Article

Music attending to linear constituent structures in tonal music

Isabel Cecilia Martínez

Abstract

This article investigates the perception of constituent linear structures of tonal musical pieces, using a divided attention paradigm combined with a click-detection technique. Two experiments were run so as to test whether the boundary of a linear constituent appears as a focal point in the perception of musical structure. In Experiment 1, musicians and non-musicians listened to open foreground prolongations in phrases with clicks located at different points of their constituent structures. Significant differences in response times were found that depended on click position in relation to the boundary; participants were faster in detecting clicks at constituent boundaries, and slower for clicks located before boundaries, with no effect of rhythmic factors. Experiment 2 used the same experimental design to explore perception of open linear foreground prolongations, with the assumption that an effect of branching (left to right, or vice versa) could orient attention differently to the boundary region. Results were similar to those of Experiment 1. Overall, the evidence supports the idea that linear constituency is a significant feature of the perception of tonal musical structure. Dominant events become cognitive reference points to which the focus of attention is allocated, and subordinate, dependent events that are associated to the former, orient expectations of continuation and/or closure.

The influence of linguistic theory on the study of music cognition is indisputable. Syntactic descriptions of aspects of tonal organization inspired by linguistic models provide potential insights about the cognition of hierarchical structure in music, and have led to the development of complex and fruitful theories of the experience of structural aspects of musical compositions of the common-practice period (see, e.g., Lerdahl & Jackendoff, 1983; Rebuschat, Rohrmeier, Hawkins, & Cross, 2012).

The feasibility of syntactic theory as a basis for exploring hypotheses about the relationships between structural features of music and their experience in cognition is anchored in the idea that music is a bona fide exemplar of the group of self-diversifying, particulate systems (Merker, 2002). Music – like language – fits Humboldt's criteria to the extent that it shares the property of making infinite use of finite resources, and the capacity to create constituent combinations whose properties may differ from the properties of any of their individual components (Abler, 1989). Similarities and differences between music and language have been extensively explored, and the potential

for syntactic analysis – mainly conceived of as applicable to the study of Western tonal music – has been evaluated in terms of its (generative) capacity to account for the combination of musical notes into musical motives by means of systems of rules, and the combination of these into musical phrases, pieces, etc. (Igoa, 2010).

Generative syntax applied to the analysis of tonal music can account for the existence of nuclear hierarchies (Lerdahl & Jackendoff, 1983), organized according to the principles of constituency (Lerdahl & Jackendoff's grouping analysis) and dependency (their time-span analysis). And recent research in the field of neuroscience suggests that despite the noticeable differences between language and music as to their domain-specific syntactic representations,

Universidad Nacional de La Plata, Buenos Aires, Argentina

Corresponding author:
Isabel Cecilia Martínez, Universidad Nacional de La Plata, Diagonal 78 no 680 La Plata 1900, Buenos Aires, Argentina.
Email: isabelmartinez@fba.unlp.edu.ar

they share neural resources that are activated during syntactic processing (Patel, 2008, p. 297), syntax thus seeming to constitute a specific point of convergence between the two domains (Patel, 2003).

Constituency and dependency in tonal music

Overall, constituency, whether in language or in music, is intrinsically related to boundary segmentations. It appears that the earliest linguistic usage of the term "constituency" in this sense can be traced to the work of Harris, Chomsky, and their collaborators (see Harris, 1952), where it refers to relationships implicated in the syntactic description of phrase structure. So-called "immediate constituent analysis" (Chomsky, 1956, p. 116) refers to the way in which, by means of formal properties of derivation, a sentence is grouped into phrases, smaller phrases, and smaller constituent phrases until the smallest unit is reached. This analysis is predicated on the assumption that linguistic-syntactic structures are hierarchical – layered like an onion and therefore susceptible to progressive segmentation (Longacre, 1960).

Discovering the words embedded in a largely continuous speech stream is one of the infant's first tasks in language acquisition; it is likely to be related to the capacity to process grammatical features of speech so as to identify word boundaries (Saffran, Newport, & Aslin, 1996). In tonal music, boundary segmentations are likely to be structured by features such as harmonic and linear progressions playing a grammatical role as syntactic markers that convey a sense of relative ending to the constituent unit (Chiappe & Schmuckler, 1997; Deliège, 1987). The experience of boundaries in musical pieces appears to be linked to the cognitive capacity to segment musical phrases, and to the perceptual sensitivity to relative degrees of tonal closure (Marvin & Brinkman, 1999).

A typical syntactic depiction of the tonal organization in music would follow the principle of compositionality in that: (i) a musical piece is divisible into parts called constituents; (ii) there are different types or categories of parts; (iii) different parts are organized in specific ways; and (iv) each part plays a specific role in the whole structure (Horton, 1999, 2001). The meaning and function of a syntactic unit are derivable from two characteristic features: precedence and dominance. Precedence deals with the temporal nature of music, requiring that some events appear earlier than others; it displays the left-to-right ordering in the concatenation of successive constituents, and defines the linear structure of the musical piece. Dominance, on the other hand, concerns the hierarchical grouping of the constituent structures; discrete events are integrated into more comprehensive syntactical units, because some of them govern the operation of others. The exemplar case is that of the head of a constituent unit; it is assigned the property to dominate over other events (Horton, 2003). Dominance and linearity can be construed as integrated, in that constituent units (motives and phrases) are composed out based on the elaboration of

some structurally significant events by other ornamental events. The hierarchical integration of both structural features that result from this process of elaboration leads to the characteristic contrapuntal, voice-leading designs that appear in countless instances of tonal pieces.

It is a hypothesis of this article that the embedded and linear features that govern the relationships of structural components in (tonal) musical compositions provide listeners with different cues to enable sensitivity to constituent boundaries. Based on the regulation of tension in the elaboration of tonality in music, and on the concomitant expectancies that tonal resolution generates, voice-leading embodies the possibility of conveying different degrees of boundary segmentation.

Dependency and linear arrangement in music

The notion of dependency is bound to the psychological presupposition that musical events in tonal pieces are experienced in terms of structural priority. The notion of structural priority poses an epistemological problem when the cognitive status of musical structure is examined. It is not clear, for example, whether dependency is informed by some inherent property of pitch events as determinants of tonal hierarchy, or whether other structural features are at work. Although it has been found that certain tonal components – such as the identity of the scale-degree – are linked to a perceptual ranking of events, playing a significant role in tonality induction (Bharucha, 1984; Krumhansl, 1979), the notion of dependency might be best informed by the characteristic linear voice-leading contours that result from the composing-out procedures in musical elaboration (Cadwallader & Gagné, 1998). In such contextual frames, structural priority might be assumed to be more an emergent feature of the compositional nature of musical constituents than a result of some abstract inherent pitch stability. This issue was highlighted by certain music analysts who emphasized the fact that tonal stability needs to be understood in a more contextual way than might be hypothesized to hold in regard to a "static" tonal hierarchy (see for example Larson, 1997, 2012). Thus, whereas inherent stability, as a property of an event within a tonal key, appears to exhibit a rather static representation of tonal pitch, contextual stability conveys a more *dynamic picture* of the organization of a piece, in the sense that it tackles linear relationships *between* pitch events within and between constituents, according to their structural roles. In delineating notions of structural priority in music, dependency appears to be the more all-encompassing structural property in the sense that it accounts for relationships that describe *which* events elaborate *which other* events. However, the linear dimension of a piece's organization is inherently linked to other pitch dimensions, such as the harmonic domain. In other words, linear procedures do not generate by themselves awareness of hierarchic differences

between events, unless structurally more important events are somehow represented at a harmonic level. Thus, the relation between the notions of dominance and subordination (constituent head-dependent events) cannot be informed exclusively by inherent pitch properties, horizontal features, or syntactic categories of tonal pitch events; more specifically, it is associated with the harmonic identity of the piece (Horton, 2003). Hierarchy, to the extent that is linearly derived, is the result of the application of linear elaboration techniques to a basic voice-leading structure; therefore, dominance–subordination relations between events are best construed as linearly-conceived phenomena.

