

# ANALYSIS, PERFORMANCE, AND TENSION PERCEPTION OF AN UNMEASURED PRELUDE FOR HARPSICHORD

---

MEGHAN GOODCHILD

McGill University, Montreal, Canada

BRUNO GINGRAS

University of Innsbruck, Innsbruck, Austria

STEPHEN MCADAMS

McGill University, Montreal, Canada

**THIS STUDY FOCUSES ON THE RELATIONSHIPS** between music analysis, performance, and tension perception. Part 1 examines harpsichordists' analyses and performances of an unmeasured prelude—a semi-improvisatory genre open to interpretive freedom. Twelve harpsichordists performed the *Prélude non mesuré* No. 7 by Louis Couperin on a harpsichord equipped with a MIDI console and submitted a formal analysis. Using a curve-fitting approach, we investigated the correspondence between analyzed segmentations and group-final lengthening. We found that harpsichordists also employed “group-final anticipation,” involving deceleration before, and acceleration through, analyzed boundaries. In Part 2, three listener groups (harpsichordists, musicians, and nonmusicians) continuously rated tension for 12 performances. In contrast to measured music, local tension peaks, rather than troughs, occurred at boundaries featuring group-final lengthening. Associations were found between global tempo and tension ratings, with significant differences among the three listener groups. Performers expressed the large-scale structure through the amount of tempo variability, which was also reflected in tension rating variability.

Received: August 21, 2014, accepted September 9, 2015.

**Key words:** performance, tension, unmeasured music, analysis, musical structure

---

**O**VER THE LAST FOUR DECADES AN INCREASING amount of research has been undertaken on musical performance, with timing measurements dominating the investigation (Gabrielsson, 2003). A large body of literature on piano performance

has found that performers convey information about the grouping structure of the piece (Clarke, 1985, 1988; Palmer, 1989, 1996a; Repp, 1992, 1995). In particular, performers tend to slow down at sectional boundaries, a process referred to as “phrase-final lengthening” or “group-final lengthening” (Bengtsson & Gabrielsson, 1980; Ohlner, 2012; Palmer, 1989; Repp, 1995; Todd, 1985). The magnitude of the tempo variation corresponds to the hierarchical importance of the boundaries, with larger *ritardandi* associated with major formal divisions of the piece (Repp, 1992; Shaffer & Todd, 1987; Todd, 1985). The *musical communication hypothesis* (Clarke, 1985; Gabrielsson, 1987; Palmer, 1989, 1996a; Penel & Drake, 1998; Repp, 1992, 1995; Shaffer, Clarke, & Todd, 1985) proposes that tempo fluctuations correspond to the performer's interpretation of musical structure and its communication to the listener. More recently, Penel and Drake (2004) reported that group-final lengthening could be attributed to motor and perceptual constraints and also to musical communication through expression of the grouping structure hierarchy.

This relationship between music-theoretical analysis and performance has also been a topic of increasing interest among music theorists. Cone's (1960) essay influenced the direction of the discussion for several decades as a one-way line of communication of analytical interpretation from theorists to performers, including work by Berry (1989) and Narmour (1988). More recently, theorists such as Lester (1995), Rink (2002), and McCreless (2009) have reversed this line of thought by advocating a collaborative view, one that values performers' insights and looks to performance as a way to illuminate analysis.

Attempting to bridge the gap between analyst and performer is problematic due to differences in terminology and artistic goals; theorists tend to analyze the musical score as an object “out of time,” whereas performers necessarily operate in the temporal dimension. Gingras, McAdams, and Schubert (2010) studied the relationship between the performer's concept of a piece both as an analyst and performer to circumvent the apparent disconnect between the language and goals of music theory and performance. In their study, organists performed J. S. Bach's “Dorian” Fugue (BWV 538)

and submitted a written analysis of the phrase structure. Gingras et al. (2010) observed that organists used group-final lengthening and the end-of-phrase *ritardando* corresponded to the hierarchical importance of major subdivisions in their written analyses. They examined whether performers whose written analyses substantially differed from other performers also emphasized unique formal aspects in their performances, but did not find a significant difference between the average size of the *ritardandi* of the performers who marked a section as a boundary and those who did not.

Since timing patterns have been found to be a major expressive device in harpsichord performance (Gingras, Asselin, & McAdams, 2013), this study investigates the link between temporal patterns and formal structure in an unmeasured prelude for harpsichord, a genre cultivated in France between 1650 and 1720. Due to its semi-improvisatory nature, the unmeasured prelude allows for considerably more interpretive freedom regarding temporal structure than measured music. Related to fantasias and toccatas, unmeasured preludes for harpsichord “go far beyond a simple introductory gesture and belong to the most ingenious and satisfying music in the keyboard repertoire” (Tilney, 1991, p. 1). Inspired by lute tablature, Louis Couperin developed a notational solution that conveys specific information about the order of pitches, but liberates the performer in terms of rhythm and tempo. Performances of an unmeasured prelude potentially display a wider range of possible structural realizations through event timing than what is found in measured music. This genre may thus shed more light on the connection between analysis and performance.

Recent research into the temporal dynamics of music listening has shown that listeners integrate musical events into a hierarchical organization, which governs their intuitions about musical tension, sense of completion, and perceptual expectations (Krumhansl, 1996). Real-time judgments have been used to test the ongoing musical experience, including segmentation of structural boundaries (Deliège, 1987) and continuous measures of musical tension (Farbood, 2012; Krumhansl, 1996; Toiviainen & Krumhansl, 2003). Segmentation provides a direct measure of parsing the musical surface; however, it does not map the build-up and decay of musical experiences. In addition, precise boundaries may not be present in all types of music, especially in music such as the unmeasured prelude, as discussed above. Musical tension and relaxation schemas have been studied as an important dimension of musical experience (Fredrickson, 1995; Krumhansl, 1996; Madsen &

Fredrickson, 1993; Nielsen, 1983) and have been linked to aspects of the musical structure, such as harmony and tonal hierarchies (Bigand & Parncutt, 1999; Lerdahl & Krumhansl, 2007). Krumhansl (1996) reported that the musical surface could be parsed into hierarchically organized units through the measurement of variations over time in the degree of experienced tension. Several musical features were found to influence tension ratings, including melodic contour, dissonance, dynamics, chromaticism, and breaks in harmonic patterns. Farbood (2012) investigated the relative contributions of several musical parameters on perceived tension in both retrospective and continuous ratings. Dynamics and pitch height were found to have a clear effect on tension in isolated contexts. In complex stimuli with the interaction of various features, the tension results were more ambiguous; however, when multiple features moved in the same direction, tension ratings were strengthened in this same direction.

The connection between musical structure, performance timing, and tension has been less studied. Krumhansl (1996) reported that peaks and rapid decreases in tension corresponded to slower tempi at section endings. Palmer (1996b) found a correlation between expressive timing and music-theoretically predicted tension (Lerdahl, 1996), with performers lengthening positions of predicted high tension more than positions of relaxation. In terms of global tempi of short phrases, listeners rated higher tension for music with a fast tempo than with a slower tempo (Ilie & Thompson, 2006). In stimuli that were designed to isolate the effects of onset frequency and tempo, a decrease in both features resulted in decreases in tension (Farbood, 2012).

Fredrickson (1997, 1999) reported that perceived tension was not greatly affected by musical expertise or familiarity with a particular piece of music. Krumhansl (1996) also showed that tension responses were uniform despite differences in music training. The relative contributions of various musical parameters to tension responses, however, have been found to vary for musicians and nonmusicians. Farbood (2012) reported that musicians were more sensitive than nonmusicians to changes in most musical features (e.g., harmony, dynamics, and onset frequency), with the exception that nonmusicians were more influenced by changes in pitch height.

Continuous response rating devices, such as spring-loaded tongs (Nielsen, 1983), the continuous response digital interface (CRDI) dial (Madsen & Fredrickson, 1993), a virtual slider or emotion space on a computer screen manipulated by a mouse (Farbood, 2012; Schubert, 2004), and a handheld mechanical slider box (McAdams,

Vines, Viellard, Smith, & Reynolds, 2004), have been used successfully to measure perceived tension or emotion ratings in previous studies. The continuous rating task is intuitive and does not require musical experience (Schubert, 2004). Although some studies have reported high inter and intrasubject correlations among subject responses (Krumhansl, 1996; Madsen & Fredrickson, 1993; Nielsen, 1983), other studies using continuous measures have found substantial variability among subjects (McAdams et al., 2004; Upham & McAdams, 2014). Given this variability, there is some concern regarding the interpretation of mean ratings. In this study, we employ several analytical approaches (to be discussed further below) in order to effectively interpret the continuous responses, to map important musical moments signaled by the listeners, and to investigate rating strategies by different listener groups.

### Present Research

The aim of the work presented here was to investigate the relationships between analysis, performance, and perception of tension. To this end, two experiments were conducted. The first experiment involved performances and written analyses by harpsichordists of an unmeasured prelude. We investigated whether there were correspondences between analysis and performance in terms of group-final lengthening and whether performers treated local-level boundaries differently than higher-level grouping boundaries. We predicted that the performances of the unmeasured prelude would have greater variability of tempo profiles among performers compared to measured music on the local level, but that there might be more correspondence among performers on higher hierarchical levels. In addition, we expected that the connections between individual analyses and strategies of tempo fluctuations would be strong given the expressive freedom of interpretation of the unmeasured prelude compared to the constraints of measured music.

For the second experiment, we conducted an exploratory perceptual study that involved collecting continuous tension ratings of three listener groups with varying levels of expertise: harpsichordists, nonharpsichordist musicians, and nonmusicians. The aims of the tension study were to map the experience of perceived tension over time, to examine the connection between performance timings and perceived tension, and to compare the ratings across the three listener groups. The study directly confronts the influence of timing variation on tension ratings, as the harpsichord has limited dynamic range (Penttinen, 2006), and other musical

features are held constant (e.g., melodic contour) for all performances.

### Experiment 1: Analyses and Performances of an Unmeasured Prelude

#### METHOD

*Piece.* The *Prélude non mesuré* No. 7 by Louis Couperin (1626-1661) was selected for this experiment as representative of the unmeasured prelude genre from the baroque harpsichord repertoire. The score is presented in Figure 1, adapted from Tilney (1991). The segmentations marked by angled lines will be discussed below. Although meter and precise rhythmic values are absent, several features of the notational system clarify the order of pitches, relative durations, and stylistic aspects for performance. Tilney (1991) explains that a single note attached to a long curved line should be held at minimum until the termination of the line, which often directs toward important goal pitches or points of the musical flow. He advises that the lines “serve to impart order to the music, preventing events from happening too fast” (Tilney, 1991, p. 6). Where several lines terminate together, the harpsichordist would release the held notes in one motion, which we term a “gestural change.” A small slur that links two pitches could be interpreted as a *port de voix*, with the first note accented, or as a rhythmically inflected pattern (Tilney, 1991). Moroney (1976) notes that, “Good prelude playing requires the invigoration of imaginative freedom, but the player must be liberally inclined to impose on the music that measurement and shape which his imagination dictates” (p. 143). Therefore, performers of the unmeasured prelude use the notational details as a point of departure for expressive interpretation.

