Psychological SCIENCE
Research, Theory, & Application in Psychology and Related Sciences
aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Predictive uncertainty underlies auditory boundary perception

Journal:	*Psychological Science*
Manuscript ID	PSCI-20-0830.R2
Manuscript Type:	Research Article
Keywords:	Entropy, Prediction, Perception, Grouping, Music

Predictive uncertainty underlies auditory boundary perception

Abstract

Anticipating the future is essential for efficient perception and action planning. Yet, the role of anticipation in event segmentation is understudied because empirical research has focused on retrospective cues such as surprise. We address this question in the context of musical phrase-boundary perception. A computational model of cognitive sequence processing was used to control the information-dynamic properties of tone sequences. In an implicit, self-paced listening task (*n*=38), undergraduates dwelled longer on tones generating high entropy (i.e., ~~low~~ high uncertainty) than those generating low entropy (i.e., ~~high~~ low uncertainty). Similarly, sequences that ended on tones generating high entropy were rated as sounding more complete (*n*=31). These entropy effects were independent of both the surprise (i.e., information content) and phrase position of target tones in the original musical stimuli. Our results indicate that events generating high entropy prospectively contribute to ~~prospective~~ segmentation processes in auditory sequence perception, independent of the properties of the subsequent event.

Statement of relevance

A significant challenge for the human perceptual system is to promote time-sensitive, context-appropriate responses by predictively processing continuous streams of complex sensory information. A large body of research shows that expectations gleaned from a lifetime of experience guide such processes, which are critical in high-risk environments like traffic or manual labor. Because most studies have focused on the degree of surprise evoked by events, there is little evidence for the role of prospective expectations in perceptual organization. Here, we control entropy in musical tone sequences by using an information-theoretic model that has been shown to reflect listeners' ~~prospective~~ predictive uncertainty. Tones that afforded relatively high uncertainty were found to draw implicit attention and influence explicit ratings of sequence completeness. Focusing attention on instances where upcoming events are statistically unconstrained could contribute to an adaptive mechanism facilitating stream segmentation that leads to efficient learning and information processing in a complex, dynamic world.

Acknowledgments

Our thanks to Dave Thompson, Carla Abawag, and Nicole D’Cunha for their assistance.

Introduction

Humans make sense of a complex, dynamic world by segmenting sequences of events into manageable units (Zacks & Swallow, 2007; Kurby & Zacks, 2008; Richmond & Zacks, 2017). Past work on segmentation has focused on retrospective cues for boundary identification, often conceptualizing group boundaries as coinciding with instances of increased relative change in stimulus features or low transition probabilities (e.g., speech: Saffran & Kirkham, 2018; action sequences: Hard et al., 2011; music: Hartmann et al., 2017; Pearce et al. 2010). However, the sophisticated prediction capabilities of the human mind (Hutchinson & Barrett, 2019) suggest that event boundaries are also anticipated ~~prospectively~~. For example, in natural conversation, turn-taking happens so rapidly that speakers likely anticipate the end of their conversation partner's sentence (Levinson, 2016). Here we investigate the role of entropy, or degree of ~~prospective~~ uncertainty about an upcoming event, in determining the perception of group boundaries in auditory sequences. We define *prediction* as the psychological processes of generating an expectation about a future event~~,~~ in terms of how likely ~~the~~ various possible outcomes are. We define *uncertainty* as the imprecision (or extent of equi-probability) of such a prediction.

Though most previous work has focused on retrospective boundary identification ~~of boundaries~~, anticipatory processing has some preliminary support. ~~Previous work has observed that w~~When self-pacing through sequential images of action sequences, participants tend to "dwell" (or pause) on perceived boundary images (Hard et al., 2011; Hard et al., 2019; Kosie & Baldwin, 2019a, 2019b). Kosie and Baldwin (2019b) proposed that this "dwell time effect" resulted from selective attention to moments of uncertainty afforded by perceiving a goal completion event. No cognitive model was devised to test this theory, however, potentially due to ~~the~~ challenges in modeling expectancy in event processing of action sequences. Indeed, one methodological drawback ~~of this methodology~~ was demonstrated by ~~the finding that~~ participants' dwelling~~ed~~ on boundary slides even when those slides were out of order, suggesting that they were responding to conceptual salience rather than to underlying expectancy dynamics (Hard et al., 2011). Cohen et al. (2007) have proposed an entropy-based segmentation model for language, but because it computes statistics from the corpus it is segmenting—including parts it has not yet seen—it does not fully capture segmentation processing in real time (Christiansen & Chater, 2016).

Because music is not only hierarchically structured (Lerdahl & Jackendoff, 1983), but also statistically well-defined, it is an ideal domain for testing psychological theories of probabilistic perception (Koelsch, Vuust, & Friston, 2019). As with non-musical sequences (Zacks et al., 2001), there is generally high inter-participant agreement regarding the location of musical phrase boundaries (Deliège, 1987; but see Pearce et al., 2010), and as with action sequences, listeners self-pacing through musical chords "dwell" on boundary chords (Kragness & Trainor, 2016, 2018). Since, however, entropy correlates strongly with phrase boundaries in music (Hansen et al., 2017), previous studies were not optimized to separate prospective effects of expectancy dynamics ~~vs.~~from effects of canonical boundary features on perceptual grouping. ~~The~~ *Information Dynamics of Music* ~~*Model*~~ (IDyOM) (Pearce, 2005) is a computational model of auditory expectation which ~~provides a means of~~ enables modelling boundary perception quantitatively using the information-theoretic concepts of entropy and information content, computed in reference to pre-existing long-term knowledge (Hansen & Pearce, 2014; Hansen et al., 2016). Entropy ~~enables~~ facilitates a test of ~~prospective~~ uncertainty as a prospective mechanism for boundary perception which can be pitted directly against information content (a measure of surprise) as a retrospective cue. For example, an individual may form a highly certain ~~prospective~~ prediction ~~for~~ about the next note in a melody but then be surprised when a different note ~~actually~~ follows. Another advantage of ~~using~~ melodic sequences is that~~, unlike images of actions,~~ any given note has little intrinsic meaning in isolation from its preceding musical context, ensuring that ~~any~~ observed effects on perception reflect the statistical structure of the sequence and not inherent features of the boundary stimulus itself. However, because uncertainty ~~processing~~ is not always available for explicit introspection (Hansen et al., 2016), implicit measures are paramount for investigating the cognitive mechanisms underlying boundary perception.

The present study used ~~the~~ IDyOM ~~model~~ to control the information-dynamic properties of melodic sequences in two experiments ~~that~~ assessing~~ed~~ the role of ~~prospective~~ ~~predictive~~ uncertainty in sequence processing. We measured participants' dwell times (Experiment 1) and explicit ratings of phrase completeness (Experiment 2) for tones that afforded high/low entropy and were phrase-beginning/phrase-ending in the melodies from which they were drawn. We predicted that tones ~~that generated~~ generating high ~~levels of prospective~~ uncertainty would lead to longer dwell times ~~(Experiment 1)~~ and higher ~~explicit~~ ratings of phrase completeness, regardless of original phrase status, ~~(Experiment 2)~~ and that this effect would be independent from ~~that of~~ retrospective surprise.

Experiment 1: Implicit Self-Pacing Task

Methods

Participants. Thirty-eight McMaster University undergraduates received psychology course credits for participating in the study (M_{age} = 19.3 years, 1 person declined to report their age, SD_{age} = 3.78, 8 men, 30 women). None of the participants were professional musicians (for more information about musical training levels, see Table S1 in SOM-R2). This sample size exceeds or corresponds to those of previous studies using this methodology to assess comparable effects (e.g., Hard et al., 2011; Kragness & Trainor, 2016, 2018). All participants were fluent in English.

Stimuli. Fifty-six monophonic stimulus sequences were selected from the soprano (i.e., highest) part in 370 four-part chorale harmonizations by Johann Sebastian Bach (Dörffel, 1875) (see SOM-R1 for details of the stimulus selection procedure). These chorale melodies are not generally known by present-day listeners in Canada. Unfamiliarity was, moreover, made more likely through complete removal of rhythmic information by granting participants control over tone durations in the self-paced dwell-time paradigm (Experiment 1) or by presenting stimuli with isochronized tone durations (Experiment 2). All chords, interference tones, and self-pacing tones were generated in MaxMSP's grand piano timbre.

Each stimulus context contained a full phrase (musical group) of seven to 17 pitches followed by the initial tone of the subsequent phrase in the original chorale melody. Tones associated with phrase beginnings and endings were unambiguously identified from notations in the musical score. This practice seems at least as objective as the reliance on trained "expert coders" to determine event boundaries in research using visual action sequences (e.g., Hard et al., 2019; Kosie & Baldwin, 2019a, 2019b). We included both phrase endings and phrase beginnings as target tones to provide a strong test of entropy's role in segmentation, controlling for compositional cues in the melodies that might signal melodic phrase endings in other ways.

Fourteen stimulus contexts were selected for each of the four experimental conditions, comprising phrase beginnings with high ("BegHi") or low entropy ("BegLo") and phrase endings with high ("EndHi") or low entropy ("EndLo"). Entropy, in this regard, quantifies the level of uncertainty governing a listener's expectations about what the pitch of the next tone following the relevant phrase beginning or phrase ending would be. Thus, Western-enculturated listeners are expected to be relatively sure about which pitch will follow the target tone in "BegLo" and "EndLo" contexts, but relatively unsure in "BegHi" and "EndHi" contexts. "Target tone", in this respect, refers to the final tone in "BegLo" and "BegHi" contexts and the penultimate tone in "EndLo" and "EndHi" contexts.

The entropy level generated by each tone in the corpus was estimated by the *Information Dynamics of Music Model* (IDyOM, version 1.3) (Pearce, 2005). This variable-order *n*-gram model uses unsupervised statistical learning to generate probability distributions governing a relevant feature of each tone in a monophonic melody. IDyOM was trained on a large dataset of 5,332 German folk songs (Schaffrath, 1995), 152 Nova Scotian songs and ballads (Creighton, 1966), and 120 English hymns (Nicholson et al., 1950)[1]. For each tone in the chorale melody, IDyOM generated a probability distribution (summing to 1) over the 44 pitch values occurring in the training corpus (i.e., MIDI pitches 45-89 corresponding to A2-F6) by combining *n*-gram models of varying order. Entropy then quantifies the shape of these probability distributions with high entropy for “flat” (relatively uniform) distributions, where there is high uncertainty about the next event, and low entropy for “spiky” (relatively nonuniform) distributions, where one or a small number of continuations are highly probable.