If the aim is to understand which aspect of pitch events' organization characterizes their syntactic status in the experience of linearly generated structures, it is necessary to create appropriate experimental conditions to test participants' sensitivities to those structural events. Linearly generated structures will be investigated under experimental situations involving attending to music. In what follows, the concept of music attending is presented and, at the end, a hypothesis of attending to temporal linearity as a syntactic feature in music is posited.

Music attending

One of the main music psychological frameworks that tackle the problem of attention to music is the theory of music attending (Boltz, 1989, 1993; Jones, 1992; Jones & Boltz, 1989). Music attending will be considered here from the point of view of its relationship to the experience of constituency and dependency in tonal music.

Music is a temporal art. How the temporal dimension affects information processing during music attending seems to be related to the event-structure that unfolds across the temporal duration of a musical piece (Brown & Boltz, 2002). If the constituent organization in music somehow orients attention, interacting with the ways in which cognitive resources are used to process the piece, it is expected that information processing will vary as a constituent unit unfolds, according to its characteristic internal organization governed by dependency constraints. In other words, the listener's sensitivity to the compositionality of a constituent unit might be an indicator of its relative relevance. A structural constituent has been assumed to consist of a complex organization in which one event (the prolonged head) is extended across the constituent unit by linear ornamental events. Interaction between feature arrangement and music perception will involve orienting attention allocation and priming the listener's expectations about the relative importance of constituent units. As a consequence of this interaction, decisions about the structural priority of events will be elicited in music attending. In so doing, attention will (i) be split across more relevant (the constituent's heads) and less relevant (the other tones of the constituent's) events, priming expectations about the propensity of the latter to extend the priority of the former;

and (ii) map those aspects of the constituent's unfolding onto appropriate control parameters for action. For example, musical information about structural priority can be used to monitor attentional effort according to the processing of the unfolding constituent. If, for example, reaction time responses were required of listeners attending to music, it would be expected that variations in their response latencies would reflect their sensitivity to the relative structural priority of the focal points around which attention is allocated.

If attention at the level of constituent structure is modelled by dependency features, it is expected that dominant events would become cognitive reference points to which the focus of attention is allocated, and subordinate, dependent events that are associated to the former would orient expectations of continuation and/or closure according to the linear constraints of the linear voice-leading elaboration. Dominance–subordinate relationships in the constituent unit might orient conscious and/or unconscious attention to the head of the constituent while subordinating the remainder of the phrasal information to the head until a new constituent head appears, requiring a re-focusing of attention allocation.

According to Jones and Boltz (1989), highly coherent events support a kind of future-oriented attention. If the linear arrangement of a constituent unit is a structural framework that provides a high degree of predictability it will orient attention in such a way as to generate expectations of how and when the constituent units are going to complete or close. Thanks to future-oriented attention, listeners will frequently anticipate not just which notes are about to come but also the temporal moment at which they are going to occur. As dynamic attending is organized around the most salient information, it will select where to allocate the focus in the moment-to-moment awareness of the ongoing musical structure. Consequently, it is expected that participants' experience and projection of structure (and hence, responses to task-demands that relate to those experiences and projections) at different locations of the constituent hierarchy will reflect such differences. The organization of events with low temporal predictability will prevent people from anticipating their future development, forcing them to attend locally to adjacent elements with the intention of organizing such non-structured information. On the other hand, future-oriented attending might support a hypothesis of constituency as a one-to-many relationship with a high level of predictability.

Assessing attention to syntax: The click technique

A group of seminal experiments in linguistics used signal-detection techniques to study linguistic information processing (Fodor & Bever, 1965; Holmes & Forster, 1970). In click-detection experiments, participants were required to detect clicks superimposed on natural language sentences at different focal points of their constituent structure, performing a simple motor task, such as pressing a key as

soon as the signal was detected. Participants' reaction times (RTs) to different click positions were assessed as measures of their sensitivity to constituency relations. Results indicated that the constituent structure of spoken sentences is represented psychologically in terms of structural proximity between events. In addition, clicks were more accurately detected and/or located at major constituent boundaries than inside the constituent unit, supporting the hypothesis that processing load is greater during the constituent unit than at the end of it. This paradigm has also been used in musical experiments. Similarities between music perception and speech processing were found when simple sequences of sounds (Gregory, 1978) or more complex musical phrases (Sloboda & Gregory, 1980; Stoffer, 1985) were used. Overall, results provide evidence of the correspondence between response latency and the structural order of the boundary where the click was located. Berent and Perfetti (1993) used the click-detection technique in an experiment that tested the online parsing of harmonic sequences. Using a divided attention paradigm, with music listening as the primary task and click detection as the secondary task, they tested differences in click-detection performance as a proxy for differences in the cognitive load required to process the sequences. Results confirmed that there was a direct relationship between increments in cognitive load and musical complexity. Moreover, increments in cognitive load were shown to be detectable in the performance of a secondary task by virtue of the participants' production of longer response latencies. Finally, the results supported the validity of the click-detection paradigm as an online method.

In summary, the use of the click technique, both in linguistics and in music, has proved to be a useful strategy to test participants' processing of constituent phrase structures, and as an online method to test the experience of constituency. However, in order to elucidate the mechanisms of attention to linear constituency in music, it is necessary to explore the ways in which priority is assessed, considering that linear voice-leading structures take the form of one-to-many relationships. In other words, attention should be conceived of as enabling listeners to properly organize relevant information, and should not be interpreted as the result of a constrained capacity to process information.

In this article, a hypothesis about constituent and linear features of musical surfaces in tonal pieces *as experienced* is proposed; results of two experiments that test online attention to those features are presented; and the relationships between syntactic hierarchy and music perception in tonal music are discussed. It is proposed that the online experience of linear constituents during music attending lies in the ability to organize the temporal unfolding of tonal structure in music. It is worth noting that, unlike most previous click-detection experiments, in the experiments that will be reported here, real, complex and multi-voiced musical fragments are used, taking account of both melodic and harmonic structure according to fundamental Schenkerian principles of voice-leading arrangement.

The assumptions

Based on evidence that (i) structural components of the input influence attentional processing (Jones, 1992; Jones & Boltz, 1989); and (ii) attentional focus does not strictly align with the temporal linearity of the input in a one-to-one relationship (Fodor & Bever, 1965; Holmes & Forster, 1970; Kaminska & Mayer, 1993; Sloboda & Gregory, 1980; Stoffer, 1985) – i.e., information about what is being heard is not continuously available but is periodically accessible, enabling the perceptual organization of a stream into integrated fragments – the following general hypotheses can be formulated:

1. Information processing is a function of the surface structure of the musical piece. Attention will vary according to the structural importance of the musical event on which the listener is focusing, and decisions about what is being heard will be related to the constituent structure of the prolongational units that comprise the piece.
2. The amount of information processed is an informative measure of the structural status of pitch events in the constituent unit. Information processing will be maximum within the constituent unit and minimum at constituent boundaries.
3. Participant's reaction time (RT) is considered a measure of the amount of information that a listener processes online.

Experiment 1: Music attending to close linear prolongations in music constituents

Hypotheses

In the present experiment, a click-detection paradigm is used with the aim of testing the following hypotheses:

1. Sensitivity to click detection will vary according to the listeners' expectations in music attending to constituent features of music.
2. Reaction time will vary according to the focal point where the click is located: RTs will be faster for clicks located at constituent boundaries, and slower for clicks located before such boundaries, within the constituent unit.