*Participants.* Twelve professional harpsichordists, seven male and five female, with an average age of 39 years (range = 21-61) participated in the study. They had played the harpsichord for an average of 22 years (range = 6-40) and were performing and residing in the Montreal area at the time of the experiment.

*Procedure.* The harpsichordists were provided with the score of the unmeasured prelude 4-6 weeks prior to the recording session. They were asked to analyze the piece, with instructions as follows: “On a copy of the score of the Unmeasured Prelude No. 7 by Louis Couperin, please indicate what you feel are the main subdivisions of the piece, by tracing vertical lines to delimit sections. Please note that there are no right or wrong answers to this question; the goal of this question is to analyze your performances in light of *your* analysis.” This relatively

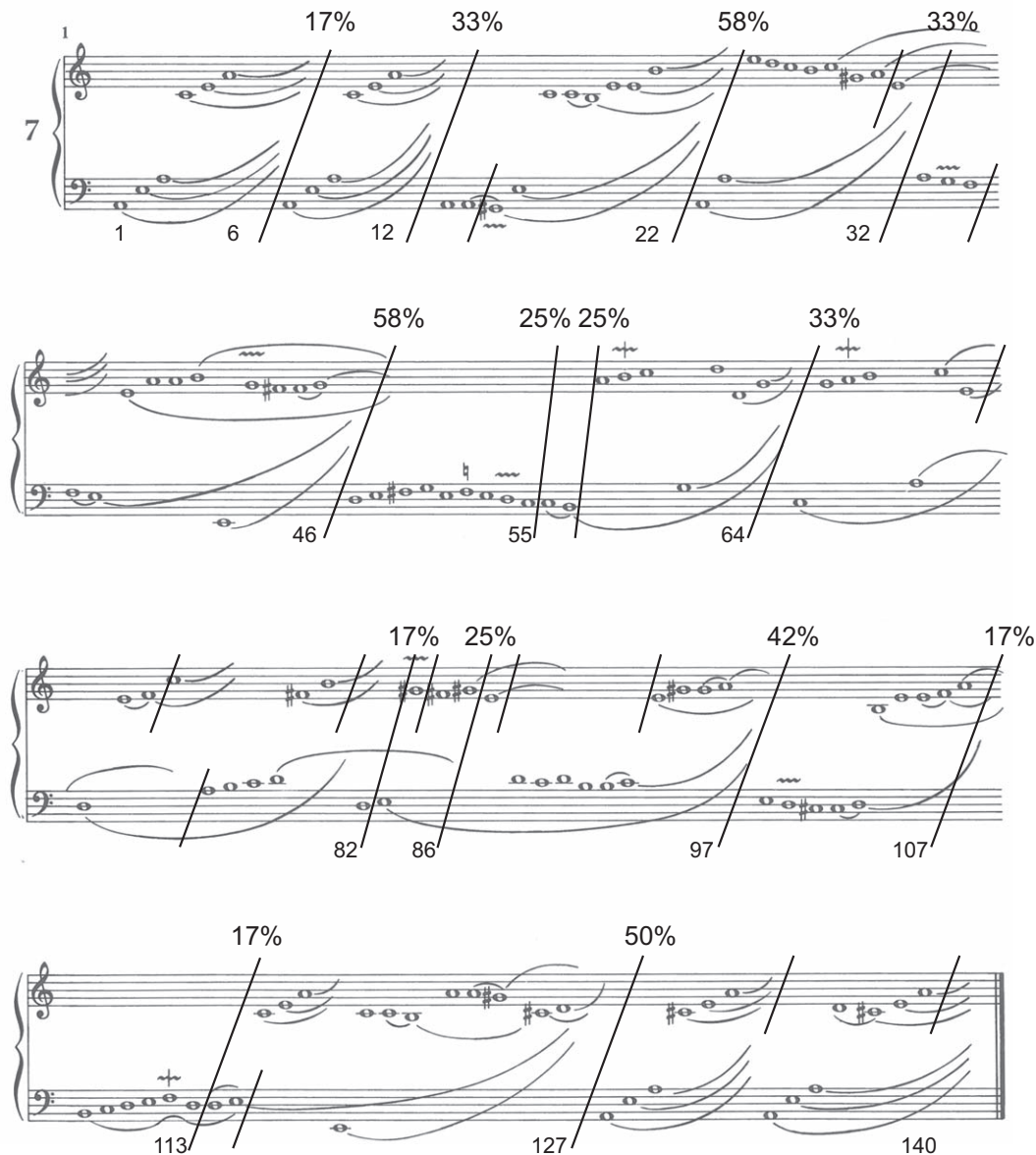


FIGURE 1. Score of *Prélude non mesuré No. 7* by Louis Couperin. Short slashes indicate the segmentation of a single harpsichordist. The percentages of agreement for other segmentations are listed above the long slashes. Event number at each boundary is indicated to the left of the long slash. Score adapted from (Tilney, 1991, p. 20). Used by permission of European American Music Distributors Company, sole U.S. and Canadian agent for Schott Music Ltd., London.

straightforward segmentation task was chosen to avoid imposing a strict music-theoretical approach, which may cause a disconnect between the language and goals of analysis and performance.

The harpsichordists performed the piece twice in a recital setting without performance directives from the experimenters. Immediately following their performances, they submitted the score with their analysis. At the same recording session, the harpsichordists also

performed two other pieces (not analyzed here), as outlined in Gingras, Asselin, and McAdams (2013). The entire session lasted approximately one hour and the participants were provided financial compensation.

Performances took place in an acoustically treated studio at CIRMMT, McGill University on an Italian-style Bigaud harpsichord with two 8-foot stops (Heugel, Paris). All performances were recorded using the same registration (the back stop). The harpsichord was equipped with

a MIDI console, which precisely measures performance parameters including note onsets and offsets, as well as key-stroke velocity. The audio signal was recorded through two omnidirectional microphones MKH 8020 (Sennheiser Electronic GmbH, Wademarck Germany), located 1 m above the resonance board and placed 25 cm apart. The audio and MIDI signals were recorded through an RME Fireface audio interface (Audio AB, Haimhausen, Germany) using Cakewalk's SONAR software (Cakewalk, Inc., Boston, MA) and were stored on a hard disk.

*Data analysis.* Note onsets and offsets were extracted from the MIDI data of the performances and matched to the score using an algorithm developed by Gingras and McAdams (2011). Note durations were determined through interonset intervals (IOI) of each notated event for each performer. Only the IOI between the onset of the event and the following event was considered. Ornaments (notated and improvised) were not included in this analysis, but will be discussed below and in Part 2.

Performed IOIs in measured music are interpreted proportionally within the metrical framework. In contrast, the performed IOIs of the unmeasured prelude were analyzed within a framework of events considered to be isochronous (i.e., without a metrical structure). We assumed that listeners of the unmeasured prelude would not be able to infer a metrical structure, but that they would be able to chunk the musical sequences into groupings based on pitch, attack points, and durations. According to Lerdahl and Jackendoff's (1983) theory, grouping sequences of events is a basic aspect of musical understanding. Listeners construct a metrical hierarchy from inferring regular patterns of events as strong and weak beats; however, the perception of metrical structure can become ambiguous when the musical cues are irregular, which we presume would be the case with the unmeasured prelude. Lerdahl and Jackendoff (1983) state, "[T]he principles of grouping structure are more universal than those of metrical structure. In fact, though all music groups into units of various kinds, some music does not have metrical structure at all, in the specific sense that the listener is unable to extrapolate from the musical signal a hierarchy of beats" (p. 18). Despite the absence of a metrical hierarchy, performers would be able to convey a grouping-structure hierarchy through the use of phenomenal accents (e.g., long notes) and structural accents (e.g., harmonic goals such as a cadence) (Lerdahl & Jackendoff, 1983). Therefore, we investigated the grouping structure as determined by performer-selected segmentations and performed durations.

## RESULTS

*Score analysis.* Out of 140 notes (hereafter called events), the harpsichordists marked 5.6 boundaries on average ( $SD = 3.7$ ). The response frequency, varying between two and 13 boundaries, indicates that the harpsichordists may have evaluated the structure at different hierarchical levels. Seven performers demarcated two or four boundaries, thereby dividing the piece into large sections. The remaining five performers analyzed on a closer, surface level with seven to 13 segmentations.

In total, the performers segmented 27 different boundaries, 13 of which were indicated by only one harpsichordist. Only seven boundaries resulted in agreement of over 30% (four or more harpsichordists) and the highest level of agreement was 58% (7 out of 12 performers). In comparison, Gingras et al. (2010) reported that organists who analyzed Bach's Dorian Fugue (222 measures and 15 cadences) marked 7 boundaries on average, with a total of 21 different subdivisions and agreement of up to 93.8%.

Figure 1 contains the score with performer-selected boundaries. Short slashes indicate the segmentation of a single harpsichordist; otherwise, the percentage of agreement is noted above the long slashes. The event number at each boundary is indicated to the left of the slash below the lower staff. In general, the harpsichordists segmented based on gestural changes (described above) or instances where changes in harmony could be implied. Boundaries with high agreement coincide with both harmonic and gestural changes (events 12, 22, 46, 64, 97, and 127). There are several cases where there is greater analytical disagreement, particularly in the interior section where the harmonic and gestural changes are less distinct. The set of 14 segmentations with agreement of 17% or more (long slashes) will be referred to as the "performers' analysis."

In addition to angled lines and slashes to indicate formal divisions, several performers used the score as a working performance copy, notating performance directions, extra ornaments, harmonic progressions, rhythmic patterns, tempo indications (e.g., *accelerando*), articulation (e.g., slurs), and formal descriptions (e.g., bridge). For example, performer 1 provided only two segmentations, but also included several downward-pointing arrows, likely to plan important goal notes as points of arrival during performance. The variety of analytical methods utilized by the performers speaks to the detailed approach of learning the piece and preparing it for performance.

It should be noted that the groupings based on the performers' analysis in the unmeasured prelude are not equivalent to those in previous studies on measured

music where the majority of structural boundaries correspond to cadences at the phrase or thematic level. Instead, the groupings occur at a more local level as chord changes or gestures, which may explain the lower agreement on the segmentations compared to Gingras et al. (2010). On the one hand, these points of articulation could be emphasized in performance through group-final lengthening. On the other hand, the linear nature of the music could also conceal a sense of arrival if the performer avoids a pause or *ritardando* in the forward flow.

Towards the end of the piece, there is an elaborated cadential formula (Ijzerman, 2011), with the only instance of root position dominant to tonic harmonic motion. Based on Lerdahl and Jackendoff's (1983) theory, the boundary at 127 would be situated at a higher level of the grouping hierarchy due to the structural accent as the goal of tonal motion. We predicted that the performers would be more likely to use group-final lengthening into event 127 compared to the local-level boundaries.