The set of 56 stimulus contexts was selected in a way that prioritized extreme high or low entropy values while ensuring that three conditions were met: First, as shown by a non-parametric Kruskal-Wallis test, all four conditions, including EndHi (Median = 2.45, IQR = 1.76), BegHi (Median = 2.69, IQR = 1.78), EndLo (Median = 2.31, IQR = 2.70), and BegLo (Median = 3.37, IQR = 1.83), were matched on information content (i.e., inverse log-probability) for the event of interest, $\chi^{2}(3) = 4.55$, $p = .208$; second, as shown by Mann-Whitney *U*-tests, EndHi (Median = 2.97, IQR = 0.08) and BegHi (Median = 3.00, IQR = 0.12) stimuli, $U = 78$, $p = .376$, as well as EndLo (Median = 1.07, IQR = 0.30) and BegLo (Median = 0.97, IQR = 0.35) stimuli, $U = 90$, $p = .734$, were matched on entropy governing the next event in the sequence. The experimenter selecting these stimuli paid no attention to any other musical features.

For the secondary analysis of all tones in the stimulus set, IC and entropy were re-estimated by re-running IDyOM with the same configuration on the final stimulus contexts. This was done because IC and entropy estimates for the initial tones in each stimulus context sometimes relied on tones from the preceding phrase in the original chorales, which was excluded from the stimuli used.

[1] For more information about the IDyOM implementation and parameters, please see SOM-R.

While unproblematic for stimulus selection based on target tones, this presented a problem for tone-level analysis. Note that due to their late position in the tone sequences, target tone entropy and IC values were identical for the two models (one used in stimulus generation and analyses of target tones, the other used in the analysis of all tones).

Procedure. The experimental procedures (for Experiment 1 and 2) received prior approval from the McMaster University Research Ethics Board and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Participants were seated facing a computer screen in a sound-attenuated room. They were instructed to press the spacebar on a computer keyboard with the pointer finger of their dominant hand to elicit the onset of each subsequent tone in the sequence. Tones decayed naturally, but were not terminated until the spacebar was pressed again to initiate the next tone. Participants were instructed to progress as quickly or slowly as they liked while listening carefully, and could not repeat previously heard tones. They were led to falsely believe that their memory for the sequences would be tested afterwards to motivate them to attend to the task (Kragness & Trainor, 2016). No other instructions regarding timing, pacing, rhythmicity, or expressivity were given. If a participant asked for further information, they were told to play through the piece in a way that would maximize their performance in the subsequent memory task.

Prior to each trial, participants saw three flashes of a fixation cross, then heard 40 50-ms tones (for a total of 2000 ms) chosen randomly on each trial from range E2 to A5 to minimize carryover from the context of the previous sequence, followed by three context-establishing chords with durations of 800, 800, and 1600 ms (Figure 1). The context-establishing chords were played in the key of the relevant melody. Throughout each trial, a circle on the screen indicated when to begin self-pacing through the melody (light green) and when to stop (dark green).

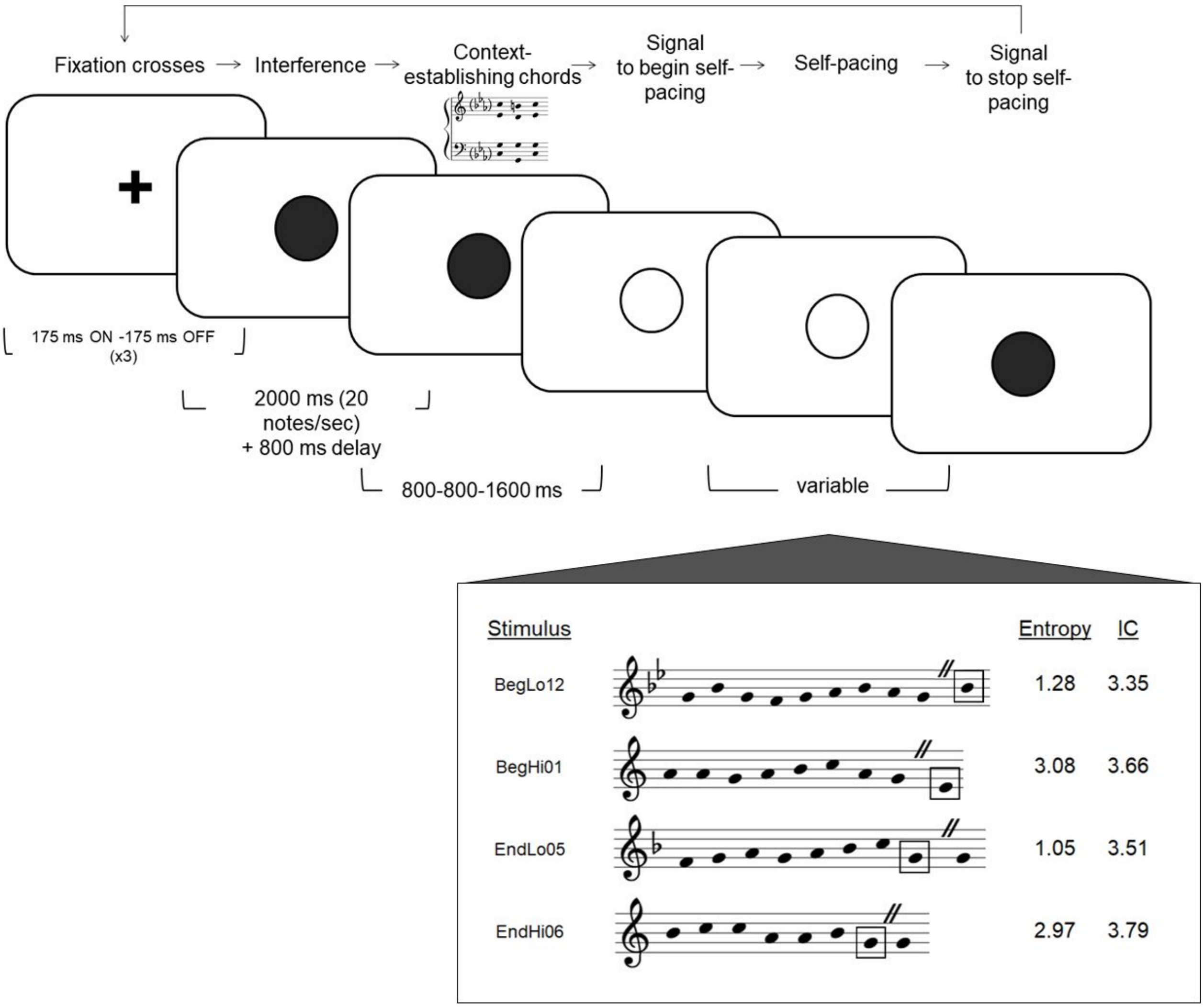
Fixation crosses
Interference
Context-establishing chords
Signal to begin self-pacing
Self-pacing
Signal to stop self-pacing
175 ms ON -175 ms OFF (x3)
2000 ms (20 notes/sec) + 800 ms delay
800-800-1600 ms
variable
Stimulus
Entropy
IC
BegLo12
1.28
3.35
BegHi01
3.08
3.66
EndLo05
1.05
3.51
EndHi06
2.97
3.79

Figure 1. Depiction of a trial from Experiment 1. In each trial, participants saw a fixation cross, followed by interference tones, then three context-establishing chords and a signal (white circle) to begin self-pacing. They then self-paced through the tone sequence until the occurrence of a stop signal (black circle). The box depicts examples of tone sequences from each condition containing target tones (boxed) generating relatively uncertain (high entropy) or relatively certain (low entropy) expectations about the pitch of the next tone, matched on IC of the current tone. The double slash indicates whether target tones were phrase beginnings (after double slash) or phrase endings (prior to double slash) in the original notation.

Data processing and statistical analysis. Despite systematic efforts to avoid duplicate stimulus contexts (e.g., multiple occurrences of a repeated phrase from a single melody or identical phrases across melodies), it was discovered after data collection that one melodic context occurred both amongst the “BegHi” and “EndHi” stimulus sets (with different target tones). Given that results did not differ substantially when excluding dwell times for these stimuli, we report statistical analyses including the full dataset here, which included 56 total tone sequences (i.e., 14 per condition).

To mitigate effects of extreme data points, a minimum dwell-time threshold of 100 ms was adopted for inclusion. Dwell times greater than 3 standard deviations above a participant’s own average (across all target and non-target dwell times) were also omitted (Kosie & Baldwin, 2019a, 2019b). These exclusion criteria eliminated an average of 1.31% of all tones and 1.70% of target tones per participant (ranging from 0-4 target tones).

For the main analysis of target tones, target dwell times were averaged by condition resulting in four condition-wise means per participant. A 2x2 repeated-measures ANOVA (including within-subjects factors boundary status and entropy) was run on target tone dwell times.

For the secondary analysis of all tones, dwell times were first log-transformed to minimize the positive skew inherent to timing data (cf. Kragness & Trainor, 2018). Subsequently, using the *lmer()* function from the *lme4* package in R (R Core Team, 2019), linear mixed-effects models were fitted with Restricted Maximum Likelihood estimates (REML). Because previous experiments have found that dwell times change systematically throughout trials (Kragness & Trainor, 2016), tone index in the sequence was always included as a predictor. Thus, whereas the null model only included tone index as a fixed effect, two further increasingly complex models added, first, the retrospective cue IC, and, second, the prospective cue entropy. Thereby, we could determine whether prospective predictive processing explained unique variance not already accounted for by retrospective surprise. Random intercepts and slopes of tone number were included for each participant. For all models, this random-effects structure produced the lowest BIC values while avoiding singular fits.

Results

Target tones. To examine the effects of boundary status (phrase-ending, phrase-beginning) and entropy (high, low), a 2x2 repeated-measures ANOVA was run on target tone dwell times. Whereas no significant interaction ($F(1,37) < 0.01$, $p = .986$, $\eta^2_p < .001$) or main effect of boundary status ($F(1,37) < 0.01$, $p = .973$, $\eta^2_p < .001$) was found, there was a significant main effect of entropy ($F(1,37) = 7.24$, $p = .011$, $\eta^2_p = .164$). Thus, as hypothesized, high-entropy target tones were generally dwelled on longer than low-entropy target tones, regardless of phrase position in the original chorale melody (Figure 2).