Method

Participants. Fifty-eight musicians and non-musicians volunteered to participate in the experiment. Professional musicians ($n = 31$, 16 females, 15 males; Mage $= 29$, $SD = 3.46$) had an average of 15 years' musical experience. Musicians were graduate professionals with a degree in music at the university or at the conservatoire; they had received at

least 10 years of instruction on a musical instrument; they had also received a mean of 6 years of instruction in music theory. Non-musicians ($n = 27$, 13 females, 14 males; Mage $= 24$, $SD = 2.45$) were university students recruited from social sciences courses at the Universidad Nacional de La Plata, Argentina. They had not taken music lessons, had no experience of playing musical instruments, or in music reading. All of them had normal hearing. All participants completed an informed consent procedure in compliance with the ethical requirements of the Universidad Nacional de La Plata.

Materials. Eleven musical excerpts, all of them belonging to the repertoire of tonal Western art music were used as stimuli (see Appendix A). All the excerpts had constituent phrases whose contrapuntal elaboration corresponded to close foreground prolongations (CFP). Foreground prolongations are linear elaborations of a single structural note (a triad chord note). In CFPs, both the structural note of the constituent unit (the constituent's head), and the last note (located at the constituent boundary, before the linear elaboration progresses to the following structural note), are the same. Foreground prolongations were taken from standard works on Schenkerian analysis (Forte & Gilbert, 1982; Salzer, 1982; Salzer & Schachter, 1989). The duration of each CFP selected (Mduration $= 4.028$ s; $SD = 1.310$) ensured that all CFPs used in the experiment fell within the span of the psychological present (Gabrielsson, 1993). The boundary area of each CFP was defined as the focal point established at the interonset interval (IOI) between the onset of the last CFP note and the onset of the subsequent note (the first note of the next CFP); (MIOI $= 365$ ms).

Stimulus pre-processing. Two click-locations were established for each CFP: (i) a click-location at the constituent boundary, located in the middle of the IOI; and (ii) a click-location before the CFP boundary, located 1 second before the last note of the CFP. It was hypothesized, however, that metric factors (i.e., a note's metrical salience) might affect the participants' responses at the boundary area. Thus, if more metrically salient notes were very close to the clicks' locations, they might elicit faster RTs. Therefore, an experimental condition in which the metrical positions of the notes at the prolongational boundary were changed was produced in order to monitor the influence of the metrical component in participants' attention. Two different metrical positions (MPs) were established at each CFP boundary: (i) MP1 – last note of CFP at a weak metrical position; (ii) MP2 – last note of CFP at a strong metrical position. Finally, melodies without clicks were included, in order to detect false alarms (DeWitt & Samuel, 1990; Stoffer, 1985), and to avoid serial position effects (Holmes & Forster, 1970), due to learned anticipations of click-locations in the course of the experiment. The changes introduced in the excerpts were made so that the original stylistic parameters were preserved, as far as possible. Thus, the editing of the fragments was "legitimate"

within the stylistic constraints of the periods from which the pieces were selected. It is noteworthy that, when explicitly asked, no participant reported any sense that anything in the fragments was stylistically incongruous.

Apparatus. The sample tones were generated in Cakewalk 9.0 MIDI sequencer through IOI's numerical writing. All the pitches were quantized and normalized. In all the samples, the sonority of the accompaniment was established at a lower level than the sonority level of the melody using the velocity function in the MIDI sequencer. The click was a digitally generated sine wave at a frequency of 1000 Hz with a duration of 7 ms. The listening level of the click was aligned with the listening level of the sample audio digital track.

Samples were reproduced using the timbre Piano 1 from the sound box Roland SC-55 mkII (bank 0-Roland GS capital tones). Participants listened to the sound sequence via headphones (Senheiser HD 435). Each sample was sent to both left and right channels but 80% of the signal was sent to the left channel and 20% to the right channel, resulting in the left channel level being 14 dB higher than the right channel. The click signal was emitted exclusively through the right channel to ensure clear audibility. The inclusion of left channel signal in the right channel was with the aim of avoiding aural fatigue. Participants' answers were produced using an Apple G4 keyboard with a serial connection (with a response latency of less than 30 ms), and participants' data were recorded using Soundforge 4.0 software.

Design

A mixed factorial repeated-measures design was used. Within-subjects factors were CFP (with strong and weak metrical positions at boundary areas), and click-location (CL) (before and at the CFP boundary). The between-subjects factor was musical experience (ME) (musicians and non-musicians).

The combination of variables totalled 66 stimuli: 11 CFPs with strong metrical positions and clicks located at constituent boundaries; 11 CFPs with weak metrical position and clicks located at constituent boundaries; 11 CFPs with strong–weak metrical positions and clicks located at constituent boundaries; 11 CFPs with clicks located before constituent boundaries; and 22 CFPs without clicks. Stimuli were randomized so as to form a different random order of presentation for each participant.

Procedure

The experiment began with a familiarization session in which participants listened to the 11 original samples of the test; each sample was repeated three subsequent times. Participants were required to listen to the pieces as naturally as possible, so as to become familiar with the main musical samples. The familiarization session was followed by a short warm-up session in which participants

received instructions on how to proceed with the required task. Participants rehearsed the click-detection task in the following way: first they listened to the sound of the click isolated, and then the click sound was presented superimposed on a musical stimulus. Subjects practised the click-detection procedure three times. During practice, they were asked to answer as quickly and precisely as possible, avoiding click-detection errors and key-pressing errors. Once the warm-up session was completed each participant was administered the test in two listening sessions with a 10-minute break between them. During the break, the participant filled in a musical experience questionnaire.

The experimental task employed a divided attention technique, with musical listening as the primary task and click-detection as the secondary task. The instructions provided to participants required them to run the two tasks simultaneously, as follows: (i) "listen carefully to the music, and press the M key as soon as you detect the click", and (ii) "recognize if the musical fragment you are currently listening to is the same as any of the fragments you heard at the opening session. Press Y or N accordingly".

The recognition task had two aims: to ensure that participants attended to the musical structure, and to distract their undivided attention from the click occurrence so as to control for the biases in responses reported in previous studies (Berent & Perfetti, 1993). Participants were tested individually. Overall, the test lasted 1 hour and 30 minutes per participant.

Results

Participants' responses were classified into hits, misses and false alarms. Only two false alarms and five misses were found. Given their insignificant number they were interpreted as missing values in subsequent analyses. Reaction times were calculated in milliseconds as the result of the interonset difference between the temporal location of the participant's response and the temporal location of the click; RTs at both click positions were averaged; RT mean responses were then compared. Results are informed in regard to: (i) the analysis of click-detection responses to clicks located before and at the CFPs boundaries; (ii) the analysis of click-detection responses relative to changes of metrical position at CFP boundaries; and (iii) the analysis of responses of recognition task.