In the following discussion, we focus on the grouping structure of the piece from the perspective of the harp-sichordists' analyses and performances, which could be informed by their understanding of harmony, melodic contour, texture, and their interactions. Therefore, we have avoided imposing a structural or harmonic analytical perspective, except for identifying the segmentation of event 127 as a point of cadential articulation.

**Expressive timing.** We predicted that the performances of the unmeasured prelude would feature more timing variation than in measured music. Previous studies on measured piano and organ music have reported high correlations for tempo patterns within and between performers (Gingras et al., 2010; Palmer, 1989; Repp, 1990). For the unmeasured prelude performances in the current study, there was a moderate mean correlation coefficient of .54 ( $SD = .18$ ,  $df = 138$ ) between all pairs of performances and a strong correlation coefficient of .87 ( $SD = .10$ ,  $df = 138$ ) between performances by the same performer.

**Note duration graphs.** Note duration graphs were completed from the IOIs of each performer to assess the correspondence between their timing profiles and their written analyses. We expected to find group-final lengthening at the performers' segmented group endings; however, the performers used expressive timing with a variety of tempo patterns. Detailed comparisons between performers 9 and 6 (hereafter P9 and P6) will further demonstrate this individuality in expressive timing. These performers were chosen because they represented the extremes of the range of performances

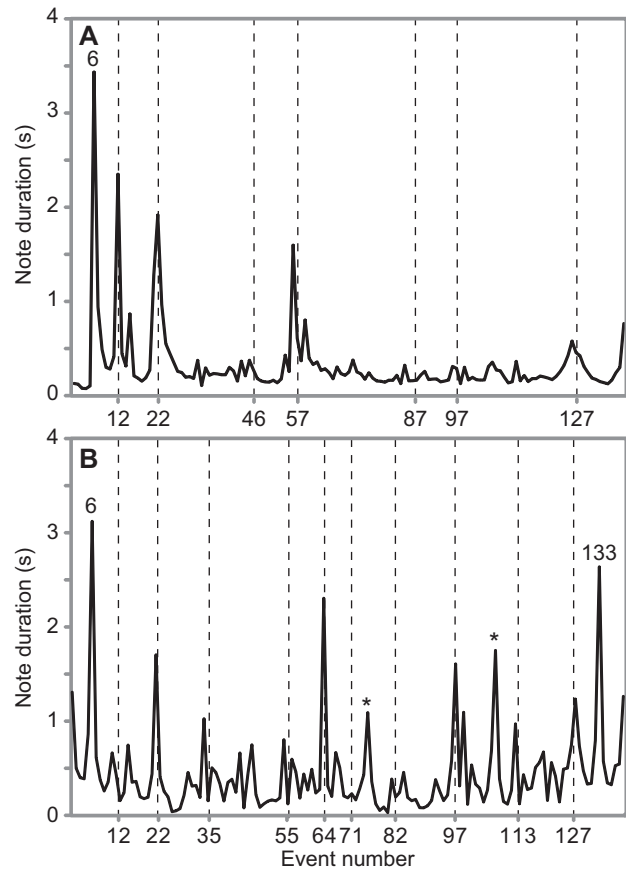


FIGURE 2. Note duration by event number for P9 (A) and P6 (B). Performer-specified boundaries (dashed-vertical lines) along the horizontal axis. Event numbers above the curve indicate boundaries from the performers' analysis discussed in the text. Asterisks indicate note duration peaks that are not associated with segmentations by P6 or the performers' analysis.

recorded in terms of mean tempo and timing variability (as explained below). Figure 2 contains the note duration graphs for performances by P9 and P6. Boundaries specified by each performer are indicated by dashed, vertical lines along the horizontal axis. Event numbers above the curve indicate pertinent boundaries derived from the "performers' analysis."

P9 performed one of the fastest base tempi of the 12 performers (0.30 s per event, 198 events per minute). The longest *ritardandi* occurred at the performer's segmented boundaries, as well as during the initial rolled arpeggio at event 6 (marked above the curve). A deliberate *ritardando* with a wide peak occurred leading up to the final cadence (event 127), indicating a protracted increase in note duration. Apart from four large decelerations into events 6, 12, 22, and 57 the performer used a relatively steady tempo throughout.

In contrast, P6 had a slower base tempo (0.49 seconds per event, 122 events per minute) and greater timing variability. Less of a direct connection exists between segmented boundaries and *ritardandi*; several large peaks are associated neither with the segmentations of the “performers’ analysis” nor analysis of this performer (e.g., events 75 and 105, marked by asterisks). In contrast, P6 slowed considerably during the second-to-last arpeggio (see peak at event 133), elaborating the post-cadential material with a larger *ritardando* than the arrival at the cadence at 127.

**Group-final lengthening.** We investigated the extent to which harpsichordists perform group-final lengthening (GFL) in the unmeasured prelude. We predicted that the performers would deploy timing patterns that reflect GFL, hereafter shortened to “GFL-reflective” following Ohriner (2012), at the end of analyzed segmentations, particularly at the point of cadential articulation. Previous studies have found that performed durational contours correspond to a parabolic shape at various structural levels, including the level of phrases (Todd, 1985), as well as local-level gestures (Repp, 1995). Therefore, we used MATLAB’s (Mathworks, version R2013b) Curve Fitting Toolbox to assess the goodness of fit of the performed durations to a parabolic (second-order quadratic) function or a linear function. First, we investigated the correspondence between GFL and the end of segments based on performer agreement. In order to facilitate this approach, a few adjustments were made to the “performers’ analysis.” Where two segmentations occurred within five or fewer events, one segmentation was chosen; therefore, event 55 was chosen over event 57 in order to divide in half the segment between events 46 and 64, and event 86 was chosen over 82 based on the higher percentage of agreement. A subdivision was included at event 133 in order to respect the repeated arpeggiations that mirror the segmentation at event 6. Finally, a boundary was included after event 140 as the last grouping of the piece. We also investigated the GFL-reflectivity of the performances by P9 and P6 for the groupings based on their individually segmented boundaries compared to the groupings based on the performers’ analysis (discussed below).

For all of the 24 performances, the last six durations for each grouping of the performers’ analysis was classified as parabolic with positive curvature or linear with a positive slope depending on the greater strength of fit ( $R^2$ ). Figure 3 contains examples of segments with strong linear and parabolic fits. Otherwise, the performed durations were classified as not GFL-reflective if there was a negative parabolic curvature, negative

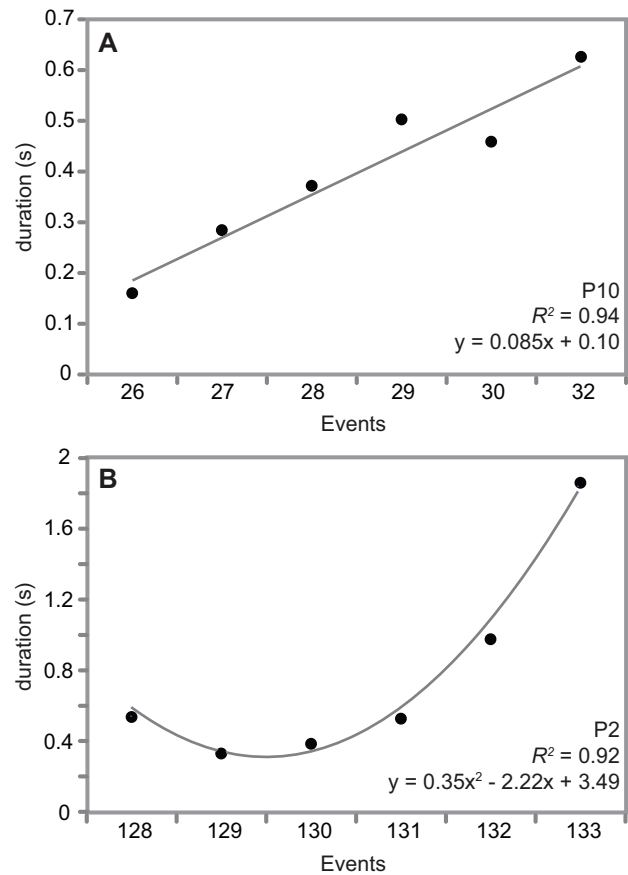


FIGURE 3. Note duration by event number with line of best fit for linear (A) and parabolic (B) functions,  $R^2$  value for quality of fit, and coefficients for equation.

linear slope, or a weak fit ( $R^2 < .20$ ). Figure 4 contains a matrix of the  $R^2$  values for quadratic or linear fits (grey to black shading). White shading indicates that the segment was not GFL-reflective. The majority of performers’ durations were GFL-reflective for the groupings at the beginning and end, particularly for the opening and closing rolled arpeggios (segments ending at events 6, 12, 133, and 140). In the interior of the piece, there were fewer fitted segments and the  $R^2$  values were lower for those that were GFL-reflective. Of particular note is that the majority of performances did not slow down into the boundary at event 127, which had 50% agreement and represents the main cadential articulation of the piece.

Looking more closely at the differences between P9 and P6, Figure 5 shows the durational contour of each segment for events 1–46. Several aspects distinguish the interpretations of the two performers. For the first three segments, P9 exhibited very deliberate lengthening of the final note in each grouping, after which point a more



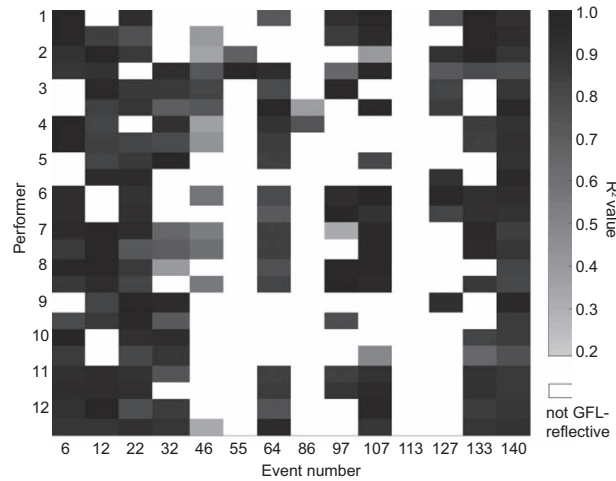


FIGURE 4. Matrix of the  $R^2$  values from the quadratic or linear fits (grey to black shading) or lack of fit (white) for 24 performances by 12 performers across 14 segments. Event numbers indicate the boundaries from the performers' analysis.

consistent and rapid tempo was established, with a slight slowing into event 32. In contrast, the first, third, and fifth segments were GFL-reflective for P6 (groupings ending at events 6, 22, and 46).

The analysis above investigated whether the durations were GFL-reflective for the boundaries of the performers' analysis. Table 1 directly evaluates the performed segments as individually analyzed by P9 and P6 (bottom

row, labeled P9 or P6) in comparison to the performers' analysis (top row, labeled "PA"). Shaded cells represent GFL-reflective segments. Text in bold and italics indicates performer-specified boundaries that differ from the performers' analysis.