We conducted post-hoc correlational analyses to examine whether participants' musical sophistication was associated with the magnitude of their dwell time effect. No significant associations were observed (see SOM-R2 for more details).

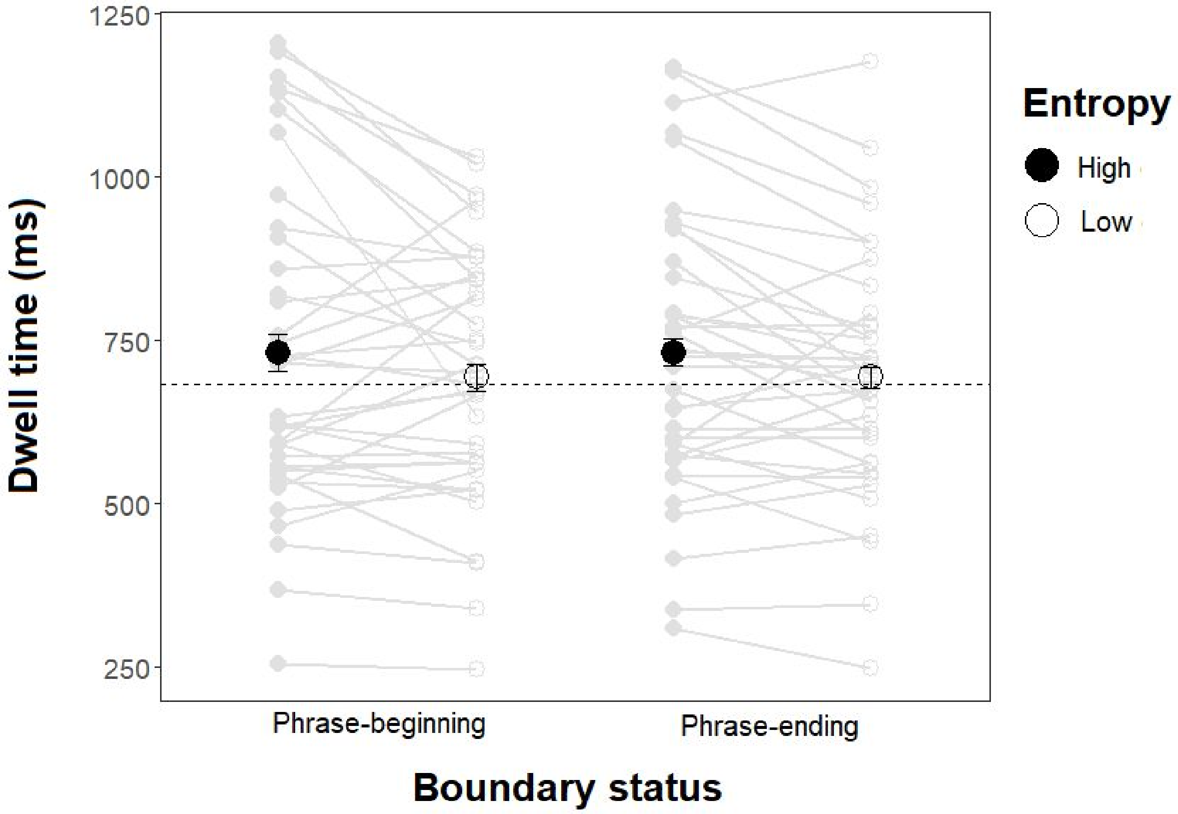

Figure 2. Dwell times (ms) for each type of target tone (BegHi, BegLo, EndHi, EndLo) in Experiment 1. The dashed line represents the average dwell time (683 ms) for non-target tones. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005). High-entropy target tones had longer dwell times than low-entropy target tones, and it made no significant difference whether target tones originated from phrase endings or phrase beginnings in the original chorale melody corpus.

All tones. If ~~prospective~~ uncertainty provides a cognitive cue for phrase segmentation, its effect on dwell times should generalize beyond the target tones occupying the extreme ranges of entropy values. Analyzing dwell times for all tones also allowed us to directly compare the effects of prospective entropy vs. retrospective information content (IC). Recall that IC was matched across target tones in the previous analysis.

Model comparisons on models refitted with Maximum Likelihood estimates found that the IC model predicted dwell times significantly better than the null model, $\chi^2(1) = 31.77$, $p < .001$. Adding entropy improved the fit significantly, $\chi^2(1) = 16.64$, $p < .001$. In the full model, log-transformed dwell times increased significantly with IC, $F(1, 19711.3) = 35.26$, $p < .001$, entropy, $F(1, 19711.2) = 16.64$, $p < .001$, and marginally non-significantly with tone index in the phrase, $F(1, 37.5) = 3.30$, $p = .077$.

Experiment 2: Explicit completeness ratings

In Experiment 1, participants dwelled longer on tones affording high-entropy continuations than on tones affording low-entropy continuations, regardless of whether they were originally phrase beginnings or endings. This suggests that when rhythmic and metrical cues are removed from the musical surface, entropic peaks in prospective pitch expectancy elicit implicit segmentation. Previous dwell-time studies have demonstrated that longer dwell times coincide with perceived boundaries (e.g., Hard et al., 2011), but Experiment 1 did not ~~provide concrete evidence~~guarantee that participants were segmenting the stimuli. Therefore, Experiment 2 was designed to provide converging evidence for effects of ~~prospective~~ prediction on segmentation using an explicit self-report measure of phrase completeness (Palmer & Krumhansl, 1987).

Methods

Participants. Thirty-one McMaster University students (not participants in Experiment 1) took part in Experiment 2. Again, none were professional musicians (see SOM-R2 for more information). This sample size exceeds those from previous studies using this methodology to assess a comparable contrast (e.g., Palmer & Krumhansl, 1987). One participant declined to report their gender and age, but among the remaining participants, the average age was 18.93 years ($SD_{\text{age}} = 2.51$

years), with 7 men and 23 women. Of the 31 participants, responses from five individuals were omitted due to uninterpretable response sheets (i.e., multiple answers for each sequence, lacking answers for certain sequences).

Stimuli. Melodic stimulus sequences were identical to those for Experiment 1, except that all notes were played with a constant duration of 400 ms. Unlike in Experiment 1, the target tone was always the final tone in the sequence.

Procedure. As in Experiment 1, the procedure took place in a sound-attenuating room. Rather than self-pacing through the sequences as in Experiment 1, participants listened to all 56 sequences in randomized order. After each sequence, participants rated how complete the sequence sounded (ranging from 1: "totally incomplete" to 7: "totally complete"). If the end of the melody was completely satisfactory, that would constitute a score of 7, but if the melody ended in a way that was implausible and unsatisfactory, that would constitute a score of 1. Participants were encouraged to use the full range of the scale.

Results

A 2x2 repeated-measures ANOVA with factors boundary status (phrase-ending, phrase-beginning) and entropy (high, low) was run on mean condition-wise ratings. Results were fully consistent with those for Experiment 1. Specifically, no significant interaction ($F(1,25) = 1.80$, $p = .192$, $\eta^2_p = .067$) nor main effect of boundary status ($F(1,25) = 0.82$, $p = .373$, $\eta^2_p = .032$) was found, whereas there was a significant main effect of entropy ($F(1, 25) = 44.11$, $p < .001$, $\eta^2_p = .638$). High-entropy target tones were rated as constituting more complete phrase endings than low-entropy target tones, regardless of phrase position in the original chorale melody (Figure 3).

Again, no significant associations with musical sophistication were observed (see SOM-R2 for more details).

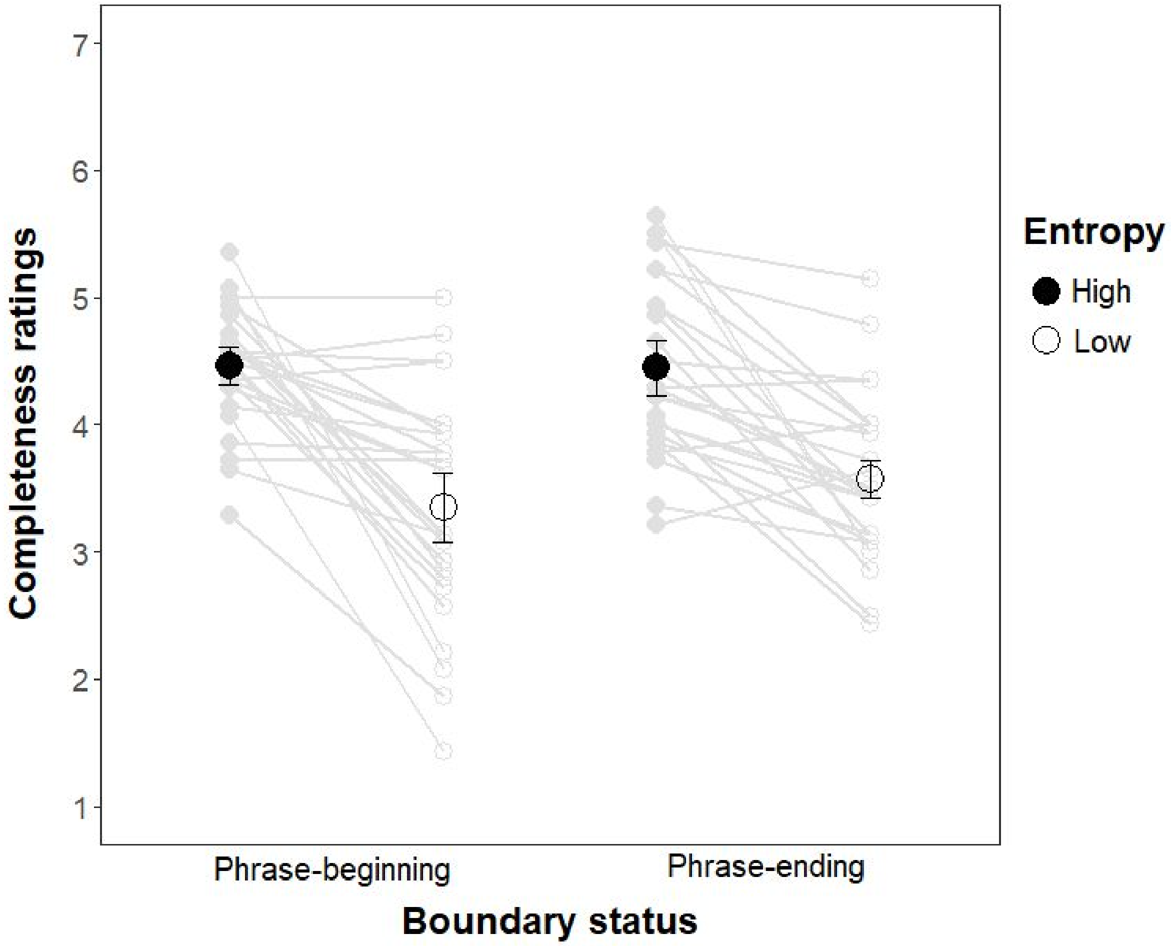

Figure 3. Completeness ratings for each type of excerpt (BegHi, BegLo, EndHi, and EndLo) in Experiment 2. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005). Stimulus sequences with final tones generating high entropy were generally deemed more complete than those generating low entropy. It made no significant difference whether tones originated from phrase beginnings or phrase endings in the original chorale melodies.

