Cognition 226 (2022) 105180

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

When did that happen? The dynamic unfolding of perceived musical narrative

Elizabeth Hellmuth Margulis [a,*], Jamal Williams [b], Rhimmon Simchy-Gross [c], J. Devin McAuley [d,**]

[a] Department of Music, Princeton University, USA
[b] Department of Psychology, Princeton University, USA
[c] Department of Psychology, Missouri Western State University, USA
[d] Department of Psychology, Michigan State University, USA

ARTICLE INFO

Keywords:
Tension
Event perception
Narrative
Music
Intersubjectivity

ABSTRACT

People readily imagine narratives in response to instrumental music. Although previous work has established that these narratives show broad intersubjectivity, it remains unclear whether these imagined stories are atemporal, or unfold systematically over the temporal extent of a musical excerpt. To investigate the dynamics of perceived musical narrative, we had participants first listen to 16 instrumental musical excerpts, which had previously been normed for factors of interest. While listening, participants continuously moved a slider to indicate their fluctuating perceptions of tension and relaxation. In a separate experimental session, participants reported the stories they imagined while listening to each excerpt, and then, while listening to the excerpts a final time, clicked a mouse to mark the time points at which they imagined new events in the ongoing imagined story. The time points of these event markings were not uniformly distributed throughout the excerpts, but were clustered at distinct moments, indicating that imagined narratives unfold in real time and entail general consensus about *when* listeners imagine events in the music. Moreover, the time points at which people tended to imagine events were correlated with the time points at which people tended to perceive salient changes in musical tension, as separately recorded within the first experimental session. The degree of alignment was greater for excerpts high in narrativity than those low in narrativity. Together, these results show that music can dynamically guide a listener's imagination and there is remarkable intersubjectivity in 'when' hear imagined story events in a piece of music.

1. Introduction

Understanding the world in terms of narrative is a ubiquitous (Nash, 2005; Rubin, 1995) and ancient proclivity, employed even by ancestral hominids (Donald, 1991). Foundational to culture (Schiffrin, De Fina, & Nylund, 2010), narrative serves as a fundamental way to organize information (Bruner, 1990; McAdams, 1993), and has been understood to subserve memory, thought, and knowledge (Schank & Abelson, 1995). Watching geometric shapes move around on a screen, people were overwhelmingly likely to describe what they saw in terms of a story (Heider & Simmel, 1944). Thus, it is perhaps unsurprising that music, too—despite a history of formalist approaches that treat music as non-referential and without a semantics (see Alperson, 2004)—can engender narrative perception.

In fact, as one possible response among a range of available modes of listening (cf. Tuuri & Eerola, 2012), people readily imagine stories in response to wordless musical excerpts (Margulis, Wong, Simchy-Gross, & McAuley, 2019). Within but not across cultures, people show striking intersubjective agreement about the stories implied by particular excerpts (McAuley, Wong, Bellaiche, & Margulis, 2021; Margulis, Wong, Turnbull, Kubit, & McAuley, 2022). These narrative imaginings do not constitute periods of haphazard mind-wandering, but rather reflect "stimulus- and context-driven episodes in which the sounds seem to shape multimedia trajectories through an imagined space" (McAuley, Wong, Mamidipaka, Phillips, & Margulis, 2021). In a narrative experience of instrumental music, the entrance of a cello melody might be

* Corresponding author at: E.H. Margulis, Department of Music, Princeton University, Princeton, NJ 08544, USA.
** Correspondence to: J.D. McAuley, Department of Psychology, Michigan State University, East Lansing, MI 48840, USA.
E-mail addresses: margulis@princeton.edu (E.H. Margulis), dmcauley@msu.edu (J.D. McAuley).

https://doi.org/10.1016/j.cognition.2022.105180
Received 24 June 2021; Received in revised form 11 May 2022; Accepted 19 May 2022
Available online 2 June 2022

heard not directly as a cello line, but rather as the surfacing of a dangerous shark. The cello melody is present, but points to an event within the ongoing narrative trajectory, with the imagined event, rather than the sound, serving as a focal point for the listener.

Previous studies reveal that narrative affordances of individual musical excerpts vary in two important ways. One, in their degree of narrativity, as measured by the percentage of people who report imagining a story while listening to a particular excerpt, and two, in the extent to which any imagined narrative grips the listener (McAuley, Wong, Mamidipaka, et al., 2021). The latter property, termed narrative engagement (NE), is measured by a four-question instrument developed and validated by Margulis et al. (2019). High NE scores indicate that the imagined story came to mind easily, was vivid, featured clear characters and events, and preoccupied the listener while they were hearing the music. These characteristics—narrativity and narrative engagement—are dissociable, such that one excerpt might trigger a narrative for a large percentage of listeners, but without engaging them deeply in this imagined story; people recognize that there's an implied narrative, but aren't gripped by it. Conversely, another excerpt might only trigger a narrative for a small percentage of listeners, but do so in an engrossing, captivating way.

A third important attribute of an individual excerpt's narrative affordances relates to story content: the degree to which it elicits similar stories in the minds of different listeners. The degree of consensus around story content can also vary independently from the degree of narrative engagement it elicits: people might imagine extremely vivid stories while listening to a particular excerpt (high narrative engagement) but their stories all might be different from one another (low agreement about story content). Conversely, people might imagine vivid stories for another excerpt (high narrative engagement) and the stories might be highly consistent from one listener to another (high agreement about story content).

Although it has been established that the stories imagined while listening to individual excerpts are often consistent across listeners within a culture (McAuley, Wong, Bellaiche, & Margulis, 2021; Margulis et al., 2022), a number of outstanding questions remain about how listeners' narrative perceptions relate to the dynamic unfolding of musical structure in time. At a minimum, a narrative involves the recounting of an event or a series of events (Abbott, 2020). How do these perceived events unfold during music listening? How do the time points of their perceived inception align or fail to align across listeners? How does an excerpt's overall narrativity influence the consistency of when listeners imagine events to begin? How do the time points of perceived events line up with salient changes along other musical and perceptual dimensions? At stake is the question of whether music guides perceived events dynamically across its time course, or whether the stories listeners imagine in response to music are atemporal. No previous studies of narrative imaginings to music have incorporated a real-time element; all previous measures were administered once people had finished listening. This methodological limitation made it difficult to ascertain what the temporal nature of narrative imaginings are like. It's possible that all the information in the excerpt was combined and employed holistically to produce a post-hoc story report. The characteristics of the excerpt, considered overall, might point to some associated referent, around which participants subsequently construct a conventional story. In this scenario, people might imagine a story in response to an excerpt without any temporal connection to the moment-by-moment unfolding of the music. The story's imagined events are concocted after the conclusion of the excerpt, by fleshing out a scene around some concrete referent implied by the music. Event clicks might then be distributed uniformly throughout the excerpt, or cluster at the end once the imagined was fully formed and the events are envisioned. This pattern of responses would be consistent with the perceptual decoupling that occurs during mind-wandering, including mind-wandering sustained during music listening (Koelsch, Bashevkin, Kristensen, Tvedt, & Jentschke, 2019; Schooler, 2011; Taruffi, Pehrs, Skouras, & Koelsch, 2017). It might also be the pattern expected from autobiographical memories sustained while listening to music—the music cues an associated memory, but the discrete events recalled within the autobiographical memory don't unfold in time with specific occurrences within the music (Belfi & Jakubowski, 2021). On the other hand, imagined stories might unfold in time bit by bit in a systematic manner, with individual events yoked to particular happenings in the music. In this scenario, the envisioning of the narrative's composite events are spurred by discrete happenings in the music itself, not by a post-excerpt narrativization of some association implied by the entirety of the excerpt.

In the visual domain, Tversky and Zacks find that visual event boundaries "correspond with identifiable physical changes" (2013, p. 83), such as actors' changes in speed or direction (Hard, Tversky, & Lang, 2006), hand movements (Zacks, Kumar, Abrams, & Mehta, 2009), or moments of unexpectedness and surprise (Hard, Meyer, & Baldwin, 2019). When participants are asked to identify event boundaries by pressing a button when they occur, they tend to select the same points in time. According to event segmentation theory (Zacks & Swallow, 2007), people automatically generate event models in short-term memory and use them to guide perceptual predictions. When the predictions fail, the models are updated with a new event representation.

Very little is known about event segmentation in a musical context. In this regard, the transition between individual movements of a piece in the Western art music tradition has been used as a model case for exploring event segmentation within naturalistic music contexts. These transitions normally involve moments of contrast—salient changes in tempo, tonality, timbre, and rhythm—and are marked by brief interpolated silent periods (Deliège, 1989). One neuroimaging investigation of perception during these transitions implicated two relevant brain networks: a ventral fronto-temporal network relevant to detecting salient events, and a dorsal fronto-parietal network relevant to information maintenance and working memory (Sridharan, Levitin, Chafe, Berger, & Menon, 2007). In narrative perceptions of music, listeners hear through these acoustic events to the dramatic situations they imply. Just as event segmentation in other domains relies both on characteristics of the observed event and personal factors such as prior knowledge, goals, and mood (Schwann & Garsoffky, 2008), the parsing of perceived musical narratives into events likely reflects both personal factors and characteristics of the sound itself.

Studies that investigate music perception as it occurs in real time, from moment to moment as the sound progresses, are also comparatively rare. The most frequently used measure involves continuously reporting fluctuations in perceived musical tension. Participants continuously move a slider while listening to indicate that the music is tensing or relaxing. Despite that little to no guidance is typically provided for participants in how to use the terms tension and relaxation, intersubject correlation tends to be high (Krumhansl, 1996; Krumhansl & Schenck, 1997; Madsen & Fredrickson, 1993; Vines, Nuzzo, & Levitin, 2005). Features such as tempo, dynamics, timbre, harmony, and melody have been shown to influence these ratings (Farbood, 2012). Given that tension is also central to narrative perception—books on fiction writing often advise aspiring authors how to build and release it across the course of their story (cf. Bell, 2011; Burroway, 2019; Stern, 2000)—this method provides an opportunity to understand whether perceived tension in music might be related to the imagined stories the music engenders. No previous study, however, has directly asked participants about their narrative perceptions of music in real time.