The analysis reveals that P9 and P6 did not perform segments with group-final lengthening for the boundaries that they personally analyzed to a greater extent than the boundaries in the performers' analysis. As shown in Table 1A, P9 did not employ group-final lengthening for the two segmentations that diverged from the performers' analysis (events 57 and 87). Although several segmentations by P9 converged with the performers' analysis, many of these were not GFL-reflective (events 46, 97, 127). Overall, only 29% of P9's analyzed segments were GFL-reflective, compared to 43% for the performers' analysis. Similarly for P6, there were more segments with group-final lengthening for the performers' analysis (64%) compared to the individual analysis (50%) (see Table 1B). For the segmentations that differ from the performers' analysis (events 35, 71, and 82), P6 employed group-final lengthening only for the segmentation at 82. This harpsichordist also emphasized the boundaries of the performers' analysis, but did not personally indicate them as segments (events 6, 46, 107, and 133).

*Group-final anticipation.* The above analyses point to a complex interplay between analyzed groupings and

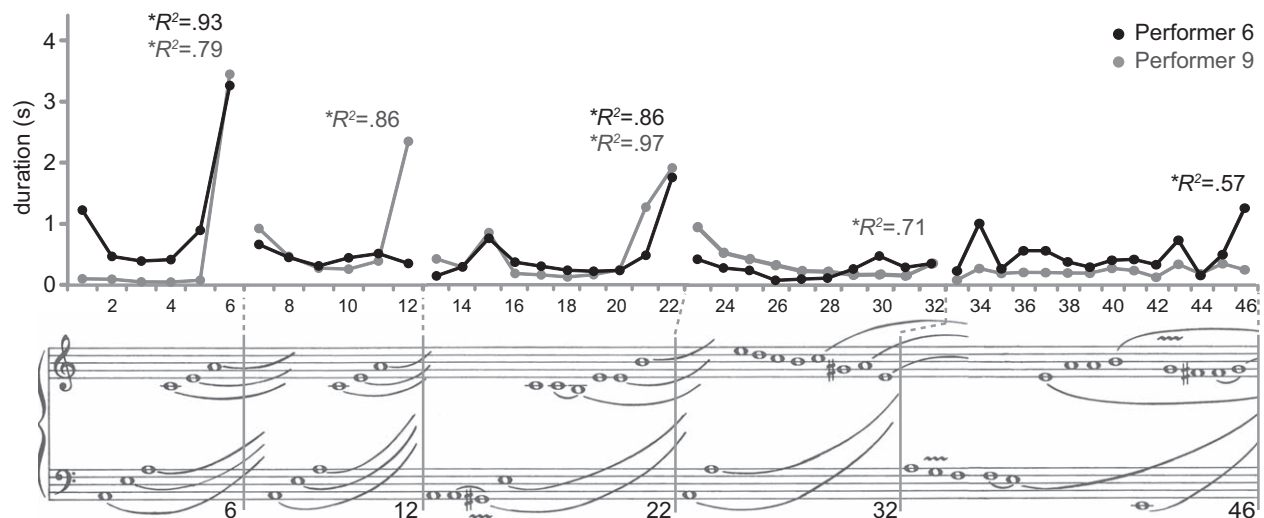


FIGURE 5. Note duration for events 1-46 for P9 (grey) and P6 (black). Grey vertical lines on the score indicate the boundaries from the performers' analysis. Asterisks indicate GFL-reflective segments (last six events considered for each grouping).  $R^2$  values indicate strength of fit for a parabolic function. Score adapted from (Tilney, 1991, p. 20). Used by permission of European American Music Distributors Company, sole U.S. and Canadian agent for Schott Music Ltd., London.



TABLE 1. Comparison of GFL-Reflective Segments for the Performers' Analysis (PA) and the Individual Analyses of P9 (A) and P6 (B).

A														
PA	6	12	22	32	46	55	64	86	97	107	113	127	133	140
P9	12	22	46	57*	87*	97	127							
B														
PA	6	12	22	32	46	55	64	86	97	107	113	127	133	140
P6	12	22	35*	55	64	71*	82*	97	113	127				

Note. Grey shading indicates GFL-reflective segments. Numbering indicates the associated event at the segmentation. Bold italic text with an asterisk indicates segmentations that differ from the performers' analysis.

performed phrases. Although performers exhibited GFL-reflectivity at certain segmentations, as shown in the durational graphs and analysis of GFL-reflective segments, many performers also slowed down before but accelerated through a boundary, a process we call "group-final anticipation." The performer anticipates the end of the grouping with the peak of the *ritardando* an event before the group end, followed by an acceleration into the following grouping. Therefore, the same understanding of grouping structure could be expressed by lengthening the final event of the grouping or by lengthening the penultimate event.

The note duration graph for P4 in Figure 6A provides an example of group-final anticipation for the final cadence at event 127. To demonstrate this more clearly, Figure 6B shows the relative event-to-event timing variation ( $\log_2$ ), with +1 and -1 representing values twice as fast and slow, respectively. The peak of the deceleration occurred several events before the cadence at event 124 (shown with an asterisk) and the peak of the acceleration actually occurred at event 127. Group-final anticipation is likely the result of the fluid nature of the unmeasured prelude and the low hierarchical level of the boundaries as suggested by harmonic or gestural changes, which may not necessitate strict markers of formal division in the same way as phrase or thematic boundaries do in measured music.

*Smoothed note durations.* In addition to the difficulty in accounting for group-final anticipation, several notated ornaments occurred a few events before segmented groupings (e.g., see Figure 1, events 43, 54, and 112). Since the duration of each event is calculated from the IOI, performed ornaments resulted in a spike in note duration. The ornaments could be understood as a written out *ritardando* before the end of the phrase, thereby imposing some constraints for performance at the end of the grouping.

The previous analyses conducted were sensitive to local-level timing variations. Using a window of four events to smooth the effects of the ornaments, the ten

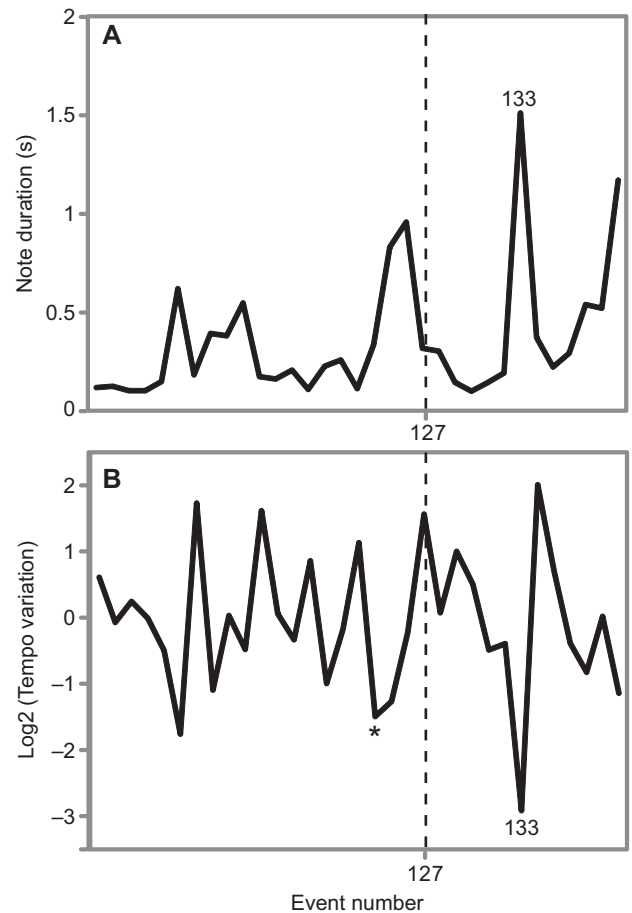


FIGURE 6. Note duration (A) and timing variation ( $\log_2$ ) (B) for events 107-140 performed by P4. Performer-specified boundary (dashed vertical line) along the horizontal axis. Numbered events above curve indicate boundaries from the performers' analysis discussed in the text. The asterisk shows the peak of deceleration at event 124.

largest *ritardandi* were calculated based on the average relative timing variation ( $\log_2$ ). Figure 7 compares the performers' smoothed *ritardandi* profiles to the performers' analysis with 25% agreement or more (black bars). The ten longest durations are indicated by open squares, while the open circles represent the relative

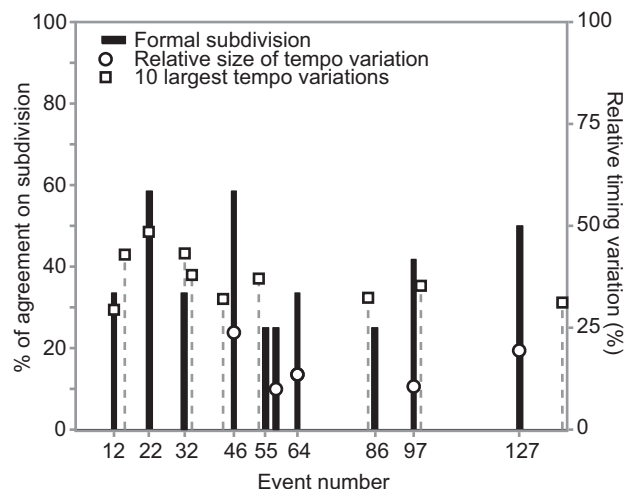


FIGURE 7. Comparison of *ritardando* profiles and formal boundaries. Open squares indicate the average relative timing variation (right scale) of the ten longest *ritardandi* using a sliding window of four events. Black bars indicate the percentage of agreement (25% or more) for a particular subdivision (left scale). The relative timing variation for the boundaries that were not associated with ten longest *ritardandi* are notated with open circles.

timing variation for the remaining boundaries. Several boundaries were directly associated with one of the 10 longest *ritardandi* (events 12, 22, 32, and 86), along with the final event of the piece. Over half of the largest tempo variations were the result of an ornament (events 15, 34, 43, 54, and 99).

Of all the analyzed boundaries, we predicted that the harpsichordists would emphasize event 127 through group-final lengthening, due to high analytical agreement and music theoretical significance as the only cadential point of articulation in the piece. However, the boundary was not one of the ten longest *ritardandi* overall. At event 127, the last note of the chord to sound (D) is the chordal seventh (a dissonance), which sets up strong expectations to resolve on C# at event 131. As discussed above, many harpsichordists connected the linear motion from events 127 to 131 with group-final anticipation, rather than a *ritardando* to mark the end of the gesture.