General Discussion

Although prediction is a fundamental component in influential theories of perceptual organization (Hutchinson & Barrett, 2019), evidence for the role of ~~prospective~~ uncertainty ~~(a prospective measure of prediction)~~ remains weak due to the empirical focus on retrospective measures of surprise (Hansen

& Pearce, 2014). Here we tested the hypothesis that uncertainty relates to boundary perception in auditory sequences, using stimuli from Western tonal music ~~in which~~with well-defined phrase boundaries ~~are well-defined~~. Sequences ~~that ended~~ ending on tones generating high-entropy expectations were perceived as more complete than those ending on tones generating low-entropy expectations (Experiment 2). This was also indicated by longer dwell times on high-entropy target tones ~~generating high entropy~~ ; ~~and,~~ indeed, across all tones in the stimulus sequences, entropy explained unique variance in dwell times not ~~already~~ accounted for by event probability (Experiment 1).

Our work raises the key question why segmentation follows peaks of ~~statistical~~ uncertainty. Christiansen and Chater's (2016) *Now-or-Never Bottleneck* posits that information ~~currently~~ in working memory needs to be processed ~~here and~~ now or be forever lost. This constraint necessitates "chunk-and-pass" processing whereby fleeting input—such as the content of music, speech, or action sequences—is quickly segmented and encoded as higher-level representational units. Following from ~~Christiansen and Chater's (2016)~~this theory, ~~it is possible that~~ events that afford high-entropy predictions may require more bits to encode and thus may require higher working memory deployment. The likelihood of exceeding memory capacity is higher after high-uncertainty events than after low-uncertainty events, ~~leading to a~~causing higher probability of "chunking" and perceiving a segment boundary.

This framework may also explain ~~the~~ previously demonstrated "dwell time" effects ~~observed in previous studies~~ (Hard et al., 2011, 2019; Kosie & Baldwin, 2019a, 2019b; Kragness & Trainor, 2016, 2018), since there is a time delay associated with segmentation and reintegration into previous knowledge. This reintegration process, however, may have a cost. Specifically, taking in new information is harder while reintegration takes place. Because the human mind aims to be one step ahead, it will attempt to balance this cost optimally. Therefore, pauses in the stimulus stream may induce a chunk to be processed even if it ends on low uncertainty (without fully exceeding working memory capacity). This may constitute one ~~of the~~ potential mechanism~~s~~ explaining why Gestalt-like principles of temporal proximity generally seem to apply to auditory sequence processing (Lerdahl & Jackendoff, 1983).

The relatively high working memory capacity required at phrase boundaries may explain previously observed *phrase-final lengthening.* Specifically, across ~~a variety of~~various languages, musical instruments, and performance contexts, speakers and performers tend to slow down at phrase endings (speech: Wightman et al., 1992; music: Palmer, 1989; Repp, 1992). While originally interpreted as a communicative gesture in music (Palmer, 1989), piano performers exhibit phrase-final lengthening even when attempting to play without expression (Penel & Drake, 1998). Combined with the observation that listeners are less prone to detect lengthening on boundary tones than within-phrase tones (Repp, 1992), ~~this led~~ Penel and Drake (1998) ~~to~~ hypothesized that perceptual biases contribute to group-final lengthening, although the source of this bias remained unspecified. ~~We propose that~~ O~~o~~ne such source could be processing constraints due to ~~predictive~~ uncertainty, which likely apply across ~~multiple~~ domains of sequential perception and production.

Here we specifically focused on modelling the uncertainty of a single feature, pitch, as a cue for phrase closure. Of course, the probabilistic characteristics of many other features (for instance, temporal, spectral, syntactic, etc.) might affect ~~the perception of~~ completeness perception. In music, these might include duration, intensity, inter-onset intervals, and performer gestures (Lerdahl & Jackendoff, 1983). Whether ~~predictive~~ uncertainty in temporal features influences musical phrase grouping remains to be tested. However, given that sensory systems prioritize anticipatory ~~processing~~ over reactive processing (Christiansen & Chater, 2016; Hutchinson & Barrett, 2019), it seems plausible that our findings should extend to the temporal domain. On the other hand, non-probabilistic and non-pitch-related features may also constrain the statistical learning giving rise to the entropy effects found here, as observed in speech segmentation (Yang, 2004). Incorporating metrical structure, previously heard motives, and limiting the number of accented tones per phrase would, for example, most likely improve the predictive power of our entropy-based model. Future work should more directly contrast the effect of anticipatory vs. adaptive cues and of probabilistic (top-down) vs. Gestalt-related (bottom-up) cues to establish their relative contribution and investigate how this may vary under different experimental conditions.

Another concern is whether IDyOM accurately reflects listener expectations. Morgan et al. (2019) found that IDyOM predictions entailed higher entropy than that computed across several participants ~~who~~ providing~~ed~~ single-tone sung continuations to melodic contexts. Task constraints likely explain this discrepancy as expectations for multiple continuations were not assessed. Furthermore, by manipulating entropy of upcoming events rather than simply analyzing the entropy of instantiated continuations, the present study differs crucially from Morgan et al. (2019). Moreover, whereas they recruited self-identified musicians, who make melodic predictions with demonstrably lower average entropy than non-musicians (Hansen & Pearce, 2014; Hansen, Vuust, & Pearce, 2016), IDyOM was configured to model expectations of the general population. At the same time, Morgan et al. (2019) made an important contribution by demonstrating a greater contribution of statistical learning than of Gestalt-based principles in predicting listener expectations. This supports IDyOM's suitability in predicting auditory boundary perception.

The finding that ~~predictive~~ uncertainty influences phrase boundary perception suggests a pertinent role for training effects. Expertise effects may be particularly prominent in the musical domain where skills and experiences differ substantially between individuals. Although some ~~previous~~ studies suggest limited effects of musical expertise on melodic segmentation processes (Palmer & Krumhansl, 1987, but see Hartmann et al., 2017), expertise levels have not always been widely sampled or manipulated systematically. The same limitation applies to the current study where no significant effects of expertise were seen (see Tables S2 and S3 in SOM-R2 for details). Yet, recent research shows that stylistic specialization results in expectations about melodic continuations that are generally lower in entropy whenever greater confidence is warranted (Hansen & Pearce, 2014; Hansen et al., 2016). The transformation of high-entropy predictions into low-entropy predictions with domain-relevant training or implicit exposure should allow musicians to perceive phrasal coherence across longer timespans. This would be consistent with observations that experts have access to more abstract and deeper levels of hierarchical structure (Chaffin & Imreh, 2002; Chi & Feltovich, 1981) which, in turn, may be associated with larger working memory capacity (Meinz & Hambrick, 2010). While awaiting sampling across more diverse expertise levels in future research, our results relating chunk size to underlying expectancy dynamics enables a novel interpretation of classical findings pertaining to expertise and working memory.

By offering an empirical challenge to the view that segmentation primarily relies on retrospective processes, the present work contributes to the emergence of an increasingly coherent model of the human mind as an eager predictive processor of sensory input. Embedded in the constant flux of time, the mind is continually forced to evaluate and recombine retrospective and prospective cues according to their immediate usefulness, and we hypothesize that sequential input in such varied domains as language, music, and visual action sequences are all subject to the constraints arising from this mental machinery.

References

Chaffin, R., & Imreh, G. (2002). Practicing perfection: piano performance as expert memory. *Psychological Science*, *13*(4), 342-349.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121-152.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. doi:10.1017/S0140525X1500031X.

Cohen, P., Adams, N., & Heeringa, B. (2007). Voting experts: an unsupervised algorithm for segmenting sequences. *Intelligent Data Analysis*, *11*(6), 607-625.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42-45.

Creighton, H. (ed.). (1966). *Songs and Ballads from Nova Scotia*. New York, NY: Dover.

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

Dörfell (ed.) (1875). *371 vierstimmige Choralgesänge von Johann Sebastian Bach* (4th ed.). Leipzig, Germany: Breitkopf & Härtel.

Hansen, N. C.,& Pearce, M. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology* 5, 1052.

Hansen, N. C., Vuust, P., & Pearce, M. (2016). "If you've got to ask, you'll never know": Style-congruent musical expertise optimises predictive auditory processing. *PLOS ONE*, *11*(10): e0163584. doi:10.1371/journal.pone.0163584

Hansen, N. C. ,Vuust, P., Pearce, M., & Huron, D. (2017, August). *Entropic Ebbs and Flows: The Expectancy Dynamics of Musical Phrases*. Paper presented at the Society for Music Perception and Cognition Meeting, San Diego, CA.

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*, *47*(1), 17-32.

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

Hartmann, M., Lartillot, O. & Toiviainen, P. (2017). Interaction features for prediction of perceptual segmentation: effects of musicianship and experimental task. *Journal of New Music Research, 46*(2), 156-174. doi:10.1080/09298215.2016.1230137

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: an emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280-291.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63-77.

Kosie, J. E., & Baldwin, D. (2019a). Attention rapidly reorganizes to naturally occurring structure in a novel activity sequence. *Cognition*, *182*, 31–44. doi:10.1016/j.cognition.2018.09.004

Kosie, J. E., & Baldwin, D. (2019b). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications*, *4*(1), 8. doi:10.1186/s41235-019-0157-4

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

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

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72-79.