Towards this end, the current study used real-time measures of both perceived tension and perceived narrative to understand how narrative event structure unfolds during music listening. In an initial experimental session, participants did not perform any tasks directly related to narrative perception, but instead were instructed simply to move a slider to indicate fluctuating perceptions of tension as musical excerpts progressed. In a second experimental session several weeks later, participants listened to musical excerpts, and afterwards provided a free response description of any story they imagined (as described in

Table 1
Musical stimuli consisted of 16 short 60s excerpts of Western instrumental classical music (High Narrativity, $n = 8$; Low Narrativity, $n = 8$). Classification of excerpts into High Narrativity and Low Narrativity categories was based on a normative story response question (SRQ) score, which measured the likelihood that the excerpt triggers a story in listeners' minds. Also shown in the table are the excerpt titles, performers, and track timing for the beginning of the 60s clips, as well as normative measures of Narrative Engagement (NE), Contrast, Topicality, Enjoyment and Familiarity (Margulis et al., 2019). High Narrativity and Low Narrativity excerpts differed in Narrativity, Narrative Engagement, and Topicality (all p's < 0.001), but did not differ in Contrast, Enjoyment or Familiarity (all p's > 0.05).

Category	ID	Title	Performer	Track timing for start of excerpt	Composer	SRQ	NE	Contrast	Topicality	Enjoy	Familiarity
High narrativity	WA4	Symphony #1 Mov. #1	Neeme Järvi + Detroit Symphony Orchestra	1:18	Still	0.98	4.20	4.67	5.33	5.05	2.50
	WA5	Egmont Overture	Zubin Mehta + Los Angeles Philharmonic	2:23	Beethoven	0.68	3.93	5.33	4.67	4.48	2.52
	WA6	Dante Symphony, S. 109 I.	Daniel Barenboim + Berlin Philharmonic Orchestra	1:12	Liszt	0.90	4.26	6.00	4.33	4.13	2.44
	WB10	Suite from Dracula: Tides	Carducci Quartet	0:00	Glass	0.75	3.88	5.00	2.00	4.08	2.06
	WC3	In the Mystic Land of Egypt	Robert Sharples + New Symphony Orchestra	0:00	Ketèlbey	0.88	4.69	2.33	5.00	4.53	2.26
	WC5	March for Military Music	Hans Preim-Bergrath + Berlin Philharmonic Wind Ensemble Martha	0:00	Beethoven	0.83	4.09	2.00	6.67	4.68	1.90
Low narrativity	WC6	Gaspard de la Nuit II.	Argerich	0:00	Ravel	0.73	3.82	1.33	4.00	3.83	1.72
	WD3	La Centra, Op. 9, No. 1, II.	Iona Brown + Academy of St. Martin in the Fields	0:00	Vivaldi	0.77	4.08	1.33	3.67	4.20	1.50
	WA13	Symphony #4, Op. 11, IV	Christopher von Dohnányi + Vienna Philharmonic	0:00	Mendelsso	0.47	3.43	5.00	3.33	3.96	2.14
	WB11	Str Quartet Op. 28 I. Mässig	Emerson String Quartet	0:00	hn	0.46	3.01	3.67	2.00	3.34	1.52
	WB14	En Blanc Et Noir I.	Martha Argerich + Steven Kovacevich	0:00	Webern	0.40	2.82	4.67	2.33	3.96	1.96
	WB7	Suite, Op. 29 I.	David Atherton + London Sinfonietta	0:00	Debussy	0.43	2.82	5.67	1.67	3.02	1.66
	WD12	Ouverture Pour le Piano	Lilya Zilberstein	0:00	Schoenber	0.50	3.37	2.33	1.67	4.95	1.98
	WD14	L. 95 No. 3 Das Lied Von	Michael Tilson Thomas + San Francisco Symphony	0:00	g Ravel	0.35	2.97	1.33	2.33	4.18	1.60
	WD1	Der Erde II. The Chanting Light of Foresight	Rova Saxophone Quartet	0:00	Mahler	0.60	3.40	2.67	2.67	3.53	1.54
	WD2	Piano Sonata #7, K. 309 II.	Maria João Pires	0:00	Riley	0.53	3.34	1.33	1.33	4.43	1.80
High						**0.81**	**4.12**	3.50	**4.46**	4.37	2.73
Low						**0.47**	**3.14**	3.33	**2.17**	3.92	2.51

Margulis et al., 2022). Then, they listened to each of the excerpts again, but this time, as each progressed, they made a mouse click whenever they perceived the inception of a new event in this imagined story. Once the excerpt ends, they filled in a box for each time point they marked, describing the event they imagined to start at the moment. This approach allowed us to track the relationship between musical features, perceived tension, and the dynamic unfolding of event structure in narrative experiences of music.

Previous studies have shown substantial intersubjective consensus on the stories implied by individual excerpts of instrumental music (McAuley, Wong, Bellaiche, & Margulis, 2021; Margulis et al., 2022). Within-culture participants could match which of two stories was plausibly generated by previous participants in response to the excerpt with 90% success rate, and approaches from natural language processing revealed that the semantic similarity between within-culture listeners at different geographic locations of narratives produced was significant for 84% of the excerpts. The paradigm used in these studies was limited to post-excerpt report, and on first account, it might seem unlikely that a more real-time method like the one used in this study (event clicks as the excerpt progressed) would reveal significant consensus in the time points within each excerpt at which participants imagine particular narrative events in their imagined story. Indeed, if particular excerpt characteristics cued generic associations which participants then, in a post-hoc process, developed into stories, we would expect people's performance on the event click task to be randomly distributed across the time course of each excerpt, varying between individuals and yielding limited patterns across participants. On the other hand, if imagined stories unfold dynamically as a person listens, with imagined events shaped by the dynamic unfolding of excerpt features, then event clicks should consistently occur at specific moments in each excerpt—the moments where acoustic features cue new perceived events in the ongoing imagined narrative. Since the tension ratings provide a kind of "ground truth" about moments when acoustic features prompt changes in affective perception, we further hypothesized that if acoustic cues drive the temporal structure of imagined narrative events, then event clicks should coalesce around moments when perceived tension changes.

The musical excerpts examined in this study varied in narrativity: the degree to which people tended to imagine a story while listening. Of interest were potential differences in intersubjectivity about "when" in a specific excerpt of music listeners' imagined narrative events. We hypothesized that higher narrativity excerpts would yield greater temporal consistency (i.e., intersubjectivity in the timing of narrative events) than lower narrativity excerpts. Excerpts also varied in two additional attributes: topicality (the degree to which patterns in the excerpt are conventionally paired with extra-musical referents) and contrast (the degree to which salient aspects of the music change across the course of the excerpt) (McAuley, Wong, Mamidipaka, et al., 2021). Given the broader correlation between change or surprise and event perception (Hard et al., 2019) as well as ideas from music theory about narrative perception (Almén, 2017), we hypothesized that contrast and topicality would influence the dynamics of event perception, but in potentially different ways. We hypothesized that greater contrast in an excerpt would trigger more perceived events (i.e., more button presses), while topicality, which captures the extent to which an excerpt uses patterns with conventional extra-musical referents, would be the stronger driver of the likelihood for people to hear narrative events at consistent temporal locations in the excerpt.

2. Methods

2.1. Participants

Sixty-nine participants ($n = 41$, female), 18–37 years of age ($M = 20.1$, $SD = 3.6$) from the University of Arkansas community took part in the first experimental session. Fifty of the participants ($n = 27$, female),

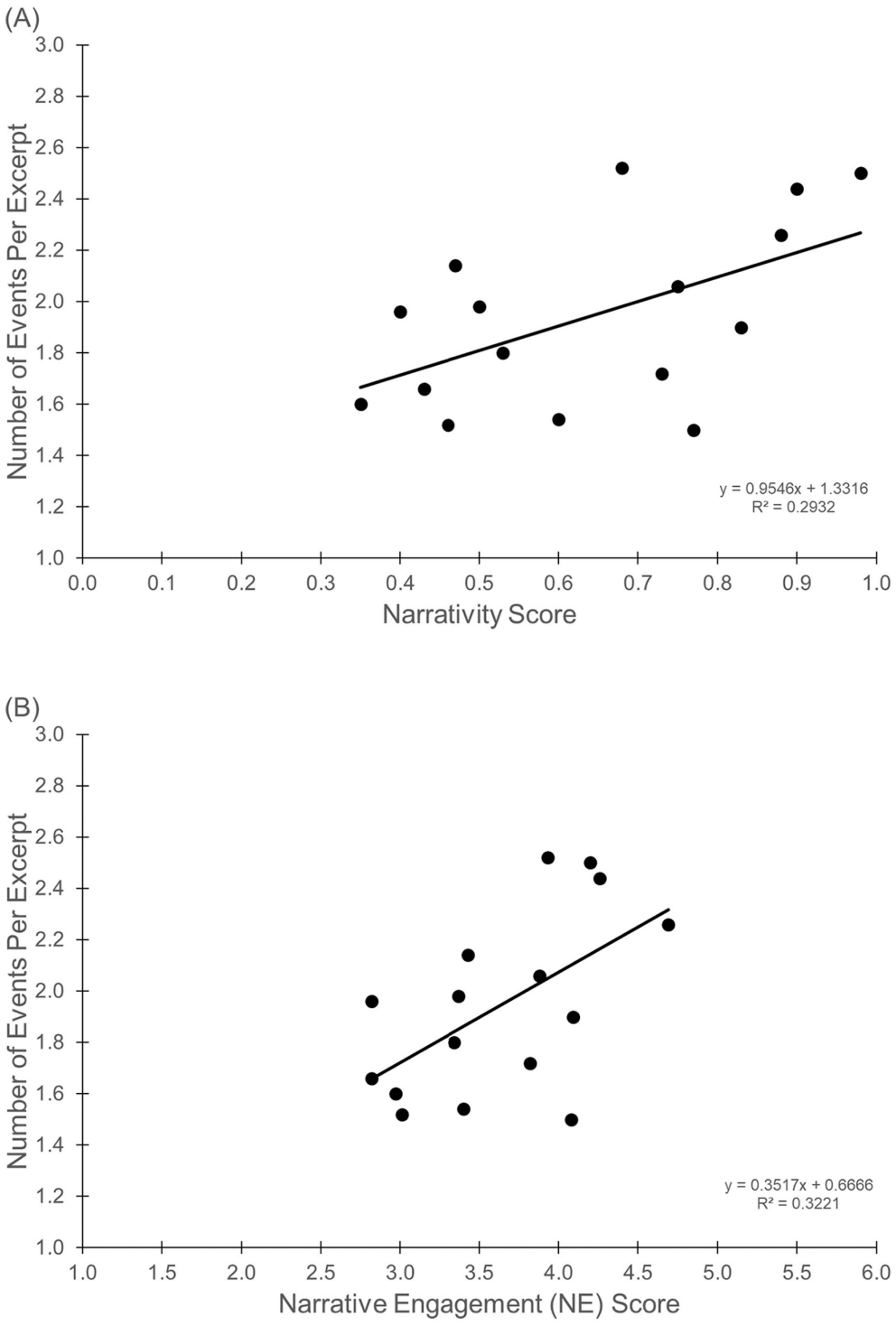

Fig. 1. The relation between normative measures of excerpt narrativity (the likelihood that listeners report imaging a story in response to the excerpt) and average number of marked narrative events per excerpt (Panel A) and between normative measures of excerpt narrative engagement and average number of marked events per excerpt (Panel B). Both narrativity and narrative engagement were positively correlated with the number of marked narrative events.