**Note duration variability.** We found that the harpsichordists conveyed a sense of the large-scale structure of the piece through the amount of variability in note duration. Figure 8 illustrates the average absolute variation in note duration, calculated as the absolute first derivative, in seconds per event. As an extension of the performers' analysis and the pattern of GFL-reflective segments, the piece was divided into two sections of nearly equal length: one including both the beginning

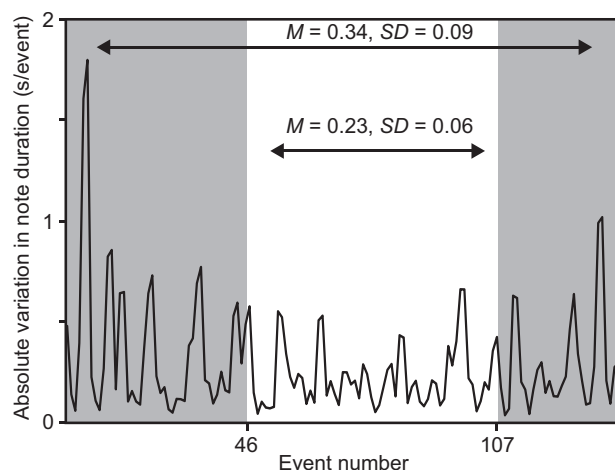


FIGURE 8. Absolute variation in note duration in seconds per event across all performers. Mean values of beginning and end compared to the interior section.

and the end (events 1-46, 108-140) and another comprising the interior section (events 47-107). The boundaries at 46 and 107 were chosen based on the high analytical agreement and also because the majority of performances were GFL-reflective. The greatest amount of timing variability occurred at the beginning and end ( $M = 0.34$ ,  $SD = 0.09$ ) and decreased for the interior section ( $M = 0.23$ ,  $SD = 0.06$ ). A paired-samples  $t$ -test across the 24 performances revealed that the variability in note duration was significantly higher for the opening and closing sections compared to the interior section  $t(23) = 8.41$ ,  $p < .001$ ,  $d = 1.72$ .

#### DISCUSSION

In general, the results indicate that the unmeasured prelude lends itself to interpretive freedom in terms of analysis and performance. Several segmentations had moderately high analytical agreement (events 22, 46, 97, and 127), but many boundaries were only identified by a single performer. The compositional structure and notational system, although semi-improvisatory, supply inherent constraints for performance; the curved lines provide some interpretive guidance for the relative durations of notes within gestures and the ornaments invite performers to highlight and dwell on certain events to varying degrees. The comparisons between contrasting performances by P9 and P6 revealed interesting similarities and dissimilarities in analysis and performance of the unmeasured prelude. In the note duration graphs, there were stark differences in the interpretations, including timing variations and the relationship between analysis and performed *ritardandi*.

By assessing the parabolic or linear fit of performed durations, we investigated the correspondence between analyzed segmentations and group-final lengthening. Performers employed group-final lengthening for the boundaries at the beginning and end to a greater extent than those in the interior section. We predicted that performers would show more GFL-reflectivity for the boundaries that they personally segmented compared to the ones that they did not segment; however, both P9 and P6 showed more correspondence for the boundaries agreed upon by the group.

Part of the discrepancy between GFL-reflectivity and analyzed segmentations is related to the variety of timing patterns employed by performers. We suggest that an alternative to the traditional group-final lengthening is group-final anticipation, which still conveys a certain understanding of grouping structure, without relying too heavily on predictable patterns of acceleration and deceleration within gestures. Clarke (1988) also discusses the ambiguity of the relationship between expressive aims and means: “the lengthening of a note can indicate that it is accented, that it finishes the structural unit at some level, or that the following (delayed) note is of structural importance” (p. 15). One example of this discrepancy is the final cadence at event 127, which had high analytical agreement and would be situated at a higher level of the grouping hierarchy due to the structural accent, but was not associated with one of the ten longest *ritardandi* on average. Performers highlighted this moment through group-final lengthening or group-final anticipation—slowing down several events before and accelerating through the resolution of the dominant seventh chord. Therefore, performers may have similar conceptions of the structure, but they express the grouping in different ways. Additionally, they may analyze the structure differently, but perform certain groupings similarly due to the inherent constraints of the gestural notation and placement of ornaments.

We found that the harpsichordists conveyed a sense of large-scale structure of the piece through the amount of timing variability. The stability of the opening and closing rolled chords may have permitted more extreme timing variations that display individual expressivity. The interior section, which features pseudo-polyphony and less distinct harmonic motion, may have required a steadier tempo, in order to underline the linear nature of the individual contrapuntal parts.

## Experiment 2: Perceived Musical Tension

Having examined the properties of the performances of the unmeasured prelude, we now turn to the perception

of those performances. For the second experiment, we conducted an exploratory study on perceived musical tension. We investigated the role of performance timing variations on tension ratings and compared the results of participants with varying levels of expertise: harpsichordists, nonharpsichordist musicians, and nonmusicians.

### METHOD

**Participants.** Fifty listeners participated in Experiment 2 (10 harpsichordists, 20 nonharpsichordist musicians, and 20 nonmusicians). All of the musicians and nonmusicians were recruited and tested in Montreal, Canada. Due to the limited number of harpsichordists in Montreal, five of the 10 harpsichordist listeners were recruited and tested in London, UK. The harpsichordist listeners, eight male and two female, had an average age of 31.6 years (range = 23–38), and had obtained at least an undergraduate degree in harpsichord performance. None of the harpsichordist listeners in Experiment 2 participated in Experiment 1 as performers. The musicians (12 male and 8 female) had studied a minimum of two years in an undergraduate degree in music and had an average age of 26.7 years (range = 20–47). The nonmusician participants (12 male and 8 female) had less than two years of music training in early childhood and had an average age of 26.6 years (range = 20–48).

**Apparatus.** The experiment took place at the Music Perception and Cognition Lab at the Schulich School of Music of McGill University (Montreal) and at Goldsmiths College (London). In both cities, participants were seated in a double-walled IAC sound-isolation chamber (IAC Acoustics, Bronx, NY): model 120-act 3 in Montreal and 1200-A in London. In Montreal, the experiment was conducted on a Macintosh G5 computer (Apple Computer, Inc., Cupertino, CA). The musical stimuli were amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA) and presented over Sennheiser HD280 Pro headphones (Sennheiser Electronic GmbH, Wademarck, Germany). In London, the experiment was conducted on a MacBook Pro 15” laptop computer (Apple Computer, Cupertino, CA). The musical stimuli were played through an Edirol UA-4FX USB external sound card (Roland Corporation, Shizuoka, Japan) and presented over Sennheiser HD202 headphones (Sennheiser Electronic GmbH, Wademarck, Germany). All participants passed an audiogram to ensure that hearing was below 20 dB HL at octave spaced frequencies (ISO 389-8, 2004; Martin & Champlin, 2000).

**Stimuli.** The stimuli used in the experiment consisted of one recording of Couperin’s *Prélude non mesuré* No.

7 by each of the 12 performers from Experiment 1. The performance selected out of the two for each performer was subjectively chosen based on the number of performance errors (wrong or missing notes) and recording quality. Two performances that were not selected for the main experiment were used for the practice trials. The stimuli were presented at a comfortable listening level that was kept constant for all of the sessions. Audio signals were sampled at 44.1 kHz with 16-bit amplitude resolution.

*Slider.* Participants used a handheld slider to continuously rate the perceived musical tension during each performance. They held the slider box on the desk or on their lap and moved the slider forward (away from themselves) to indicate higher perceived tension and backward (toward themselves) to indicate less perceived tension. The slider box was equipped with a rubber band to provide feedback through physical resistance. As the slider is moved in the direction of more tension, the participant increasingly felt more resistance to the movement. In pilot studies, participants reported that the slider was an intuitive method of rating that obviates the need for visual attention to the device.

*Procedure.* The first part of the experiment consisted of the practice trials, during which time the experimenter was present to answer questions. Participants were told that the practice pieces were representative of the performances in the main part of the experiment. The data from the practice stage were not analyzed. The main experiment consisted of one performance by each of the 12 performers, interspersed with 15 seconds of birdsong and nature sounds as a short break. The order of performances was randomized for each participant.

The computer interface, programmed with the PsiExp computer environment (Smith, 1995), provided instructions on the screen and allowed participants to advance at their own pace through the trials by clicking the mouse on a button on-screen. Participants began each trial with the slider at the bottom (no tension) and were asked to use the full range of the slider for at least one performance during the whole experiment. They were warned about the possibility of minor sound glitches and performance errors and were asked to ignore these issues. Once the experimental session ended, the participants completed a questionnaire and received financial compensation. The entire session lasted approximately 30 min.

Previous studies have avoided specific definitions of tension (Farbood, 2012; Fredrickson, 1997, 1999; Krumhansl, 1996; Madsen & Fredrickson, 1993; Silvey, 2011). Given the wide range of expertise of the

participants, no explicit definition of musical tension was provided.

*Data collection and analysis.* The slider was connected to an AtoMIC Pro analog-to-MIDI converter (Fléty, 2002), which converted the slider position to a 7-bit value and sent MIDI timings to the controlling computer with a maximum response latency of 10 ms. For each participant, a log file recorded the tension slider ratings continuously over time for each performance. The matched score of the MIDI performance data was used to establish a correspondence between performed durations and the tension values, which were averaged over the duration of each score event. Due to the force exerted by the elastic band when stretched at maximum, a block was inserted into the slider, resulting in half the scale of MIDI ratings between 54 and 127. These results were normalized to a range between 0 and 100 for each participant.

## RESULTS

*Mean tension ratings by event.* Graphs of the mean tension ratings by event were produced for each performer, with separate curves for each group of participants (harpsichordists, musicians, and nonmusicians). Mean ratings provide a measure of central tendency for each event for each listener group. The variability of tension ratings is discussed below. Returning to the two performers discussed in Experiment 1, as an example, P9 and P6 received the highest and some of the lowest mean tension ratings, respectively. Figure 9 contains the responses of the three listener groups and also maps the note durations of the single performance by P9 and P6. The mean group tension responses to P9 correspond quite closely to one another, with some individual differences. Generally, there is one initial tension arch from the beginning until around event 64, followed by a larger arch after 64 until the end; the first tension arch aligns with the large *ritardando* peaks for the first three gestures in the first half of the piece, and the second tension curve corresponds to the steadier tempo throughout the second half of the piece.