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

Levinson, S. C. (2016). Turn-taking in human communication: origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6-14. doi:10.1016/j.tics.2015.10.010

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of working memory capacity. *Psychological Science*, *21*(7), 914-919.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23-34.

Nicholson, S., Knight, G. H., and Bower, J. D. (Ed.). (1950). *Ancient and Modern Revised*. Suffolk, UK: William Clowes and Sons.

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

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 116.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (Doctoral dissertation). City University, London, UK. Retrieved from https://openaccess.city.ac.uk/id/eprint/8459/1/

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

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: a psychological segmentation model. *Psychological Research*, *61*(1), 12-32.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Repp, B. H. (1992). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition*, *44*(3), 241-281.

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: event models from perception to action. *Trends in Cognitive Sciences*, 21(12), 962-980.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181-203. doi:10.1146/annurev-psych-122216-011805

Schaffrath, H. (1995). *The Essen Folksong Collection in the Humdrum Kern Format* (D. Huron, Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities. Retrieved from https://kern.humdrum.org/cgi-bin/browse?l=essen/europa/deutschl

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

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451-456.

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

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29-58.

Predictive uncertainty underlies auditory boundary perception

Abstract

Anticipating the future is essential for efficient perception and action planning. Yet, the role of anticipation in event segmentation is understudied because empirical research has focused on retrospective cues such as surprise. We address this question in the context of musical phrase-boundary perception. A computational model of cognitive sequence processing was used to control the information-dynamic properties of tone sequences. In an implicit, self-paced listening task (*n*=38), undergraduates dwelled longer on tones generating high entropy (i.e., high uncertainty) than those generating low entropy (i.e., low uncertainty). Similarly, sequences that ended on tones generating high entropy were rated as sounding more complete (*n*=31). These entropy effects were independent of both the surprise (i.e., information content) and phrase position of target tones in the original musical stimuli. Our results indicate that events generating high entropy prospectively contribute to segmentation processes in auditory sequence perception, independent of the properties of the subsequent event.

Statement of relevance

A significant challenge for the human perceptual system is to promote time-sensitive, context-appropriate responses by predictively processing continuous streams of complex sensory information. A large body of research shows that expectations gleaned from a lifetime of experience guide such processes, which are critical in high-risk environments like traffic or manual labor. Because most studies have focused on the degree of surprise evoked by events, there is little evidence for the role of prospective expectations in perceptual organization. Here, we control entropy in musical tone sequences by using an information-theoretic model that has been shown to reflect listeners' predictive uncertainty. Tones that afforded relatively high uncertainty were found to draw implicit attention and influence explicit ratings of sequence completeness. Focusing attention on instances where upcoming events are statistically unconstrained could contribute to an adaptive mechanism facilitating stream segmentation that leads to efficient learning and information processing in a complex, dynamic world.

Acknowledgments

Our thanks to Dave Thompson, Carla Abawag, and Nicole D'Cunha for their assistance.

Introduction

Humans make sense of a complex, dynamic world by segmenting sequences of events into manageable units (Zacks & Swallow, 2007; Kurby & Zacks, 2008; Richmond & Zacks, 2017). Past work on segmentation has focused on retrospective cues for boundary identification, often conceptualizing group boundaries as coinciding with instances of increased relative change in stimulus features or low transition probabilities (e.g., speech: Saffran & Kirkham, 2018; action sequences: Hard et al., 2011; music: Hartmann et al., 2017; Pearce et al. 2010). However, the sophisticated prediction capabilities of the human mind (Hutchinson & Barrett, 2019) suggest that event boundaries are also anticipated. For example, in natural conversation, turn-taking happens so rapidly that speakers likely anticipate the end of their conversation partner's sentence (Levinson, 2016). Here we investigate the role of entropy, or degree of uncertainty about an upcoming event, in determining the perception of group boundaries in auditory sequences. We define *prediction* as the psychological processes of generating an expectation about a future event in terms of how likely various possible outcomes are. We define *uncertainty* as the imprecision (or extent of equi-probability) of such a prediction.

Though most previous work has focused on retrospective boundary identification, anticipatory processing has some preliminary support. When self-pacing through sequential images of action sequences, participants tend to "dwell" (or pause) on perceived boundary images (Hard et al., 2011; Hard et al., 2019; Kosie & Baldwin, 2019a, 2019b). Kosie and Baldwin (2019b) proposed that this "dwell time effect" resulted from selective attention to moments of uncertainty afforded by perceiving a goal completion event. No cognitive model was devised to test this theory, however, potentially due to challenges in modeling expectancy in event processing of action sequences. Indeed, one methodological drawback was demonstrated by participants' dwelling on boundary slides even when those slides were out of order, suggesting that they were responding to conceptual salience rather than to underlying expectancy dynamics (Hard et al., 2011). Cohen et al. (2007) have proposed an entropy-based segmentation model for language, but because it computes statistics from the corpus it is segmenting—including parts it has not yet seen—it does not fully capture segmentation processing in real time (Christiansen & Chater, 2016).

Because music is not only hierarchically structured (Lerdahl & Jackendoff, 1983), but also statistically well-defined, it is an ideal domain for testing psychological theories of probabilistic perception (Koelsch, Vuust, & Friston, 2019). As with non-musical sequences (Zacks et al., 2001), there is generally high inter-participant agreement regarding the location of musical phrase boundaries (Deliège, 1987; but see Pearce et al., 2010), and as with action sequences, listeners self-pacing through musical chords "dwell" on boundary chords (Kragness & Trainor, 2016, 2018). Since, however, entropy correlates strongly with phrase boundaries in music (Hansen et al., 2017), previous studies were not optimized to separate prospective effects of expectancy dynamics from effects of canonical boundary features on perceptual grouping. *Information Dynamics of Music* (IDyOM) (Pearce, 2005) is a computational model of auditory expectation which enables modelling boundary perception quantitatively using the information-theoretic concepts of entropy and information content, computed in reference to pre-existing long-term knowledge (Hansen & Pearce, 2014; Hansen et al., 2016). Entropy facilitates a test of uncertainty as a prospective mechanism for boundary perception which can be pitted directly against information content (a measure of surprise) as a retrospective cue. For example, an individual may form a highly certain prediction about the next note in a melody but then be surprised when a different note follows. Another advantage of melodic sequences is that any given note has little intrinsic meaning in isolation from its preceding musical context, ensuring that observed effects on perception reflect the statistical structure of the sequence and not inherent features of the boundary stimulus itself. However, because uncertainty is not always available for explicit introspection (Hansen et al., 2016), implicit measures are paramount for investigating the cognitive mechanisms underlying boundary perception.

The present study used IDyOM to control the information-dynamic properties of melodic sequences in two experiments assessing the role of uncertainty in sequence processing. We measured participants' dwell times (Experiment 1) and explicit ratings of phrase completeness (Experiment 2) for tones that afforded high/low entropy and were phrase- beginning/phrase-ending in the melodies from which they were drawn. We predicted that tones generating high uncertainty would lead to longer dwell times and higher ratings of phrase completeness, regardless of original phrase status, and that this effect would be independent from retrospective surprise.

Experiment 1: Implicit Self-Pacing Task

Methods

Participants. Thirty-eight McMaster University undergraduates received psychology course credits for participating in the study (M_{age} = 19.3 years, 1 person declined to report their age, SD_{age} = 3.78, 8 men, 30 women). None of the participants were professional musicians (for more information about musical training levels, see Table S1 in SOM-R2). This sample size exceeds or corresponds to those of previous studies using this methodology to assess comparable effects (e.g., Hard et al., 2011; Kragness & Trainor, 2016, 2018). All participants were fluent in English.

Stimuli. Fifty-six monophonic stimulus sequences were selected from the soprano (i.e., highest) part in 370 four-part chorale harmonizations by Johann Sebastian Bach (Dörffel, 1875) (see SOM-R1 for details of the stimulus selection procedure). These chorale melodies are not generally known by present-day listeners in Canada. Unfamiliarity was, moreover, made more likely through complete removal of rhythmic information by granting participants control over tone durations in the self-paced dwell-time paradigm (Experiment 1) or by presenting stimuli with isochronized tone durations (Experiment 2). All chords, interference tones, and self-pacing tones were generated in MaxMSP's grand piano timbre.

Each stimulus context contained a full phrase (musical group) of seven to 17 pitches followed by the initial tone of the subsequent phrase in the original chorale melody. Tones associated with phrase beginnings and endings were unambiguously identified from notations in the musical score. This practice seems at least as objective as the reliance on trained "expert coders" to determine event boundaries in research using visual action sequences (e.g., Hard et al., 2019; Kosie & Baldwin, 2019a, 2019b). We included both phrase endings and phrase beginnings as target tones to provide a strong test of entropy's role in segmentation, controlling for compositional cues in the melodies that might signal melodic phrase endings in other ways.

Fourteen stimulus contexts were selected for each of the four experimental conditions, comprising phrase beginnings with high ("BegHi") or low entropy ("BegLo") and phrase endings with high ("EndHi") or low entropy ("EndLo"). Entropy, in this regard, quantifies the level of uncertainty governing a listener's expectations about what the pitch of the next tone following the relevant phrase beginning or phrase ending would be. Thus, Western-enculturated listeners are expected to be relatively sure about which pitch will follow the target tone in "BegLo" and "EndLo" contexts, but relatively unsure in "BegHi" and "EndHi" contexts. "Target tone", in this respect, refers to the final tone in "BegLo" and "BegHi" contexts and the penultimate tone in "EndLo" and "EndHi" contexts.

The entropy level generated by each tone in the corpus was estimated by the *Information Dynamics of Music Model* (IDyOM, version 1.3) (Pearce, 2005). This variable-order *n*-gram model uses unsupervised statistical learning to generate probability distributions governing a relevant feature of each tone in a monophonic melody. IDyOM was trained on a large dataset of 5,332 German folk songs (Schaffrath, 1995), 152 Nova Scotian songs and ballads (Creighton, 1966), and 120

English hymns (Nicholson et al., 1950)[1]. For each tone in the chorale melody, IDyOM generated a probability distribution (summing to 1) over the 44 pitch values occurring in the training corpus (i.e., MIDI pitches 45-89 corresponding to A2-F6) by combining *n*-gram models of varying order. Entropy then quantifies the shape of these probability distributions with high entropy for "flat" (relatively uniform) distributions, where there is high uncertainty about the next event, and low entropy for "spiky" (relatively nonuniform) distributions, where one or a small number of continuations are highly probable.