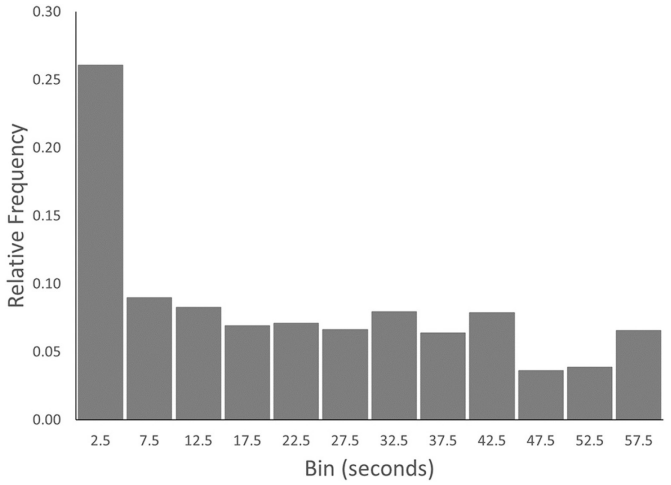

Fig. 2. Relative frequency histogram of the time points of all marked events (mouse clicks) combining all excerpts and participants.

18–37 years of age (M=20.2, SD =5.9) returned for a second experimental session several weeks later. Of the fifty participants who completed both experimental sessions, 50% reported no music training, while 50% report having some formal music training. Years of formal music training ranged between 0.5 and 13 years (M= 5.3, SD= 3.6 years). All participants reported normal hearing. The study was approved by the University of Arkansas Institutional Review Board and informed consent was obtained from all participants. This study was not preregistered.

2.2. Materials

Stimuli were sixteen 60s instrumental musical excerpts selected from a set of 128 60s excerpts for which we have normative measures from 321 participants in Arkansas and 141 participants in Dimen of enjoyment, familiarity, narrativity (the likelihood that the excerpt generates a story in listeners' minds), and narrative engagement (how vividly any imagined story engages the listener), as well as ratings of topicality and musical contrast from 3 expert music theorists blind to the hypotheses of the study (Margulis et al., 2019); see Table 1. Topicality captures the degree to which an excerpt makes use of sonic patterns conventionally associated with particular extra-musical referents (e.g. descending arpeggios for a dream sequence), and contrast captures the degree to which an excerpt features salient changes (rather than consistency and continuation) in dynamics, pitch, tempo, timbre, rhythm, and articulation. All excerpts were drawn from commercially available recordings of wordless Western classical music, beginning at a part of the track that permitted extraction of a one-minute stretch with a clearly segmented beginning and ending point. The sixteen excerpts were classified into two categories based on their narrativity score from Margulis et al. (2019). Eight excerpts were High Narrativity excerpts that are frequently heard in terms of a narrative, while eight excerpts were Low Narrativity excerpts that are less frequently heard in terms of a narrative. Narrativity was operationally defined as the likelihood that the excerpt triggered a story in listeners' minds based on normative responses to the story response question for each excerpt (Did you imagine a story when listening to the excerpt) in Margulis et al. (2019). Based on the normative data, the likelihood that listeners imagined stories was significantly greater for High Narrativity excerpts (M = 0.81, SD = 0.10) than for Low Narrativity excerpts (M = 0.47, SD = 0.08), $t(14) = 7.6$, $p < 0.001$, Cohen's d =3.82. Narrative Engagement scores and expert ratings of Topicality were also significantly greater for High Narrativity excerpts (NE: M = 4.12, SD = 0.28; Topicality: M = 4.46, SD = 1.37) than for Low Narrativity excerpts (NE: M = 3.14, SD = 0.27; Topicality, M = 2.17, SD = 0.64), $t(14) = 7.2$, $p < 0.001$, Cohen's d = 3.58, and $t(14)$ = 4.32, $p < 0.001$, Cohen's d =2.16, respectively. Across the excerpts, Narrativity, Narrative Engagement, and Topicality were positively correlated (Narrativity v. NE, r=0.94, $p < 0.001$; Narrativity v. Topicality, r=0.76, $p < 0.001$; NE v. Topicality, r=0.77, $p < 0.001$). High Narrativity and Low Narrativity excerpts did not differ in enjoyment, familiarity, or expert ratings of Contrast (all p's > 0.05). Across the excerpts, Contrast was not correlated with Narrativity, NE, or Topicality (all p's > 0.83). Data and materials are available upon request.

2.3. Equipment

Participants listened to the stimuli through binaural, around-ear Sennheiser HD600 headphones, and made responses using a using a computer keyboard, mouse, and DirectIN Rotary Controller (PCB v2014) while individually seated at a Dell OptiPlex 7010 desktop computer (22-in. Dell P2212H monitor; Windows 7) in a 4-ft 4-ft WhisperRoom Sound Isolation Enclosure (MDL 4848E/ENV. The experiment was presented using MediaLab and DirectRT software (Version 2016.1.108).

2.4. Procedure

Participants completed two experimental sessions separated by several weeks at the Music Cognition Lab at the University of Arkansas. They completed two tasks in each session. The first task in Session 1—which was the focus of a separate study and not reported here—was to estimate the durations of each of the sixteen musical excerpts (8 high narrativity, 8 low narrativity). Next, participants listened to same sixteen excerpts again in a random order and made continuous judgments of musical tension while listening. Tension was defined to participants as the anticipation music creates in a listener's mind for relaxation or release. While listening to each musical excerpt, participants turned a dial to the right to indicate that the excerpt was increasing in tension, and turned the dial to the left to indicate that the except was decreasing in tension. The dial was reset to the lowest value (all the way to the left) before each trial, meaning that all tension ratings started from the zero point before the start of the excerpt. Finally, participants completed a brief demographic questionnaire. Session 1 lasted approximately 50 min.

In the second session, which occurred several weeks later, participants listened to the 16 musical excerpts (8 High Narrativity and 8 Low Narrativity) in a random order, with each excerpt played twice in succession. On the first hearing, participants were asked to type a detailed description of any story they had imagined while listening into an open-ended text box after the excerpt concluded. On the second hearing, participants were asked to recall the story they had imagined on the first hearing, listen to the same excerpt again, and click a mouse as the music progressed to indicate the moments at which each event occurred in their imagined story. If they did not imagine a story during the first hearing, they were instructed to not mark any events (not produce any mouse clicks) during the second hearing. Once the excerpt concluded, participants typed a detailed description into an open-ended text box for each event they had imagined at each time point clicked. Session 2 lasted approximately 50 min. The results reported below focus on the quantitative data, with the free-response, qualitative data examined in a separate manuscript.

3. Results

The presentation of the results is in two sections. The first section examines the dynamics of online perception of imagined narrative events during music listening, as measured by participants' use of mouse clicks to mark time points in the music where each imagined story event began. We were interested in (1) the overall number of story events indicated for High and Low Narrativity excerpts, (2) the temporal