Analysis of click-detection responses. An 11 (musical fragments) x 2 (click-locations) x 2 (musical experience) mixed repeated-measures ANOVA was run. Within-subjects' factors were musical fragments (MF), and click-locations (CL) (before and at the boundary); the between-subjects factor was musical experience (ME) (musicians and non-musicians). Factor MF was not significant, meaning that there were no differences in the pattern of responses between the different musical fragments. Therefore, data were collapsed to run a 2 (CL) x 2 (ME) repeated-measures ANOVA (see Figure 1). Click-location was significant
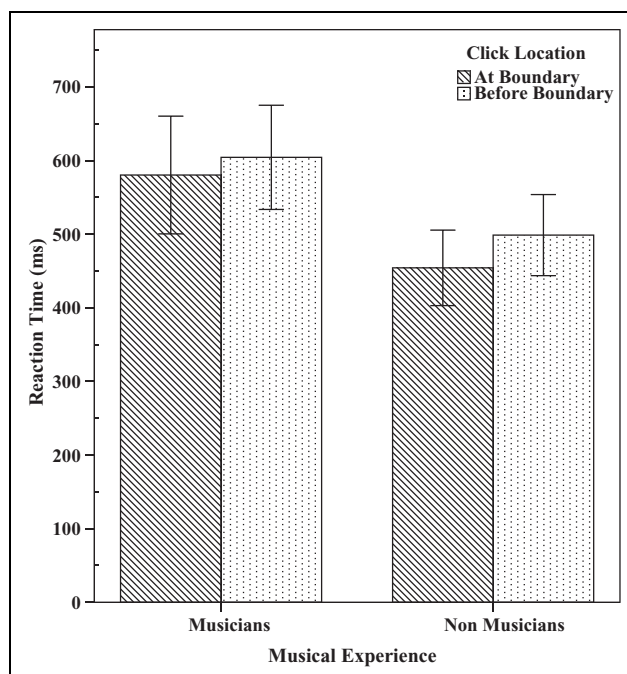
Figure 1. Experiment 1 reaction time mean responses for factors click-location (at the boundary, before the boundary) and musical experience (musicians, non-musicians).

(F[1, 56] = 30.16; $p < .001$): RTs of both musicians and non-musicians were faster in detecting the click signal at the boundary, and slower in detecting the click located before the boundary. Factor ME was also significant (F[1, 56] = 6.13; $p < .016$). Curiously, non-musicians were faster than musicians. This result will be discussed later.

Interaction between CL and ME was not significant. Despite the differences in RTs, both musicians and non-musicians showed the same pattern of behaviour at both click-locations: clicks at boundary locations always elicited a quicker response, independently of the musical example.

Analysis of responses when metrical position changes at the CFP boundary. To study the effect of the change in the metrical position at CFPs boundaries, only those samples with clicks at boundary locations and a change of metrical position were analysed. The musical samples had the last note of the CFP unit either at weak (as in the original musical piece) or at strong (modified version) metrical position. The mean RTs at both metrical positions were compared. A repeated-measures ANOVA did not show significant differences. Therefore, the metrical positions at CFPs boundaries – stressed or unstressed – had no effect on the participants' sensitivity to clicks located at that focal point.

Analysis of the recognition task. The accuracy of melody recognition between musicians and non-musicians was compared. Correct and incorrect responses were counted (Table 1). As expected, musicians were more accurate than non-musicians ($\chi^2 = 296.06$; $p < .001$).

Table 1. Experiment 1 melody recognition accuracy: Comparison between musicians and non-musicians.

Responses	Musicians	Non-musicians	Total
Correct	2,100	1,287	3,387
Incorrect	628	1,089	1,717
Total	2,728	2,376	5,104

Discussion

The aim of Experiment 1 was to estimate the participants' sensitivity to click occurrences at boundary locations of linear voice-leading arrangements to see (i) whether the boundary appears as an attentional focal point in the music's linear arrangement; and (ii) whether it was possible to ascribe to the experience of music constituency the notion of structural priority in connection with dependency criteria linked to the composing out of linear processes. Behavioural evidence of the listener's sensitivity to CFPs boundaries during music attending was found. Reaction times were significantly quicker in responses to clicks located at CFPs boundaries.

Results contribute to supporting the psychological evidence of the CFP boundary as a focal point in music attending. The way in which participants react when they are processing musical information at different focal points of a constituent unit agrees with the idea of CFP *as a constituent percept*. Reaction times also accounted for a sense of fulfilment experienced at the focal point of the constituent boundary, which confirms previous evidence concerning the tendency of structural units to preserve their integrity by resisting interruptions (Fodor & Bever, 1965). Closed foreground prolongations, as far as they conform to dependency criteria, can be understood as contrapuntal surface descriptions of the constituent function in tonal musical structure.

The click-detection technique has proven to be a useful tool to test sensitivity to constituency in music. Despite the similarity demonstrated by musicians and non-musicians in the processing of the two focal points at the constituent units, differences in their patterns of behaviour were found in other respects. Musicians were both more accurate and slower than non-musicians. A more precise recognition activity entails paying more attention to the surface details while listening to the unfolding of the piece; this, in turn, makes the processing load heavier. The qualitative difference in the participants' aural analysis might account for the delays found in their RTs. Nevertheless, the resulting association between recognition accuracy and processing speed requires further investigation.

Evidence of the incidence of metrical factors in RT was *not* found. This result supports theoretical interpretations that consider metrical and linear aspects of musical structure to be – to some extent – independent components (Schachter, 1999). The evidence collected here indicates that, during short-term music attending, participants seem to focus more on the continuity of the linear aspects of pitch organization (Schenker, [1935] 1979) than on a parsing procedure based uniquely on attention to the time-span organization of rhythmic-metric information (Lerdahl & Jackendoff, 1983, ch. 7).

Experiment 2: Music attending to open linear prolongations in music constituents

Experiment 2 aimed to test listeners' sensitivity during music attending to voice-leading arrangements that presented *open* prolongations (Larson, 1997; Lerdahl, 1997). It was assumed that results of Experiment 1 – using only closed foreground prolongations – might be due to a strict compliance with a left-to-right ordering in the concatenation of the dominant-linear constituent component, leaving out of consideration other linear arrangements such as open prolongations.

Unlike closed linear arrangements, open foreground prolongations present a voice-leading unfolding of the fundamental chord notes, using suffix and/or affix melodic prolongations, that is, a particular type of voice-leading arrangement in which a structural tone of the fundamental line is elaborated through embellishment tones that are placed before and/or after the structural tone (see Cadwallader & Gagné, 1998; Forte & Gilbert, 1982; Larson, 1997).

The experiment was designed to test a possible constituent function of these open prolongations during an aural perception activity that involved close attention to music. A constituent function, analogous to that of syntactic units in speech, was also assigned to these hierarchical units. If similar results were found in music attending to open prolongations, the structural priority assigned to the constituent head in musical experience could be extended so as to encompass more complexity in musical processing.

The assumptions

The assumptions were similar to those of Experiment 1. Thus, (i) information processing is a function of the hierarchical structure of the musical piece; (ii) information processing is maximum within the constituent unit and minimum at the unit's boundary; and (iii) information processing may be assessed in terms of RT, given that attention varies according to the structural importance of the focal point on which the click is placed.

Hypothesis

It was hypothesized that RT would vary according to the focal point where the click was located: RTs for clicks placed within a boundary between two successive units would be faster than for those placed one second before the boundary. As there was no effect of the salience of metrical units on the results of Experiment 1, rhythmical salience was ignored as a variable.

Method

Participants. Sixty-one participants volunteered to take part in the experiment. Thirty were non-musicians ($n = 30$, 18 males, 16 females; Mage $= 23.4$, $SD = 2.83$), and 31 were musicians ($n = 31$, 15 males, 16 females; Mage $= 29.7$, $SD = 2.45$). They were all resident in the Buenos Aires area. Musicians were graduate professionals with a degree in music at the university or at the conservatoire; they had received at least 10 years of instruction on a musical instrument; they had also received a mean of 6 years of instruction in music theory. Non-musicians were university students recruited from social sciences courses at the Universidad Nacional de La Plata, Argentina. They had not taken music lessons, had no experience of playing musical instruments, or in music reading. All of them had normal hearing. All participants completed an informed consent procedure in compliance with the ethical requirements of the Universidad Nacional de La Plata.

