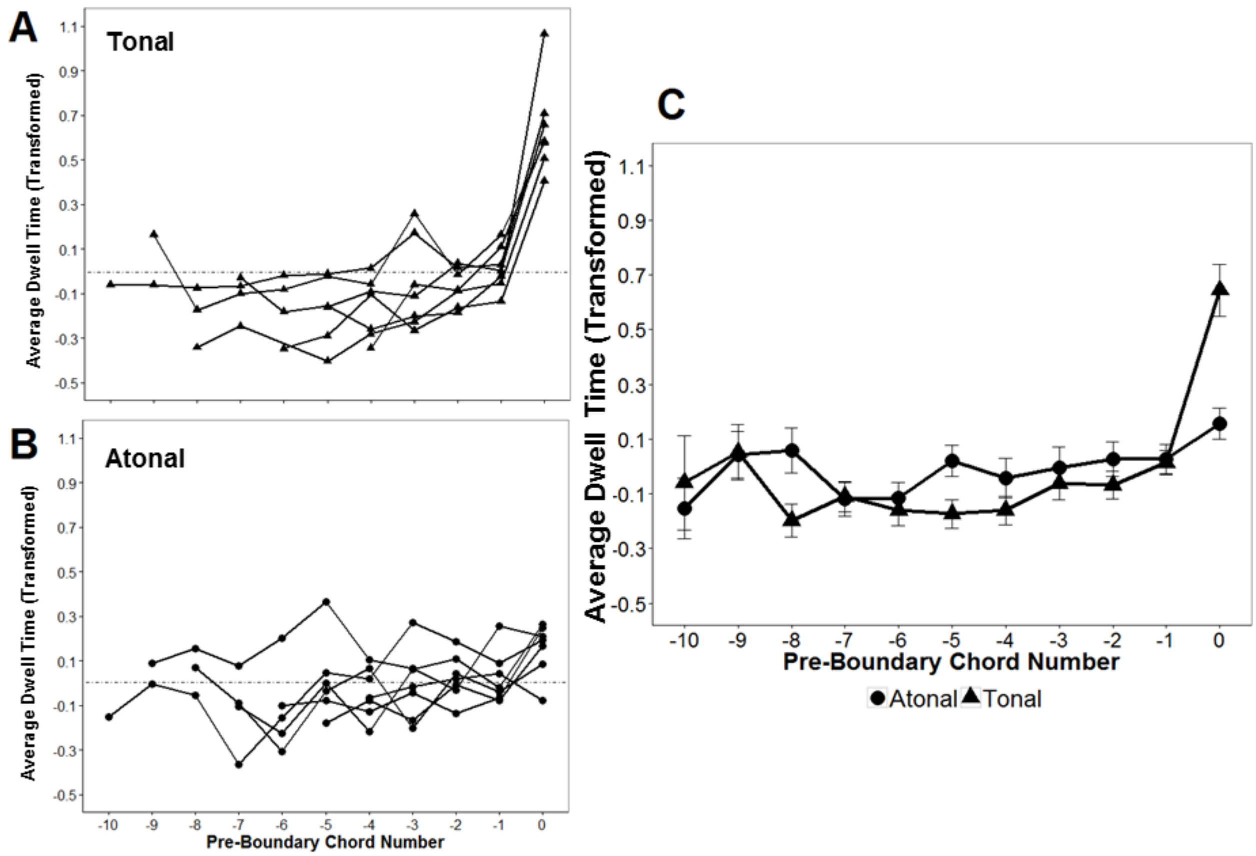

Figure 5. (A) and (B) represent Tonal and Atonal dwell time profiles, respectively, for Experiment 2. Because the meter was random, the x-axis represents the chord number before the boundary chord (0). Each of the 6 lines in each figure represents the average of phrases of a particular length (5 to 11 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.

1

To test whether boundary dwell times were different from non-boundary dwell times, transformed 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 5). 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.

