

Hierarchical Beat Perception Develops Throughout Childhood and Adolescence and Is Enhanced in Those With Musical Training

Jessica E. Nave-Blodgett, Joel S. Snyder, and Erin E. Hannon
University of Nevada, Las Vegas

Most music is temporally organized within a metrical hierarchy, having nested periodic patterns that give rise to the experience of stronger (downbeat) and weaker (upbeat) events. Musical meter presumably makes it possible to dance, sing, and play instruments in synchrony with others. It is nevertheless unclear whether or not listeners perceive multiple levels of periodicity simultaneously, and if they do, when and how they learn to do this. We tested children, adolescents, and musically trained and untrained adults with a new meter perception task. We presented excerpts of human-performed music paired with metronomes that matched or mismatched the metrical structure of the music at 2 hierarchical levels (beat and measure), and asked listeners to provide a rating of fit of metronome and music. Fit ratings suggested that adults with and without musical training were sensitive to both levels of meter simultaneously, but ratings were more strongly influenced by beat-level than by measure-level synchrony. Sensitivity to two simultaneous levels of meter was not evident in children or adolescents. Sensitivity to the beat alone was apparent in the youngest children and increased with age, whereas sensitivity to the measure alone was not present in younger children (5- to 8-year-olds). These findings suggest a prolonged period of development and refinement of hierarchical beat perception and surprisingly weak overall ability to attend to 2 beat levels at the same time across all ages.

Keywords: beat perception, development, enculturation, meter perception, music

Synchronous dancing, marching, swaying, head bobbing, and hand clapping are everyday musical behaviors that have been observed across human cultures (Brown & Jordania, 2013). These ubiquitous and spontaneous behaviors appear to be effortless, yet they entail many levels of perceptual processing, from segmenting continuous acoustic information into discrete events, to perceptually inferring an isochronous beat from the patterning of those events, to determining when and how we should move within that temporal framework. Rhythmic structures and behaviors are char-

acterized by tremendous cross-cultural diversity (Clayton, 2000; Toussaint, 2013), suggesting that childhood acquisition may play a key role in shaping these abilities. Nevertheless, relatively little is currently known about when and how children learn to perceive musical beat and meter, two interrelated structures that enable dancing and moving in time with music.

Musical *beat* can be defined as the quasi-periodic pattern of perceptually salient points in time (Lerdahl & Jackendoff, 1983, 1985; Lester, 1986). Listeners are presumed to infer beats at periodic positions in the music when events are more likely to occur, or when events are longer, louder, or at the boundaries of a melody (Desain & Honing, 2003; Hannon, Snyder, Eerola, & Krumhansl, 2004; Jones & Pfordresher, 1997; Large & Palmer, 2002; Palmer & Krumhansl, 1990; Repp, 2010). When listeners are asked to perform simple periodic movements to music by tapping a finger or drumming, their movements tend to correspond to the composer-notated beats, with good interlistener agreement (Drake, Jones, & Baruch, 2000; Drake, Penel, & Bigand, 2000; Snyder & Krumhansl, 2001; Toiviainen & Snyder, 2003; van Noorden & Moelants, 1999; for a review, see Repp & Su, 2013). Listeners can even infer beats from simple, acoustically uniform sequences of events such as a metronome, a phenomenon known as the “tick-tock effect” (e.g., Bolton, 1894; Brochard, Abecasis, Potter, Ragot, & Drake, 2003).

A range of approaches have yielded robust evidence that beat is perceptually salient to adult listeners. For example, adults have greater difficulty reproducing and discriminating rhythms that induce only a weak sense of beat than rhythms that induce a strong sense of beat (Essens & Povel, 1985; Grahn & Brett, 2007, 2009; Povel & Essens, 1985; Povel, 1981). This “beat-based advantage”

This article was published Online First August 27, 2020.

Jessica E. Nave-Blodgett, Joel S. Snyder, and Erin E. Hannon, Department of Psychology, University of Nevada, Las Vegas.

The authors thank Andrew Freeman for his expertise and assistance in the data analytic plan. This research was supported in part by a University of Nevada, Las Vegas Foundation Board of Trustees Fellowship and a Barrick Graduate Fellowship awarded to Jessica E. Nave-Blodgett.

Portions of the results from this article have previously been presented at the following conferences: the Society for the Perception of Music Cognition biennial conference in Nashville, Tennessee (August 2015) and in San Diego, California (August 2017); the International Conference on Music Perception and Cognition in San Francisco, California (July 2016); the New England Sequencing and Timing forum in Amherst, Massachusetts (March 2015) and Storrs, Connecticut (March 2017); and the Mid-Winter Meeting of the Association for Research in Otolaryngology in San Diego, California (February 2018).

Correspondence concerning this article should be addressed to Jessica E. Nave-Blodgett, Department of Psychology, University of Nevada, Las Vegas, 4505 South Maryland Parkway, MS 5030, Las Vegas, NV 89154. E-mail: naveblod@unlv.nevada.edu or jes.nave@gmail.com

is associated with activation of particular motor areas of the brain and is thought to reflect the perceptual underpinnings of movement to music (Grahn, 2012). Listeners readily match a metronome or click track to the beat of a musical stimulus (Fujii & Schlaug, 2013; Hannon et al., 2004; Iversen & Patel, 2008; Leow, Parrott, & Grahn, 2014; Müllensiefen, Gingras, Musil, & Stewart, 2014), providing the clearest evidence that listeners perceive the beat in naturalistic musical listening. Similarly, listeners are faster and better at detecting oddball or deviant (“syncopated”) events at strong than at weak beat positions in a rhythmic sequence (Ladinig, Honing, Haden, & Winkler, 2009), and their brain responses are larger for events, deviants, or omissions that occur on than off the beat (Fujioka, Ross, & Trainor, 2015; Geiser, Sandmann, Jäncke, & Meyer, 2010; Geiser, Ziegler, Jancke, & Meyer, 2009; Iversen, Repp, & Patel, 2009; Snyder & Large, 2005; Vuust et al., 2005). When listeners are instructed to imagine a specific beat structure within a metronome-like stimulus (e.g., imagine a beat every two or three events), their cortical neural responses to otherwise identical sounds vary by the imagined beat condition (Celma-Miralles, de Menezes, & Toro, 2016; Fujioka, Zendel, & Ross, 2010; Iversen et al., 2009; Nozaradan, Peretz, Missal, & Mouraux, 2011; Paul, Sederberg, & Feth, 2015).

Although the primary beat is most salient, it is often multiplied or subdivided to give rise to other levels of perceived periodic structure, called *meter*. Perceiving meter, by definition, entails grasping multiple hierarchical levels of beat simultaneously (Lerdahl & Jackendoff, 1985; London, 2002, 2012). Composers often specify the meter of a musical piece using time signature notation, which indicates how many beats occur per measure. For example, 4/4 time has four beats per measure, and the first beat or downbeat of the measure is presumably perceived as stronger or more salient than other beats because it represents the convergence of at least two levels of the metrical hierarchy (see Figure 1).

Like beat, meter is a musical universal (Savage, Brown, Sakai, & Currie, 2015) and it is critical for coordinating complex patterns of movement across multiple individuals. For example, in many types of group folk and social partner dance, specific movements (e.g., side steps or back steps) occur on specific beats within the measure, and all participants begin moving together, often on the downbeat of the measure. Movement analysis of trained dancers across a variety of genres and dance forms (e.g., Samba, Charleston, Chacarera) suggests that distinct body movements correspond to distinct levels of the metrical hierarchy, for example with distal limb movements marking the primary beat and torso or hip movements marking higher (slower) levels of the meter, such as half-measures or measures (Burger, London, Thompson, & Toiviainen, 2018; Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2014; Naveda et al., 2015; Naveda & Leman, 2010; Toiviainen, Luck, &

Thompson, 2010). Thus, natural musical movement often reflects not only the beat, but also the metrical structure of music.

The evidence reviewed above clearly indicates that listeners’ perception differs for events that do or do not align with a beat. However, it is not clear whether listeners perceptually distinguish between stronger versus weaker beat positions based on their place in the metrical hierarchy (e.g., the first beat of the measure), especially when listeners lack formal music training. For listeners to perceive this hierarchical aspect of meter, they must perceive at least two levels of beat simultaneously, not just differentiate between patterns of beats at different tempi (e.g., whether a metronome in 3/4 or 6/8 fits the music better, without knowing where each measure begins). Studies that investigate perception of meter (not just beat) often depend on musically trained participants who can actively imagine or impose a metrical structure on sounds, or they report that only trained musicians are sensitive to metrical structure (Geiser et al., 2009, 2010; Jongsma, Desain, & Honing, 2004; Palmer & Krumhansl, 1990; but see Ladinig et al., 2009). Because these studies often examine meter perception using simple melodies (Jongsma et al., 2004; Palmer & Krumhansl, 1990) or relatively simple monotonic rhythms (Fujioka et al., 2010; Geiser et al., 2009, 2010; Iversen et al., 2009; Ladinig et al., 2009; Snyder & Large, 2005), the salience of meter in these studies may be limited, at least for untrained listeners. There is some evidence that neural responses are earlier and stronger to events in metrically strong versus weak locations in simple melodies even for individuals with little musical training (Fitzroy & Sanders, 2015), but it is unclear whether neural responses correspond to a perceptual difference or reflect preperceptual stimulus processing. Thus, it may be necessary to use rich, complex musical stimuli and ecologically valid listening situations while listeners make perceptual judgments to demonstrate perceptual sensitivity to simultaneous levels of meter among those with limited musical listening experience or training.

A major question is when perceptual sensitivity to both beat and meter emerge during development. Dancing and other synchronous musical behaviors are readily observable even among adults who have no formal music training (Bégl et al., 2017; Repp & Su, 2013; Sowiński & Dalla Bella, 2013). However, unlike adults, young children do not move in precise synchrony with music or other rhythmic stimuli. Although spontaneous movement to music can be observed as early as infancy, most infants’ movements are not synchronized with music, nor are they rhythmically regular for more than a few seconds at a time (Fujii et al., 2014; Ilari, 2015; Zentner & ErOLA, 2010). Even once children are capable of self-propelled walking, they do not move in precise synchrony with music (ErOLA, Luck, & Toiviainen, 2006), and they struggle to drum in synchrony with metronomes and other humans, with a substantial proportion of young children moving randomly (Kirschner & Ilari, 2014; Kirschner & Tomasello, 2009; Woodruff Carr, Tierney, White-Schwoch, & Kraus, 2016).

Synchronous musical movement appears to develop rather gradually over the course of childhood and adolescence. Four-year-olds synchronize with simple isochronous rhythms more accurately than 2½-year-olds, although neither age group performs anywhere near adult-level accuracy (Provasi & Bobin-Bègue, 2003). Synchronization and reproduction accuracy for isochronous metronomes and complex rhythms, and accuracy of tapping the beat of musical rhythms increase throughout childhood (Drake, 1993;

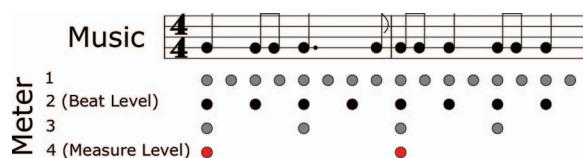


Figure 1. An illustration of the theoretical hierarchical relationships among events in rhythms and the periodic beats experienced as musical meter. See the online article for the color version of this figure.

Drake et al., 2000; McAuley, Jones, Holub, Johnston, & Miller, 2006). The gradual improvement in synchronization accuracy continues between 10 and 17 years of age, particularly for adolescents with music training, but even 16- and 17-year-olds synchronize less accurately with isochronous metronomes than adults (Braun Janzen, Thompson, & Ranvaud, 2014; Thompson, White-Schwoch, Tierney, & Kraus, 2015). Teens who are more accurate synchronizers also exhibit more faithful encoding of sounds in the auditory brainstem (Tierney & Kraus, 2013) than less accurate synchronizers. However, the available evidence suggests a protracted developmental timeline during which precise, accurate motoric entrainment to the beat of simple rhythms is not attained until early adulthood and may continue into middle adulthood (McAuley et al., 2006; Thompson et al., 2015). Moreover, many of the synchronization studies use isochronous metronomes as their synchronization stimulus, where each event is equivalent to the beat. This may not tell the full story of the development of motor or perceptual entrainment to beat of complex musical rhythms, in which the location of the beat must be inferred and is not equivalent to every note onset.

The protracted developmental trajectory for musical entrainment could reflect slowly developing beat and meter processing in general, or it could arise from a lack of precise motor coordination during childhood. In the latter case, we might expect to observe perceptual sensitivity to beat and meter before children can synchronize to music. However, studies using perceptual paradigms with children tell a mixed story. In support of early and robust beat perception, several behavioral findings suggest that sensitivity to beat appears to be present shortly after birth. Four-month-old infants discriminate metronomes that have a 15% tempo difference (Baruch & Drake, 1997), but only at certain base tempos. After 7-month-olds are habituated to a set of rhythms that induce a particular beat, they listen longer to novel rhythms that adults perceive to conflict with the previously established beat than to new rhythms with the same beat (Hannon & Johnson, 2005). Infants are also better at detecting pitch and rhythm changes to conventional rhythms that induce a strong beat than to those that induce a weak beat (Trehub & Hannon, 2009).