Materials. Eleven musical fragments belonging to the repertoire of tonal Western art music were used. All the excerpts had constituent phrases whose contrapuntal elaboration corresponded to voice-leading unfoldings of the notes of the fundamental line using prefix and/or suffix open foreground prolongations (OFP) (see Appendix B). The OFPs were taken from standard works on Schenkerian analysis (Forte & Gilbert, 1982; Salzer, 1982; Salzer & Schachter, 1989). Given the higher complexity of the voice-leading elaborations, the durations of the OFPs used in Experiment 2 were greater than those of the CFPs used in Experiment 1 (Mduration $= 12.28$ secs; $SD = 4.74$). The boundary area of each OFP was defined as the focal point established at the interonset interval (IOI) between the onset of the last OFP note and the onset of the subsequent note (MIOI $= 365$ ms). Two click-locations were established for each OFP: (i) a click-location at the constituent boundary, located in the middle of the IOI; and (ii) a click-location before the OFP boundary, located 1 second before the last note of the OFP.

Apparatus and stimulus pre-processing. The MIDI musical sequences were manually generated in Reaper v.05. In order to enhance the ecological validity of the stimuli, in all the samples the sonority of the accompaniment was established at a lower level than the sonority level of the melody using the velocity function in the MIDI sequencer. The musical fragments were reproduced using the sample instrument Steinway Real Piano in Kontakt 5. The click was a digitally generated sine wave at a frequency of 1000 Hz with a duration of 7 ms. The listening level of the click was aligned to the listening level of the sample audio digital track. Participants used M70x Pro headphones to listen to the musical samples. Musical samples were sent to both left and right channels, with 75% of the total signal in the right channel, and 25% in the left. The click signal was emitted exclusively through the right channel to ensure clear audibility. This was intended to avoid audio fatigue affecting the participants'

responses. Data from the participants' answers were obtained as in Experiment 1, using an Apple G4 keyboard with a serial connection (with a response latency of less than 30 ms), and recorded using Reaper v.5 software.

Procedure

The procedure was the same as the one followed in Experiment 1. During the familiarization session participants listened to the 11 original samples of the test. Five "dummy" extracts were inserted so that even if musicians recognized all the pieces, since they would not hear them all in the latter stage of the experiment this might serve to prevent inattention and ensure that their concentration level remained high. Each melodic fragment was played three times without clicks. The participants were instructed to listen carefully to the music in order to internalize it.

The warm-up session was identical to that of Experiment 1. Following the warm-up session, each participant was administered the test in two listening sessions with a 10-minute break between them. During the break, each participant filled in a musical experience questionnaire.

Again, the experimental task employed a divided attention technique, with musical listening as the primary task and click-detection as the secondary task. Instructions provided to participants were the same as in Experiment 1. Participants were required to perform two tasks simultaneously: (i) to hear the click, and press a button on a keyboard as quickly as possible after hearing it; and (ii) to determine whether they heard the extract in the initial familiarization session, pressing the corresponding button. This choice was made during a gap of 5 s between each playing. As before, the recognition task had two aims: to ensure that participants attended to the musical structure, and to distract their undivided attention from the click occurrence, as this would mean that their reaction time responses would be biased for attending uniquely to the signal. Participants were tested individually. Overall, the test lasted 1 hour and 20 minutes per participant.

Experimental design

A mixed factorial repeated-measures design was used. Within-subjects factors were open foreground prolongation (OFP), and click-location (CL) (before and at the OFP boundary). The between-subjects factor was musical experience (ME) (musicians and non-musicians).

The combination of variables totalled 33 stimuli: 11 OFPs with clicks located at constituent boundaries; 11 OFPs with clicks located before constituent boundaries; and 11 OFPs without clicks. Stimuli were randomized so as to form a different random order of presentation for each participant.

Results

Analysis of click-detection responses. Participants' responses were classified into hits, misses and false alarms. Only two

false alarms and one miss were found, hence they were interpreted as missing values in subsequent analyses. Reaction times (RTs) were calculated in milliseconds as the result of the interonset difference between the temporal location of the participant's response and the temporal location of the click. Reaction times at both click positions were averaged; RT mean responses were then compared.

An 11 (musical fragments) x 2 (click-locations) x 2 (musical experience) mixed repeated-measures ANOVA was run. Within-subjects factors were musical fragments (MF), and click-locations (CL) (before and at the boundary); the between-subjects factor was musical experience (ME) (musicians and non-musicians). Factor MF was not significant, showing no differences in the pattern of responses between the different musical fragments. Therefore, data were collapsed to run a 2 (CL) x 2 (ME) mixed repeated-measures ANOVA (see Figure 2). Click-location was significant (F[1, 59] = 25.03; p < .001): RTs of both musicians and non-musicians were faster in detecting clicks located at the boundaries, and slower in detecting clicks located before the boundaries. Factor ME was also significant (F[1, 59] = 4.57; p < .03). Again, non-musicians were faster than musicians, as in Experiment 1.

Interaction between CL and ME was not significant. Despite the differences in RTs, both musicians and non-musicians showed the same pattern of behaviour at both click-locations, indicating that clicks at the boundary locations always elicited a quicker response, independent of the musical examples.

Analysis of the recognition task. In the anaylsis of the recognition task, the accuracy of melody recognition between musicians and non-musicians was compared. Procedure and results were similar to those of Experiment 1 (Table 2). As expected, musicians were more accurate than non-musicians (χ^2 = 83.60; p < .001).

Discussion

The aim of Experiment 2 was to estimate whether the boundaries of voice-leading arrangements – linked to composing-out processes at the structural level of open foreground prolongations – appeared as attentional focal points to participants during music perception. Results contribute to supporting the psychological evidence of the OFP boundary as a focal point in music attending. Results also confirm that the structural priority assigned to the constituent head in the linear arrangement can be extended so as to encompass more complexity in music attending.

Reaction times were significantly faster in responses to clicks located at OFPs boundaries. The way participants react when they are processing musical information at different focal points of a constituent unit agrees with the idea of the OFP as a constituent percept. Reaction times also accounted for a sense of fulfilment experienced at the focal point of the constituent boundary.
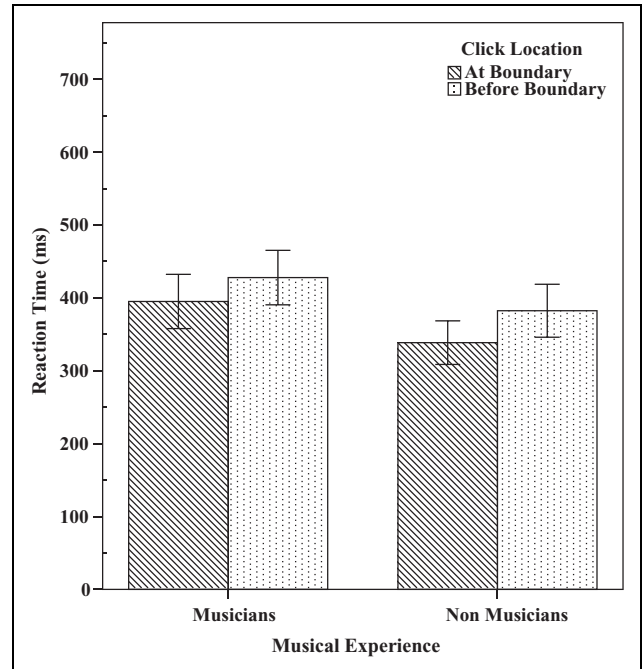
Figure 2. Experiment 2 reaction time mean responses for factors click-location (at the boundary, before the boundary) and musical experience (musicians, non-musicians).

Table 2. Experiment 2 melody recognition accuracy: Comparison between musicians and non-musicians.