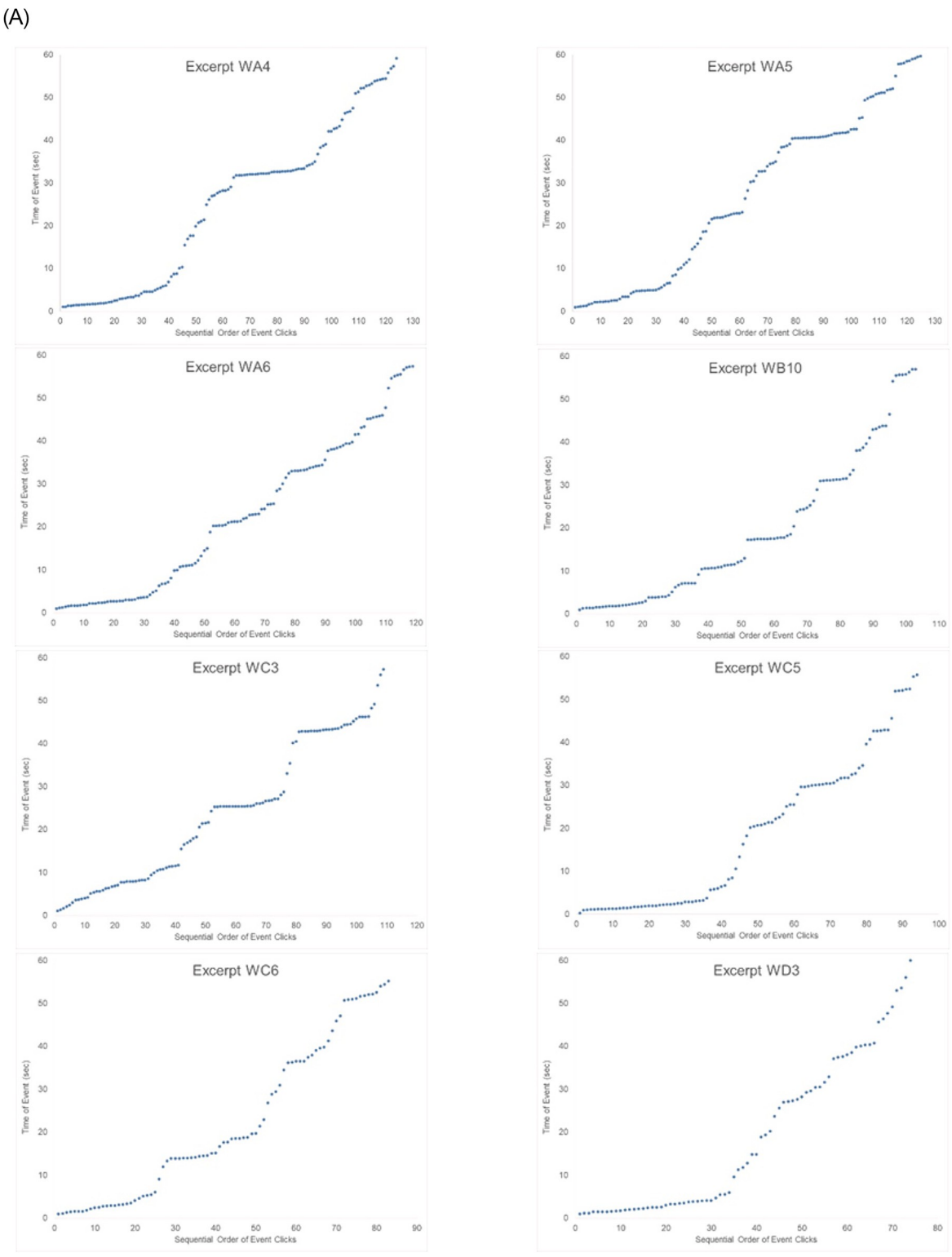

Fig. 3. Temporal distribution of marked story events for High Narrativity excerpts (Panel A) and Low Narrativity excerpts (Panel B). The time points of all mouse (event) clicks for a musical excerpt across participants were first sorted according to whenintime each event click occurred. The sorted event-click data is shown with the temporal order of the event clicks plotted on the x-axis and the time of the event on the y-axis. If participants tend to mark an event as occurring at approximately the same time point during listening to an excerpt, then the points cluster on the y-axis and appear in the graph visually as horizontal line segments.

(B)

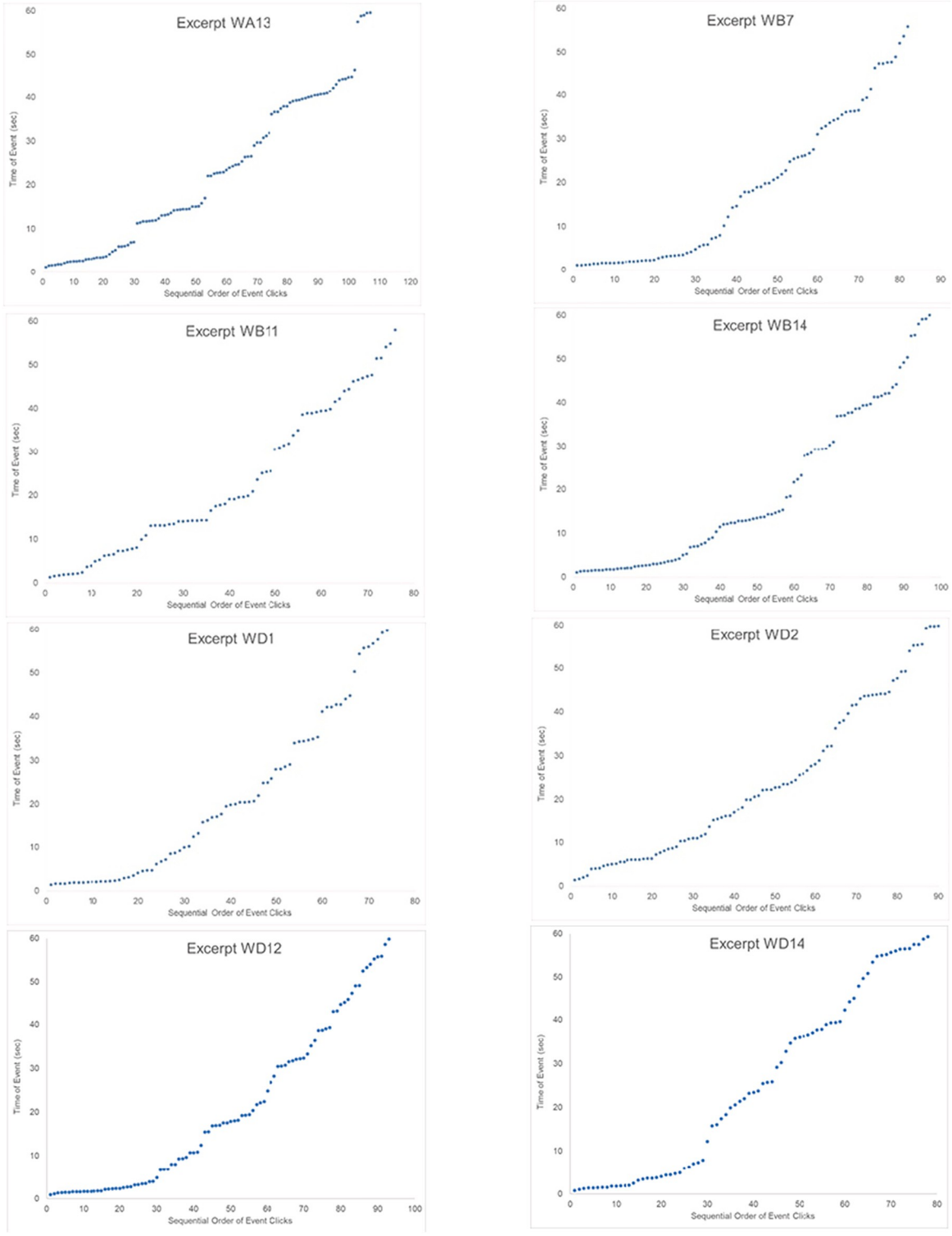

Fig. 3. (*continued*).

distribution (and consistency) of those events, and (3) the relation between the dynamics of event perception to normative measures of narrative engagement, topicality, musical contrast, enjoyment and familiarity. The latter measures of enjoyment and familiarity, equated across High Narrativity and Low Narrativity excerpts, were included to rule out their contribution to participants' perception of story events during music listening. The second section examines continuous ratings of musical tension for the High Narrativity and Low Narrativity excerpts and their relation to the dynamics of online event perception.

3.1. Dynamics of event perception

Across the sixteen excerpts, participants identified (via mouse clicks) between zero and five story events while listening to an excerpt during the on-line event perception task (M = 1.94 events per excerpt, SD = 0.94). Supporting the idea that excerpts with higher narrativity would engender a higher number of imagined events, participants indicated that there were more story events (by making more mouse clicks) during the online event perception task when listening to the High Narrativity excerpts (M = 2.11 events per excerpt, SD = 0.97) than when listening to the Low Narrativity excerpts (M = 1.77 events per excerpt, SD = 1.04), $t(49) = 3.09$, $p = 0.003$, Cohen's d = 0.44). Moreover, the average number of marked events (mouse clicks) per excerpt was positively correlated with the excerpt's narrativity score from Margulis et al. (2019; the likelihood that the excerpt generates a story in listeners' minds), $r(14) = 0.54$, $p = 0.03$, and was also positively correlated with the normative narrative engagement (NE) score for the excerpt, $r(14) = 0.57$, $p = 0.02$; see Fig. 1. The average number of marked events was also positively correlated with Contrast, $r(14) = 0.57$, $p = 0.02$, and Topicality, $r(14) = 0.50$, $p = 0.05$.

With respect to the timing of a participant's mouse clicks, participants did not tend to identify events in quick succession; instead, their marked events were distributed over the temporal extent of each excerpt. The minimum time between successively marked events was >2.5 s for 72% of participants, while an additional 18% marked only one event within 2.5 s of a preceding event and another 6% marked only 2 or 3 events within 2.5 s of a preceding event across all sixteen excerpts. Thus, for 96% of participants, mouse clicks were distributed over the time course of each excerpt. The final 4% of participants (≠ 2) made approximately 50% of their mouse clicks within 2.5 s of a preceding event, which was very atypical. Of note, these two participants made most of their mouse clicks at the start of each excerpt.

Fig. 2 shows a relative frequency histogram of the time points of all marked events (mouse clicks) combining all excerpts and participants. Two features of the overall temporal distribution of marked events stand out. First, there was a general tendency for participants to identify narrative events (make mouse clicks) at the beginning of each excerpt. This was true even when excluding the two participants that tended to identify narrative events only at the start of each excerpt. Second, after the start of the excerpts, there was a tendency across excerpts for participants to produce a relatively uniform distribution of marked events over the temporal extent of the excerpts. Thus, when responses to all excerpts are combined, participants do not show any general structure in the temporal distribution of their clicks across the temporal extent of a piece of music, other than the tendency to mark an event at the beginning; aside from that first click, marked events had a relatively uniform distribution over the temporal extent of the excerpt.

However, when we considered the temporal distribution of marked events for each excerpt separately, rich excerpt-specific temporal structure emerged. As an initial consideration of whether there was shared consensus about 'when' narrative events occurred in each excerpt, Fig. 3 shows the sorted time points of marked events (mouse clicks) for each excerpt with the temporal order of the mouse clicks on the x-axis and the time of the marked event on the y-axis; Panels A and B show the resulting graphs for the eight High Narrativity and Low Narrativity excerpts, respectively. If participants mark an event as occurring at approximately the same time point while listening to an excerpt, then the points cluster on the y-axis and appear in these by-excerpt graphs visually as horizontal line segments. The length of these line segments reflects the inter-subject consistency with which participants heard an event around that time point; longer (and flatter) line segments reflect greater temporal consistency across participants (intersubjectivity) in the timing of a perceived event than shorter (or more positively sloped) line segments. A diagonal line without any step-like shape indicates littletono consistency in the time points at which listeners marked events for that excerpt.

Visual inspection of Fig. 3 reveals a wide range of step patterns, indicating that participants agreed in many instances 'when' narrative events were occurring in the music. Some excerpts, such as WA4, have only a few steps, each of which are relatively long, indicating broad consensus around a small number of time points at which events were imagined to occur. Others, such as WB10, show a series of more numerous shorter steps, indicating that there was a larger number of discrete, agreed-upon time points at which participants tended to imagine events. Others still, such as WD2, show little-to-no step-like pattern in the temporal distribution of events, indicating little agreement across participants about the time points in the music at which they imagined events while listening.