Investigations of infants' neural responses to musical rhythms suggest that the basic perceptual elements necessary for beat and meter perception may be present early in development. For example, newborn infants exhibit larger mismatch negativity brain responses to event omissions that occur on strong than on weak positions in a drum sequence (Winkler, Háden, Ladinig, Sziller, & Honing, 2009). However, these responses could reflect other aspects of the stimulus (e.g., number of omitted instruments) that were confounded with beat position. A few months after birth, infants exhibit mismatch negativities in response to early or late events in metronome-like patterns (Otte et al., 2013), suggesting a strong sensitivity to temporal regularity. When presented with repeating rhythms, 7- and 15-month-old infants' brain responses contained energy at the beat and measure frequencies of ambiguous and nonambiguous rhythms (Cirelli, Spinelli, Nozaradan, & Trainor, 2016). However, similar responses to rhythmic stimuli have been observed in anesthetized rodents, suggesting such responses in humans at least partially reflect stimulus-driven rather than perceptual processes (Rajendran, Harper, Garcia-Lazaro, Lesica, & Schnupp, 2017). Without a perceptual task, it is impossible to determine whether the infants perceived multiple levels of

meter in the rhythms or whether their neural responses simply reflect low-level processing of the energy content of the stimulus.

In support of the idea that meter perception develops gradually, there is considerable evidence that beat perception changes over the course of development because of listening experience. Whereas 12-month-old North American infants were only able to discriminate familiar, simple-meter rhythms with a regular beat, 6-month-old infants successfully detected disruptions to rhythms with either familiar (simple and regular beat) or unfamiliar (complex and irregular beat) metrical structures (Hannon & Trehub, 2005a; Hannon & Trehub, 2005b). However, after two weeks of at-home exposure to children's music containing complex meters with an irregular beat, children from 12 months to 7 years detected disruptions of complex-meter rhythms, whereas older children and adults remained sensitive only to disruptions of simple-meter rhythms (Hannon & Trehub, 2005b; Hannon, Vanden Bosch der Nederlanden, & Tichko, 2012). By 4–8 months of age, infants prefer music with metrical structures consistent with their native culture (Soley & Hannon, 2010). By 7 months, North American infants already show stronger preferences for and more accurate discrimination of music with duple meter, which in Western music is more common than music with triple meter (Bergeson & Trehub, 2006). This advantage for duple over triple meter is further enhanced by participation in music classes where duple meter songs are more prevalent (Gerry, Faux, & Trainor, 2010; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012). By 5 years of age, North American children are better at determining if a metronome matches the beat of a piece of music in simple than in complex meter (Einarson & Trainor, 2015, 2016); yet even then, their accuracy levels do not approach that of adults in similar tasks (Iversen & Patel, 2008; Müllensiefen et al., 2014; Puyjarinet, Bégel, Lopez, Dellacherie, & Dalla Bella, 2017). This evidence suggests that something about listeners' perception of musical temporal structure is changing over the course of childhood, presumably in response to culture-specific listening experience, although it is unclear what musical structures are driving children's performance in these tasks.

To our knowledge, no study has systematically examined the development of both beat and meter perception using natural, ecologically valid musical stimuli and tasks. The current study therefore had two goals. First, we asked whether listeners are sensitive to both musical beat and meter, such that beats at multiple time scales are experienced as stronger or weaker than others based on their position in the metrical hierarchy. We compared individuals with and without formal musical training to examine whether the ability to simultaneously perceive multiple levels of metrical structure relies on intense musical training or whether it is present in individuals without such training. Second, we examined how and when beat and meter perception develop from childhood through adulthood.

Participants performed a simple task in which they provided a rating of fit between rich, complex ballroom dance music and a metronome that reflects two beat levels of the meter. We used this task to examine beat and meter perception among adults with and without formal music training (Experiments 1A and 1B) and in typically developing children and adolescents (Experiment 2A and 2B). We chose ballroom dance music for our stimuli because ballroom music is composed with the explicit intention of eliciting and supporting synchronized movement to multiple beat levels,

and as such it contains myriad cues to multiple levels of different metrical structures (e.g., waltz vs. march).

Experiment 1A

In Experiment 1A we compared perceptual judgments made by adults with and without formal music training who were asked to rate the fit between musical excerpts paired with metronomes that matched or mismatched multiple levels of metrical structure (i.e., at the beat and measure levels). In Experiment 1B we examined adult judgments of fit in response to metronomes that only matched or mismatched at the measure level.

Method

Participants. To compare performance on the meter perception task as a function of musical experience, we recruited participants with low or no musical training (*nonmusicians*) and those with substantial musical training (*musicians*). The nonmusicians were recruited from the undergraduate psychology student subject pool of the University of Nevada, Las Vegas, and comprised 32 (17 female) students who received credit in psychology classes at the University of Nevada, Las Vegas for their participation. Nonmusicians had fewer than five years of instrumental or vocal musical training or performance ($M = 1.06$, $SD = 1.47$), and 19 of 32 reported having no formal music training at all. There were also 28 (17 female) musicians who had five or more years of formal musical training ($M = 15.46$, $SD = 11.38$), a common minimum threshold of training for individuals to be considered musicians (Skoe & Kraus, 2013). Most of the musician participants were recruited from the university's music department and the surrounding community and were compensated for their participation with entry into a gift card raffle. Six musicians were recruited through the psychology department's subject pool. Given the potential relevance of dance experience for perceiving beat and meter especially in ballroom dance music, we also collected information

about dance training, which ranged from 0–17 years (see Table 1 for full demographic information).

We did not perform auditory pure-tone hearing screenings on participants, but all participants self-reported normal hearing and no history of hearing problems. All participants were conversationally or functionally competent in English, having learned English from birth ($n = 51$) or during early childhood ($n = 9$, $M_{\text{age learned}} = 4.6$ years old, $SD = 2.1$, range = 2–8 years). All experimenter interactions with participants, informed consent, and text instructions were in English. All participants verbally confirmed that they understood the instructions and the task. This research was approved by the Social/Behavioral Institutional Review Board of the University of Nevada, Las Vegas.

Power analysis. In this experiment, we wished to detect, if present, significant interactions between metrical levels (beat and measure) within participants or interactions between metrical levels and musicianship between participants (i.e., beat by musical training or measure by musical training), as in previous studies (Palmer & Krumhansl, 1990). Prior studies using a similar method to ours (e.g., the Beat Alignment Task; Iversen & Patel, 2008) have reported effect sizes for beat-level matching ranging from $d = 1.27$ to 1.70 (Leow et al., 2014; Müllensiefen et al., 2014). Previous tasks only manipulated the beat-level matching between the metronome and music; we assumed that measure-level matching would have a smaller effect than that of beat-level matching and estimated our effect sizes more conservatively.

Our analytical approach was a two-level, participant-clustered multilevel model, where level 1 was the 96 trials clustered within the same individual, and level 2 was participants. We used the "Optimal Design Plus Empirical Evidence" program (Raudenbush & Liu, 2000; Spybrook et al., 2011) for power analyses in multilevel modeling to estimate our necessary sample size. We modeled the required sample size (number of participants, L2) to achieve a power of 0.8 to detect effect sizes between $d = 0.15$ and $d = 0.80$ for fixed effects at Level 1 or cross-level interactions, and assumed a possible variance of effect sizes between 0.05 ($SD_{\text{effect size}} = \pm$

Table 1
Experiment 1A and 1B Participant Demographic Information

Characteristic	Experiment 1A			Experiment 1B
	Full sample	Musicians	Nonmusicians	Full sample
<i>n</i> (<i>n</i> females)	60 (34)	28 (17)	32 (17)	36 (19)
Age in years	24.92 (9.67)	29.54 (12.12)	20.88 (3.75)	19.7 (5.5)
Age range	18–57	18–57	18–35	18–42
Parental education > 4-year college degree (% of sample)	35 (58%)	21 (78%)	14 (44%)	12 (33%)
<i>n</i> Hispanic (% of sample)	9 (15%)	2 (7%)	7 (22%)	12 (33%)
<i>n</i> L1 other than English (% of sample)	9 (15%)	1 (4%)	8 (25%)	14 (39%)
<i>n</i> bilingual (% of sample)	20 (33%)	10 (37%)	10 (31%)	14 (39%)
<i>n</i> participants with musical training (% of sample)	40 (66%)	28 (100%)	13 (39%)	14 (39%)
Years musical training	7.78 (10.63)	15.46 (11.39)	1.06 (1.47)	1.42 (2.8)
Range of musical training	0.5–42.0	5.0–42.0	0.5–4.0	0.25–10
Age started musical training	10.51 (3.44)	9.82 (3.22)	12.00 (3.54)	12.07 (4.92)
<i>n</i> participants with dance training (% of sample)	15 (25%)	8 (30%)	7 (21%)	9 (25%)
Years dance training	1.28 (3.24)	1.41 (2.96)	1.17 (3.50)	1.06 (2.58)
Range of dance training	0.40–17.0	0.40–14.0	0.5–17.0	0.5–12.5
Hours music listening/week	15.82 (12.13)	16.57 (14.41)	15.16 (9.90)	15.08 (14.22)

Note. All values are means and standard deviations unless otherwise specified. Years musical training and years dance training include all participants in the sample or subgroup. Ranges of musical and dance training and age started musical training only include participants with relevant training. All values are based on participant self-report.

0.20) and 0.10 ($SD_{\text{effect size}} = \pm 0.30$). At $d = 0.15$ and variance of 0.10 (the most conservative estimate), a sample size of 50 participants was required for 0.8 power, and for larger effect sizes ($d = 0.2, 0.8$) between 15 and 25 participants were sufficient to reach 0.8 power. Our sample size of 60 individuals with 96 trials each was determined to be adequate for assessing small-to-moderate fixed effects at level 1 and cross-level interactions between level 1 and level 2 factors. Other approaches to estimating sample size for multilevel model designs suggest a 30/30 rule with a minimum of 30 groups (participants) and 30 members (trials/measurements) in each group (Hox, 1995, 2010) to achieve reasonable power to detect fixed effects in multilevel modeling, which our sample size (60/96) exceeded.

Stimuli. We selected six pieces of expressively performed ballroom dance music, half in triple meter (3/4) and half in duple meter (4/4) at slow (89 BPM), medium (104 BPM), and fast tempos (124 BPM; compact disk: “Ballroom Dance Music,” The Swiss Ballroom Orchestra, Blaricum CD Company, 1999). We selected tempo-matched pairs of duple and triple pieces, and further equated tempo by using the “Change Tempo” function in Audacity(R) (Version 2.1.2; Audacity Team, 2020) to set paired pieces to the same tempo (without altering the pitch of the music). For each of the six pieces, we selected four five-measure excerpts, resulting in segments 7–14 s in duration depending on tempo.

For each piece, we created metronome tracks that matched or mismatched at the main beat level and the measure level. The temporal positions of each beat and measure were identified using the Bar and Beat Tracker VAMP plugin (Davies & Plumbley, 2007; Stark, Davies, & Plumbley, 2009) in Audacity, and further adjusted by the first author who is a trained musician. Next, we created metronomes (“Generate Click-Track” command in Audacity), consisting of 10-ms sine-wave clicks, with a higher pitched click marking the first beat of each measure (MIDI pitch 92: G#6, 1661.2Hz), and lower-pitched clicks marking all other beats (MIDI pitch 80: G#5, 830.6Hz), a pattern that is typical in many commercially available metronomes. Thus, depending on the meter, the higher-pitched click occurred once every three or four clicks, replacing a lower-pitched click. Across four conditions, the met-

ronome was matched to the music at the level of beat, the measure, both, or neither (see Figure 2). Fully matching metronomes had lower- and higher-pitch clicks that matched the beat and measure of the music, respectively (Figure 2A). When the metronome matched the beat but not the measure, clicks occurred on every beat, but higher-pitch clicks did not consistently align with the first beat of each measure (i.e., they occurred every three beats instead of every four or vice versa, wrapping in and out of phase with the measure of the music; Figure 2B). When the metronome matched the measure but not the beat, the higher-pitch clicks occurred regularly with the first beat of every measure, but the intervening clicks did not match beats in the music (i.e., three clicks were presented asynchronously to four beats of the music, or vice versa; Figure 2C). Fully mismatching metronomes were 6% faster in tempo than the musical excerpt, meaning neither beat- nor measure-level clicks matched beats or measures in the music, and the measure length was swapped (duple for triple and vice versa Figure 2D). The 6% tempo increase is above the threshold for listeners’ ability to detect tempo differences between sequences of isochronous metronomes (Drake & Botte, 1993). Table 2 shows the asynchronies that result from each of the conditions in Figure 2 at the level of the beat and measure.

The metronome and music were generated in stereo, with the metronome in one ear channel and the musical excerpt in the other channel. This dichotic presentation of the music and the metronome was expected to aid participants in perceptually separating the two auditory channels (Hartmann & Johnson, 1991). Half of the stimuli had the metronome in the left channel, and half had the metronome in the right channel. Participants listened to 96 different music/metronome combinations throughout the experiment (four trials of each of six musical pieces across four metronome conditions). All stimulus materials are available on the Open Science Framework Repository at <https://osf.io/z2ab6/>.

Apparatus and procedure. Participants performed the rating task on a PC desktop computer (Dell, Round Rock, TX) running Windows 7 (Microsoft Corporation, Seattle, WA). Participants listened to the stimuli through over-the-ear noise-attenuating headphones (approximately 20 dB SPL passive noise attenuation;



Figure 2. Alignment of the metronome tracks to the musical excerpts. Measure-level, higher-pitch clicks are indicated by red circles, regular beat-level, lower-pitch clicks by smaller black circles. Vertical alignment between the musical note and the circle indicates onset synchrony between the musical beat and the metronome beat; misalignment indicates asynchronous temporal onsets for the beat of the music and the beat of the metronome. (A) Beat- and measure-level clicks of the metronome match both the beat and measure positions in the music. (B) Beat-level clicks match the beat of the music while measure-level clicks mismatch the measure of the music, although measure-level clicks still occur on a beat position. (C) Measure-level clicks match the downbeat of the music, but beat-level clicks mismatch the beat of the music. (D) Neither beat- nor measure-level clicks align with beat or measure positions in the music (fully asynchronous onsets). See the online article for the color version of this figure.