For P6, the ratings of the three listener groups generally resemble a plateau throughout. One striking difference between the tension ratings of P9 and P6 are the responses to the final rolled arpeggios. For P9, the tension ratings gradually decrease after event 127 for harpsichordist and musician groups, whereas the nonmusicians' ratings increase after event 133. For P6, the ratings drop considerably after event 127, but all three groups show a slight increase after event 133. The differences between the tension profiles may reflect the

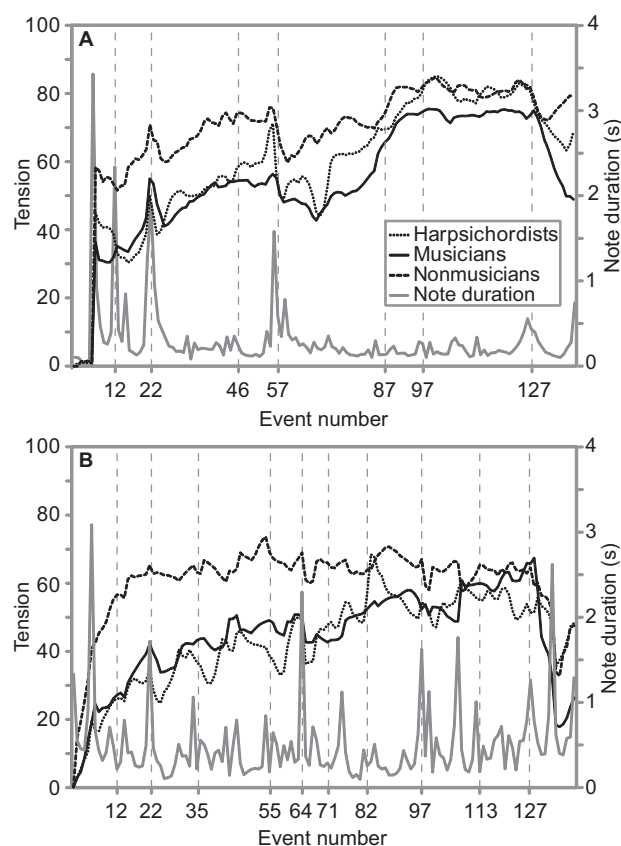


FIGURE 9. Note duration (right scale) and mean tension ratings (left scale) by listener group for P9 (A) and P6 (B). Performer-specified boundaries (dashed-vertical lines) along the horizontal axis.

different cadential profiles of the performers (as discussed above); P9 has a large *ritardando* at the cadence, whereas P6 emphasizes the post-cadential material (rolled arpeggios).

Krumhansl (1996) found that tension peaks followed by sharp decreases in tension occurred at section boundaries with hierarchical importance. For the unmeasured prelude, the tension profiles are shifted; tension peaks occur directly at the segmentations, with decreases in tension occurring at the onset of the next grouping. Consider, for example, the tension peaks for P9 and P6 that correspond directly to note duration peaks at boundaries, followed by a tension drop for the following segment (see Figure 9A, events 22, 55, 127; Figure 9B, events 22, 64, 97). These results also differ from Farbood (2012), who reported that a decrease in tempo and onset density resulted in decreases in tension ratings in simple stimuli.

**Boundary expression and perception case study.** In order to investigate the correspondence between analysis, performance, and tension perception, we examined the

analyses and timing profiles of each performer and the tension response profiles by participant group for a single boundary. Event 46 was chosen for this case study due its analytical importance (58.3% performer agreement). Table 2 shows the results of this analysis.

As shown in the “Segmented Boundary” column, seven out of 12 performers segmented at event 46. To assess which harpsichordists emphasize this boundary through performance, we calculated whether one of their ten longest durations occurs at this point. Four harpsichordists (3, 4, 7, and 12) exhibit a marked *ritardando* at event 46. In total, two of the seven performers who segmented the boundary also expressed it as one of their ten longest durations. P4 and P12 articulated this boundary with a *ritardando* (one of their ten longest), but they did not segment in their formal analysis. P2, P6, and P11 did not indicate event 46 as a boundary in their analysis or performance. Next, we calculated whether the grouping is GFL-reflective, which takes into account the last six durations before event 46 using the curve-fitting approach described in Part 1. Out of the 12 performances, eight were GFL-reflective, but only four harpsichordists (P1, P3, P7, and P10) who segmented the boundary also performed group-final lengthening at this point.

Given that local tension peaks occurred at durational peaks in the unmeasured prelude, the tension profiles for each listener group were also analyzed using the curve-fitting approach outlined above in relation to the timing data. Through this method, we determined whether an increase in tension occurred in relation to the boundary (i.e., a group-final tension increase). The same procedure was employed as outlined above for the group-final lengthening of performance durations. Listeners responded with tension increases to performers 2, 3, 4, 7, 10, and 11 (tension increase by one listener group) and particularly to 5, 6, 9, and 12 (two or more listener groups) to a greater extent than to performers 1 and 8 (no listener groups). This analysis demonstrates the difficulty in directly assessing the connection between analysis, performance, and tension judgments for a particular boundary. In addition, the results reveal the complexity in attempting to model the tension response of the three listener groups regarding a particular boundary, as no clear patterns emerge.

**Associations between timing and tension.** Means were calculated from the normalized continuous rating data for each participant for each of the 12 performances. A mixed-design ANOVA tested the differences in global mean tension ratings for all 12 performers (within-subjects variable) and the listener groups

TABLE 2. Analysis, Performance, and Tension Ratings for the Boundary at Event 46.

Performer	Segmented Boundary	Top 10 longest duration	GFL-reflective	Tension increase		
				Harp	Mus	Non
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
<b>Totals</b>	7	4	8	4	7	4

Note. Shaded cells indicate that the boundary was segmented, was one of the top 10 longest durations, was GFL-reflective, and had a tension increase for that performer. Harp = harpsichordist, Mus = musician, Non = nonmusician.

TABLE 3. Regression of 1st Derivative of Tension Ratings and 1st Derivative of Tempo.

Lag (# of events) Significance	0 $p < .05$					1 $p < .05$				
Performers	2	5	6	7	9	4	5	7	8	9
$R^2$	.04	.04	.08	.12	.11	.03	.09	.04	.04	.38
$F$	6.26	5.29	11.9	16.9	17.4	4.48	13.2	6.04	5.71	82.0
df			(1,138)					(1,137)		

(between-subjects variable). The main effect of performer was statistically significant,  $F(7.32, 344) = 6.12$ ,  $p < .001$ ,  $\eta_p^2 = .12$ ,  $\varepsilon = .67$ . Post hoc analyses with a Bonferroni adjustment revealed that mean ratings of tension for P9 were significantly higher than those for performers 2, 6, 7, 10, and 12 ( $p < .001$ ).

The between-subjects factor of training was also significant,  $F(1, 47) = 965.75$ ,  $p < .001$ ,  $\eta_p^2 = .95$ . Post hoc analyses using Tukey's HSD indicated that the nonmusicians' global mean tension ratings were significantly higher than both the musicians' ratings ( $p = .001$ , mean difference = 14.7) and the harpsichordists' ratings ( $p = .009$ , mean difference = 13.6).

One hypothesis to explain the difference in tension ratings among performers is the average tempo, as measured in events per second. P9 had the highest ( $M = 0.61$ ) and P6 had one of the lowest ( $M = 0.50$ ) average tension ratings, respectively. We investigated the relationship between average tempo and average tension ratings for each performer. A Pearson's correlation revealed a strong positive association,  $r(10) = .83$ ,  $p = .001$ , with average tempo explaining 68.6% of the variation in tension ratings. The results indicate that as the average tempo increases, the average tension rating also increases.

Another hypothesis to explain the difference in tension ratings among performers is the amount of timing variability. As previously discussed, P9 had a faster and steadier tempo, particularly after event 64, where the tension ratings increased considerably (see Figure 8A). P6 had more timing variability and the tension ratings plateau throughout (see Figure 8B). To measure the variability of timing and tension variability, we calculated the derivative, which measures the rate of change. We performed a regression analysis of the first (discrete) derivative of tension ratings on first (discrete) derivative of note duration over time (140 events) with lags of 0 to 3 events (see Table 3). For the lags of 0 and 1, seven of twelve performers reached significance ( $p < .05$ ); however, the model provided low  $R^2$  values overall. The model worked best for P9, with 38% of the variance explained for a lag of 1. For the models that reached significance, the lag time was minimal, suggesting that listeners reacted fairly quickly to changes in note duration.

In Part 1, performance timings were calculated in relation to score events in order to compare across performances. One potential drawback of this approach is that it ignores ornaments. To incorporate the effects of



notated and improvised ornaments on tension ratings, we calculated the average onset density as the total number of onsets performed (including ornaments) divided by the total duration of each performance. We also determined the average event density as the number of notated score events (140) divided by the total duration of each performance. Pearson correlation coefficients were computed among the three variables of average tension, average onset density, and average event density. The results of these correlational analyses indicated that all three variables were positively correlated with one another. The bivariate correlation between average tension and average onset density was  $r(11) = .79, p = .002$ . The bivariate correlation between average tension and average event density was  $r(11) = .82, p = .001$ . Finally, the bivariate correlation between average onset density and average event density was  $r(11) = .99, p = .001$ . A partial correlation was then computed between average tension and average onset density, controlling for average event density. A negative, nonsignificant relationship was found,  $r(9) = -.41, p = .21$ . This result indicates that although there was a significant positive correlation between average tension and onset density, the relationship does not hold when the effects of ornaments on tension are isolated from those of event density.

*Listener response profiles.* In order to investigate the differences between listeners, we mapped the responses for all 12 performances on a single graph for each listener. Visual inspection revealed distinct response profiles for each listener group. As shown in Figure 10A, a typical harpsichordist listener had a wide range of tension curves; tension ratings for some performances reach one or both extremes of the tension range, suggesting that the harpsichordist listeners responded to subtle nuances of each performance. Figure 10B illustrates characteristic tension ratings for a musician listener, with a general climb to the cadence at event 127 and a pronounced drop after this point. The general shape that emerges across performances suggests a strong influence of the global structure. As shown in Figure 10C, a typical nonmusician listener reached a plateau after about 20 events, and many performances reach the upper-end of the range. These profiles help to explain the significant differences mentioned previously in global mean tension ratings between the nonmusicians and the other two listener groups, particularly the high mean tension ratings for the nonmusician listeners across all performances and the wider range between the highest and lowest mean ratings for the harpsichordists (range = 21.3) compared to the musicians (range = 10.2).