Responses	Musicians	Non-musicians	Total
Correct	856	686	1,542
Incorrect	157	268	425
Total	1,013	954	1,967

Results provided further behavioural evidence of the listeners' sensitivity to the structural linearity of music constituents. Musicians and non-musicians behaved in a similar way when they experienced constituency in OFPs in connection to composing-out linear processes that involve the use of both structural priority and dependency criteria. Open foreground prolongations, in so far as they fit the above-mentioned criteria, can be understood as contrapuntal surface descriptions of the constituent function in tonal musical structure.

Musicians were both more accurate and slower than non-musicians, as in Experiment 1. It is likely (as was already posited) that the qualitative difference in the participants' aural analysis, and the more precise recognition activity that makes the processing load heavier, might account for the delays found in their RTs. As before, the resulting association between recognition accuracy and processing speed requires further investigation. Nevertheless, participants' behaviour is consistent with the findings of many experiments which have not shown significant differences between musicians and non-musicians when

they are required to perform tasks linked to this kind of basic musical knowledge (Bigand, 1990, 1994).

Finally, click-detection methodology has proven to be a useful tool to test sensitivity to constituency in music.

General discussion

The present research aimed to investigate the relationships between linear and constituent features of tonal musical pieces *as experienced*. Based on previous evidence indicating that, during music attending, information about what is being heard enables the perceptual organization of a stream into integrated structural components, it was assumed that (i) if information processing is a function of the linear constituent structure of a musical phrase, attention will vary according to the structural importance of the musical event on which the listener is focusing; (ii) attention will be maximum within the constituent unit, and minimum at constituent boundaries; and (iii) a participant's reaction time will be a reliable measure of the amount of information that a listener processes online. In Experiment 1, constituent phrases containing closed foreground prolongations were used, and the participants' sensitivity to clicks located before and at the constituent boundaries was measured. In addition, the metrical component of the linear arrangement was manipulated in order to control for potential rhythmic effects on music attending to linear constituency. In Experiment 2, similar methods were used to test participants' sensitivity to constituent phrases containing open foreground prolongations, that is, more complex linear voice-leading elaborations. The need to obtain behavioural evidence of the participants' ability to organize the temporal unfolding of tonal structure in music was central to the purposes of both experiments. It is worth noting that, unlike most previous click-detection studies, in the experiments reported here, real, complex and multi-voiced musical fragments were used, taking account of both melodic and harmonic structure according to fundamental Schenkerian principles of voice-leading arrangement. Behavioural evidence of the listeners' sensitivity to CFPs, and to OFP boundaries during music attending was found. Reaction times were significantly quicker in responses to clicks located at boundaries of both linear constituent units. Results contribute to supporting the psychological evidence of linear voice-leading arrangements as constituent focal points in music attending.

In this article, an effect of prolongational structure rather than metrical structure on the experience of constituent structure in common-practice period music was found. When the metrical component of the prolongational boundary was manipulated by altering the metrical position (and thus metrical stress or weight) of the boundary note, the experimental isolation of tonal and metrical factors by means of the manipulation of the note's metrical position at the boundary location showed that metrical factors had no signficant effect on RTs. Although substitutions of the constituent's events might produce alterations in certain aspects of the constituent's intrinsic organization and may elicit, as a consequence, variations in the listener's response, the fact that the change in the metrical position at the boundary location did not affect RTs supports the assumption that listeners focus their attention more on the tonal pitch characteristics rather than on the metrical components of the constituent organization when they are processing the voice-leading component of musical structure. This apparent independence of pitch and metrical components that occurs when prolongational aspects of the unfolding tonal structure are mentally processed supports the hypotheses of several musicological theories (Forte & Gilbert, 1982; Salzer, 1982; Salzer & Schachter, 1989; Schenker [1935] 1979) about the primacy of the pitch component in the phenomenal experience of the underlying linearity of tonal structure.

Music attending seems, notwithstanding, to be related to time estimation, that is, to the way in which temporal intervals are phenomenologically filled in (Jones & Boltz, 1989). According to the results of Experiments 1 and 2, the click-detection task did not interfere with the online sense of temporal continuity of the voice-leading unfolding. In tonal music, the unfolding of a sequence of pitch events in the course of a given constituent generates expectations about the events' continuation at different focal points, and a relative sense of fulfilment at the boundary. As long as music attending took place, the boundary of a prolongational unit naturally appeared to the perceiver as a focal point at which a relative sense of completion was reached. Reaction times accounted for the participants' sensitivity to such structural features.

According to Epstein (1995) the sense of time in music seems to be dual: on the one hand, it is related to the chronometric time experience; on the other, it seems to be integral to the intrinsic configuration of a particular piece of music. Both temporal dimensions shape musical time, and are assumed to be available when music is experienced. It is at the core of voice-leading theory that the temporal unfolding of the linear structure conveys meaning in a narrative sense, being more akin to Epstein's postulated second dimension of the sense of time.

In tonal music, arrangements of musical phrases must be conceived of as having beginnings and endings. A hypothesis concerning the relationships between constituents and linear arrangements can be stated as long as the latter satisfy dependency criteria. But in so doing, specific conditions need to be set up in order to assign to the unfolding of the voice-leading a propensity to activate a constituent parsing mode. A study on the effects of phrasing in melody recognition tested the function of the boundary of musical sequences in guiding melodic parsing during perception. As in previous psycholinguistic studies, and also in line with the results reported here, Chiappe and Schmuckler (1997) showed that the surface phrase plays a distinct functional role in structuring memory for musical passages. The authors claim to have extended previous studies by using

"naturalistic" melodic fragments taken from Schubert songs as stimuli, instead of using artificial melodies. However, two potentially problematic issues can be identified in Chiappe and Schmuckler's study, concerning (i) the scope of the musical meaning emerging from the propensity of the unfolding voice-leading to activate a constituent-dependent parsing mode, and (ii) the ecological validity of the materials and task involved. In the first place, musical sequences were presented unaccompanied, that is, without harmonic texture, imposing a higher degree of ambiguity on the inferrable voice-leadings than would have been the case were the melodies to have been presented in harmonized form. In the second place, the experimental task employed a probe technique that involved participants' production of goodness-of-fit judgements based on comparisons of the original melodies with small melodic motives that – in spite of having been extracted from the original melodies – appeared here absent of the musical meaning provided by the context where they were located in the original phrase. On the contrary, in the two experiments reported in this article, we were able to test real instances of the voice-leading construction (closed and open foreground prolongations) using musical materials and tasks that preserved the "natural" context involved in the online listening experience of music.

Theoretical descriptions of the continuous linearity that results from voice-leading unfoldings have generally neglected consideration of the determinants of constituent structure, relegating these to the domain of musical form. However, it is unlikely that, in linear voice-leading structures, constituent boundaries will necessarily be delimited by the same structural markers that appear in formal phrases. For example, variations in the arrangement of the linear events might produce alterations in certain structural aspects of the constituent's intrinsic organization and may elicit, as a consequence, variations in the listener's attending responses. Moreover, it is not clear to what extent those alterations are driven either by the pitch or by the time structure of the constituent phrase, or by a combination of both. Prince and colleagues ran a series of experiments with the aim of exploring the extent to which pitch and time are integrated or independent dimensions in perceptual and cognitive processing (Prince, Thompson, & Schmuckler, 2009). They investigated the effect of temporal variation in perception of pitch structure using the probe tone technique in perceptual and cognitive tasks so as to explore the way structural attributes of tonality and metre affect pitch-time relations. The experiments involved goodnes-of-fit comparisons between tonal contexts and probe tones that varied in pitch class and temporal position; they also used temporal and pitch speed classifications tasks. Overall, they found an asymmetry in music attention, with pitch being the salient attribute that shaped the perception and cognition of time. Employing a variant of their previous experiments, Prince, Schmuckler, and Thompson (2009) tested participants in perceptual and cognitive tasks that required