Table 2
Average Asynchrony Between Metronome Clicks and Musical Beat and Measure Downbeats

Condition	Metrical level	Asynchrony (ms)	Asynchrony (% IOI)
Beat matching/measure matching	Beat	0.18 (0.03)	.03%
	Measure	0.18 (0.06)	<.01%
Beat matching/measure mismatching	Beat	0.18 (0.03)	.03%
	Measure	482.71 (93.02)	24%
Beat mismatching/measure matching	Beat	187.81 (22.45)	33%
	Measure	0.20 (0.07)	<.01%
Beat mismatching/measure mismatching	Beat	145.6	25%
	Measure	509.69 (96.96)	25%

Note. Asynchronies are reported in milliseconds with mean (standard deviation) and as percent interonset interval (IOI), for each metrical level (beat or measure). Beat level asynchronies reflect the time between a beat-level metronome click and the nearest beat in the music, and measure level asynchronies reflect the time between a higher pitch metronome click and the nearest measure downbeat in the music.

Sennheiser 280 Pro, Sennheiser Corporation, Old Lyme, CT). All stimuli were presented to participants at approximately 70 dB SPL, well above normal hearing thresholds but not so loud as to be uncomfortable. A custom program written in Presentation software (Neurobehavioral Systems, Palo Alto, CA) controlled the stimulus presentation and response collection.

The on-screen instructions stated the following:

On each trial, you will hear of clip of music in one ear and a click track in the other ear . . . Your task is to determine how well the click track matches the music you hear. You will listen to the entire clip of music and then be asked to rate how well the click track matched the music on a scale of 1 through 4, where 1 means *not well at all* and 4 means *very well*.

The prompt to respond (“How well did the sounds fit together?” with the 4-point scale) did not appear until the end of the stimulus ensuring participants heard the entire stimulus. The next trial began after participants responded or after five seconds had elapsed.

The experiment was preceded by a practice session. Each participant first encountered three demonstration trials: one fully matching metronome, one beat-matching/measure-mismatching metronome, and one fully mismatching metronome. Each demonstration sound file was accompanied by text stating how other listeners might rate the metronome, for example, “Most people would think this click-track matched the music very well” for a fully matching metronome, and so forth. Following the demonstration trials, participants completed three practice trials in which they listened to three additional musical excerpts paired with a metronome (fully matching, beat matching/measure mismatching, and fully mismatching) and rated the fit of the metronome to the music just as in the experimental trials. They received feedback using the same wording as in the demonstration trials regardless of their answers. After participants completed the practice, they could ask the experimenter for clarification or they could begin the experimental trials. The experiment contained four blocks of 24 trials each, presented by the computer in a random order for each participant. Between every block of trials, participants had the opportunity to take a short break while remaining seated at the computer.

After completing the experiment, participants filled out a brief demographic questionnaire that surveyed their musical, dance,

language, cultural background, and hearing status. The entire experiment lasted approximately 30 min, including informed consent, practice session, experiment, and demographic questionnaire.

Data analysis. Multilevel modeling (MLM) analyses allow for the simultaneous comparison of within-participant experimental manipulations and between-participants individual differences along with cross-level interactions (Baayen, Davidson, & Bates, 2008; Hox, 2010; Raudenbush & Bryk, 2002).¹ We investigated the effect of beat- and measure-level matching metronomes on participants’ ratings of fit while controlling for tempo and musical meter within participants, and controlling for years of musical training, years of dance training, and hours of musical listening between participants. Our primary hypothesis was that beat- and measure-level matching would significantly alter participants’ ratings of fit. We hypothesized that participants’ perception of measure-level matching would be moderated by beat-level matching, and thus expected a Beat \times Measure interaction between the two factors. We also hypothesized that musical training would make participants more sensitive to beat- and measure-level matching. Thus, we created three cross-level interaction terms: Beat \times Musical Training, Measure \times Musical Training, and the three-way interaction of Beat \times Measure \times Musical Training.

Because there was no prior research that used MLM to examine meter perception, we performed the analyses in a hierarchical model-building manner. In all models, the dependent variable was participants’ rating of fit between the metronome and music in each trial. First, we tested an unconditional (baseline) model, equivalent to a one-way ANOVA with participant (subject) as the random effect. In Model 1, we entered the within-subjects experimental variables of Beat (matching or mismatching), Measure (matching or mismatching), Tempo (slow, medium, and fast), and musical Meter (duple or triple), along with the interaction term of Beat \times Measure, as well as the between-participants individual difference variables of Musical Training (in years), Dance Training (years), and Music Listening (hours per week). We included random slopes for

¹ ANOVA-based power analyses, results, and figures from all four experiments can be found on the OSF page as supplemental material and are not substantially different from the MLM results reported here.

Beat, Measure, and Meter. In Model 2, we entered the cross-level interaction terms of Beat \times Musical Training, Measure \times Musical Training, and Beat \times Measure \times Musical Training to test the hypothesis that musical training affects how participants perceive beat and measure information in music.

All within-participants (level 1) predictors were dichotomous variables. Beat- and measure-level matching was coded as “0” for mismatching and “1” for matching. Tempo was separated and dummy-coded into two dichotomous variables, slow and fast (with the medium/104 BPM tempo taken as reference, when the values for both slow and fast were zero). Meter was coded with duple meter (4/4) as “0” and triple meter (3/4) as “1.” All between-participants (level 2) predictors were grand-mean centered: the grand mean of all participants for each variable (music training/dance training/music listening) was calculated, and then subtracted from each participant’s score on the relevant variable. We did not standardize between-participants predictors.

We treated all missing data as data missing completely at random (MCAR), as there was no increased likelihood of missing a response as a function of missing a previous response (Rubin, 1987). Each participant experienced 96 trials in the experiment, giving a total of 96 possible ratings of fit (the dependent variable). However, as the experiment moved on automatically to the next trial after 5 s with no response, there was a small number of missing trials across the entire dataset (11 trials; $<0.01\%$). Seven individuals had at least one missed trial, and the maximum number of missing trials was three (3% of an individual’s total). There were no missing values for within-participants factors. Most between-participants (Level 2) data were present, but one participant did not provide years of musical training, one participant did not provide years of dance training, and eight participants did not provide hours of music listening information. We did not impute or estimate missing values either for ratings of fit (dependent variable) or the between-participants predictors.

All statistical analyses were performed using IBM SPSS Statistics Version 20 (IBM Corporation, Armonk, NY) using the MIXED command. All models used restricted maximum likelihood estimation (REML). Our covariance matrix was unstructured, and we allowed for covariances between random slopes and intercepts. In all models, we used Satterthwaite approximations to estimate degrees of freedom and obtain two-tailed p values for each predictor and interaction term (Satterthwaite, 1946). All models presented in the article successfully converged. Syntax for all models and datasheets are available on the companion OSF page for this article.

Results and Discussion

Ratings of fit are presented in Figure 3, for the whole sample (3A) and for the musician and nonmusician participants separately (3B and 3C). MLM results,² including coefficients, estimated variances, and ICC, are reported in Table 3. In the baseline (intercept-only) model, most of the variance in ratings was attributable to within-subjects variance (1.34, 95% CI [.29, 1.39]) and a small portion was attributable to between-subjects variance (0.05, 95% CI [0.03, 0.08]). Approximately 96% of the unexplained variance was within participants (experimental effects) and 4% of

the variance in ratings was between participants (individual differences).

Hypothesis 1: Beat- and measure-level information influence ratings of fit. Model 1 examined the effect of the within-subjects (Level 1) experimental factors of Beat, Measure, Tempo, Meter, and the Beat \times Measure interactions on participants’ ratings while controlling for Musical Training, Dance Training, and Music Listening as fixed effects, with random slopes for Beat, Measure, Meter, and Beat \times Measure.

There were significant fixed effects of Beat, Tempo, and the Beat \times Measure interaction. Participants rated beat-matching metronomes significantly higher than beat-mismatching metronomes when controlling for all other factors ($\gamma = 1.52$, 95% CI [1.39, 1.64]). Measure-level matching did not significantly alter ratings of fit after controlling for other factors. However, the interaction between Beat \times Measure was statistically significant (Figure 3A). We performed a simple slopes test (Preacher, Curran, & Bauer, 2006) examining difference in ratings for beat-mismatching and beat-matching metronomes depending on measure-level alignment while controlling for all other factors. When the beat of the metronome mismatched the music, participants’ ratings did not differ for measure-mismatching (1.70, 95% CI [1.60, 1.80]) or measure-matching metronomes (1.61, 95% CI [1.47, 1.75]; $\omega = -0.10$, 95% CI [-0.23, 0.04]). By contrast, when the beat of the metronome matched the music, participants gave higher ratings when the measure also matched (3.55, 95% CI [3.45, 3.64]) than when the measure did not match (3.22, 95% CI [3.10, 3.34]; $\omega = 0.33$, 95% CI [0.23, 0.42]). The significantly higher ratings for metronomes that matched both the beat and measure of the music compared with metronomes that only matched at the beat result suggests participants could perceive beat- and measure-level information simultaneously in the music and metronome. However, the lack of a difference in ratings when the beat did not match may indicate that listeners may have difficulty detecting measure-level mismatching when the beat also mismatches.

Participants gave slightly higher ratings of fit to faster (124 BPM) musical excerpts compared with moderate (104 BPM) musical excerpts, regardless of beat- or measure-level matching ($\gamma = 0.11$, 95% CI [0.07, 0.15]). After controlling for all within-participant variables, none of the between-participants variables (musical training, dance training, music listening) had a statistically significant main effect. Random effects indicated that participants who gave higher ratings overall had smaller differences between their ratings for mismatching and matching metronomes compared with participants who had lower overall ratings ($r_{\text{beat,intercept}} = -.22$; $r_{\text{measure,intercept}} = -.48$), which likely reflects ceiling effects in use of the rating scale.

In Model 1, the addition of the within-subjects predictors accounted for approximately 64% of the within-subjects variance relative to the random-effects ANOVA baseline model. The pro-

² We performed the MLM on the full sample and excluding the subset of non-native English speakers ($n = 9$). There were no changes in the direction/sign of the results as a result of excluding the non-native English speakers, so we report the results from the full sample.

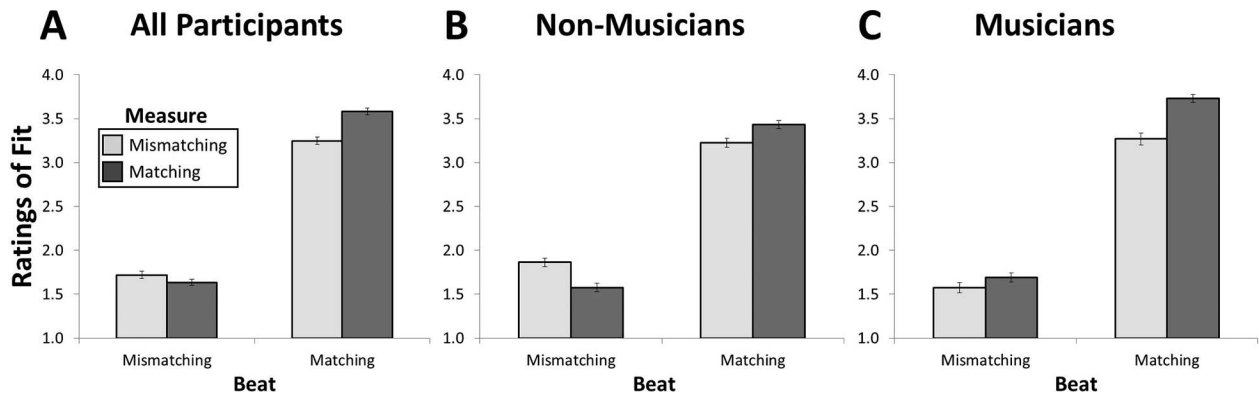


Figure 3. Ratings of fit between metronome and music by beat- and measure-level alignment between the metronome and the music. (A) Entire sample ($n = 60$). (B) Nonmusicians (participants with <5 years musical training) group only ($n = 32$). (C) Musicians (participants with ≥ 5 years musical training) group only ($n = 28$). Error bars are within-subject error bars (Cousineau, 2005) and represent 1 standard error above and below the mean.

portion of between-subjects variability increased to 20%, and the within-subjects variability decreased to 80%.

Hypothesis 2: Music training affects beat- and measure-level perception. In Model 2 we assessed the hypothesis that musical training interacts with participants' perception of beat- and measure-level matching between the metronome and music. We added the cross-level interactions among beat, measure, and musical training to the model. All previous significant fixed effects of Beat, Tempo, and Beat \times Measure remained statistically significant with only minor changes to the estimated parameters (see Table 3). With the addition of the cross-level interactions, the fixed

effect of Musical Training became statistically significant. Controlling for all other variables, for every year of musical training above the sample mean (7.78 years), participants' mean ratings decreased slightly ($\gamma = -0.016$, 95% CI $[-0.025, -0.009]$).

All three cross-level interactions also had significant main effects, supporting the hypothesis that musical training moderates the perception of beat and measure. The Beat \times Musical Training interaction was similar to the Measure \times Musical Training interaction: for every year of musical training above the sample mean, participants gave slightly higher ratings of fit to beat-matching and measure-matching metronomes than to nonmatching metronomes (see Table 4). A simple slopes test of the Beat \times Musical Training interaction demonstrated that the difference in ratings between beat-mismatching and beat-matching metronomes was greater in participants with more musical training than those with less (see Table 4). The simple slopes test of the Measure \times Musical Training interaction demonstrated that participants with little-to-no musical training gave significantly lower ratings to measure-matching than measure-mismatching metronomes, whereas participants with more musical training did not differ in their ratings (see Table 4).