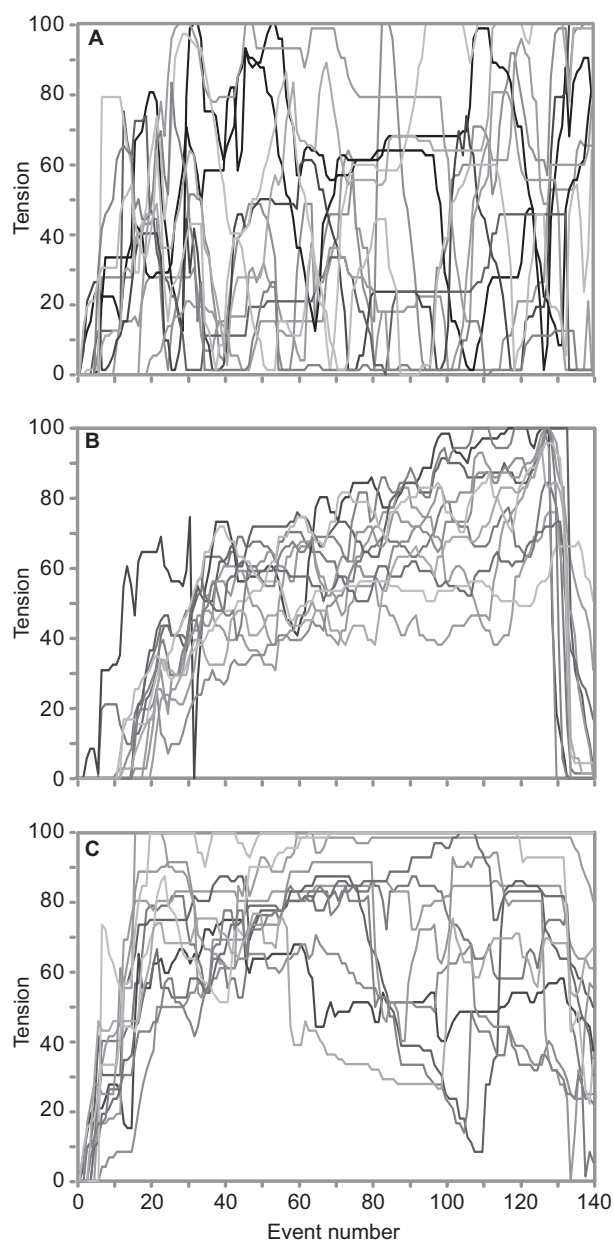


FIGURE 10. Tension ratings by typical listeners for all 12 performances: A) harpsichordist, B) musician, C) nonmusician

*Tension variability.* We investigated whether the variability in tension ratings is related to the large-scale variability in performed note durations. In the same manner as the performance data presented in Part 1, the absolute (discrete) first derivative of tension ratings was calculated for each participant by event and averaged across performers (see Figure 10). For the averaged absolute tension variability for each performer, the tension variability at the beginning and end ( $M = 0.41$ ,

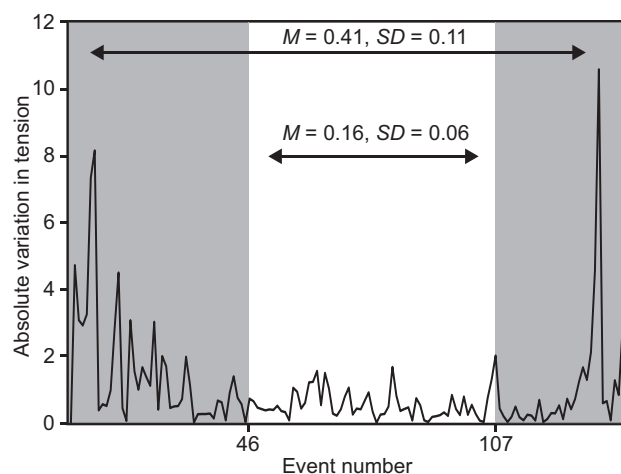


FIGURE 11. Absolute variation in tension ratings per event across all performers. Mean values of beginning and end compared to the interior section.

$SD = 0.11$ ) was significantly higher than within the interior section ( $M = 0.16$ ,  $SD = 0.06$ ),  $t(11) = 6.96$ ,  $p < .001$ ,  $d = 2.01$ .

It might be argued that the tension variability that occurs at the beginning may be exaggerated due to the constraints of the experimental procedure, which required that participants begin with the slider at the bottom. Examining Figures 8 and 11, the largest spikes in absolute tension variability at the beginning and end were not connected to the initial or final motions of the slider, but directly reflect the large spikes in tempo variability in the treatment of the rolled arpeggios into events 6 and 133. In other words, the wide range of performance interpretations of the initial and final rolled arpeggios appear to influence the variability of tension ratings at these moments. Many of the peaks in tension variability are associated with boundaries of the performers' analysis, suggesting that the group-final lengthening or acceleration at those important structural markers may influence differences in perceived tension. The lowest points of tension variability occur in the middle of a gesture or phrase, events that are less likely to be dramatically emphasized through performance, which would result in smoother or less extreme movements of the slider.

#### DISCUSSION

The tension graphs for each performer indicate a surprising difference in the experience of tension at the end of groupings in the unmeasured prelude compared to the ends of phrases in measured music; listeners

rated higher tension at phrase endings where group-final lengthening occurs. The differences in interpretation by P9 and P6 resulted in different tension profiles of the three listener groups. Ratings for P6 generally resemble a plateau throughout, whereas two large tension arches occurred for P9. The differences related to mean tempo and timing variability that were observed between P9 and P6 in Experiment 1 were reflected in the tension results of the regression analyses; a moderate association between timing variability and tension variability was found, particularly for the model of P9, and significant correlations were found between mean tempo and mean tension ratings, as well as mean onset density and mean tension ratings.

The case study at event 46 attempted to quantify differences in the performance and perception of a boundary with high analytical agreement. No clear association between analysis, performance and tension judgments was found, a result that emphasizes the lack of a one-to-one correspondence between boundary segmentation, GFL-reflectivity, and tension responses. This analysis highlights the importance of considering the entire listening context, not just localized moments. To consider this connection further, Gingras, Pearce, Goodchild, Dean, Wiggins, and McAdams (2016) employ time series analysis techniques to directly confront the chain of influence from score events to performance to tension perception for the unmeasured prelude.

Several different approaches were used to investigate the differences in tension ratings among the three listener groups. The mixed-design ANOVA revealed that the nonmusicians' mean ratings were significantly higher than those for the musicians and harpsichordists. Similarly, Frederickson (1999) reported higher ratings for less experienced listeners, which was interpreted as less discrimination of the stimulus. Visual inspection of each listener's tension ratings for all 12 performers uncovered trends in responses depending on expertise group. The general plateau for a typical nonmusician listener across different performances explains the higher mean tension ratings compared to musicians and harpsichordists. A typical harpsichordist listener produced a variety of tension curves, often reaching both extremes of the tension scale within a single performance. The wide range of mean tension ratings also suggests that harpsichordists responded more to specific performance features compared to the musicians and nonmusicians. The common shape of the response profiles across performances and the narrower range of mean tension ratings across performances suggest that musicians are strongly influenced by the inherent musical structure.

## General Discussion

This study focused on performances of an unmeasured prelude for harpsichord as an ecologically valid object of study, which allowed for the direct exploration of the connections between expressive timing, analysis, and tension perception. As predicted, analyses and performances of the unmeasured prelude were more varied than those found in measured music, with lower agreement on segmentations and lower timing correlations among performances. We expected that the interpretive freedom provided by the unmeasured structure would aid in uncovering connections between analysis and performance, particularly GFL-reflectivity of segmented groupings. In contrast to Gingras et al. (2010), who observed that the group-final lengthening of organists corresponded to the hierarchical importance of major subdivisions in their analyses, the harpsichordists did not slow down to a greater extent with group-final lengthening at the cadence, which would be situated at a higher level of the grouping hierarchy as a structural accent. One explanation is that a variety of tempo patterns were discovered; in addition to group-final lengthening, the performers also slowed down before and accelerated through the analyzed boundaries, a process we describe as group-final anticipation. Gingras et al. (2010) did not find a significant difference between the average size of the *ritardandi* of the organists who marked a section as a boundary and those who did not. Similarly, harpsichordists 9 and 6 showed more GFL-reflectivity for the boundaries of the performers' analysis compared to their own segmentations.

The results of the tension study confirm that the boundaries of the unmeasured prelude do not operate in the same way as in measured music as reported by Krumhansl (1996) and Farbood (2012). Local peaks of tension, rather than decreases, occur at boundaries with group-final lengthening. However, the association between longer durations and tension peaks does not hold on the large-scale level; the results of the analyses on the associations between timing and tension indicated that faster mean tempi correlated with higher mean tension ratings, which relates to the findings of Ilie and Thompson (2006). In addition, a moderate link was found between timing variability and tension variability.

Verbal responses after the tension study assist in interpreting these findings. One harpsichordist listener noted that "those who brought out the harmony by lingering on a dissonance or delayed before something expected created more tension." Musicians commented on a balance between slowing for certain moments and

a consistent tempo, indicating a connection between tension and performances that were "well-balanced, with good use of time and waves of intensity" and performances that displayed "a variety of tempo patterns, [tempo] rubato, flourishes, and prolonged dissonances." Several nonmusicians noted that they rated higher tension for performances that connected the tempo changes with certain musical moments, such as the "arch shape of the melody" and "faster notes at the climax." In many comments there was a subtle connection between perceived tension and listening engagement, even suggesting that they preferred the performances that they had rated higher in tension. These observations and the results of the perceptual study suggest that the tension ratings were affected by the manner in which performers are able to connect their timing variations to inherent aspects of the score, such as emphasizing certain dissonances or gestural changes, while at the same time ensuring that they maintain a swift pace. For example, it is likely that listeners did not hear a connection between musical features and many expressive timing changes by P6. There are several moments that are quite idiosyncratic, such as the acceleration through event 12 (neither group-final lengthening nor group-final anticipation) and extreme note duration peaks that do not correspond to segmentations. These expressive timing variations may have seemed more random compared to P9's emphasis of a few structural moments, for example. These divergences in interpretation, in addition to overall mean tempo, help to explain the perception of tension on local and large-scale levels.

The performances of the unmeasured prelude conveyed a sense of global structure through the amount of timing variability. The performers showed the most interpretive diversity at the beginning and end of the piece during the stable arpeggios and maintained the steadiest tempo during the more unstable, interior section. The tension data also reflected the large-scale structure, with the largest variability in tension ratings at the beginning and end, and the lowest variability in the middle section. These results suggest an intriguing connection between the inherent structure of the piece, variability in timing, and variability in tension ratings.

This study is an important addition to the research on performance and tension perception, as it reveals the complexities involved in the interplay between individual expression, performance limitations, and tension perception of different listener groups. Performers agree on aspects of the structure, but may express this understanding in different ways. Furthermore, performers have individual impressions of the music, but the gestural and harmonic structure may impose certain constraints on

performance. Our results indicate that the perception of tension for these differences in interpretation is influenced by the musical context, the underlying tempo, and the expertise of the listener.

We used a curve-fitting approach to assess the parabolic or linear fit of the performed durations at the performers' analyzed segmentations. We found that performers used group-final lengthening to a greater extent for the groupings at beginning and end of the piece compared to the interior section and the cadence. This method was also useful in assessing whether tension ratings increased at a particular boundary. In the future, the curve-fitting approach could be applied with a moving window to map the locations of the highest goodness-of-fit segments for parabolic or linear decelerations, as well as accelerations. These results could reveal further insights into strategies of timing in relation to the inherent musical structure.

Activity analysis, currently being developed by Upham and McAdams (2014), is an analytical approach that focuses on particular types of events in continuous responses (e.g., increases in tension ratings above a certain threshold). It considers the coincidence of a given activity over multiple responses to the same stimulus and statistically evaluates response coordination. In the future, we plan on employing activity analysis to assess response coordination for the three listener groups and to pinpoint statistically significant moments of local activity for increasing or decreasing tension ratings. This type of analysis would assist in disentangling tension increases related to performer-specific timing variations in contrast to those related to inherent aspects of the score (e.g., tension increases for musicians at the cadence across performances).