comparisons between tones preceding and following a tonal context. The experimental design involved variation of the context of judgement (tonal versus atonal) and also the temporal location of the targets tones. Findings were similar to the previous studies, indicating that, in tonal music perception, the focus seems to be on pitch structure while the temporal predictability of the target tone influences detection of changes in the target. Prince (2011; 2014) tested participants' selective attention to changes in the dimensional structure of pitch and time: using melodies in which the order of the pitches was randomly permuted, participants (musicians and non-musicians) were required either to ignore the pitch or the time component and, conversely, to attend to both at the same time. Similar patterns of responses were found in both groups, presumably due to pitch or time selective attention, and also in respect of pitch-time integration when attending simultaneously to both dimensions. Nevertheless, even in the selective attention task, the irrelevant dimension influenced participants' responses. Prince and Pfordresher (2012), in turn, tested pitch and time dimensions using perception and performance tasks. They found similar results, supporting a hypothesis of pitch salience bound to the tonal quality of the pitch sequences. In all the experiments reported, the researchers kept in mind the hypothesis that Western listeners had developed critical sensitivities to tonal pitch structures through constant and repeated exposure to tonal musical pieces as the most parsimonious explanation for the dimensional perceptual asymmetry. Given that the question of whether pitch was more salient than time was pervasive in those studies, Prince and Schmuckler (2014) examined the alignment of pitch and time information in the music of the common-practice period, analysing the frequency of occurrence of the scale degrees, and the temporal positions of the notes of 365 representative pieces of tonal music. These analyses were correlated to perceptual measures of tonal (Krumhansl & Kessler, 1982) and metric (Palmer & Krumhansl, 1990) hierarchies. They found that tone distributions correlated positively with the tonal and metric hierarchies as tested by Krumhansl and Kessler, and Palmer and Krumhansl, respectively. However, Prince and Schmuckler's (2014) results also noted that tonal hierarchies varied across levels of metric stability, showing a theoretical discrepancy with Prince and colleagues' previous experimental findings.

It is self-evident that composers of Western tonal music used pitch classes and temporal positions coherently in their compositions. But, as stated elsewhere (Martínez, 2008), although it is apparent that in tonal music both the frequency of occurrence of events – involving statistical counting of pitch frequency and their duration – is perhaps the key factor that facilitates hierarchical processing, the way in which pitch and rhythm operate relies more on the dynamics of music temporal unfolding than on more static hierarchical representation of tonality and metre. The distinction between the tonal hierarchy and the event

hierarchy (Bharucha, 1984) must be recognized; the latter is elicited in response to the frequency of occurrence and temporal positioning of specific pitch events of a specific musical composition, and it contains information relative to the function of each pitch event as it appears linearly elaborated – that is, embedded in the constituent-dependent arrangement – throughout the piece. One is inclined to think, then, that the knowledge emergent from the internalization of an underlying structure implies more than implicit knowledge in the form of an invariant statistical distribution of pitch-class tonal stability. Tonal and event hierarchies are psychological principles that provide a necessary but not sufficient explanation of the problem of how the underlying tonal unfolding is cognitively experienced. As to constituency and dependency as anchors of the voice-leading relationships, our results suggest that dependency relations would perform significant roles in contextual linear arrangements in which structural priority emerges more as a result of the compositionality of the tonal organization than because of some abstract inherent stability of the pitch events.

Previous research has helped to elucidate some of the ways in which pitch-time dimensions combine in music perception. However, that research has tended to adopt a very reduced psychological approach to the idea of integration of pitch and rhythm in music perception – indeed, treating pitch and rhythm as intrinsically separate dimensions rather than as dimensions that can be decomposed out of the complex musical stimulus. In most of the research, only artificial melodies are used as experimental materials, with a consequent deficit in ecological validity. The type of constituency-dependency approach that was adopted in the experiments that were reported here, together with the tasks undertaken by the participants, reflect, on the one hand, the complexities of tonal musical structure, and preserve, on the other, naturalistic listening conditions.

Coming back to the parallels between music and language noted at the outset of this article, it is interesting to consider what neural processes might be implicated in the segmentation process both in language and in music. Recent work by Knösche and colleagues (2005) on boundary segmentation suggests that patterns of neural activation do not reflect the detection of boundaries themselves, but instead account for the operation of processes that are needed to guide the ongoing focus of attention during parsing *between* phrases. This can be a clue as to the ways attention to linearity can provide a sense of completion during parsing at boundary regions. According to Neuhaus, Knösche, and Friederici (2006) musicians and non-musicians differ when they process the acoustic musical signal, suggesting different bottom-up perceptual strategies (musicians' attention to tonal-harmonic structure versus non-musicians' sensitivity to discontinuity in the melodic input) in the processing of boundary markers in music. In our experiments, we did not find differences between these groups in low-level attending processes. Glushko,

Steinhauer, DePriest, and Koelsch (2016) studied further correspondences in the neural cognitive phrasing mechanisms in both language and music, finding strong resemblances at the onset of the boundary area in both domains, and reporting also that musical expertise improves the processing of prosodic phrases in language. These results increase our knowledge of the neurophysiological correlates of phrasing in music, but still do not provide a comprehensive answer to the question of whether these correlates mirror those of prosodic phrasing in language. Notwithstanding the potential interest of this research in terms of thinking about the neural processes that might be implicated in the segmentation process, however, again, the stimuli that tended to be used in these latter studies perhaps lack sufficient sophistication to be capable of providing clues as to the neural correlates of the experience of linearity in music, in terms of the complexity of constituency-dependency of the voice-leading compositional arrangements of musical pieces.

Overall, theories that have emerged from cognitive science and linguistics understand reductional representations as the result of processes deployed by an ideal competent listener. However, few specifics are provided about the ways in which those processes are instantiated in the course of music's temporal unfolding. In the experiments reported here the emphasis was placed in testing sensitivity to linearity in online listening contexts. Music unfolding prompts expectations about incoming information, and attention is therefore related to the way temporal intervals are filled in as the musical piece progresses. Click-detection proved useful in fulfiling this goal; online detection of extraneous signals located in the constituent unit, and the immediate production of mechanical responses seem not to interrupt the ongoing listening process, simultaneously providing indirect evidence of the cognitive status of different events in the musical organization. Differences between musicians and non-musicians were not found in respect of sensitivity to linear structures, confirming previous findings about commonalities in music cognition when the cognitive processes underpinning sensitivity to tonal-harmonic relations of listeners with different degrees of musical expertise are investigated at a fundamental level of cognitive activity.

Declaration of conflicting interests

Funding

Peer review

Frank Russo, Ryerson University, Psychology.
Michael Thorpe, University of Roehampton, Psychology.

References

Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, *12*, 1–13.

Berent, I., & Perfetti, Ch. A. (1993). An on-line method in studying music parsing. *Cognition*, *46*, 203–222.

Bharucha, J. (1984). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. *Journal of Experimental Psychology: General*, *113*(3), 421–425.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music*, *18*, 45–59.

Bigand, E. (1994). Contributions de la musique aux recherches sur la cognition auditive humain. In S. Mc Adams & E. Bigand (Eds.), *Penser les sons: Psychologie cognitive de l'audition* (pp. 249–298) Paris, France: Presses Universitaires de France.

Boltz, M. (1989). Perceiving the end: Effects of tonal relationships on melodic completion. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 749–761.

Boltz, M. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, *53*(6), 585–600.

Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 600–615.

Cadwallader, A., & Gagné, D. (1998). *Analysis of tonal music: A Schenkerian approach*. New York, NY: Oxford University.

Chiappe, P., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review*, *4*(2), 254–259.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113–124.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, *4*(4), 325–359.

DeWitt, L., & Samuel, A. (1990). The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology*, *119*(2), 123–144.