The test of the three-way interaction (Dawson & Richter, 2006) between Beat, Measure, and Musical Training demonstrated that for metronomes that mismatched at the level of the beat, musicians (average of ≥ 15 years music training) did not differ in their ratings between measure-mismatching and measure-matching metronomes ($\omega = 0.090$, $p = .249$), whereas nonmusicians (average of ≤ 1 year) gave significantly lower ratings of fit to measure-matching metronomes than to measure-mismatching metronomes ($\omega = -0.26$, $p = .001$; slope difference = 0.35, $t(59) = 4.11$, $p < .001$). This suggests that musicians and nonmusicians perceived measure-level matching differently when the beat of the metronome did not match the music. In contrast, when the beat of the metronome matched the music, both musicians ($\omega = 0.40$, $p < .001$) and nonmusicians ($\omega = 0.26$, $p < .001$) gave significantly higher ratings of fit to measure-matching over measure-mismatching metronomes, although the increase (slope) was slightly higher for musicians over nonmusicians (slope difference: 0.14, $t(59) = 2.15$, $p = .037$). This latter finding strengthens the argument that musical training is not necessary to perceive multiple

Table 3

Results of Multilevel Models in Experiment 1A Investigating Differences in Ratings of Fit of Metronome to Music

Parameter	Baseline	Model 1	Model 2
Fixed effects			
Rating of fit (intercept)	2.544**	1.705**	1.705**
Within-subjects factors (Level 1)			
Beat		1.518**	1.518**
Measure		-0.097	-0.097
Slower tempo		0.022	0.022
Faster tempo		0.110**	0.110**
Musical meter		-0.043	-0.043
Beat \times Measure		0.423**	0.423**
Beat \times Musical Training			0.021**
Measure \times Musical Training			0.024**
Beat \times Measure \times Musical Training			-0.015*
Between-subjects factors (Level 2)			
Musical training		-0.0004	-0.017**
Dance training		0.004	0.004
Music listening		0.002	0.002
Random effects			
Within-subjects (residual)	1.341**	0.482**	0.482**
Between-subjects (Intercept)	0.053**	0.121**	0.087**
Intraclass correlation (ICC)	0.0377	0.201	0.153

Note. All coefficients are unstandardized. Between-participants (Level 2) variables are grand-mean centered. Complete covariance matrix available on OSF as supplemental material.

* $p < .05$. ** $p < .01$.

Table 4
Simple Slopes Tests From Experiment 1A of Beat \times Musical Training and Measure \times Musical Training

Factor	Intercept	95% CI	Slope	95% CI
Beat \times Musical Training				
Nonmusicians (1.1 y)	1.82*	[1.59, 2.04]	1.38**	[1.25, 1.51]
Sample average (7.8 y)	1.70*	[1.62, 1.79]	1.52**	[1.41, 1.63]
Musicians (15.5 y)	1.57*	[1.49, 1.66]	1.68**	[1.54, 1.81]
Measure \times Musical Training				
Nonmusicians	1.82*	[1.72, 1.92]	-0.260**	[-0.406, -0.115]
Sample average	1.70*	[1.62, 1.79]	-0.097	[-0.219, 0.026]
Musicians	1.57*	[1.47, 1.68]	0.090	[-0.062, 0.242]

Note. All coefficients are unstandardized. Each of the lines estimates the slope and intercept for musical training at 1.1 years (nonmusician subgroup), 7.8 years (overall sample average), and 15.5 years (musician subgroup). Ratings of fit are on a 1 (*low*) to 4 (*high*) scale. Intercept value indicates rating of fit for mismatching (beat or measure, as appropriate) metronome. Significant simple slope values indicate that the change in ratings of fit (e.g. from beat-mismatching to beat-matching metronomes) after controlling for all other variables is statistically significant.

* $p < .05$. ** $p < .01$.

levels of musical meter simultaneously, as all participants gave significantly higher ratings to fully matching metronomes over beat-matching/measure-mismatching metronomes regardless of their amount of musical training.

The addition of the cross-level interaction terms in Model 2 accounted for approximately 28% of the unexplained variance between-subjects in Model 1. After fitting Model 2, approximately 15% of the remaining unexplained variance in scores was between-participants, and 85% within-participants.

Summary. Adult listeners, regardless of their amount of formal musical training, gave the highest ratings of fit to metronomes that matched the music at two metrical levels: the beat level and the measure level. This is compelling evidence that listeners perceive multiple levels of metrical structure simultaneously while listening to music. However, listeners did not give equal weighting to both metrical levels; ratings were far more swayed by beat-level than by measure-level matching.

It is unclear why listeners with lower amounts of formal musical training gave higher ratings to fully mismatching metronomes compared with beat-mismatching/measure-matching metronomes. Perhaps in this condition, nonmusicians were perceptually overwhelmed by asynchronies at the beat level, which were in fact greater in the measure-matching than in the measure-mismatching condition (see Table 2). Beat-level asynchronies may have been so salient to nonmusicians that they failed to notice synchrony or asynchrony at the measure level. If this was the case, nonmusicians should have no difficulty distinguishing measure-matching from measure-mismatching metronomes when beat-level information is removed entirely from the metronomes. To confirm this, we conducted a second, modified version of the experiment in which we presented nonmusicians with metronomes that matched or mismatched the music only at the level of the measure.

Experiment 1B

Method

Participants. We recruited a new sample of 36 (19 female) undergraduate students to participate in this experiment. None of

these individuals participated in the previous experiment. We collected information about previous music and dance instruction from participants as in Experiment 1A (see Table 1 for full demographic information). All participants were functionally and conversationally fluent in English; they were either native speakers ($n = 23$) or learned English during childhood ($n = 13$, $M_{\text{age learned}} = 5.3$ years old, $SD = 2.75$ years, range = 2–12 years). As in Experiment 1A, all experimenter-participant interactions, test materials, demographic questionnaires, and instructions were in English. All participants verbally confirmed that they understood the instructions and the task. We did not perform pure-tone audiometry hearing tests on participants, but all participants self-reported normal hearing and no history of long-term hearing deficits in a demographic questionnaire. This research was approved by the Social/Behavioral Institutional Review Board of the University of Nevada, Las Vegas.

Power analysis. Given the large effect of beat ($\gamma = 1.52$) and the moderate effect of the interaction between beat and measure ($\gamma = 0.42$) in Experiment 1A, we estimated that the effect of measure-level information when beat-level information was removed would likely fall between an estimated $d = 0.20, 0.50$. As such, our sample size of 36 participants with 96 trials per participant was sufficient to reach a power of 0.8 to detect an effect size of at least $d = 0.2$ with a variance of 0.10 (Raudenbush & Liu, 2000; Spybrook et al., 2011), and was larger than the suggested 30/30 rule (Hox, 1995, 2010).

Stimuli and procedure. The experimental procedure, program, and instructions were the same as described above, except for in this experiment we silenced the beat-level clicks of the metronomes, leaving only the measure-level clicks (see Figure 4). This resulted in three conditions: (a) a fully matching condition, in which the click matched the measure-level downbeat of the music (corresponding to both measure-matching conditions in Experiment 1A; Figure 4A), (b) a measure mismatching condition in which the click fell on a beat in the music other than the downbeat of the measure and thus was period-mismatched (corresponding to the beat-matching/measure-mismatching condition of Experiment 1A; Figure 4B), and (c) a fully mismatching condition, in which

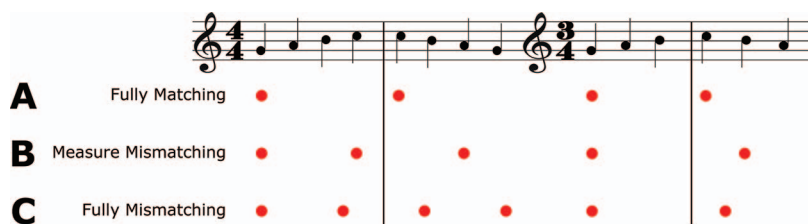


Figure 4. Alignment of the metronome tracks to the musical excerpts in Experiment 1B. (A) Fully Matching metronomes had clicks that aligned with the downbeat of each measure. (B) Measure Mismatching metronomes had clicks that did not always fall on the downbeat of the measure, but always fell on a beat of the music. (C) Fully Mismatching metronomes had clicks that did not fall on either beat or measure downbeat locations (were fully asynchronous with the music). All metronome sounds were higher pitched clicks (1661.2Hz) indicating measure-level downbeats. See the online article for the color version of this figure.

the measure-level click did not fall regularly on any beat in the music (corresponding to the beat- and measure-mismatching condition of Experiment 1A; Figure 4C). Thus, while clicks in the measure-mismatching condition always aligned with a musical beat (just not the downbeat of the measure), clicks in the fully mismatching condition almost never aligned with any musical beats. Measure-level asynchronies between the metronome and music (see Table 2) were minimal in the fully matching condition ($M_{asynch} = 0.18$ ms, $SD = 0.06$ ms; .03% IOI), but they were similarly high for both the measure-mismatching condition ($M_{asynch} = 482.71$ ms, $SD = 93.02$ ms; 24% IOI) and the fully mismatching condition ($M_{asynch} = 509.69$ ms, $SD = 96.96$ ms; 25% IOI).

Data analysis. We used MLM to examine the effects of measure-level metronome alignment on participants' ratings while controlling for tempo and meter differences within-participants, and effects of musical training, dance training, and music listening between-participants. Our primary hypothesis was that fit ratings would reflect participants' sensitivity to the measure-level alignment of the metronome to the music. We entered Measure (fully matching, measure mismatching, and fully mismatching), Tempo (slow, medium, and fast), and Meter (duple and triple) as within-subject variables and Musical Training, Dance Training, and Music Listening as between-subjects variables, with random slopes for Measure and Meter. As musical training was related to sensitivity to both beat- and measure-level information in Experiment 1A, we hypothesized we would find effects of music training here too. We tested this hypothesis with a cross-level interaction between Measure \times Musical Training. We compared the full model to a baseline model (one-way random effects ANOVA).

Measure-level alignment was separated and dummy-coded into two dichotomous variables – “Fully Matching” and “Fully Mismatching” (the “Measure Mismatching” condition was used as reference, when the values for the other two conditions were zero). All other within- and between-participants predictors were coded and treated the same as in Experiment 1A. There were a small number of missing ratings across the entire dataset (32; <1%). A total of nine participants had at least one missing response, with one individual who failed to respond in 18 trials (19% of their total); all other participants missed between one and four trials. There were no missing predictors at any level. We did not impute or estimate missing values for ratings for missed trials. All statis-

tical analyses were performed using the same program and parameters as Experiment 1A. All models presented successfully converged. Syntax for models and datasheets are available on the OSF page.

Results and Discussion

Participants' ratings for each metronome condition are shown in Figure 5. MLM results³ are reported in Table 5. As in Experiment 1A, most of the variance in ratings in the baseline model was attributable to within-subjects variance (1.22, 95% CI [1.16, 1.28]) and a small portion of the variance was attributable to between-subjects variance (0.05, 95% CI [0.03, 0.09]). Approximately 96% of the variance in ratings was within participants (experimental effects), and 4% of the variance in ratings was between participants (individual differences).

Hypothesis 1: Measure-level information affects ratings of fit. There were significant fixed effects of Measure and Tempo. Compared with measure-mismatching metronomes, fully matching metronomes received higher ratings ($\gamma = 0.77$, 95% CI [0.57, 0.96]), and fully mismatching metronomes received lower ratings ($\gamma = -0.74$, 95% CI [-0.89, -0.61]), after controlling for all other factors. Participants' significantly higher ratings for fully matching metronomes over measure mismatching metronomes strongly suggests that participants are sensitive to measure-level information. If participants attended only to beat-level information, they should have rated these two metronomes similarly because both align with beats in the music. Instead, they gave higher ratings to the metronome that aligned not just with beats but specifically with the downbeats of each measure.

Participants gave higher ratings of fit to slower tempo ($\gamma = 0.23$; 95% CI [0.16, 0.30]) and faster tempo ($\gamma = 0.20$; 95% CI [0.13, 0.27]) excerpts than to medium tempo excerpts, regardless of metronome alignment. Neither the meter of the musical excerpt nor participants' musical training, dance training, and music listening significantly impacted participants' ratings.

Random effects for fully matching metronomes indicated that participants that gave higher overall ratings did not rate fully

³ As in Experiment 1A, we performed the MLM twice: once including the $n = 14$ non-English L1 participants, and once without. There were no changes in the direction/sign of the results as a result of excluding the non-native English speakers: we report the results from the full sample.

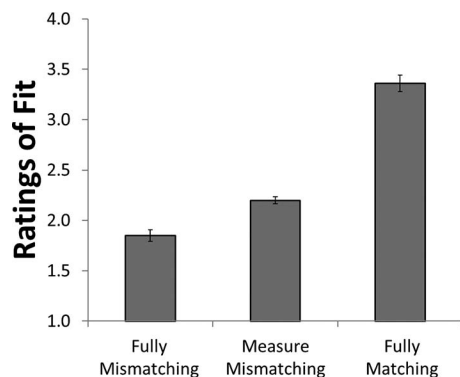


Figure 5. Ratings of fit between measure-only metronome and music as a function of the measure-level alignment between the metronome and music. Visualization collapsed over the variables of tempo and meter. Error bars are within-subject standard error (Cousineau, 2005).

matching metronomes as highly as participants with lower overall ratings ($r_{\text{fully matching, intercept}} = -.56$). We did not find a significant random effects relationship for fully mismatching metronomes.

Hypothesis 2: Musical training affects measure-level perception. Unlike in Experiment 1A, where we reported interactions between music training and sensitivity to both beat and measure, we did not find evidence to support the hypothesis that music training influences ratings when only measure-level information is available. All participants differentiated between all three metronomes in Experiment 1B, perhaps because the present task was simpler in that it did not require listeners to simultaneously track both beat- and measure-level information.

The full model, with the added within-participants predictors, accounted for approximately 67% of the unexplained variance in ratings of fit within-subjects relative to the random-effects ANOVA baseline model. The between-subjects variability increased to 19%, and the within-subjects variability decreased to 81% in the full model.