To identify the time points of identified narrative events that showed temporal consensus across participants, we first counted the number of mouse clicks by distinct participants in a sliding 2.5 s window that started at the beginning of the excerpt and then was shifted in 0.1 s increments until the end of the excerpt. Fig. 4 shows the resulting distributions for the High Narrativity excerpts (Panel A) and Low Narrativity excerpts (Panel B). Peaks in these graphs represent clusters of mouse clicks by different participants. A temporally-consistent marked event was operationally defined as a local peak having greater than five mouse clicks within the 2.5 s window with no higher peak within a 2.5 window of the local peak. For each excerpt, we measured the number of consistent events by counting the local peaks that met the above criteria and their degree of temporal consistency by the number of mouse clicks associated with each local peak. There were, on average, the same *number* of temporally consistent events identified for the High Narrativity excerpts (M = 4.75, SD = 1.39) and the Low Narrativity excerpts (M = 4.25, SD = 0.89), $t(14) = 0.86$, $p = 0.41$, Cohen's $d = 0.42$ (see Fig. 5A), but the average *degree* of temporal consistency of these marked events was significantly greater for High Narrativity excerpts (M= 15.1, SD = 2.6) than for the Low Narrativity excerpts (M = 11.5, SD = 1.7), $t(14) = 3.31$, $p = 0.005$, Cohen's $d = 2.4$ (see Fig. 5B).[1]

Finally, we examined the relationship between the number and degree of temporally consistent events with normative measures of narrative engagement, topicality, contrast, enjoyment and familiarity (see Table 2). This analysis revealed that the *number* of temporally consistent story events in each excerpt was associated with contrast ratings (i.e., excerpts with more musical contrast produced a greater *number* of temporally consistent marked events), $r(14) = 0.55$, $p = 0.027$, but was not related to any of the other measures (all p's > 0.13). On the other hand, the mean degree of temporal consistency of the marked events (with higher scores indicating greater consistency/consensus about *when*) was reliably associated with Narrativity, $r(14) = 0.71$, $p = 0.002$, Narrative Engagement, $r(14) = 0.74$, $p = 0.001$, Topicality, $r(14) = 0.73$, $p = 0.001$, and Enjoyment, $r(14) = 0.54$, $p < 0.05$, but not with the other measures. Thus, excerpts that were more likely to trigger stories in listeners' minds and were more narratively engaging led to more consistent temporal marking of events across participants,

[1] A similar pattern was observed for different choices of threshold over an intermediate range. However, given the greater degree of temporal consistency of marked events for High Narrativity excerpts compared to Low Narrativity excerpts, a *sufficiently* high threshold would also yield differences in the number of temporally consistent events across High and Low Narrativity excerpts.

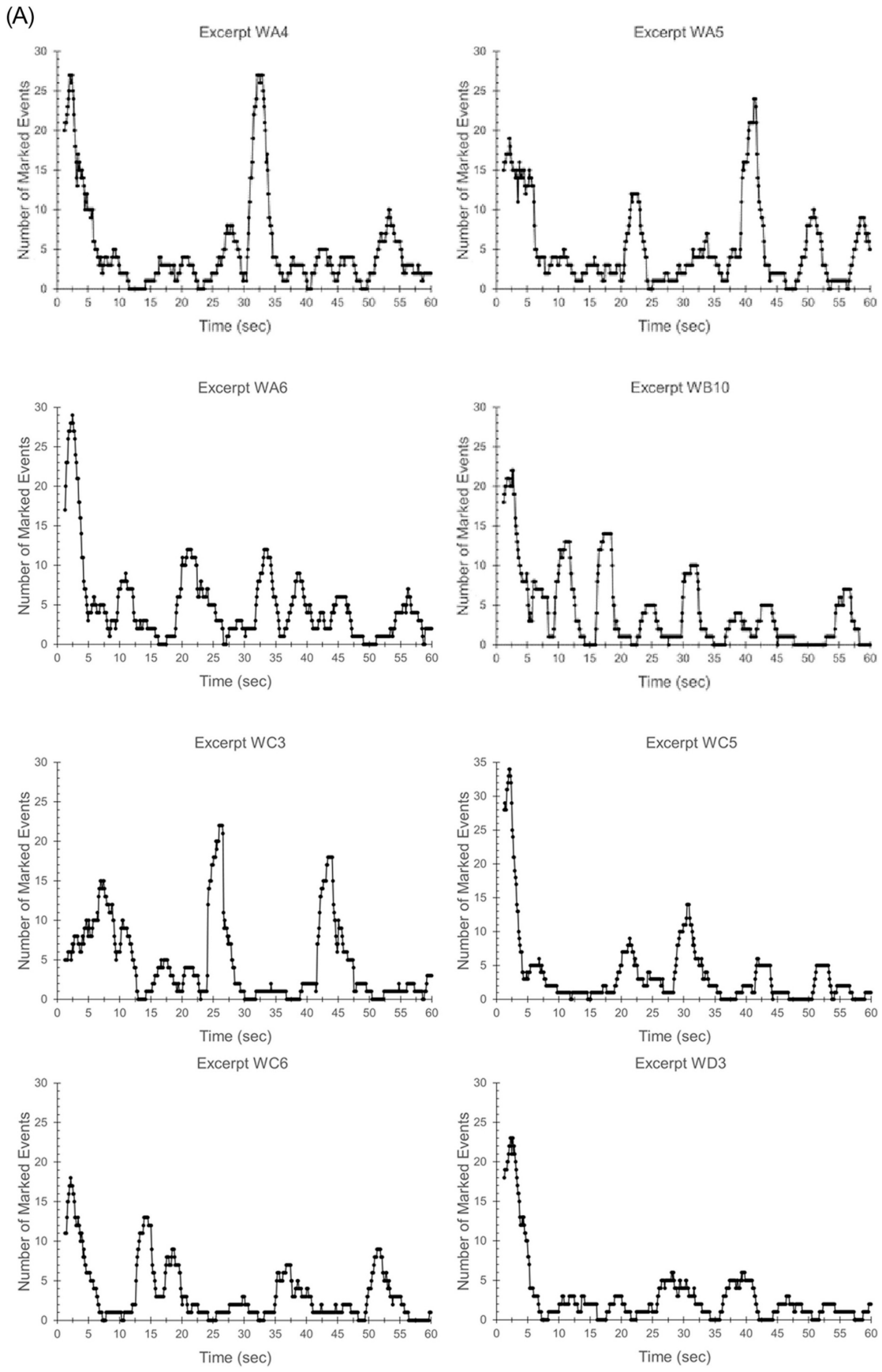

Fig. 4. Temporal consistency in marked story events across the temporal extent of each excerpt. Panel A shows High Narrativity excerpts and Panel B show Low Narrativity excerpts.

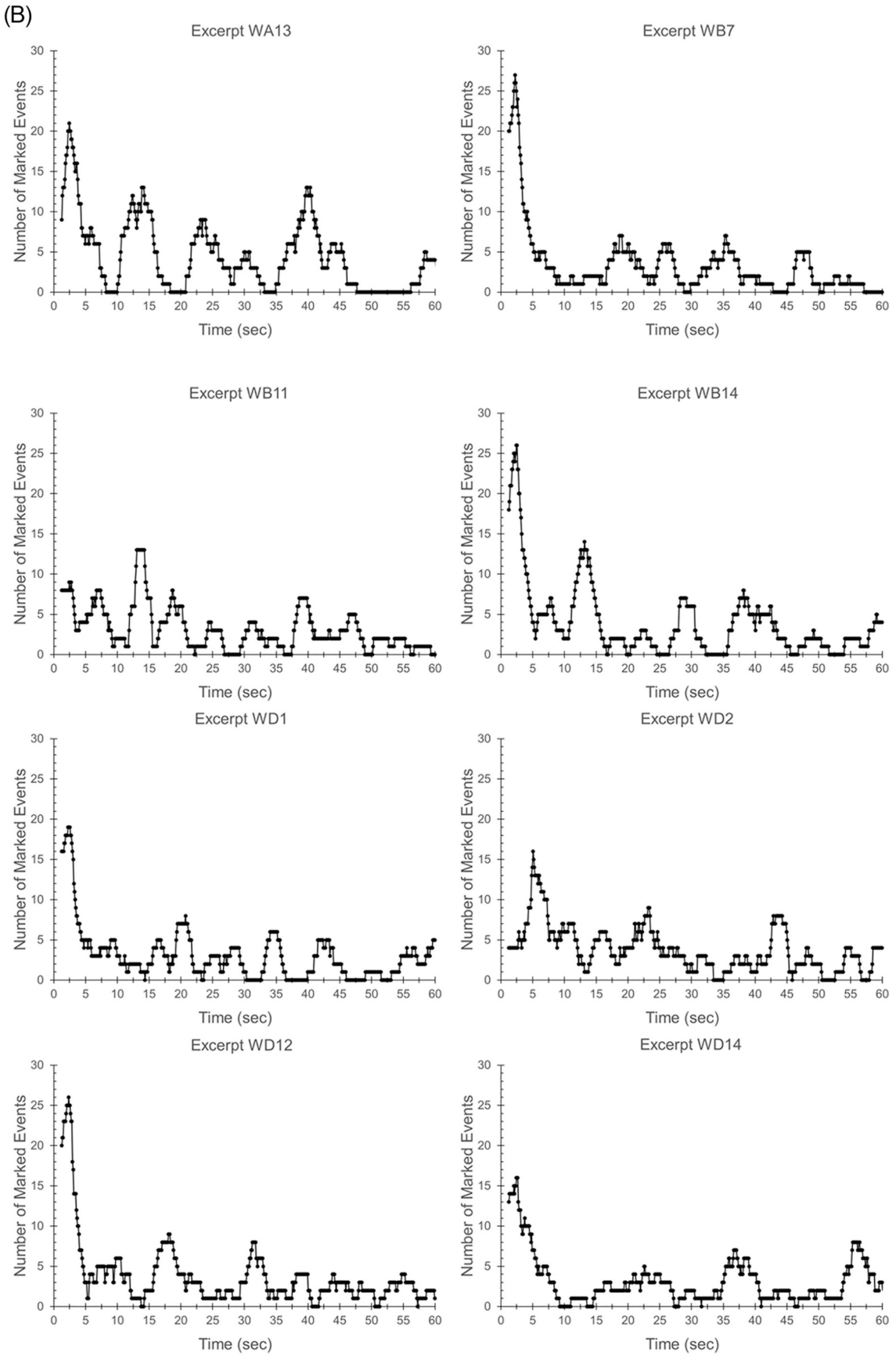

Fig. 4. (*continued*).

but not necessarily to have *more* perceived story events.