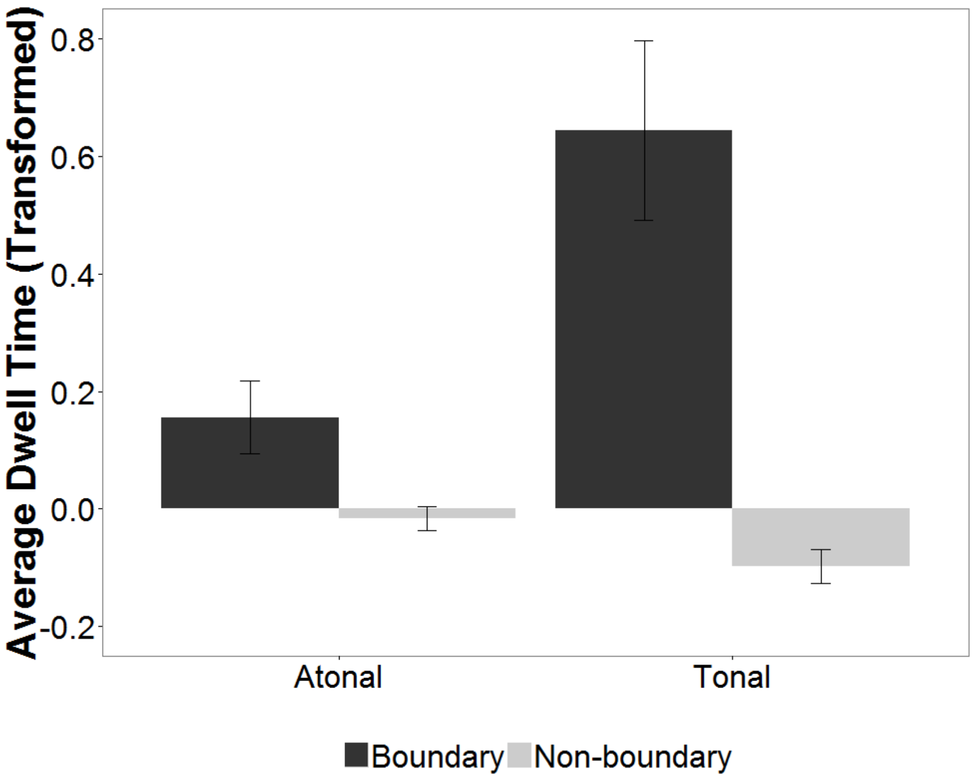

Figure 6. Mean dwell times binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard errors of the mean. Participants 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 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 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 ($M_{AB} = -0.035$, $SD_{AB} = 0.206$, $M_{AN} = 0.041$, $SD_{AN} = 0.070$), $t(16) = -1.360$, $p = .904$.

[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.

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 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 ($r_\tau(15) = -.122$, $p = .522$).

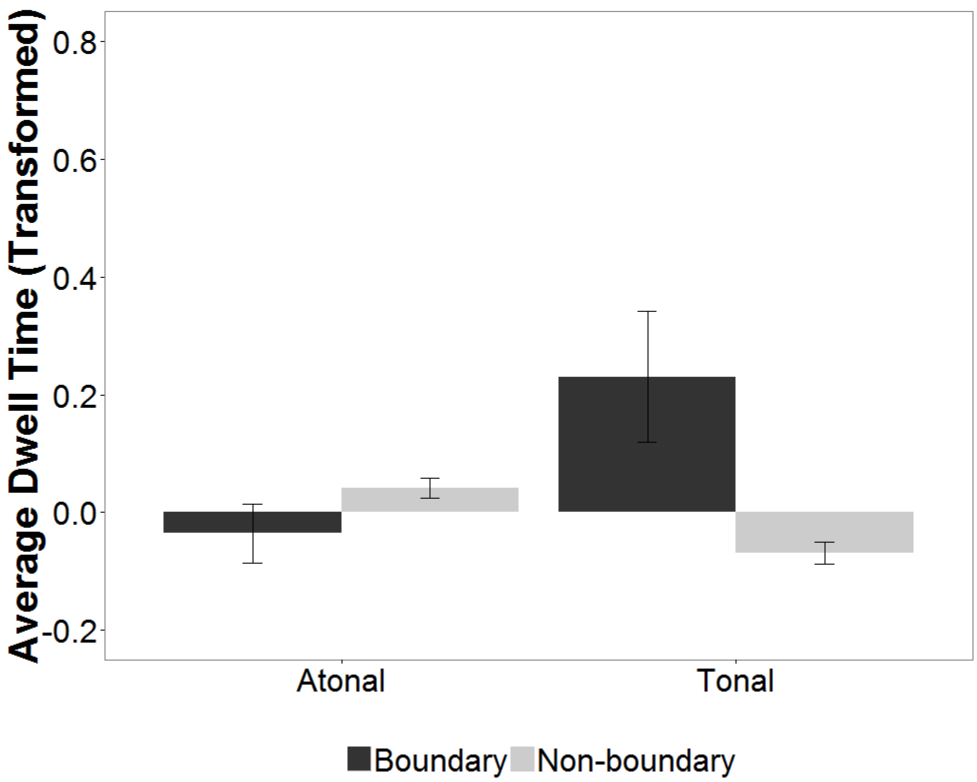

Figure 7. Mean dwell times for non-musicians binned by tonality and boundary status. Bars represent means of transformed dwell times for each trial type (it should be noted that there were many more raw data points for non-boundary chords than boundary chords). Error bars represent standard error of the mean. Participants dwelled longer on boundaries than non-boundaries in the tonal condition, but not the atonal condition.

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)*

Table 1. Averages for untransformed dwell times in milliseconds (including the first tonal and first atonal blocks) with standard deviations in parentheses.

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.

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. doi: 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. doi: 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. doi: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. doi: 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. 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.

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

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

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics, 14*(1), 37-40. doi; 3758/BF03198614

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

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

Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23*(27), 9240-9245. doi: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. doi: 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. doi: 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. doi: 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. doi: 10.1162/089892900562183

Koelsch, S., & Jentschke, S. (2008). Short-term effects of processing musical syntax: An ERP study. *Brain Research, 1212,* 55-62. doi: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. doi: 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. doi: 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. doi: 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. doi: 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. doi: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. doi: 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. doi: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. doi: 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. doi:10.1162/089892906775990642

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

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

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology*, *24*, 157-178. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1037/h0081052

Smith, N.A., & Trainor, L.J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distractors. *Infancy, 16,* 655-668. doi: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. doi: 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. doi: 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. doi: 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. doi: 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. doi:10.3758/BF03194732

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

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

Zacks, J.M., & Swallow, K.M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80-84. doi: 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. doi: 10.1126/science.1238414