The results of Experiment 1B clearly demonstrate that when beat-level information is removed from metronomes, all listeners—regardless of musical training—readily distinguish measure-matching from measure-mismatching metronomes. In fact, sensitivity to measure-level matching appears to be greater when beat-level information is left out, as suggested by the greater difference in ratings between measure-mismatching and fully matching metronomes in Experiment 1B as compared with the corresponding metronomes in 1A. This does not necessarily indicate that listeners can attend to beat- and measure-level information simultaneously in all situations, given that they might have simply shifted their attention from the beat level to the measure level when appropriate. However, the significant difference between the measure-mismatching and fully mismatching conditions in Experiment 1B is not predicted by the roughly equivalent measure-level asynchrony for these two conditions (see Table 2). It instead suggests that at the very least, participants in Experiment 1B were simultaneously sensitive to the presence of synchrony or asynchrony at the beat level, and they decided a beat-level match was a better fit than a complete mismatch.

This result also stands in contrast to the unexpected trend in Experiment 1A for individuals with lower amounts of formal

musical training to give lower fit ratings to measure-matching/beat-mismatching metronomes than to fully mismatching metronomes, which reversed in Experiment 1B when beat-level asynchronies were removed from the metronome. It is possible that listeners with less musical training may have difficulty perceiving two levels of meter simultaneously in situations with a conflict between beat and measure. Most Western music does not contain situations in which beat- and measure-level information conflict, at least not for longer than a few seconds (London, 2012). Therefore, nonmusicians would be less likely to encounter this situation in naturalistic musical listening situations as compared with musically trained individuals. Music training may also influence general cognitive abilities, such as working memory, cognitive flexibility, or executive function (Bialystok & Depape, 2009; Moreno et al., 2011; Zuk, Benjamin, Kenyon, & Gaab, 2014), making it easier for individuals with more musical training to resolve apparent conflicts between information at different metrical levels.

Experiment 2A

Experiments 1A and 1B provided evidence that listeners are sensitive to both the beat and measure level of musical meter and that this ability does not require musical training. Experiment 2A examined when this sensitivity emerges during childhood. Because some studies have suggested that neonates and young infants respond to disruptions or changes of beat in rhythmic sequences (Hannon & Johnson, 2005; Hannon & Trehub, 2005a; Winkler et al., 2009), it is often assumed that sensitivity to metrical perception emerges very early in life. This view is undermined by evidence that exposure to and familiarity with a musical idiom enhances sensitivity to metrical disruptions (Hannon, Soley, & Levine, 2011; Hannon, Soley, & Ullal, 2012; Hannon & Trehub, 2005a, 2005b; Hannon et al., 2012; Ullal-Gupta, Hannon, & Snyder, 2014). Such evidence suggests that meter perception could arise

Table 5
Results of Multilevel Models in Experiment 1B Investigating Differences in Ratings of Fit of Metronome to Music

Parameter	Baseline	Full model
Fixed effects		
Rating of fit (intercept)	2.792**	2.459**
Within-subjects factors (Level 1)		
Measure—fully matching		0.766**
Measure—fully mismatching		−0.746**
Slower tempo		0.229**
Faster tempo		0.200**
Musical meter		−0.017
Fully Matching × Musical Training		0.063
Fully Mismatching × Musical Training		−0.029
Between-subjects factors (Level 2)		
Years musical training		−0.019
Years dance training		0.016
Hours music listening		0.001
Random effects		
Within-subjects (Residual)	1.218**	0.664**
Between-subjects (Intercept)	0.052**	0.152**
Intraclass correlation (ICC)	0.041	0.187

Note. All between-participant coefficients are unstandardized, and all between-participant predictors are grand-mean centered.

** $p < .01$.

gradually over childhood as listeners acquire nuanced representations of the structure of music in their culture. In Experiment 2A, we used our paradigm cross-sectionally with young listeners at a range of ages, to examine the developmental trajectory of meter perception.

In Experiment 2B, we compared children's and adolescents' judgments of fit in response to metronome/music pairings that only matched or mismatched at the measure level. Like Experiment 1B with adults, we asked whether children and adolescents can perceive measure-level information when metronomes contained no beat-level information.

Method

Participants. For Experiment 2A, we recruited a sample of typically developing children and adolescents, ranging from 5 to 17 years of age ($N = 154$, 91 female), distributed across five contiguous age groups of roughly equal size. Caregivers of all children and adolescents reported that children had no known hearing impairments and were either in good health or had no more than a minor cold or illness on the day of testing. Because our above findings with adults suggested that music training is not necessary for meter perception, we did not recruit separate samples of musically trained and untrained children. Recruiting from the general child population, only 18 children (of 154) had more than 3 years of music training, and only 21 had more than 3 years of dance training. Full demographic and background information is provided in Table 6. As a token of appreciation, children and adolescents received a toy and t-shirt after their session at the lab.

An additional eight participants began the task but were excluded from data analysis due to unwillingness to complete the experiment ($n = 6$) or failure to follow instructions (i.e., giving the same response throughout the entire experiment and/or admitting that they ignored instructions; $n = 2$). This research was approved by the Social/Behavioral Institutional Review Board of the University of Nevada, Las Vegas.

Power analysis. We wished to investigate whether children's and adolescents' sensitivity to beat- and/or measure-level information related to their age. This could manifest as two-way (i.e., Beat \times Age or Measure \times Age) or three-way (Beat \times Measure \times Age) interactions. Younger children perform worse compared with adults on similar metronome-judging tasks (Einarson & Trainor, 2016; Puyjarinet et al., 2017); thus, we expected a smaller effect size than in the adult sample in Experiment 1A. Previous effect sizes reported for tasks adapted from the Beat Alignment Test (Iversen & Patel, 2008) ranged from $d = 0.68$ to 1.14 (Einarson & Trainor, 2015, 2016). We modeled our two-level participant-clustered model and found that to achieve a power of 0.8 to detect a fixed effect with an effect size of $d = 0.15$, given a conservative estimate of a low effect size ($r^2 = 0.05$) for the Level 2 covariate of age, we needed approximately 150 individuals with 48 trials each, or approximately 80 individuals with 96 trials each as a total sample size (Raudenbush & Liu, 2000; Spybrook et al., 2011). For larger effect sizes (both d at Level 1 and r^2 at Level 2), considerably fewer participants were necessary. Our sample size of 154 participants (92 with 48 trials each and 62 with 96 trials each) met this threshold for sufficient power and was above the 30/30 rule suggested for multilevel model research (Hox, 1995, 2010).

Table 6
Experiment 2A Participant Demographic Information

Characteristic	Full sample	5- to 6-year-olds	7- to 8-year-olds	9- to 10-year-olds	11- to 13-year-olds	14- to 17-year-olds
n (females)	154 (91)	30 (14)	31 (15)	31 (22)	31 (21)	31 (19)
Age in years	10.38 (3.54)	5.83 (0.46)	7.91 (0.62)	9.99 (0.56)	12.23 (0.94)	15.79 (1.18)
Age range	5.08–17.92	5.08–6.83	7.08–8.83	9.00–10.83	11.00–13.83	14.08–17.92
Age in months	124.56 (42.53)	70.00 (5.57)	95.00 (7.43)	119.84 (6.69)	146.71 (11.29)	189.48 (14.17)
Mother with 4-year college degree or higher (% of sample)	76 (49%)	13 (43%)	16 (52%)	12 (39%)	14 (45%)	21 (68%)
n Hispanic (% of sample)	44 (29%)	5 (17%)	9 (29%)	10 (33%)	11 (35%)	9 (29%)
n bilingual (% of sample)	47 (31%)	9 (30%)	12 (39%)	7 (23%)	7 (23%)	12 (39%)
n participants with musical training (% of sample)	64 (42%)	7 (23%)	8 (26%)	13 (42%)	16 (52%)	20 (65%)
Years musical training	1.14 (2.03)	0.3 (7)	0.59 (1.28)	1.04 (1.68)	1.95 (2.09)	1.66 (3.13)
Range of musical training	0.10–12.00	0.50–3.00	0.33–5.00	0.15–6.00	0.10–8.00	1.00–12.00
n participants with dance training (% of sample)	69 (45%)	11 (37%)	11 (35%)	18 (58%)	16 (52%)	13 (42%)
Years dance training	1.49 (2.47)	0.46 (0.92)	0.82 (1.52)	1.57 (2.12)	1.96 (2.94)	2.45 (3.47)
Range of dance training	0.05–12.00	0.05–3.50	0.15–6.00	0.15–8.00	1.00–12.00	0.50–10.00
Hours music listening per week	12.15 (14.39)	6.73 (6.59)	6.88 (7.95)	8.63 (10.33)	15.81 (15.88)	19.98 (23.24)
WASI-II Combined t scores	101.86 (13.27)	97.59 (13.37)	104.72 (15.05)	102.10 (11.45)	103.10 (11.64)	101.74 (14.35)

Note. WASI-II = Wechsler Abbreviated Scale of Intelligence. All values are means and standard deviations unless otherwise specified. Years musical training and years dance training include all participants. Ranges of musical and dance training contain only participants with relevant training. All values based on caregiver report of child information.

Stimuli. Stimuli were identical to Experiment 1A. To keep the study developmentally appropriate, the length of the experiment was shortened for the youngest groups. Younger children (5–10 years) heard 48 metronome/music pairings, whereas older children/adolescents (11–17 years) heard 96 metronome/music pairings.

Procedure. Participants were tested individually on an iMac computer (Apple, Inc., Cupertino, CA) booted into Windows 7 (Microsoft Corporation, Seattle, WA) running Presentation Software. The computer was equipped with a Cedrus RB-830 response pad (Cedrus Corporation, San Pedro, CA), a desktop keyboard, two desktop speakers, and either child-sized headphones (Kidz Gear, El Dorado Hills, CA) or adult-sized headphones (Sony MDR 7506 headphones, Sony Electronics Inc., Park Ridge, NJ), which were used as appropriate to the participant's head size.

Procedures for the task were the same as in Experiment 1A, with the following modifications for younger children ages 5–10. To make the task more engaging, the experimenter told each participant that a fictional band made entirely of bugs (insects) was searching for a new drummer to join their band. Each participant was read the following instructions by the experimenter, which were also presented on the screen: "Bugsy [the main bug character] needs your help to judge how good the performances are. He is asking you to be an audition judge, and let him know how good the drummers play along with his songs. [...] To be an audition judge, you will listen to the song and the drummer together, and then tell Bugsy how well the drummer played along with Bugsy's song."

After each experimental trial, children were prompted to rate the match between the metronome and the music with the text "How well did the drummer match Bugsy's song?" Participants ages 5–10 used the response box to give their responses on four colored buttons (red, yellow, green, and blue), with each color corresponding to a rating from "1" to "4," as in Experiment 1A. Red responses indicated *not very well* matching, and blue responses indicated *well matching*, as was visible on the screen and explained to children by the experimenter. This scale was additionally illustrated with a color-matched "thumbs up" or "thumbs down" system visible on the computer screen, where red (1) was marked as "two thumbs down," yellow (2) as "one thumb down," green (3) as "one thumb up," and blue (4) as "two thumbs up." This explanation (colors and symbols) was visible every time the participant rated the match between music and metronome.

Because the younger participants only received 48 trials instead of 96, the test trials were divided into three blocks of 16 trials each. Children also received six practice trials instead of three (two fully matching, two beat-matching measure-mismatching, and two fully mismatching), with the option to repeat the trials if the child did not understand the task after the first practice session.

Adolescents ages 11–17 performed the task exactly as adults (four blocks of 24 trials each), with the same instructions and practice trials used in Experiment 1A. They used the numbers 1 through 4 on the keyboard for their responses, the same as adults. While the participant (child or adolescent) completed the metronome matching task, the caregiver(s)/parent(s) of participants completed a demographic form about the child/adolescent's developmental, music, dance, and language history.

To account for the possibility that variations in performance on the task were attributable to developmental or individual differ-

ences in cognitive ability rather than beat perception, we performed a brief assessment of IQ of all child and adolescent participants. All participants completed the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler & Hsiao-Pin, 2011). The two subtests measure verbal and nonverbal intelligence, respectively, and taken together they yield an estimate of general cognitive ability (IQ). In the vocabulary subtest, participants define a series of words, whereas in the matrix reasoning subtest, participants select an image from a set of options that best completes a larger, incomplete pattern. Raw scores on the WASI-II subtests were converted to standardized scores based on the published age norms. Because the WASI-II is only normed to age 6, for participants younger than 6 years old we used the youngest published norms. The WASI-II was administered by the experimenter or another research assistant, which took approximately 10–15 min per participant. WASI-II data were not available for six participants who completed the behavioral task, owing to experimenter error ($n = 2$) or refusal to participate or answer questions ($n = 4$).

Data analysis. We used MLM to examine the effects of beat- and measure-level matching metronomes on children's and adolescents' ratings of fit, along with the effects of age and musical training on beat- and measure-level perception. Our primary hypothesis was that age would relate to young listeners' perception of the beat- and measure-level information in the metronome and music. In Model 1 we entered the within-subjects variables of Beat (matching or mismatching), Measure (matching or mismatching), Tempo (slow, medium, or fast), Meter (duple or triple); the within-subjects interaction of Beat \times Measure; the between-subjects variables of Age (in months), Musical Training, Dance Training, Music Listening, and Intelligence; and the cross-level interaction terms of Beat \times Age, Measure \times Age, and Beat \times Measure \times Age. Our secondary hypothesis was that musical training would interact with children's and adolescents' perception of beat- and measure-level information in music, as it did in adults in Experiment 1A. In Model 2 we added the Beat \times Musical Training, Measure \times Musical Training, and Beat \times Measure \times Musical Training cross-level interactions into the model.