3.2. Tension and narrative dynamics

Fig. 6 shows the aggregate tension profiles from the continuous tension-rating task (averaged across participants) plotted as a function of time for the High Narrativity excerpts (Panel A) and the Low Narrativity excerpts (Panel B). First, we asked how reliable (consistent) participants' generated tension curves were by calculating the intra-correlation coefficients (ICCs) across participants for each excerpt. ICC estimates of inter-rater-reliability were calculated using a 2-way random-effects model for mean raters assessing consistency across raters (Koo & Li, 2016). Consistency was measured instead of absolute agreement because we were interested in whether different participants would produce similar tension profiles for each excerpt, not whether participants turned the knob to exactly the same tension value at each time point. Mean-rater ICCs for the tension ratings were very high for all excerpts (M = 0.98, SD = 0.02, Range = 0.93–0.99), indicating excellent reliability.

Overlaid on the tension curves for each excerpt are red circles that indicate the time points of temporally consistent events from the event perception task. The size of the circles represents the degree of temporal consistency. The smallest circle indicates that the marked event exceeded the minimum threshold for the operational definition of a temporal consistent event (> 5 marked events within the 2.5 s window). The medium size circle means that the marked event exceeded a threshold that was twice as large (> 10 marked events within the 2.5 s window) and the largest sized circle means that the marked event exceeded a threshold that was four times as large (> 20 marked events within the 2.5 s window). Visual inspection of the graphs revealed for many of the excerpts a surprisingly high degree of alignment between the time points of consistently marked events and abrupt changes in tension (corresponding to local minima, local maxima, or major inflection points in the tension curves).

To quantify the degree of alignment between temporally-consistent marked events and these features of the tension curves, we delineated each excerpt into twenty-four 2.5 s bins. For each bin, we coded (yes or no) to indicate (1) whether there was a red circle indicating a temporally consistent event and (2) whether the tension curve was at a local minima, local maxima, or major inflection point. We then determined the number of hits (number of bins in which an event was marked that corresponded to a local minima, maxima, or inflection point) and the number of false alarms (number of bins in which an event was marked and there was not at local minima, maxima, or inflection point). A proportion correct (PC) score was then calculated from the hit rate (H) and false alarm rate (F) as follows: PC = ½ [H + (1 – F)]. Overall, PC was significantly greater than chance (M = 0.70, SD = 0.11), $t(15) = 7.36$, $p < 0.001$, Cohen's d = 1.84, and the degree of alignment was also somewhat greater for High Narrativity excerpts (M = 0.74, SD = 0.11) than Low Narrativity excerpts (M = 0.66, SD = 0.09); the PC difference, however, between the two excerpt categories was not significant, $t(14) = 1.65$, $p = 0.12$, Cohen's $d = 0.83$.

In general, PC varied over a fairly broad range across excerpts (PC Range: 0.43–0.86). To further investigate the factors that might influence the degree of match between the time points of marked events and points of change in the tension curves across excerpts, the two categories of excerpt were combined and a correlational analysis was conducted between PC scores and normative measures of Narrativity, Narrative Engagement, Topicality and Contrast for each of the excerpts. Supporting the trend we observed for High Narrativity excerpts to produce higher PC values than Low Narrativity excerpts, PC was positively correlated with the normative measure of excerpt Narrativity, $r(14) = 0.56$, $p = 0.02$. PC was also positively correlated with Topicality, $r(14) = 0.54$, p = 0.03, but not Contrast, $r(14)$ = 0.23, p = 0.39. It is further notable that PC was not correlated with either excerpt Enjoyment, $r(14) = 0.06$, $p = 0.83$, or Familiarity, $r(14) = 0.03$, $p = 0.92$.

Overall, these analyses show that listeners tend to hear story events at points in the music that have an abrupt change in tension for excerpts to which participants readily narrativize. For example, excerpt WA4 starts with a clarinet passage; approximately 33 s in, a lush melody enters in the violin, ushering in a period of textural thickening and dynamic intensification that peaks and begins to subside around 52 s. At each of these inflection points—when the tension begins to sharply increase at around 33 s and when it begins to subside around 52 s—participants tended to report the onset of a new imagined event in their ongoing imagined narrative. Similarly, for excerpt WA6, participants frequently reported imagining a new event 20 s in, when a new, heavily accented chromatic ascent begins, and around 32 s in, when that ascent concludes and a new, smoother one begins. Both moments also serve as inflection points for perceived tension. In excerpt WC5, people tended to frequently imagine events starting around 21 s in, when a new, louder, rhythmically insistent passage enters from the brass and tension starts to increase rapidly, and at 30 s in, when a smoother, quieter melody begins in the winds, and tension rapidly drops. In each of these examples, salient moments in the music are linked both with changes in perceived tension and the imagination of a new event in listeners' ongoing imagined stories.

4. Discussion

The current study investigated the dynamics of listeners' narrative perceptions of wordless instrumental music using a real-time task: listeners clicked a mouse to mark the time points at which they imagined new narrative events as the music unfolded. There were three main questions. First, of primary interest, was whether listeners would show intersubjectivity about 'when' narrative events occurred over the temporal extent of each excerpt, or whether narrative perceptions of music would be atemporal (i.e., listeners imagine stories in response to the music, but not as an unfolding sequence of events that is time-locked to particular points in the music, exhibiting little to no intersubjectivity about 'when'). Second, we were interested in the relationship between story dynamics and musical excerpt narrativity (the likelihood that a story is generated in listeners' minds), as well as between story dynamics and normative measures of excerpt narrative engagement, topicality, contrast, enjoyment and familiarity. Finally, we were interested in the question of whether listeners would tend to hear story events at time points in the music that featured changes in musical tension.

There were four main findings. First, for the real-time event marking task, the number of generated mouse clicks (marked story events) was greater for High Narrativity excerpts than for Low Narrativity excerpts, as well as greater for more narratively engaging than less narratively engaging excerpts, supporting the validity of the distinction of the two categories of excerpt. Second, with respect to the primary question of interest, *when* listeners imagine story events is not randomly distributed throughout each excerpt, but rather shows substantial intersubjectivity about 'when' – with event markings clustered at specific time points within each excerpt. Moreover, the extent of clustering at particular points in the music was greater for the High Narrativity excerpts than for the Low Narrativity excerpts, as well as for excerpts that were more narratively engaging than less narratively engaging. This structure only emerges within individual excerpts; as Fig. 2 shows, when responses to all stimuli are pooled, event clicks are relatively uniformly distributed throughout the 60s period, aside from a concentration of clicks right at the beginning. The fact that event clicks tend to coincide temporally for excerpts that are experienced more narratively suggests that the more a listener is absorbed in the ongoing imagined story implied by a piece, the more discrete acoustic cues drive the temporal structure of their imagined narrative. In the analysis of temporal consistency, a third finding emerged that suggests a potential dissociation between the number of temporally consistent event markings and the degree of consistency (i.e., strength of clustering) among them. The *number* of story events was robustly correlated with normative ratings of contrast,

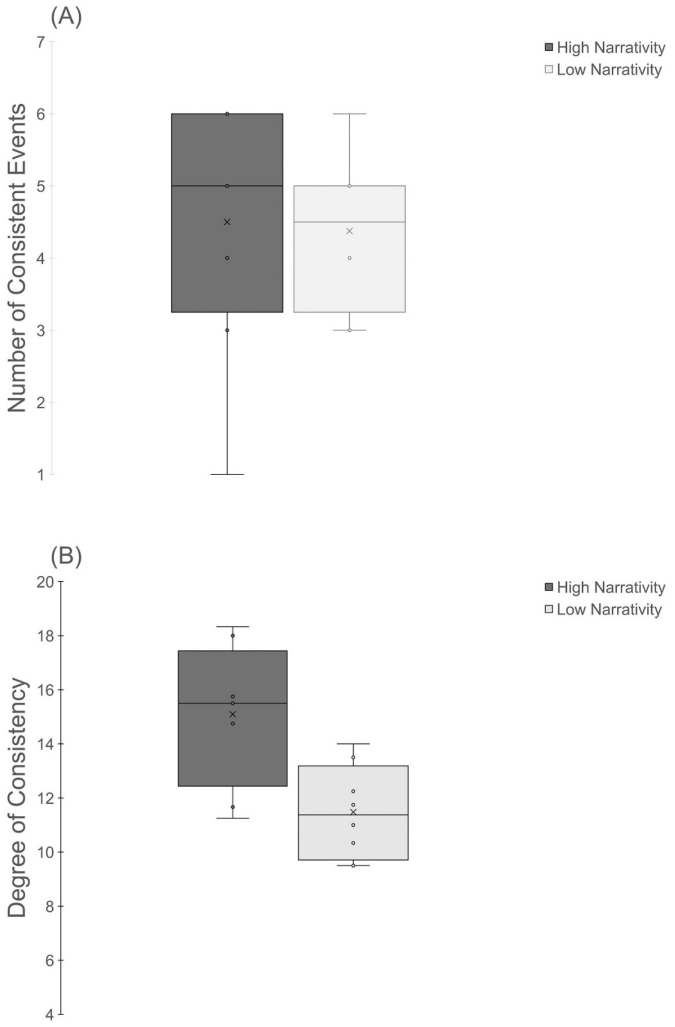

Fig. 5. Box and whisker plots showing number of consistently marked events (Panel A) and degree of temporal consistency (Panel B) for High Narrativity and Low Narrativity excerpts. For degree of temporal consistency, lower values indicate greater temporal consistency. The number of events did not significantly differ between High Narrativity and Low Narrativity excerpts ($p = 0.41$), but marked events showed greater temporal consistency for High Narrativity excerpts than Low Narrativity excerpts ($p = 0.005$).

but not topicality, whereas the *degree* of consistency was robustly correlated with topicality, but not contrast. Fourth, we found a high degree of alignment between the time points of consistently marked events and abrupt changes in tension (corresponding to local minima, local maxima, or major inflection points in the tension curves) for many of the excerpts. Moreover, the strength of alignment was positively correlated with the normative measures of excerpt narrativity and topicality, but not contrast, suggesting that only when excerpts vividly absorb listeners in an imagined narrative do acoustic cues tightly drive both imagined and affective responses in a temporally structured way.