The set of 56 stimulus contexts was selected in a way that prioritized extreme high or low entropy values while ensuring that three conditions were met: First, as shown by a non-parametric Kruskal-Wallis test, all four conditions, including EndHi (Median = 2.45, IQR = 1.76), BegHi (Median = 2.69, IQR = 1.78), EndLo (Median = 2.31, IQR = 2.70), and BegLo (Median = 3.37, IQR = 1.83), were matched on information content (i.e., inverse log-probability) for the event of interest, $\chi^{2}(3) = 4.55$, $p = .208$; second, as shown by Mann-Whitney *U*-tests, EndHi (Median = 2.97, IQR = 0.08) and BegHi (Median = 3.00, IQR = 0.12) stimuli, $U = 78$, $p = .376$, as well as EndLo (Median = 1.07, IQR = 0.30) and BegLo (Median = 0.97, IQR = 0.35) stimuli, $U = 90$, $p = .734$, were matched on entropy governing the next event in the sequence. The experimenter selecting these stimuli paid no attention to any other musical features.

For the secondary analysis of all tones in the stimulus set, IC and entropy were re-estimated by re-running IDyOM with the same configuration on the final stimulus contexts. This was done because IC and entropy estimates for the initial tones in each stimulus context sometimes relied on tones from the preceding phrase in the original chorales, which was excluded from the stimuli used. While unproblematic for stimulus selection based on target tones, this presented a problem for tone-level analysis. Note that due to their late position in the tone sequences, target tone entropy and IC

[1] For more information about the IDyOM implementation and parameters, please see SOM-R.

values were identical for the two models (one used in stimulus generation and analyses of target tones, the other used in the analysis of all tones).

Procedure. The experimental procedures (for Experiment 1 and 2) received prior approval from the McMaster University Research Ethics Board and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Participants were seated facing a computer screen in a sound-attenuated room. They were instructed to press the spacebar on a computer keyboard with the pointer finger of their dominant hand to elicit the onset of each subsequent tone in the sequence. Tones decayed naturally, but were not terminated until the spacebar was pressed again to initiate the next tone. Participants were instructed to progress as quickly or slowly as they liked while listening carefully, and could not repeat previously heard tones. They were led to falsely believe that their memory for the sequences would be tested afterwards to motivate them to attend to the task (Kragness & Trainor, 2016). No other instructions regarding timing, pacing, rhythmicity, or expressivity were given. If a participant asked for further information, they were told to play through the piece in a way that would maximize their performance in the subsequent memory task.

Prior to each trial, participants saw three flashes of a fixation cross, then heard 40 50-ms tones (for a total of 2000 ms) chosen randomly on each trial from range E2 to A5 to minimize carryover from the context of the previous sequence, followed by three context-establishing chords with durations of 800, 800, and 1600 ms (Figure 1). The context-establishing chords were played in the key of the relevant melody. Throughout each trial, a circle on the screen indicated when to begin self-pacing through the melody (light green) and when to stop (dark green).

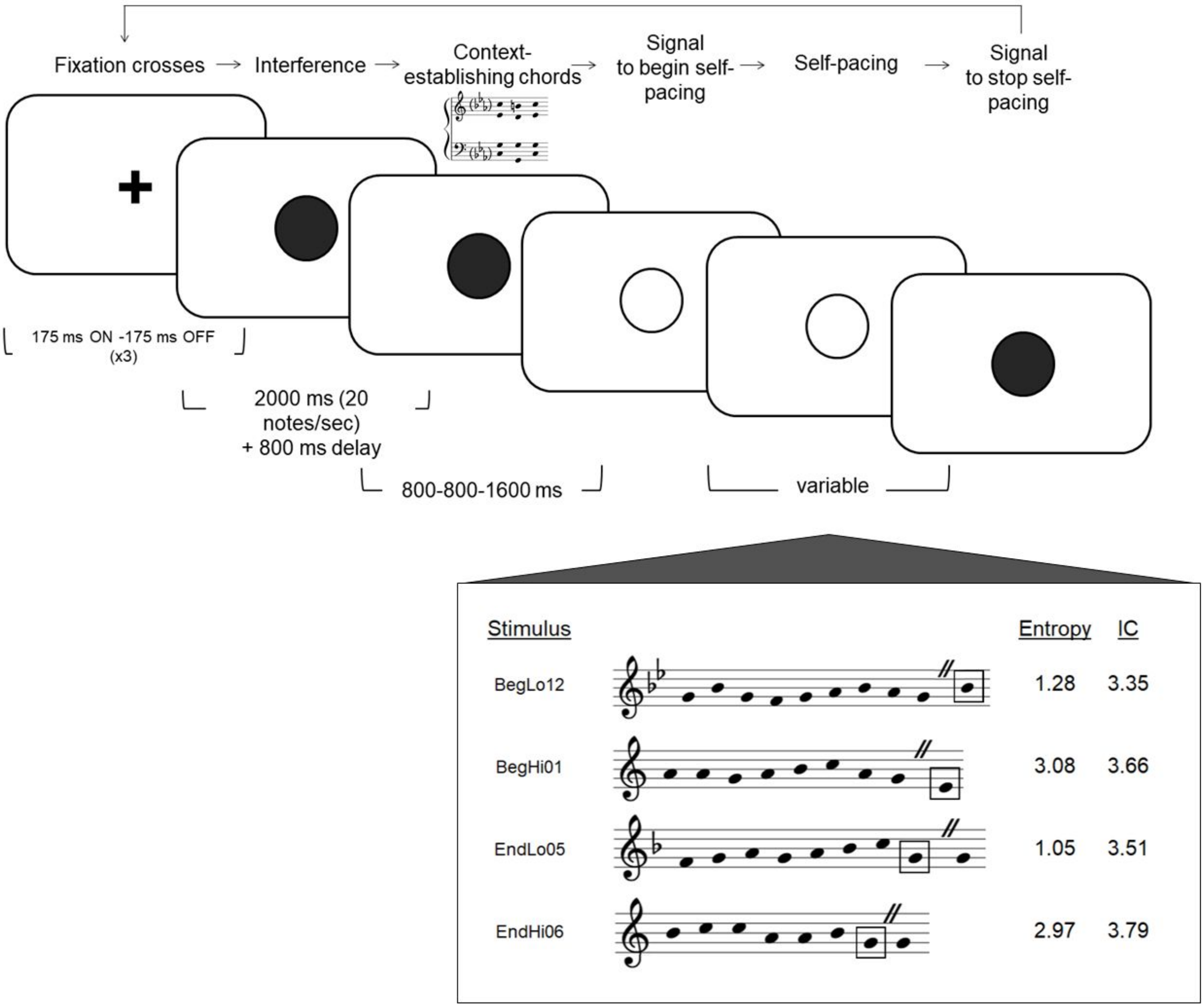
Fixation crosses
Interference
Context-establishing chords
Signal to begin self-pacing
Self-pacing
Signal to stop self-pacing
175 ms ON -175 ms OFF (x3)
2000 ms (20 notes/sec) + 800 ms delay
800-800-1600 ms
variable
Stimulus
Entropy
IC
BegLo12
1.28
3.35
BegHi01
3.08
3.66
EndLo05
1.05
3.51
EndHi06
2.97
3.79

Figure 1. Depiction of a trial from Experiment 1. In each trial, participants saw a fixation cross, followed by interference tones, then three context-establishing chords and a signal (white circle) to begin self-pacing. They then self-paced through the tone sequence until the occurrence of a stop signal (black circle). The box depicts examples of tone sequences from each condition containing target tones (boxed) generating relatively uncertain (high entropy) or relatively certain (low entropy) expectations about the pitch of the next tone, matched on IC of the current tone. The double slash indicates whether target tones were phrase beginnings (after double slash) or phrase endings (prior to double slash) in the original notation.

Data processing and statistical analysis. Despite systematic efforts to avoid duplicate stimulus contexts (e.g., multiple occurrences of a repeated phrase from a single melody or identical phrases across melodies), it was discovered after data collection that one melodic context occurred both amongst the "BegHi" and "EndHi" stimulus sets (with different target tones). Given that results did not differ substantially when excluding dwell times for these stimuli, we report statistical analyses including the full dataset here, which included 56 total tone sequences (i.e., 14 per condition).

To mitigate effects of extreme data points, a minimum dwell-time threshold of 100 ms was adopted for inclusion. Dwell times greater than 3 standard deviations above a participant's own average (across all target and non-target dwell times) were also omitted (Kosie & Baldwin, 2019a, 2019b). These exclusion criteria eliminated an average of 1.31% of all tones and 1.70% of target tones per participant (ranging from 0-4 target tones).

For the main analysis of target tones, target dwell times were averaged by condition resulting in four condition-wise means per participant. A 2x2 repeated-measures ANOVA (including within-subjects factors boundary status and entropy) was run on target tone dwell times.

For the secondary analysis of all tones, dwell times were first log-transformed to minimize the positive skew inherent to timing data (cf. Kragness & Trainor, 2018). Subsequently, using the *lmer()* function from the *lme4* package in R (R Core Team, 2019), linear mixed-effects models were fitted with Restricted Maximum Likelihood estimates (REML). Because previous experiments have found that dwell times change systematically throughout trials (Kragness & Trainor, 2016), tone index in the sequence was always included as a predictor. Thus, whereas the null model only included tone index as a fixed effect, two further increasingly complex models added, first, the retrospective cue IC, and, second, the prospective cue entropy. Thereby, we could determine whether prospective predictive processing explained unique variance not already accounted for by retrospective surprise. Random intercepts and slopes of tone number were included for each participant. For all models, this random-effects structure produced the lowest BIC values while avoiding singular fits.

Results

Target tones. To examine the effects of boundary status (phrase-ending, phrase-beginning) and entropy (high, low), a 2x2 repeated-measures ANOVA was run on target tone dwell times. Whereas no significant interaction ($F(1,37) < 0.01$, $p = .986$, $\eta^2_p < .001$) or main effect of boundary status ($F(1,37) < 0.01$, $p = .973$, $\eta^2_p < .001$) was found, there was a significant main effect of entropy ($F(1,37) = 7.24$, $p = .011$, $\eta^2_p = .164$). Thus, as hypothesized, high-entropy target tones were generally dwelled on longer than low-entropy target tones, regardless of phrase position in the original chorale melody (Figure 2).