Within-subjects (Level 1) predictors were dichotomous and used the same coding scheme as Experiment 1A. Between-subjects (Level 2) predictors were unstandardized and grand-mean centered. Participants' ages were rounded to the nearest whole month from the date of participation. We treated missing data as MCAR. Some ratings were missing (10; <1%). A total of six participants had at least one missing response, with participants failing to respond in 1–4 trials. We did not have values for years of musical training for two participants, years of dance training for two participants, hours of music listening for 26 participants, and IQ score estimates for five participants. We did not interpolate or mean-substitute values. All statistical analyses were performed using the same program and parameters as prior experiments. All models presented successfully converged. Syntax for models and datasheets are available on the OSF page.

Results and Discussion

MLM results are reported in Table 7 and ratings by metronome condition are displayed in Figure 6. The random-effects ANOVA baseline model partitioned the variance in ratings as mostly attrib-

Table 7
Results of Multilevel Models in Experiment 2A Investigating Differences in Ratings of Fit of Metronome to Music

Parameter	Baseline	Model 1	Model 2
Fixed effects			
Rating of fit (intercept)	2.69**	2.44**	2.44**
Within-subjects factors (level 1)			
Beat		0.739**	0.726**
Measure		-0.651**	-0.650**
Slower tempo		0.025	0.025
Faster tempo		0.074**	0.074**
Musical meter		-0.059	-0.059
Beat × Measure		0.326**	0.325**
Beat × Age		0.008**	0.006**
Measure × Age		-0.008**	-0.007**
Beat × Measure × Age		<.001	-0.001
Beat × Musical Training			0.087**
Measure × Musical Training			-0.059
Beat × Measure × Musical Training			0.059
Between-subjects factors (Level 2)			
Age (in months)		-0.007**	-0.006**
IQ		0.003	0.003
Musical training		0.023	-0.027
Dance training		-0.022	-0.022
Music listening		-0.002	-0.001
Random effects			
Within-subjects (residual)	1.202**	0.943**	0.943**
Between-subjects (intercept)	0.086**	0.128**	0.120**
Intraclass correlation (ICC)	0.066	0.120	0.113

Note. All between-participant coefficients are unstandardized, and all between-participant predictors are grand-mean centered.

** $p < .01$.

utable to within-participants factors (1.20; 95% CI [1.17, 1.24]), with a smaller portion of variance attributable to between-participants factors (0.09; 95% CI [0.06, 0.11]). Approximately 93% of the variance in ratings was within-participants (experimental effects) and 7% between-participants (individual differences).

Hypothesis 1: Age affects beat- and measure-level perception. Model 1 tested the hypothesis that age interacted with young listeners' perception of beat- and measure-level information in the music and metronomes. There were significant fixed effects of Beat, Measure, the Beat × Measure interaction, Tempo, and participant Age, and significant Beat × Age and Measure × Age interactions. Children and adolescents gave beat-matching metronomes higher ratings than beat-mismatching metronomes ($\gamma = 0.74$, 95% CI [0.63, 0.85]), and gave measure-matching metronomes lower ratings than measure-mismatching metronomes ($\gamma = -0.65$, 95% CI [-0.82, -0.49]).

The Beat × Measure interaction was also significant. A simple-slopes test examining these variables while controlling for all other factors demonstrated when the beat of the metronome mismatched the music, all children gave significantly lower ratings to measure-matching metronomes (1.78, 95% CI [1.62, 1.95]) than to measure-mismatching metronomes (2.44, 95% CI [2.35, 2.52]; $\omega = -0.65$, 95% CI [-0.82, -0.49]). This was similar to adults, who also gave lower ratings to beat-mismatching/measure-matching metronomes, likely because of the greater mismatch between the beat of the metronome and music in this condition. However, unlike adults, when controlling for all other factors, children gave lower ratings to measure-matching metronomes

(2.85, 95% CI [2.78, 2.92]) than to measure-mismatching metronomes (3.18, 95% CI [3.09, 3.26]) when the beat of the metronome matched the music ($\omega = -0.33$, 95% CI [-0.40, -0.26]). As shown in Figure 6, this trend is not evident from overall means, where ratings appear comparable for the two beat-matching metronomes. This result thus underscores the importance of other factors in the model—especially the interactions with age—for predicting variation in children's ratings.

As hypothesized, the Beat × Age and Measure × Age interactions were both significant. A simple slopes test demonstrated that all children rated beat-matching metronomes as better fitting than beat-mismatching metronomes, and this difference increased with age (see Table 8), suggesting that sensitivity to beat-level information develops with age. The Measure × Age interaction effect was in the opposite direction. The simple-slopes test demonstrated that all children rated measure-matching metronomes as worse fitting than measure-mismatching metronomes, and the difference in ratings increased with age (see Table 8). This may once again reflect an increasing sensitivity to the beat-level mismatching in the beat-mismatching/measure-matching metronome condition, rather than a "true" effect of measure (we investigate this in Experiment 2B, see below). The Beat × Measure × Age interaction was not statistically significant.

All children also gave slightly higher ratings to metronomes paired with faster (124 BPM) music compared with moderate tempo music regardless of metronome alignment, controlling for all other factors ($\gamma = 0.07$, 95% CI [0.02, 0.13]). Age was the only significant between-participants factor: for every month in age above the grand mean (124.56 months; 10.4 years), children's average ratings slightly decreased ($\gamma = -0.006$, 95% CI [-0.008, -0.004]), regardless of metronome alignment to the music.

After controlling for all within-participant variables, the between-participants factors of musical training, dance training, music listening, and IQ did not significantly impact participants' ratings. Random effects of beat were the same as Experiment 1A: younger listeners with higher overall ratings gave lower ratings of fit to beat-matching metronomes compared with younger listeners that had lower overall ratings ($r_{\text{beat,intercept}} = -.61$). In Model 1, the addition of the within-participants predictors accounted for approximately 21% of the variance relative to the random-effects ANOVA baseline model. The proportion of unaccounted-for within-participants variability decreased to 88% and the proportion of unaccounted-for between-participants variability increased to 12%.

Hypothesis 2: Musical training affects beat- and measure-level perception while controlling for age effects. In Model 2 we tested the hypothesis that musical training affects the perception of beat- and measure-level information in children and adolescents above the impact of age, through the addition of the cross-level interaction terms of Beat × Musical Training, Measure × Musical Training, and Beat × Measure × Musical Training. The pattern of significance for the previously added fixed effects did not change and the coefficients remained similar to those estimated in Model 1 (see Table 7). Of the added interaction terms, only Beat × Musical Training was statistically significant. A simple slopes test demonstrated that amount of musical training did not significantly relate to how participants rated beat-mismatching metronomes, but ratings of beat-matching metronomes were significantly higher in children with more musical

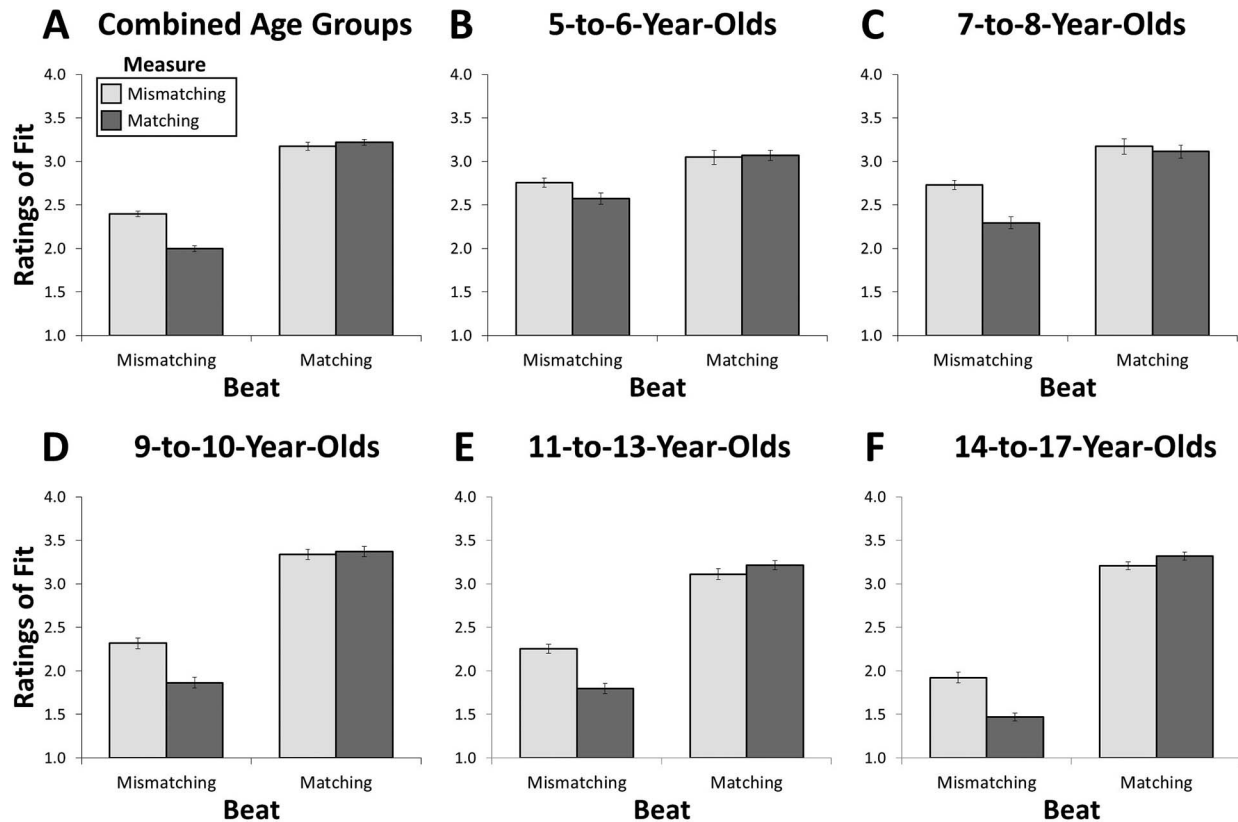


Figure 6. Children and adolescents' ratings of metronome to music as a function of metronome alignment. (A) combined sample ($n = 154$), (B) 5- to 6-year-olds ($n = 30$), (C) 7- to 8-year-olds ($n = 31$), (D) 9- to 10-year-olds ($n = 31$), (E) 11- to 13-year-olds ($n = 31$), (F) 14- to 17-year-olds ($n = 31$). All error bars represent within-subject error (Cousineau, 2005).

training ($+1$ SD from the mean; Table 8). Musical training did not significantly relate to how participants rated measure-level differences between metronomes.

Given metronomes that matched or mismatched rich, ecologically valid examples of ballroom dance music, children and adolescents ages 5–17 were able to determine when the beat of the metronome matched or mismatched the music. These results indicate that even the youngest children could perceive musical beat, and their performance gradually increased with age. It appears, based on visual inspection (see Figure 6), that with age, participants increasingly tended to give higher ratings to metronomes that matched at both the beat and measure level than to metronomes that just matched at the beat level, like adults. However, there was considerable heterogeneity within ratings in the older ages: it is likely that meter perception emerges gradually in late adolescence, and likely at individually differing rates based on other factors. The current results support the idea that the ability to perceive multiple levels of metrical structure in music develops slowly, emerging some time in late adolescence, and does not reach full maturity until young adulthood.

As a group, children in the study seemed not to use measure-level matching in their ratings. On one hand, unlike adults, children showed no evidence of sensitivity to measure-level information in beat-matching conditions when holding all other factors

constant (and age did clearly relate to differences in perception). On the other hand, in beat-mismatching conditions, children—like adults—gave higher ratings to measure-mismatching than measure-matching metronomes. As in Experiment 1A, this could arise because in the two beat-mismatching conditions, there was greater beat-level asynchrony in the measure-matching than measure-mismatching versions (see Table 2). Experiment 1B showed that when beat-level information was removed, adults responded more robustly to measure-level information. Perhaps the children and adolescents in our study were sensitive to measure-level information but simply devoted their attention to beat-level synchrony. If this is the case, then children should be able to detect measure-level synchrony when beat-level information is removed.

Alternatively, younger children might not pick up on measure-level synchrony at all. Children might have general difficulty tracking slower temporal information, as suggested by findings that young children have faster preferred tempos and greater difficulty synchronizing to music at slower (higher) metrical levels (Drake, Jones, et al., 2000; McAuley et al., 2006). If this is the case, then even when beat-level information is removed, younger children's ratings should remain unaffected by measure-level synchrony. To resolve these questions, we conducted a version of Experiment 1B with children and adolescents ages 5 to 13 years to

Table 8
Simple Slopes Tests From Experiment 2A of Beat \times Age, Measure \times Age, and Beat \times Musical Training

Factor	Intercept	95% CI	Slope	95% CI
Beat \times Age				
5- to 6-year-olds	2.83**	[2.69, 2.97]	0.309**	[0.13, 0.49]
7- to 8-year-olds	2.65**	[2.55, 2.76]	0.506**	[0.38, 0.64]
9- to 10-year-olds	2.47**	[2.38, 2.56]	0.701**	[0.59, 0.81]
11- to 13-year-olds	2.28**	[2.18, 2.37]	0.913**	[0.79, 1.03]
14- to 17-year-olds	1.97**	[1.80, 2.13]	1.249**	[1.05, 1.45]
Measure \times Age				
5- to 6-year-olds	2.83**	[2.69, 2.97]	-0.216**	[-0.37, -0.06]
7- to 8-year-olds	2.65**	[2.55, 2.76]	-0.416**	[-0.52, -0.32]
9- to 10-year-olds	2.47**	[2.38, 2.56]	-0.614**	[-0.76, -0.46]
11- to 13-year-olds	2.28**	[2.18, 2.37]	-0.828**	[-1.08, -0.57]
14- to 17-year-olds	1.97**	[1.80, 2.13]	-1.169**	[-1.61, -0.73]
Beat \times Musical Training				
No music training	2.47**	[2.37, 2.57]	0.627**	[0.51, 0.75]
Sample average	2.44**	[2.36, 2.53]	0.726**	[0.63, 0.83]
+1 SD music training	2.37**	[2.27, 2.50]	0.901**	[0.76, 1.04]

Note. All coefficients are unstandardized. Intercept value indicates rating of fit for mismatching (beat or measure, as appropriate) metronome. Age group tests are performed using the mean age of participants in that age group, and music training tests using the mean and standard deviations of the sample (see Table 6).