Together, these findings are the first to demonstrate that the inception of discrete events in listeners' ongoing imagined narratives are not atemporal, but show substantial intersubjectivity about 'when' in the music specific narrative events occur. In a classic study, people stared at a supposedly blank wall and sustained what felt like imaginings. These imaginings were actually shown, however, to be guided by the subconscious perception of faint images (Perky, 1910). The current study brings a new perspective to the relationship between perception and imagination, showing that the dramatic imaginings people sustain while listening to music actually reveal the influence of dynamic, moment-to-moment changes in the music. If the specific musical excerpts had not been driving the time course of event perception, people should have reported events at temporal intervals that were randomly distributed across the temporal extent of the excerpts, but this is not what they did. Instead, people tended to consistently report events at specific, but different, time points in each musical excerpt. This finding reveals that narrative events imagined in music are not atemporal, but are shaped by the music in systematic ways as the music unfolds in time, demonstrating that perception can guide imagination dynamically in time.

Table 2
Pearson correlations between number and degree of temporally consistent events in each excerpt and normative measures of narrative engagement (NE), contrast, topicality, enjoyment and familiarity (* = $p < 0.05$; ** = $p < 0.01$).

Temporally consistent events	NE	Contrast	Topicality	Enjoyment	Familiarity
Number	−0.09	**0.55***	0.14	0.002	0.40
Degree of consistency	**0.74****	0.41	**0.73****	**0.54***	0.26

Event perception is an important aspect of narrative processing; the more that things seem to happen, the more narratively a particular scenario is understood (László, 2008). The current study reveals that people register more discrete events for high-narrativity excerpts—that is, for excerpts to which many people imagine a story when listening—and for more narratively engaging excerpts—that is, the ones for which the imagined stories tended to be most vivid and engrossing. The positive correlations observed between our real-time measure of event perception and normative ratings of narrativity and narrative engagement for each musical excerpt further bolster the idea that narrative imaginings to music are robust phenomena that unfold dynamically across the time course of listening and are mediated systematically by both how likely the music is to trigger a story in listeners minds and how narratively engaging that story is.

Music theorists postulate that salient changes in the music (moments of contrast) can rarely be made sense of by listeners in purely musical terms; such changes often compel listeners to understand them in terms of the imagined actions of an implied intentional agent (Almén, 2017). Indeed, participants in this study did mark more discrete events for higher contrast excerpts: the more frequently and noticeably the music switched from loud to soft, low to high, smooth to detached, etc. the more frequently participants clicked to report that they were imagining new discrete events in their ongoing imagined story.

But the degree to which these imagined events coalesce in time (degree of temporal consistency) does not appear to depend on contrast, but on topicality—an excerpt's tendency to draw upon patterns conventionally associated with an extra-musical referent (e.g. descending arpeggios on a harp that frequently connote entrance into a dreamlike state). People's event markings clustered more at specific time points for high topicality excerpts. When conventional referents exist within an excerpt, people are likelier to share a sense of *when* within that excerpt imagined events begin. Together, the relationship between number of events and contrast on the one hand and between degree of agreement at event time points and topicality on the other reinforces the notion that narrative responses to music vary along multiple dimensions. One type of musical feature (e.g. contrast) may inspire people to imagine a particularly eventful narrative, but another (e.g. topicality) may drive agreement about the moments that events occur regardless of how many are imagined. This delineation echoes the separation between the dimensions of narrativity, narrative engagement, and consensus that emerged in our previous studies (Margulis et al., 2019; McAuley, Wong, Bellaiche, & Margulis, 2021).

Although the contents of the story descriptions are reported elsewhere, taking a closer look at an illustrative example can help situate the results reported here. Excerpt WA6 was drawn from the first movement of Liszt's Dante Symphony, S. 109, and featured significant contrast

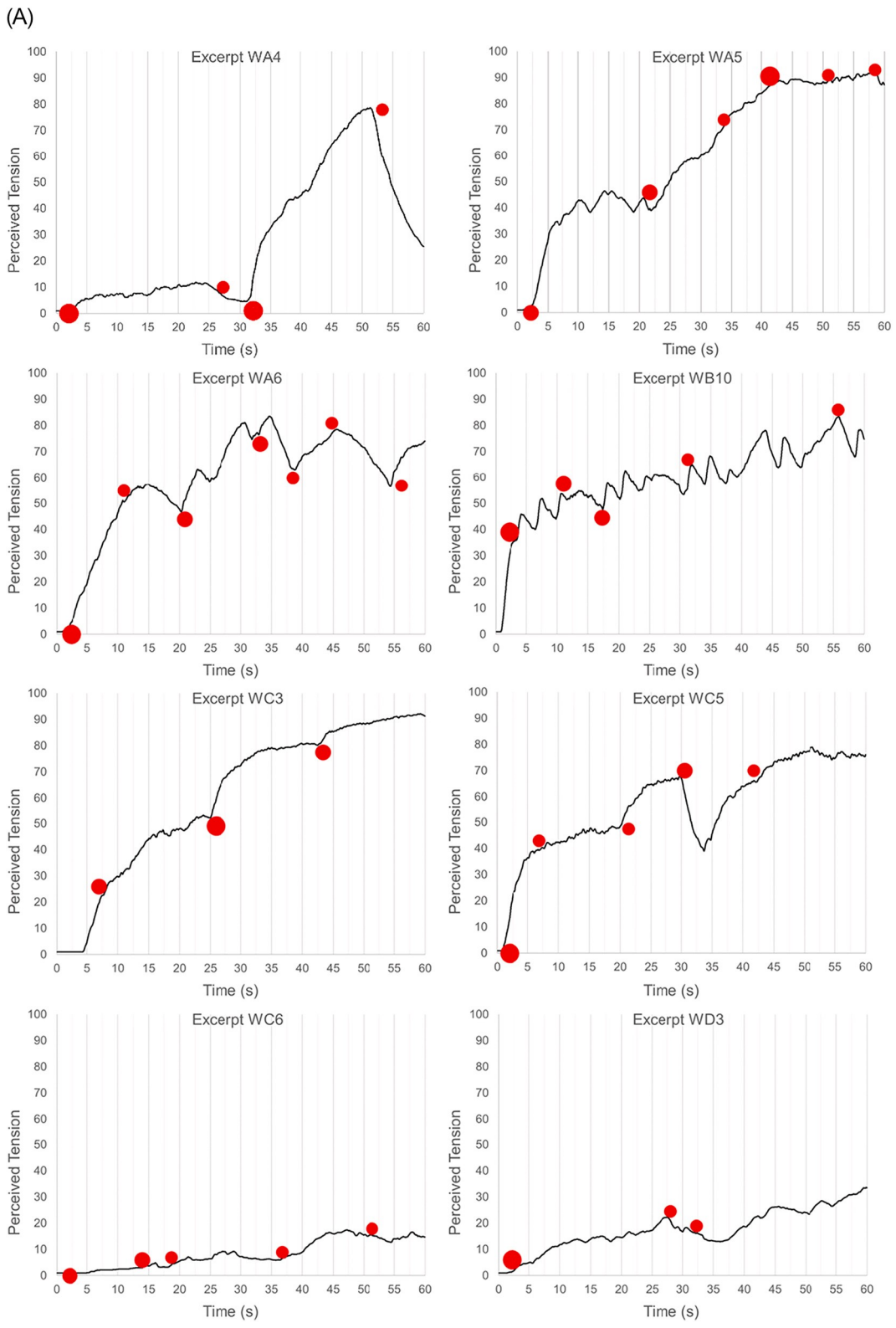

Fig. 6. Continuous tension ratings for High Narrativity excerpts (Panel A) and Low Narrativity excerpts (Panel B). The red circles indicate the time points of temporally consistent events from the event perception task. The size of the circles represents the degree of temporal consistency. The smallest circle indicates that the marked event exceeded the minimum threshold for the operational definition of a temporal consistent event (> 5 marked events within the 2.5 s window). The medium size circle means that the marked event exceeded a threshold that was twice as large (> 10 marked events within the 2.5 s window) and the largest sized circle means that the marked event exceeded a threshold that was four times as large (> 20 marked events within the 2.5 s window). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