We conducted post-hoc correlational analyses to examine whether participants' musical sophistication was associated with the magnitude of their dwell time effect. No significant associations were observed (see SOM-R2 for more details).

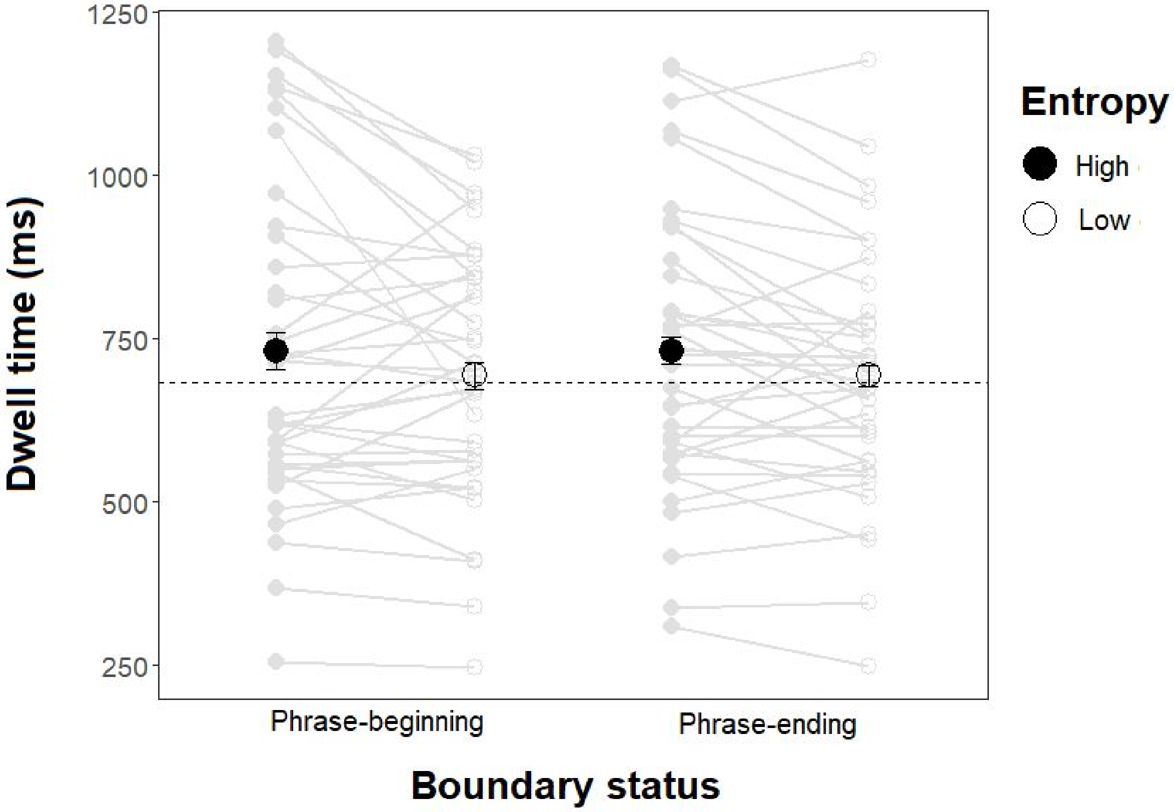

Figure 2. Dwell times (ms) for each type of target tone (BegHi, BegLo, EndHi, EndLo) in Experiment 1. The dashed line represents the average dwell time (683 ms) for non-target tones. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005). High-entropy target tones had longer dwell times than low-entropy target tones, and it made no significant difference whether target tones originated from phrase endings or phrase beginnings in the original chorale melody corpus.

All tones. If uncertainty provides a cognitive cue for phrase segmentation, its effect on dwell times should generalize beyond the target tones occupying the extreme ranges of entropy values. Analyzing dwell times for all tones also allowed us to directly compare the effects of prospective entropy vs. retrospective information content (IC). Recall that IC was matched across target tones in the previous analysis.

Model comparisons on models refitted with Maximum Likelihood estimates found that the IC model predicted dwell times significantly better than the null model, $\chi^2(1) = 31.77$, $p < .001$. Adding entropy improved the fit significantly, $\chi^2(1) = 16.64$, $p < .001$. In the full model, log-transformed dwell times increased significantly with IC, $F(1, 19711.3) = 35.26$, $p < .001$, entropy, $F(1, 19711.2) = 16.64$, $p < .001$, and marginally non-significantly with tone index in the phrase, $F(1, 37.5) = 3.30$, $p = .077$.

Experiment 2: Explicit completeness ratings

In Experiment 1, participants dwelled longer on tones affording high-entropy continuations than on tones affording low-entropy continuations, regardless of whether they were originally phrase beginnings or endings. This suggests that when rhythmic and metrical cues are removed from the musical surface, entropic peaks in prospective pitch expectancy elicit implicit segmentation. Previous dwell-time studies have demonstrated that longer dwell times coincide with perceived boundaries (e.g., Hard et al., 2011), but Experiment 1 did not guarantee that participants were segmenting the stimuli. Therefore, Experiment 2 was designed to provide converging evidence for effects of prediction on segmentation using an explicit self-report measure of phrase completeness (Palmer & Krumhansl, 1987).

Methods

Participants. Thirty-one McMaster University students (not participants in Experiment 1) took part in Experiment 2. Again, none were professional musicians (see SOM-R2 for more information). This sample size exceeds those from previous studies using this methodology to assess a comparable contrast (e.g., Palmer & Krumhansl, 1987). One participant declined to report their gender and age, but among the remaining participants, the average age was 18.93 years ($SD_{\text{age}} = 2.51$

years), with 7 men and 23 women. Of the 31 participants, responses from five individuals were omitted due to uninterpretable response sheets (i.e., multiple answers for each sequence, lacking answers for certain sequences).

Stimuli. Melodic stimulus sequences were identical to those for Experiment 1, except that all notes were played with a constant duration of 400 ms. Unlike in Experiment 1, the target tone was always the final tone in the sequence.

Procedure. As in Experiment 1, the procedure took place in a sound-attenuating room. Rather than self-pacing through the sequences as in Experiment 1, participants listened to all 56 sequences in randomized order. After each sequence, participants rated how complete the sequence sounded (ranging from 1: "totally incomplete" to 7: "totally complete"). If the end of the melody was completely satisfactory, that would constitute a score of 7, but if the melody ended in a way that was implausible and unsatisfactory, that would constitute a score of 1. Participants were encouraged to use the full range of the scale.

Results

A 2x2 repeated-measures ANOVA with factors boundary status (phrase-ending, phrase-beginning) and entropy (high, low) was run on mean condition-wise ratings. Results were fully consistent with those for Experiment 1. Specifically, no significant interaction (F(1,25) = 1.80, p = .192, η^2_p = .067) nor main effect of boundary status (F(1,25) = 0.82, p = .373, η^2_p = .032) was found, whereas there was a significant main effect of entropy (F(1, 25) = 44.11, p < .001, η^2_p = .638). High-entropy target tones were rated as constituting more complete phrase endings than low-entropy target tones, regardless of phrase position in the original chorale melody (Figure 3).

Again, no significant associations with musical sophistication were observed (see SOM-R2 for more details).

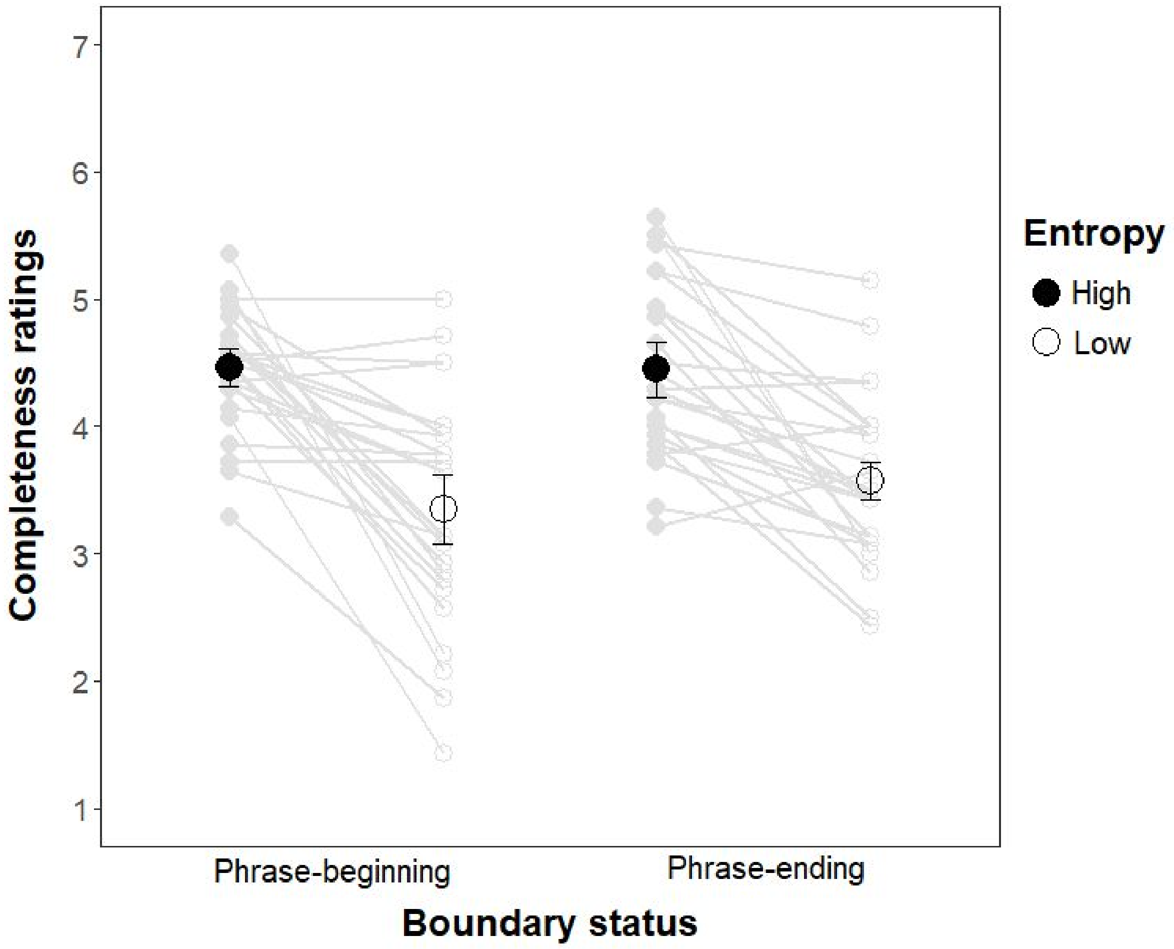

Figure 3. Completeness ratings for each type of excerpt (BegHi, BegLo, EndHi, and EndLo) in Experiment 2. Error bars represent within-subject 95% confidence intervals (Cousineau, 2005). Stimulus sequences with final tones generating high entropy were generally deemed more complete than those generating low entropy. It made no significant difference whether tones originated from phrase beginnings or phrase endings in the original chorale melodies.