** $p < .01$.

investigate whether they could distinguish measure-matching from measure-mismatching metronomes when the beat-level of the metronome was removed.

Experiment 2B

Method

Participants. We recruited 42 (25 female) child and adolescent participants between the ages of 5–13 years, spread across four age groups, who did not participate in Experiment 2A. Demographic information is provided in Table 9, separated by age group. Families were recruited from the Las Vegas area using fliers and advertisements for participating in scientific studies related to the development of music and language. Caregivers of

all children and adolescents reported that children were normal-hearing and had no known hearing impairments and were either in good health or had no more than a minor cold or illness on the day of testing. As a token of appreciation for participating in the study, children and adolescents received a toy and t-shirt after their session at the lab. One additional participant started the task but withdrew from participation prior to completing the practice session of the experiment; their data are not included. This research was approved by the Social/Behavioral Institutional Review Board of the University of Nevada, Las Vegas.

Power analysis. In Experiment 1B, the effect of measure-level information for adults was much stronger than in Experiment 1A ($\gamma = 0.76$ vs. $\gamma = -0.10$), suggesting that when beat-level information is removed from the metronome, the measure-level

Table 9
Experiment 2B Participant Demographic Information

Characteristic	Full sample	5- to 6-year-olds	7- to 8-year-olds	9- to 10-year-olds	11- to 13-year-olds
<i>n</i> (<i>n</i> females)	42 (25)	12 (6)	11 (9)	9 (4)	10 (6)
Age in years	9.09 (2.44)	6.18 (0.55)	8.31 (0.67)	10.33 (0.56)	12.48 (0.89)
Age range in years	5.25–13.25	5.25–6.91	7.17–8.91	9.08–10.66	11.0–13.25
Age in months	109.02 (29.30)	74.92 (6.57)	99.82 (7.99)	120.44 (6.77)	149.80 (10.65)
Mother with 4-year college degree or higher (% of sample)	22 (52%)	6 (50%)	7 (63%)	4 (44%)	5 (50%)
<i>n</i> Hispanic (% of sample)	14 (33%)	4 (33%)	2 (17%)	4 (44%)	4 (40%)
<i>n</i> bilingual (% of sample)	23 (55%)	4 (33%)	9 (73%)	4 (44%)	6 (60%)
<i>n</i> participants with musical training (% of sample)	17 (40%)	3 (25%)	4 (36%)	4 (44%)	6 (60%)
Years musical training	1.00 (1.62)	0.58 (1.44)	0.50 (0.74)	1.25 (2.20)	1.80 (1.75)
Range of musical training	0.5–6.0	1.0–5.0	0.75–2.0	0.50–6.0	2.0–4.0
<i>n</i> participants with dance training (% of sample)	24 (57%)	5 (45%)	7 (63%)	5 (56%)	7 (70%)
Years dance training	2.55 (3.15)	1.83 (2.25)	2.14 (2.45)	1.89 (2.47)	4.40 (4.67)
Range of dance training	1.0–12.0	1.0–6.0	1.0–8.0	1.0–6.0	1.0–12.0
Hours music listening per week	6.95 (6.90)	9.64 (11.46)	5.27 (3.17)	4.25 (2.91)	8.25 (3.88)

Note. All values are means and standard deviations unless otherwise specified. Years musical training and years dance training include all participants. Ranges of musical and dance training contain only participants with relevant training. All values based on caregiver report of child information.

information is far more salient to listeners. In Experiment 2A, the Beat \times Age and Measure \times Age interactions were statistically significant. Thus, we sought to detect an interaction between age and measure perception (when metronomes contained no beat-level information) and assumed that the effect size should be of at least moderate size. Assuming a fixed effect size of $d = .25$ for an interaction between age group and measure, we needed a minimum of 35 participants with 48 trials each to detect such an effect with a power of 0.8 (Raudenbush & Liu, 2000; Spybrook et al., 2011). Our sample size met this minimum (42/48) and was above the 30/30 rule (Hox, 1995, 2010).

Stimuli. The stimuli were identical to those in Experiment 1B: the beat-level clicks of the metronome were silenced, leaving only the measure-level clicks (see Figure 4). As in Experiment 2A, to keep the task developmentally appropriate for the younger participants, the number of music/metronome pairings was half (48) those in Experiment 1B.

Procedure. All participants were tested individually on an iMac computer booted into Windows 7 running Presentation software. The participant wore either child-sized (Kidz Gear) or adult-sized (Sony MDR 7506) headphones depending on what was most appropriate for the child or adolescent's head size and comfort.

All participants (ages 5–14) completed the same task. The experimental procedure, program, and task were identical to the program used by the 5- to 10-year-old participants in Experiment 2A, with a minor change to the “story” portion of the task. Participants now judged how well a single bug’s “drumming” matched the music instead of multiple bugs. The instructions were otherwise the same as in Experiment 2A: on each trial, participants were asked “how well did Frances’ [the bug drummer] drumming match Bugsy’s song?” The response box, labels, and computer images were the same as in Experiment 2A.

Data analysis. Using MLM, we examined the effects of measure-level alignment between the metronome and music on younger listeners’ ratings of fit while controlling for tempo and meter differences within participants, and the effects of age, musical training, dance training, and music listening between participants. Our primary hypothesis was that age would interact with children’s and adolescents’ perception of the fit between the metronomes to the music, even controlling for other within- and between-participants factors. We fit a single model with Measure (fully matching, measure mismatching, and fully mismatching), Tempo (slow, medium, and fast), and Meter (duple or triple) as within-subject variables; Age (in months), Musical Training, Dance Training, and Music Listening as between-subjects variables; and cross-level interactions of Measure \times Age and Measure \times Musical Training. The dependent variable was participants’ ratings of fit. We compared this model with a baseline model. All level 1 predictors were dichotomous and coded as in Experiment 1B. Level 2 predictors were grand-mean centered and unstandardized.

Missing data were treated the same as in previous experiments. There were no missing response data and no missing within-participants (level 1) predictors. We did not have values for hours of music listening for four participants and did not replace missing data. All statistical analyses were performed using the same program and parameters as previous experiments. All models presented successfully converged. Syntax and datasheets for all models are available on the OSF page.

Results and Discussion

MLM results are reported in Table 10, and ratings of fit for each metronome condition are presented in Figure 7. In the random-effects ANOVA baseline model, most of the variance in ratings of fit was within-participants (1.00; 95% CI [0.94, 1.06]) and a smaller portion was between-participants (0.13; 95% CI [0.08, 0.22]). Of the unexplained variance in ratings of fit, approximately 88% of the variance was within-participants (experimental factors) and 12% was between-participants (individual differences).

There were significant fixed effects of Measure and Age, and significant Measure \times Age and Measure \times Musical Training interactions. Compared with measure mismatching metronomes, children as a group gave significantly higher ratings to fully matching metronomes ($\gamma = 0.24$, 95% CI [0.09, 0.38]), and significantly lower ratings to fully mismatching metronomes ($\gamma = -0.41$, 95% CI [-0.59, -0.24]). Age related to lower ratings of fit, regardless of metronome alignment: older children gave lower average ratings than younger children ($\gamma = -0.006$, 95% CI [-0.012, -0.0003]).

The Measure \times Age interaction suggested that sensitivity to measure alignment changed with age. The difference in ratings of fully matching versus fully mismatching metronomes was larger in older than younger children. A simple slopes test demonstrated that 5- to 8-year-olds’ ratings did not significantly differentiate measure mismatching from fully matching metronomes, whereas 9- to 13-year-olds, like adults in Experiment 1B, gave significantly higher ratings to fully matching than to measure mismatching metronomes (see Table 11). This result suggests that children in the youngest groups were not sensitive to measure-level matching even when beat-level information was removed. Five- to 6-year-

Table 10
Results of Multilevel Models in Experiment 2B Investigating Differences in Ratings of Fit of Metronome to Music

Parameter	Baseline	Full model
Fixed effects		
Rating of fit (Intercept)	2.96**	2.94**
Within-subjects factors (Level 1)		
Measure—fully matching		0.237**
Measure—fully mismatching		-0.413**
Slower tempo		0.099
Faster tempo		0.012
Musical meter		-0.044
Fully Matching Measure \times Age		0.013**
Fully Mismatching Measure \times Age		-0.012**
Fully Matching Measure \times Musical Training		-0.101*
Fully Mismatching Measure \times Musical Training		-0.041
Between-subjects factors (Level 2)		
Age (months)		-0.006*
Musical training		0.062
Dance training		0.046
Music listening		-0.011
Random effects		
Within-subjects (residual)	1.00**	0.789**
Between-subjects (intercept)	0.135**	0.160**
Intraclass correlation (ICC)	0.119	0.169

Note. All between-participant coefficients are unstandardized, and all between-participant predictors are grand-mean centered.

* $p < .05$. ** $p < .01$.

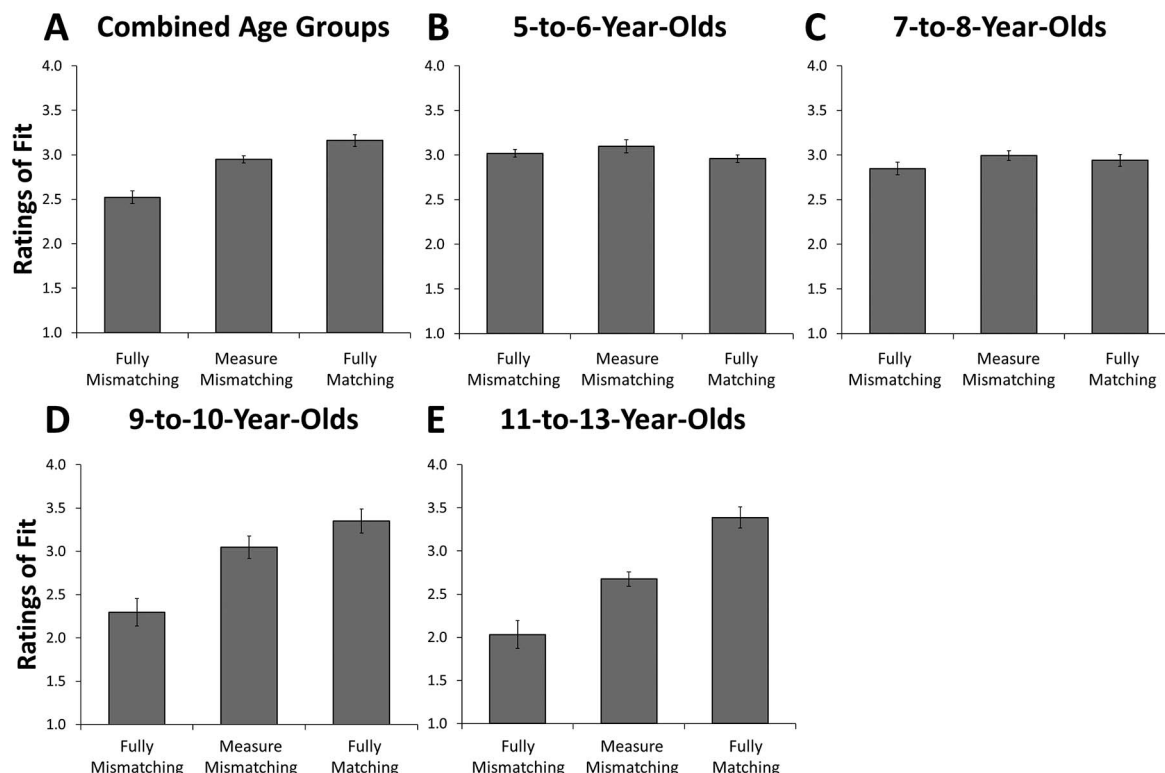


Figure 7. Children and Adolescents' ratings of fit between measure-only metronomes and music separated by metronome condition. (A) full sample ($n = 42$), (B) 5- to 6-year-olds ($n = 12$), (C) 7- to 8-year-olds ($n = 11$), (D) 9- to 10-year-olds ($n = 9$), (E) 11- to 13-year-olds ($n = 10$). Error bars indicate within-subject standard error (Cousineau, 2005).

olds did not differ in their ratings of fit between fully mismatching and measure mismatching metronomes, whereas all other age groups gave significantly lower ratings of fit to fully mismatching metronomes (see Table 11). Thus, the youngest participants' ratings did not differ across the three metronome conditions.

The Measure \times Musical Training interaction was significant for fully matching metronomes, but not fully mismatching metronomes. A simple slopes test demonstrated that the difference in ratings for fully matching versus measure mismatching metronomes was in fact larger for children with more musical training ($+1$ SD above the mean) than for children with less musical training (see Table 11). This significant effect, while seemingly paradoxical, is perhaps consistent with the Beat \times Musical Training interaction in Experiment 2A: If musical training is related to enhanced attention to beat-level information, children with more musical training may have given more weight to any beat-level match than matching on the exact downbeat.

There were no significant random-effects relationships in the model. In the full model, the addition of the within-participants predictors accounted for approximately 21% of the unexplained variance in ratings within-participants, relative to the random-effects ANOVA baseline model. The between-participants variability increased to 17% and the within-participants variability decreased to 83% in the full model.