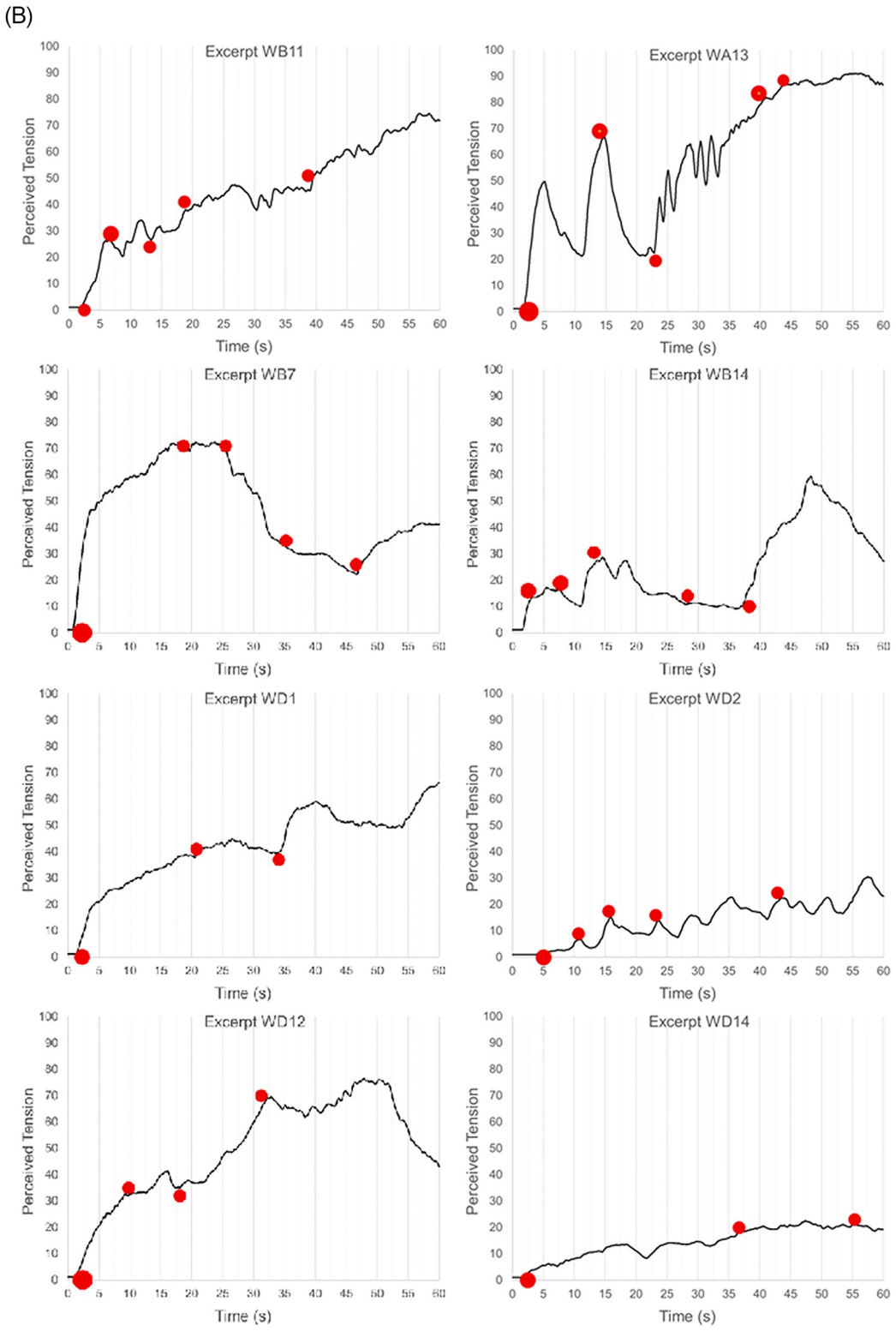

Fig. 6. (*continued*).

along many musical dimensions, with individual passages recurring after contrasting material intervenes. The story descriptions people provided after listening to the entire excerpt tended to articulate themes of threat and pursuit or chase, but they also tended to describe a series of repeated attempts and dodges that map onto the recurrence heard in the music itself. Story after story describes someone chasing or pursuing someone else (a cat chasing a mouse, a shark stalking a boat, a murderer pursuing their victim, a pirate corralling their captives), with repeated occurrences of pouncing and evasion. Descriptions of the discrete events imagined at particular time points confirmed that these narrative repetitions were imagined at moments where passages in the music repeated; moreover, the series of sub-arches in the tension graph align

with the individual episodes of perceived chasing and evading.

Similarly, in another High Narrativity excerpt, WA5, from Beethoven's Egmont Overture, narrative imaginings seem to relate tightly to perceived tension and musical structure. The excerpt begins with a few seconds of introductory material; it's not until this passage culminates on a loud, accented point of arrival that many participants indicate the beginning of a perceived narrative event. This is noteworthy because the beginning of the excerpt constituted the most frequently indicated time point for an imagined event onset in most excerpts; the delay here indicates that listeners were sensitive to the introductory nature of the initial passage. It's at this same moment that perceived tension in the excerpt beings to shoot upwards. The point at which perceived tension reaches its zenith and begins to level off, around 41 s, marks another time point at which many participants imagined a narrative event. Musically, this is the point at which the opening theme returns loudly and with a full orchestral texture after a long crescendo. Stories to this excerpt tended to feature tales of increasing menace, with the event recounted at 41 s constituting the moment of peak threat: a witch bursting into a wedding and cursing everyone involved; a boat tipping and sending its passengers overboard after a prolonged storm; a massive beast appearing; the enemies gaining the upper hand in the battle.

This relationship between perceived tension and perceived events occurred despite that participants were asked to indicate their fluctuating perceptions of musical tension and relaxation as the excerpts progressed during the first experimental session, before participants were asked to perform any task related to imagining stories (which did not occur until the second experimental session several weeks later). Although this tension-rating task has frequently been used in the literature (see Farbood, 2012 for an overview), it has not previously been understood in terms of drama or narrative (although see Jackendoff & Lerdahl, 2006 for ideas about potential connections between the two). Here, however, the tension profiles people produced in response to individual excerpts before they were asked to think explicitly in terms of narrative relate in concrete ways to how they understood the excerpt narratively at a subsequent session temporally removed from the first by several weeks. High-narrativity excerpts (excerpts in which a high percentage of people tended to imagine a narrative while listening) elicited higher average tension ratings, and furthermore, the time points at which people tended to indicate events occurring during a separate experimental session weeks later consistently corresponded to time points at which they reported salient changes in tension. These relationships suggest that perceived tension, long held to be a fundamental aspect of affective responses to music (Krumhansl, 2002), actually intertwine deeply with its narrative implications. Whether perceived tension provides a dramatic arc along which imagined stories are made to hang, or whether imagined stories themselves mediate and partially determine the perceived tension in any individual excerpt remains a topic for future research. Studies could use implicit tasks (e.g., the lexical decision task) to assess the real-time presence of narrative associations, or could examine the relationship between tension and other types of musical imaginings such as visualizations and autobiographical memories, or could use neuroimaging to tease apart the relevant neural substrates.

Taken together, the present study's findings are the first, as far as we are aware, to demonstrate that music consistently shapes imagined narratives (i.e., when listeners perceive events to happen) as the music unfolds in time with the degree of intersubjectivity about when a narrative event occured mediated by both the music's narrativity (i.e., the tendency to trigger a story in listeners minds) and the degree of narrative engagement. Moreover, the time points of marked events shows remarkable alignment for many of the excerpts examined here with points of abrupt tension change in the music. This work, thus, provides a new method for understanding how imagination is shaped by perception in temporal contexts. Innumerable questions for future inquiry fall out of this work. To what degree are imaginative episodes that feel highly subjective reliably guided by changing stimuli in the perceptual environment? Do other types of musical imaginings such as autobiographical memories show similarly shared temporal structure? How does the shared temporal orientation to these significant narrative moments connect to larger processes involving synchrony and social bonding (Kirschner & Tomasello, 2010; Savage et al., 2021)? The present study suggests that music is a particularly useful domain for exploring the relationship between imagination and perception in a rich, temporal context.

CRediT authorship contribution statement

Elizabeth Hellmuth Margulis: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Jamal Williams**: Formal analysis, Visualization. **Rhimmon Simchy-Gross**: Software, Formal analysis. **J. Devin McAuley**: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition.

Acknowledgements

This research was supported by the Division of Behavioral and Cognitive Sciences of the National Science Foundation, Award Numbers 1734025 (PI: EHM) and 1734063 (PI: JDM). The authors thank the members of the TAP Lab at Michigan State University, the members of the Music Cognition Lab at the University of Arkansas, and the members of the Music Cognition Lab at Princeton University for their many helpful comments on this work.

References

Abbott, H. P. (2020). *The Cambridge introduction to narrative*. Cambridge University Press.
Almén, B. (2017). *A theory of musical narrative*. Indiana University Press.
Alperson, P. (2004). The philosophy of music: Formalism and beyond. In P. Kivy (Ed.), *The Blackwell guide to aesthetics*. Blackwell Publishing.
Belfi, A. M., & Jakubowski, K. (2021). Music and autobiographical memory. *Music & Science, 4*.
Bell, J. S. (2011). *Elements of fiction writing: Conflict and suspense*. Writer's Digest Books.
Bruner, J. (1990). *Acts of meaning*. Harvard University Press.
Burroway, J. (2019). *Writing fiction: A guide to narrative craft*. The University of Chicago Press.
Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review, 4*, 213–230.
Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception, 29*(4), 387–428.
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., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34*(6), 1221–1235.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259.
Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*, 33–72.
Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior, 31*(5), 354–364.
Koelsch, S., Bashevkin, T., Kristensen, J., Tvedt, J., & Jentschke, S. (2019). *Scientific Reports, 9*, 10317.
Koo, T., & Li, M. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155–163.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*(3), 401–432.
Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11*(2), 45–50.
Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's divertimento no. 15. *Musicae Scientiae, 1*(1), 63–85.
László, J. (2008). *The science of stories: An introduction to narrative psychology*. Taylor & Francis.
Madsen, C. K., & Fredrickson, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface. *Journal of Music Therapy, 30*(1), 46–63.
Margulis, E. H., Wong, P. C. M., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications, 5*(1), 146.
Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded

intersubjectivity. *Proceedings of the National Academy of Sciences, 119*(4). e2110406119.
McAdams, D. P. (1993). *The stories we live by: personal myths and the making of the self*. Guilford Publications.
McAuley, J. D., Wong, P. C. M., Bellaiche, L., & Margulis, E. H. (2021). What drives narrative engagement with music? *Music Perception, 38*(5), 509–521.
McAuley, J. D., Wong, P. C. M., Mamidipaka, A., Phillips, N., & Margulis, E. H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition, 212*, 104712.
Nash, C. (2005). *Narrative in culture: The uses of storytelling in the sciences, philosophy and literature*. Taylor & Francis.
Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology, 21*(3), 422–452.
Rubin, D. C. (1995). *Stories about stories* (p. 164). Lawrence Erlbaum Associates, Inc..
Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences, 44*(E59), 1–36.
Schank, R. C., & Abelson, R. P. (1995). So all knowledge isn't stories? In R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 227–234). Hillsdale, NJ: Lawrence Erlbaum.
Schiffrin, D., De Fina, A., & Nylund, A. (2010). *Telling stories: Language, narrative, and social life*. Georgetown University Press.
Schooler, J. W. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences, 15*, 319–326.
Schwann, S., & Garsoffky, B. (2008). In J. M. Zacks, & T. F. Shipley (Eds.), *Understanding events: From perception to action* (pp. 391–414). Oxford University Press.
Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural dynamics of event segmentation in music: Converging evidence for dissociable ventral and dorsal networks. *Neuron, 55*(3), 521–532.
Stern, J. (2000). *Making shapely fiction*. W.W. Norton.
Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of sad and happy music on mind-wandering and the default mode network. *Scientific Reports, 7*, 14396.
Tuuri, K., & Eerola, T. (2012). Formulating a revised taxonomy for modes of listening. *Journal of New Music Research, 41*, 137–152.
Tversky, B., & Zacks, J. M. (2013). Event perception. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 83–94).
Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception, 23*(2), 137–152.
Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition, 112*(2), 201–216.
Zacks, J. M., & Swallow, K. M. (2007). Event Segmentation. *Current Directions in Psychological Science, 16*(2), 80–84.