Epstein, D. (1995). *Shaping time: Music, the brain, and performance*. New York, NY: Schirmer.

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*, 414–420.

Forte, A., & Gilbert, S. E. (1982). *Introduction to Schenkerian analysis*. New York, NY: W. W. Norton & Company, Inc.

Gabrielsson, A. (1993). The complexities of rhythm. In T. J. Tighe & W. J. Dowling (Eds.), *Psychology and music: The understanding of melody and rhythm* (pp. 93–120). Hillsdale, NJ: Lawrence Erlbaum Associates.

Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *PLoS ONE*, *11*(5), e0155300.

Gregory, A. H. (1978). Perception of clicks in music. *Perception & Psychophysics*, *24*, 171–174.

Harris, Z. S. (1952). Discourse analysis. *Language*, *28*(1), 1–30.

Holmes, V. M., & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception & Psychophysics*, *7*, 297–301.

Horton, T. (1999). *Some formal problems with schenkerian representations of tonal structure*. (Unpublished MPhil thesis). University of Cambridge, UK.

Horton, T. (2001). The compositionality of tonal structures: A generative approach to the notion of musical meaning. *Musicae Scientiae*, *5*(2), 131–159.

Horton, T. (2003). *The formal structure of tonal theory* (Unpublished doctoral thesis). Faculty of Music, University of Cambridge, UK.

Igoa, J. M. (2010). Sobre las relaciones entre la msica y lenguaje. *Epistemus*, *1*, 97–125.

Jones, M. R. (1992). Attending to musical events. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 91–110). Washington, DC: American Psychological Association.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Kaminska, Z., & Mayer, P. (1993). Transformation, migration and restoration: Shades of illusion in the perception of music. *Contemporary Music Review*, *9*(1–2), 151–161.

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.

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

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368.

Larson, S. (1997). The problem of prolongation in tonal music: Terminology, perception, and expressive meaning. *Journal of Music Theory*, *41*(1), 101–136.

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Bloomington, IN: Indiana University.

Lerdahl, F. (1997). Issues in prolongational theory: A response to Larson. *Journal of Music Theory*, *41*(1), 141–155.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Longacre, R. E. (1960). *String constituent analysis.* Language, *36(1) Part 1*, 63–88.

Martínez, I. C. (2008). *The cognitive reality of prolongational structures in tonal music* (Doctoral dissertation). Retrieved from https://pure.roehampton.ac.uk/portal/en/studentthesis/search.html

Marvin, E. W., & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception*, *16*(4), 389–407.

Merker, B. (2002). Music: The missing Humboldt system. *Musicae Scientiae*, *1*(1), 3–21.

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.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 728–741.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University.

Prince, J. B. (2011). The integration of stimulus dimensions in the perception of music. *The Quarterly Journal of Experimental Psychology*, *64*(11), 2125–2152.

Prince, J. B. (2014). Pitch structure, but not selective attention, affects accent weightings in metrical grouping. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2073–2090.

Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, *141*(2), 184–198.

Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception*, *31*(3), 254–270.

Prince, J., Schmuckler, M., & Thompson, W. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition*, *37*(3), 368–381.

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1598–1617.

Rebuschat, P., Rohrmeier, M., Hawkins, J. A., & Cross, I. (Eds.). (2012). *Language and music as cognitive systems*. Oxford, England: Oxford University Press.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Salzer, F. (1982). *Structural hearing: Tonal coherence in music.* New York, NY: Dover.

Salzer, F., & Schachter, C. (1989). *Counterpoint in composition*. New York, NY: Columbia University.

Schachter, C. (1999). Durational reduction. In J. N. Strauss (Ed.), *Unfoldings: Essays in Schenkerian theory and analysis* (pp. 54–78). New York, NY: Oxford University Press.

Schenker, H. ([1935] 1979). *Free composition* [Der freie Satz] (E. Oster, Trans.). New York, NY: Schirmer.

Sloboda, J. A., & Gregory, A. H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology*, *34*, 274–280.

Stoffer, T. (1985). Representation of phrase structure in the perception of music. *Music Perception*, *3*(2), 191–220.

Appendix A

Examples A1 to A11: The samples represent closed foreground prolongations (CFP) extracted from the whole musical fragments that were used in the test. In all the samples, (a) indicates click-locations at the prolongational boundary, and last CFP note at weak metrical position; (b) indicates click-locations before the prolongational boundary and last CFP note at weak metrical position; and (c) indicates click-locations at the prolongational boundary and last CFP note at strong metrical position. (In Experiment 1, participants listened to only 1 click per musical fragment in each trial.)

Example A.1. Mozart, Fantasy in D, KV 397, Adagio, mm 12–13.
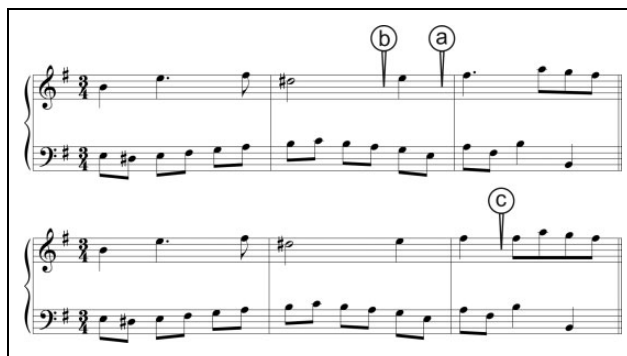
Example A.2. Haendel, Sonata in E, Op. 1, no. 1a, II (Andante), mm 1–4.

Example A.3. Mozart, Rondó in F, KV 494, mm 95–96.

Example A.4. Bach, Little Prelude in C, mm 1–4.

Example A.5. Chopin, Mazurka Op. 41, no. 3, mm 5–6.

Example A.6. Beethoven, Piano Sonata Op. 27, no. 2, Adagio sostenuto, mm 3–4.

Example A.7. Beethoven, Piano Sonata Op.13, II, mm 37–40.

Example A.8. Schubert, Piano Sonata in B flat, D. 960, IV, mm 122–123.

Example A.9. Mozart, Rondó in F, KV 494, mm 97–98.

Example A.10. Schubert, Piano Sonata in B flat, D. 960, IV, mm 124–125.

Example A.11. Mozart, Fantasy in D, KV 397, mm 14–15.

Appendix B

Examples B.1 to B.11. The samples represent open foreground prolongations (OFP) extracted from the whole musical fragments that were used in the test. In all the samples, prefix/suffix prolongations are marked in brackets. Clicks located at suffix and/or prefix boundaries are indicated by an icon. (In Experiment 2, participants listened to only 1 click per musical fragment in each trial. Only clicks located at boundaries of suffix and/or affix prolongations are indicated in examples B.1 to B.11.)

Example B.1. Schubert, Impromptu Op. 90, no. 3, mm 1–4.

Example B2. Beethoven, Piano Sonata Op. 10, no. 1, II, mm 95–97.

Example B.3. Mendelssohn, Song without words Op. 85, no. 1, mm 1–5.

Example B.4. Mozart, Piano Sonata KV 545, II, mm 17–18.

Example B.5. Mendelssohn, Song without words Op. 102, no. 4, mm 4–8.

Example B.6. Mozart, Fantasy in D, KV 397, Adagio, mm 12–13.

Example B.10. Mozart, Piano sonata, KV 545, II, mm 17–18.

Example B.7. Schubert, Impromptu Op. 142, no. 3, mm 5–8.

Example B.8. Schubert, Impromptu Op. 90, no. 3, mm 1–4.

Example B.11. Mendelssohn, Song without words Op. 102, no. 4, mm 4–8.

Example B.9. Mendelssohn, Song without words Op. 85, no. 1, mm 1–5.