General Discussion

Although prediction is a fundamental component in influential theories of perceptual organization (Hutchinson & Barrett, 2019), evidence for the role of uncertainty remains weak due to the empirical focus on retrospective measures of surprise (Hansen & Pearce, 2014). Here we tested the hypothesis that uncertainty relates to boundary perception in auditory sequences, using stimuli from Western tonal music with well-defined phrase boundaries. Sequences ending on tones generating high-entropy expectations were perceived as more complete than those ending on tones generating low-entropy expectations (Experiment 2). This was also indicated by longer dwell times on high-entropy target tones; indeed, across all tones in the stimulus sequences, entropy explained unique variance in dwell times not accounted for by event probability (Experiment 1).

Our work raises the key question why segmentation follows peaks of uncertainty. Christiansen and Chater's (2016) *Now-or-Never Bottleneck* posits that information in working memory needs to be processed now or be forever lost. This constraint necessitates "chunk-and-pass" processing whereby fleeting input—such as the content of music, speech, or action sequences—is quickly segmented and encoded as higher-level representational units. Following from this theory, events that afford high-entropy predictions may require more bits to encode and thus may require higher working memory deployment. The likelihood of exceeding memory capacity is higher after high-uncertainty events than after low-uncertainty events, causing higher probability of "chunking" and perceiving a segment boundary.

This framework may also explain previously demonstrated "dwell time" effects (Hard et al., 2011, 2019; Kosie & Baldwin, 2019a, 2019b; Kragness & Trainor, 2016, 2018), since there is a time delay associated with segmentation and reintegration into previous knowledge. This reintegration process, however, may have a cost. Specifically, taking in new information is harder while reintegration takes place. Because the human mind aims to be one step ahead, it will attempt to balance this cost optimally. Therefore, pauses in the stimulus stream may induce a chunk to be processed even if it ends on low uncertainty (without fully exceeding working memory capacity). This may constitute one potential mechanism explaining why Gestalt-like principles of temporal proximity generally seem to apply to auditory sequence processing (Lerdahl & Jackendoff, 1983).

The relatively high working memory capacity required at phrase boundaries may explain previously observed *phrase-final lengthening*. Specifically, across various languages, musical instruments, and performance contexts, speakers and performers tend to slow down at phrase endings (speech: Wightman et al., 1992; music: Palmer, 1989; Repp, 1992). While originally interpreted as a communicative gesture in music (Palmer, 1989), piano performers exhibit phrase-final lengthening even when attempting to play without expression (Penel & Drake, 1998). Combined with the observation that listeners are less prone to detect lengthening on boundary tones than within-phrase tones (Repp, 1992), Penel and Drake (1998) hypothesized that perceptual biases contribute to group-final lengthening, although the source of this bias remained unspecified. One such source could be processing constraints due to uncertainty, which likely apply across domains of sequential perception and production.

Here we specifically focused on modelling the uncertainty of a single feature, pitch, as a cue for phrase closure. Of course, the probabilistic characteristics of many other features (for instance, temporal, spectral, syntactic, etc.) might affect completeness perception. In music, these might include duration, intensity, inter-onset intervals, and performer gestures (Lerdahl & Jackendoff, 1983). Whether uncertainty in temporal features influences musical phrase grouping remains to be tested. However, given that sensory systems prioritize anticipatory over reactive processing (Christiansen & Chater, 2016; Hutchinson & Barrett, 2019), it seems plausible that our findings should extend to the temporal domain. On the other hand, non-probabilistic and non-pitch-related features may also constrain the statistical learning giving rise to the entropy effects found here, as observed in speech segmentation (Yang, 2004). Incorporating metrical structure, previously heard motives, and limiting the number of accented tones per phrase would, for example, most likely improve the predictive power of our entropy-based model. Future work should more directly contrast the effect of anticipatory vs. adaptive cues and of probabilistic (top-down) vs. Gestalt-related (bottom-up) cues to establish their relative contribution and investigate how this may vary under different experimental conditions.

Another concern is whether IDyOM accurately reflects listener expectations. Morgan et al. (2019) found that IDyOM predictions entailed higher entropy than that computed across several participants providing single-tone sung continuations to melodic contexts. Task constraints likely explain this discrepancy as expectations for multiple continuations were not assessed. Furthermore, by manipulating entropy of upcoming events rather than simply analyzing the entropy of instantiated continuations, the present study differs crucially from Morgan et al. (2019). Moreover, whereas they recruited self-identified musicians, who make melodic predictions with demonstrably lower average entropy than non-musicians (Hansen & Pearce, 2014; Hansen, Vuust, & Pearce, 2016), IDyOM was configured to model expectations of the general population. At the same time, Morgan et al. (2019) made an important contribution by demonstrating a greater contribution of statistical learning than of Gestalt-based principles in predicting listener expectations. This supports IDyOM's suitability in predicting auditory boundary perception.

The finding that uncertainty influences phrase boundary perception suggests a pertinent role for training effects. Expertise effects may be particularly prominent in the musical domain where skills and experiences differ substantially between individuals. Although some studies suggest limited effects of musical expertise on melodic segmentation processes (Palmer & Krumhansl, 1987, but see Hartmann et al., 2017), expertise levels have not always been widely sampled or manipulated systematically. The same limitation applies to the current study where no significant effects of expertise were seen (see Tables S2 and S3 in SOM-R2 for details). Yet, recent research shows that stylistic specialization results in expectations about melodic continuations that are generally lower in entropy whenever greater confidence is warranted (Hansen & Pearce, 2014; Hansen et al., 2016). The transformation of high-entropy predictions into low-entropy predictions with domain-relevant training or implicit exposure should allow musicians to perceive phrasal coherence across longer timespans. This would be consistent with observations that experts have access to more abstract and deeper levels of hierarchical structure (Chaffin & Imreh, 2002; Chi & Feltovich, 1981) which, in turn, may be associated with larger working memory capacity (Meinz & Hambrick, 2010). While awaiting sampling across more diverse expertise levels in future research, our results relating chunk size to underlying expectancy dynamics enables a novel interpretation of classical findings pertaining to expertise and working memory.

By offering an empirical challenge to the view that segmentation primarily relies on retrospective processes, the present work contributes to the emergence of an increasingly coherent model of the human mind as an eager predictive processor of sensory input. Embedded in the constant flux of time, the mind is continually forced to evaluate and recombine retrospective and prospective cues according to their immediate usefulness, and we hypothesize that sequential input in such varied domains as language, music, and visual action sequences are all subject to the constraints arising from this mental machinery.

References

Chaffin, R., & Imreh, G. (2002). Practicing perfection: piano performance as expert memory. *Psychological Science*, *13*(4), 342-349.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121-152.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. doi:10.1017/S0140525X1500031X.

Cohen, P., Adams, N., & Heeringa, B. (2007). Voting experts: an unsupervised algorithm for segmenting sequences. *Intelligent Data Analysis*, *11*(6), 607-625.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42-45.

Creighton, H. (ed.). (1966). *Songs and Ballads from Nova Scotia*. New York, NY: Dover.

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

Dörfell (ed.) (1875). *371 vierstimmige Choralgesänge von Johann Sebastian Bach* (4th ed.). Leipzig, Germany: Breitkopf & Härtel.

Hansen, N. C.,& Pearce, M. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology* 5, 1052.

Hansen, N. C., Vuust, P., & Pearce, M. (2016). "If you've got to ask, you'll never know": Style-congruent musical expertise optimises predictive auditory processing. *PLOS ONE*, *11*(10): e0163584. doi:10.1371/journal.pone.0163584

Hansen, N. C. ,Vuust, P., Pearce, M., & Huron, D. (2017, August). *Entropic Ebbs and Flows: The Expectancy Dynamics of Musical Phrases*. Paper presented at the Society for Music Perception and Cognition Meeting, San Diego, CA.

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*, *47*(1), 17-32.

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

Hartmann, M., Lartillot, O. & Toiviainen, P. (2017). Interaction features for prediction of perceptual segmentation: effects of musicianship and experimental task. *Journal of New Music Research, 46*(2), 156-174. doi:10.1080/09298215.2016.1230137

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: an emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280-291.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63-77.

Kosie, J. E., & Baldwin, D. (2019a). Attention rapidly reorganizes to naturally occurring structure in a novel activity sequence. *Cognition*, *182*, 31–44. doi:10.1016/j.cognition.2018.09.004

Kosie, J. E., & Baldwin, D. (2019b). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications*, *4*(1), 8. doi:10.1186/s41235-019-0157-4

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

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

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72-79.

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

Levinson, S. C. (2016). Turn-taking in human communication: origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6-14. doi:10.1016/j.tics.2015.10.010

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of working memory capacity. *Psychological Science*, *21*(7), 914-919.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23-34.

Nicholson, S., Knight, G. H., and Bower, J. D. (Ed.). (1950). *Ancient and Modern Revised*. Suffolk, UK: William Clowes and Sons.

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

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 116.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (Doctoral dissertation). City University, London, UK. Retrieved from https://openaccess.city.ac.uk/id/eprint/8459/1/

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

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: a psychological segmentation model. *Psychological Research*, *61*(1), 12-32.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Repp, B. H. (1992). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition*, *44*(3), 241-281.

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: event models from perception to action. *Trends in Cognitive Sciences*, 21(12), 962-980.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181-203. doi:10.1146/annurev-psych-122216-011805

Schaffrath, H. (1995). *The Essen Folksong Collection in the Humdrum Kern Format* (D. Huron, Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities. Retrieved from https://kern.humdrum.org/cgi-bin/browse?l=essen/europa/deutschl

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

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451-456.

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

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29-58.