Ornaments, often notated several events before segmentations, pose considerable problems for interpretation, because performers often lengthen the events prior to and during the ornament but not the final event of the grouping. The perception of tempo fluctuations has been found to depend on the musical context; for example, *ritardandi* are more difficult to detect at the end of musical units, causing a warping of experienced time (Repp, 1992, 1996). Therefore, ornaments may influence the perception of tension in different contexts. Our analysis of average onset density attempted to incorporate both notated and improvised ornaments by calculating the total number of onsets in relation to the duration of the piece. The isolated effect of average onset density on average tension ratings, controlling for the average density of notated events, resulted in a negative and nonsignificant correlation. Based on this result, we speculate that the ornaments may be

perceived as a moment of stasis that embellishes and prolongs the notated event, but does not drive the tension ratings over and above the expressive timings of notated events. A systematic evaluation of ornaments, including the location of improvised ornaments in other unmeasured preludes, could provide insight into improvisatory expressive details and the resulting tension judgments.

One limitation of this study is that we did not analyze the harmonic structure or account for harmony in our analyses. Although a harmonic reduction has been proposed for the unmeasured prelude (Ijzerman, 2011), we argue that harmonic interpretations vary across performers and therefore harmony cannot be assumed to be a constant. Additionally, the perception of harmony in the unmeasured prelude is likely influenced by expressive performance parameters. In order to investigate this further, a potential study could investigate the manipulation of timing and articulation patterns, such as sustained pitches, and their effect on tension perception in isolated contexts.

The experimental procedures described in this study could be used to investigate other unmeasured preludes in order to assess how performers respond to score constraints in various contexts (e.g., gestures and cadential patterns). Additionally, the findings could be relevant for other improvisatory music and metrically ambiguous music, such as Gregorian chant, the *alap* (opening section) of North Indian raga, free jazz, and certain types of contemporary music. In particular, our curve-fitting approach to investigate grouping structure could contribute to research on segmentation in improvisation (Dean & Bailes, 2014; Dean, Bailes, & Drummond, 2014). Although there are considerable challenges studying this "beautiful but puzzling music" (Tilney, 1991, p. 1), the unmeasured prelude reveals intriguing insights into performance timing and tension perception.

### Author Note

This research was supported by a SSHRC postdoctoral fellowship to Bruno Gingras, and a grant from NSERC (RGPIN 321774-10) and a Canada Research Chair awarded to Stephen McAdams. The MIDI harpsichord used in this study was generously loaned to us by Alain Poirier, the former director of the Conservatoire National Supérieur de Musique et de Danse de Paris. We wish to thank Pierre-Yves Asselin, Bennett Smith, Julien Boissinot and Harold Kilianski (CIRMMT) for their technical assistance, McGill Sound Recording for permission to use their equipment, and Martha de

Francisco for her expertise with sound recording. Lauren Stewart provided the use of the audiometric booth at Goldsmiths (London). Many thanks to Ryan Ouckama for his assistance with data processing and Mitch Ohriner for his helpful suggestions on a previous version of the manuscript.

A preliminary analysis of the empirical data in this article was presented at the 11<sup>th</sup> International Conference on Music Perception and Cognition (ICMPC11),

Seattle, Washington in August 2010 and reported in the conference proceedings.

We wish to thank the European American Music Distributors Company for the permission to reproduce the score from (Tilney, 1991).

*Correspondence concerning this article should be addressed to* Meghan Goodchild, Schulich School of Music, 555 Sherbrooke Street West, Montreal, Quebec, H3A 1E3. E-mail: meghan.goodchild@mail.mcgill.ca

## References

- BENGTSSON, I., & GABRIELSSON, A. (1980). Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology*, 21, 257-268.
- BERRY, W. (1989). *Musical structure and performance*. New Haven, CT: Yale University Press.
- BIGAND, E., & PARNCUTT, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, 62, 237-254.
- CLARKE, E. F. (1985). Structure and expression in music performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London, UK: Academic Press.
- CLARKE, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1-26). Oxford, UK: Oxford University Press.
- CONE, E. T. (1960). Analysis today. *The Musical Quarterly*, 46(2), 172-188.
- DEAN, R. T., & BAILES, F. (2014). Cognitive processes in musical improvisation. In G. E. Lewis & B. Piekut (Eds.), *The Oxford handbook of critical improvisation studies*. Oxford, UK: Oxford University Press.
- DEAN, R. T., BAILES, F., & DRUMMOND, J. (2014). Generative structures in improvisation: Computational segmentation of keyboard performances. *Journal of New Music Research*, 43(2), 224-236.
- DELIÈGE, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, 4, 325-360.
- FARBOOD, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, 29, 387-428.
- FLÉTY, E. (2002). AtoMIC Pro: A multiple sensor acquisition device. *Proceedings of the 2002 Conference on New Interfaces for Musical Expression (NIME 2002)* (pp. 96-101). Dublin, Ireland: MediaLabEurope.
- FREDRICKSON, W. E. (1995). A comparison of perceived musical tension and aesthetic response. *Psychology of Music*, 23(1), 81-87.
- FREDRICKSON, W. E. (1997). Elementary, middle, and high school student perceptions of tension in music. *Journal of Research in Music Education*, 45(2), 626-635.
- FREDRICKSON, W. E. (1999). Effect of musical performance on perception of tension in Gustav Holst's First Suite in E-Flat. *Journal of Research in Music Education*, 47(1), 44-52.
- GABRIELSSON, A. (1987). Once again: The theme from Mozart's piano sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm, Sweden: Royal Swedish Academy of Music.
- GABRIELSSON, A. (2003). Music performance research at the millennium. *Psychology of Music*, 31(3), 221-272.
- GINGRAS, B., ASSELIN, P., & MCADAMS, S. (2013). Individuality in harpsichord performance: Disentangling performer- and piece-specific influences on interpretive choices. *Frontiers in Psychology*, 4(895), 1-16.
- GINGRAS, B., & MCADAMS, S. (2011). Improved score-performance matching using both structural and temporal information from MIDI recordings. *Journal of New Music Research*, 40(1), 43-57.
- GINGRAS, B., MCADAMS, S., & SCHUBERT, P. (2010). The performer as analyst: A case study of J. S. Bach's "Dorian" Fugue (BWV 538). In C. Utz (Ed.), *Music Theory and Interdisciplinarity - 8th Congress of the Gesellschaft für Musiktheorie Graz 2008* (pp. 305-318). Saarbrücken, Germany: Pfau-Verlag.
- GINGRAS, B., PEARCE, M. T., GOODCHILD, M., DEAN, R. T., WIGGINS, G., & MCADAMS, S. (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 594-609.
- IJZERMAN, J. (2011). Louis Couperin's Préludes non mesurés - Unmeasured? *Dutch Journal of Music Theory*, 16(2), 112-123.
- ILIE, G., & THOMPSON, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23, 319-330.
- KRUMHANS, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, 13, 401-432.
- LERDAHL, F. (1996). Calculating tonal tension. *Music Perception*, 13, 319-363.

- LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- LERDAHL, F., & KRUMHANS, C. L. (2007). Modeling tonal tension. *Music Perception*, 24, 329-366.
- LESTER, J. (1995). Performance and analysis: Interaction and interpretation. In F. Lerdahl & R. Jackendoff (Eds.), *The practice of performance: Studies in musical interpretation* (pp. 197-216). Cambridge, MA: MIT Press.
- MADSEN, C. K., & FREDRICKSON, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy*, 30(1), 46-63.
- MARTIN, F. N., & CHAMPLIN, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, 11, 64-66.
- MCADAMS, S., VINES, B. W., VIELLARD, S., SMITH, B. K., & REYNOLDS, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception*, 22, 297-350.
- MCCRELESS, P. (2009). Analysis and performance: A counterexample? *Dutch Journal of Music Theory*, 14(1), 1-16.
- MORONEY, D. (1976). The performance of unmeasured harpsichord preludes. *Early Music*, 4(2), 143-151.
- NARMOUR, E. (1988). On the relationship of analytical theory to performance and interpretations. In L. B. Meyer, E. Narmour, & R. A. Solie (Eds.), *Explorations in music, the arts, and ideas* (pp. 317-340). Stuyvesant, NY: Pendragon.
- NIELSEN, F. V. (1983). *Oplevelse af musikalsk spænding* [The experience of musical tension]. Copenhagen: Akademisk Forlag.
- OHNER, M. S. (2012). Grouping hierarchy and trajectories of pacing in performances of Chopin's Mazurkas. *Music Theory Online*, 18(1), 1-16. Retrieved from <http://mtosmt.org/issues/mto.12.18.1/mto.12.18.1.ohner.php>
- PALMER, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331-346.
- PALMER, C. (1996a). On the assignment of structure in music performance. *Music Perception*, 14, 23-56.
- PALMER, C. (1996b). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13, 433-453.
- PENEL, A., & DRAKE, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research-Psychologische Forschung*, 61(1), 12-32.
- PENEL, A., & DRAKE, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception and Psychophysics*, 66, 545-562.
- PENTTINEN, H. (2006). On the dynamics of the harpsichord and its synthesis. In V. Verfaillie (Ed.), *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)* (pp. 115-120). Montreal, Canada: ICDAE. Retrieved from [http://www.dafx.ca/proceedings/papers/p\\_115.pdf](http://www.dafx.ca/proceedings/papers/p_115.pdf)
- REPP, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America*, 88, 622-641.
- REPP, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America*, 92, 227-260.
- REPP, B. H. (1995). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98, 141-160.
- REPP, B. H. (1996). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14, 161-183.
- RINK, J. (2002). *Musical performance: A guide to understanding*. Cambridge, UK: Cambridge University Press.
- SCHUBERT, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, 21, 561-585.
- SHAFFER, L. H., CLARKE, E. F., & TODD, N. P. N. (1985). Metre and rhythm in piano playing. *Cognition*, 20(1), 61-77.
- SHAFFER, L. H., & TODD, N. P. N. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139-152). Stockholm, Sweden: Royal Swedish Academy of Music.
- SILVEY, B. A. (2011). The effects of orchestration on musicians' and nonmusicians' perception of musical tension. *Research and Issues in Music Education*, 9(1). Retrieved from <http://www.stthomas.edu/rimeonline/vol9/silveyfinal.htm>
- SMITH, B. K. (1995). An environment for psychoacoustic experimentation using the IRCAM musical workstation. In *Proceedings of the Society for Music Perception and Cognition Conference*. Berkeley, CA: University of California.
- TILNEY, C. (Ed.). (1991). *The art of the French unmeasured prelude* (Vol. 2). (H. Lerch, Trans.). London, UK: Schott.
- TODD, N. (1985). A model of expressive timing in tonal music. *Music Perception*, 33, 33-57.
- TOIVAINEN, P., & KRUMHANS, C. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32, 741-766.
- UPHAM, F., & MCADAMS, S. (2014). *Activity analysis and coordination in continuous responses to music*. Manuscript submitted for publication.