In summary, when metronomes presented only measure-level information without beat-level information, older children showed evidence of distinguishing between metronomes that matched or mis-

matched the music at the measure-level, but younger children did not. This is remarkable given that the present task could be seen as measuring beat perception but at a slower tempo—the tempo of the measure. Young children's failure in this task is consistent with the notion that they have difficulty tracking slower tempos in general, and it is inconsistent with the interpretation that they are sensitive to measure-level information but simply devote more attention to beat-level information. Even listeners 7 to 8 years of age, who distinguished between fully mismatching and fully matching metronomes, could not distinguish between metronomes that fell on the downbeat and those that fell on another beat.

Together, the results of Experiment 2A and 2B demonstrate that beat perception develops gradually, and that the ability to perceive slower metrical levels (i.e., the measure level) in music takes an even more gradual and protracted developmental course. Although listeners 9–13 years of age were sensitive to measure-level information in the absence of beat-level information in Experiment 2B, listeners of this age did not consistently use both beat- and measure-level information simultaneously in Experiment 2A. This suggests that the ability to perceive multiple levels of musical meter simultaneously is slow to develop and continues to improve into young adulthood.

General Discussion

The two main goals motivating our study were to determine whether listeners without formal music training are sensitive to

Table 11
Simple Slopes Tests From Experiment 2B of Measure \times Age and Measure \times Musical Training

Factor	Intercept	95% CI	Slope	95% CI
Fully Matching Measure \times Age				
5- to 6-year-olds	3.15*	[2.91, 3.40]	-0.20	[-0.42, 0.02]
7- to 8-year-olds	3.00*	[2.83, 3.17]	0.12	[-0.03, 0.27]
9- to 10-year-olds	2.87*	[2.68, 3.05]	0.38*	[0.22, 0.54]
11- to 13-year-olds	2.94*	[2.77, 3.11]	0.24*	[0.09, 0.38]
Fully Mismatching Measure \times Age				
5- to 6-year-olds	3.15*	[2.91, 3.40]	0.001	[-0.26, 0.26]
7- to 8-year-olds	3.00*	[2.83, 3.17]	-0.30*	[-0.47, -0.13]
9- to 10-year-olds	2.87*	[2.68, 3.05]	-0.55*	[-0.74, -0.36]
11- to 13-year-olds	2.68*	[2.38, 2.98]	-0.91*	[-1.23, -0.58]
Fully Matching Measure \times Musical Training				
No music training	2.88*	[2.67, 3.08]	0.34*	[0.17, 0.51]
Sample average	2.94*	[2.77, 3.11]	0.24*	[0.09, 0.38]
+1 SD music training	3.04*	[2.81, 3.27]	0.07	[-0.12, 0.27]

Note. All coefficients are unstandardized. Intercept indicates rating mismatching metronome. Significant simple slope values indicate that the change in ratings of fit (e.g. mismatching to matching metronomes) after controlling for all other variables is statistically significant. Age tests are performed using the mean age for that age group and musical training using the sample average and standard deviation (see Table 9).

* $p < .01$.

multiple metrical levels in music and to investigate how beat and multiple levels of meter perception develop through childhood into adulthood. To do this, we used rich, complex recordings of ballroom dance music containing myriad cues to meter and representing an ecologically valid approach to testing meter perception. Our findings are novel in showing that (a) adult listeners can perceive multiple levels of musical meter simultaneously and do not require formal musical training to do so and (b) prior to age 18, children and adolescents can only attend to one level of metrical structure at a time, and preferentially perceive the beat level rather than the measure level. Moreover, even sensitivity to a single level of beat—at least as measured in our task—appears to increase gradually with age, such that even 14- to 17-year-olds exhibited less sensitivity than did young adults (18+ years).

While adults demonstrated sensitivity to both beat and measure information simultaneously, beat-level information had a much larger overall impact on their responses than did measure-level information. This was shown by larger main effects of beat than measure in Experiments 1A and 2A. Similarly, nonmusicians and children gave lower fit ratings to conditions that had more beat-level asynchrony, even when there was synchrony at the measure-level. This was not due to an inability to perceive measure-level information: Experiment 1B demonstrated the ability to perceive measure-level information in isolation does not depend on formal musical training, since nonmusicians gave higher ratings of fit to metronomes (with only measure-level information) that fell on the measure-level downbeat than to metronomes that fell on other beat-level locations or off-beat locations.

Experiment 2B demonstrated that the ability to perceive slower metrical levels takes a more protracted course than perception of faster levels. Younger children were not capable of perceiving measure-level fit between metronome and music, children ages 9 to 10 years used beat-level information to perform the task, and only the younger adolescents could successfully attend to the measure-level information. Yet even though younger adolescents could attend to a slower level of beat (the measure) in isolation,

they did not show strong evidence of attending to two levels of meter simultaneously in Experiment 2A. This suggests that while children and adolescents can perceive different beat frequencies in music, they are not able to track multiple levels simultaneously. Thus, the ability to perceive the metrical hierarchy of music likely does not develop until very late in adolescence. Even when the ability to perceive multiple levels of meter emerges, however, listeners still attend preferentially to beat-level information.

Greater attention to the beat level than to the measure level aligns with evidence from other tasks investigating meter perception in adults. Studies that recorded neural responses to metrically structured stimuli found the strongest responses at the beat level and considerably smaller responses at slower metrical levels (Nozaradan et al., 2011, 2012). Palmer and Krumhansl (1990) reported that in a metrical probe tone task, both musicians and nonmusicians gave higher ratings to on-beat than off-beat probes, but only musicians' probe tone ratings reflected nuanced differences between higher and lower levels of the metrical hierarchy.

A similar beat-perception study using a metronome-matching task reported roughly 60% accuracy among 5-year-olds for the beat level (Einarson & Trainor, 2016), suggesting that young children struggle to accurately perceive the beat in music. Slowly developing beat perception might partly explain why very young children do not move their bodies in time with music even when explicitly asked to do so (Eerola et al., 2006) and why children's synchronization and rhythm reproduction abilities are far below adult levels but improve with increasing age (Drake, 1993; Drake et al., 2000; McAuley et al., 2006; Provasi & Bobin-Bègue, 2003). This evidence undermines the notion that young children have fully developed perception of beat and meter but are simply limited by immature motor systems (Winkler et al., 2009). Indeed, the development of the motor system could play an interactive role with the development of musical meter perception. Studies with adults performing rhythm-based perception tasks found activation in motor areas of the brain (Chen, Penhune, & Zatorre, 2008; Grahn & Brett, 2007; Marchant & Driver, 2013; Sakai et al., 1999;

Schubotz, Friederici, & von Cramon, 2000) and in neural activity bands linked to motor movement (Fujioka et al., 2010, 2015; Iversen et al., 2009; Snyder & Large, 2005), even when the task was perceptual and did not include a large motor component. Thus, children's still-developing motor systems could gate or otherwise contribute to the slow development of meter perception, with the ability to perceive and produce increasingly complex rhythmic structures developing in tandem. The available evidence suggests general timing abilities undergo gradual developmental changes that influence both perception and production of temporal musical structures.

What might explain such a gradual developmental trajectory for acquiring sensitivity to musical meter? One possibility is that slower levels of metrical structure are incompatible with children's optimal rate of event tracking, which is generally faster than is observed among adults. For example, in two-interval temporal bisection tasks, children under age 9 tend to overestimate auditory intervals (Lustig & Meck, 2011; McCormack, Brown, Maylor, Darby, & Green, 1999). Similarly, children's perceptually preferred tempos are correlated with their spontaneous motor tempos, with gradually increasing preferred IOIs of ~300 ms at age 4–5 to ~600 ms in adulthood (Drake, Jones, et al., 2000; McAuley et al., 2006). Preferred tempo might constrain the rate at which children can accurately synchronize to an auditory stimulus, as well as their success at continuing to produce that target tempo in silence; for young children, synchronization variability increases dramatically at nonpreferred tempos, and continuation tapping reliably drifts toward preferred tempo (McAuley et al., 2006). Given that the tempos of musical excerpts in our study had beat-level IOIs ranging from 484–674 ms (89–124 BPM), all beat-level IOIs were at least slightly slower than the preferred tempo for the youngest group (as suggested by McAuley et al., 2006), and the corresponding measure-level IOIs, which ranged from 1,452–2,696 ms, were far outside young children's optimal tempo range. This could explain why the youngest two age groups appeared to be insensitive to measure-level synchrony even when no other information was available.

This account is consistent with oscillator models of timing, which have been successfully used to model empirical rhythm production and perception data and which provide a framework for understanding beat and meter perception (for a review, see Large & Snyder, 2009). Oscillator models are inspired in part by Dynamic Attending Theory (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999), which assumes that attention entrains to temporal frequencies in rhythmic stimuli. This attentional oscillation can adapt its period and phase to match rhythmic information present in the environment; however, entrainment is not as strong when the period of an external stimulus is too different (larger or smaller) from the intrinsic period of the oscillator (e.g., the preferred tempo). Importantly, the range of tempos over which entrainment can occur appears to widen with age, such that adults can produce and synchronize to a much wider range of tempos than can children (Drake et al., 2000; McAuley et al., 2006). Thus, children's faster preferred tempos and restricted flexibility to adapt to periodicities outside those intrinsic periods may have limited their sensitivity to measure-level and even slower beat-level information in the present study. Future studies could therefore be designed with beat (and measure) tempi that are considerably faster.

Listeners' perception and production of multiple levels of musical meter is also influenced by acquired, top-down knowledge of the structures that predominate in a listener's culture. Gradual acquisition of musical meter is generally consistent with recent findings suggesting cognitive skills and abilities such as language learning, working memory, crystallized intelligence, and even other aspects of music perception continue to develop slowly through adolescence and do not peak until adulthood (Hartshorne & Germine, 2015; Hartshorne, Tenenbaum, & Pinker, 2018; Weiss, Cirelli, McDermott, & Trehub, 2020). Moreover, we know that mastery of a language provides access to increasingly abstract structures that occur at very slow time scales, such that entrained brain responses to language stimuli can be observed at exclusively faster levels in non-native adult speakers but at frequencies as slow as 1 Hz in native speakers (Ding et al., 2017).

Similarly, adults are much more accurate at synchronizing to culturally familiar than unfamiliar music, and they can synchronize with a wider range of metrical levels in culturally familiar music (Cameron, Bentley, & Grahn, 2015; Drake & Ben El Heni, 2003). Adults are less accurate at synchronizing to culturally unfamiliar meters, even when the target is isochronous (Ullal-Gupta et al., 2014). Numerous studies suggest that in both rhythm perception and production tasks, listeners systematically distort rhythmic patterns to fit into culturally familiar metrical templates (Hannon & Trehub, 2005a, 2005b; Hannon et al., 2012; Jacoby & McDermott, 2017; Repp, London, & Keller, 2013; Snyder, Hannon, Large, & Christiansen, 2006). In a metronome-matching task similar to this one, young children were also better able to match a metronome to culturally familiar music than to culturally unfamiliar music (Einarson & Trainor, 2015, 2016). Although sufficiently slow optimal/preferred tempos and wide entrainment regions are probably a prerequisite for perceiving multiple levels of musical meter, it is also possible that acquisition of culture-specific knowledge of musical meter helps listeners to overcome temporal processing constraints that are otherwise observed in tasks using simple stimuli such as metronomes. Future research is needed to examine the extent to which acquisition of culture-specific metrical knowledge interacts with performance on other basic temporal tasks.

Our adult samples in Experiments 1A and 1B did include several nonnative English (English L2) speakers: our research sample was drawn from a linguistically and ethnically diverse university.⁴ In the interest of including a representative sample in our region, we chose not to exclude non-native English speakers. We had no a priori reason to expect native language to influence performance in our task, and our stimuli did not contain complex metrical structures, which could give rise to culture-specific differences in metrical perception (Hannon et al., 2012; Hannon & Trehub, 2005a). We nevertheless conducted analyses with and without non-native English speakers as reported, and obtained the same results either way. Future research comparing the effects of linguistic background on meter perception could specifically recruit larger English L2 samples and compare factors like age of English acquisition and L1 background on meter perception.

We found no evidence that children's general cognitive abilities drive changes or individual differences in sensitivity to metrical

⁴ In 2018, 57% of attending undergraduates at this institution identified as belonging to a racial or ethnic minority in the United States.

structure. There was no relationship between children's and adolescents' IQ scores and their performance in our task. This finding is consistent with prior research, which has reported no association between measures of (verbal) intelligence and beat perception and production (Einarson & Trainor, 2015, 2016; Woodruff Carr et al., 2014) or only a small relationship (McAuley et al., 2006). Other studies have reported that beat-related perceptual processing in young children predicts reading readiness (Ozernov-Palchik, Wolf, & Patel, 2018), and auditory working memory (Einarson & Trainor, 2016; Kraus, Strait, & Parbery-Clark, 2012; Strait, Hornickel, & Kraus, 2011; Woodruff Carr et al., 2014). It is important to acknowledge that small sample sizes and task differences may drive some of the observed variability between studies. For example, synchronization tasks may have different working memory and attentional demands than rhythm discrimination tasks, and different cognitive measures (verbal IQ tests, reading readiness) may involve different timing demands (see Hannon, Nave-Blodgett, & Nave, 2018). We did not assess language abilities or basic auditory processing abilities in either our adult or developmental samples, so we cannot make any definitive conclusions about the relationships between these abilities and meter perception in music.

There is a growing body of research connecting specific aspects of rhythmic processing and rhythm perception to language abilities, especially in developmental disorders such as dyslexia or specific language impairments. Those with developmental dyslexia are less accurate and more variable when tapping to isochronous metronomes as compared with typically developing age-matched samples (Overy, Nicolson, Fawcett, & Clarke, 2003; Thomson, Fryer, Maltby, & Goswami, 2006; Thomson & Goswami, 2008; Wolff, 2002). These same language-impaired populations also have difficulty distinguishing when an accented beat in an otherwise isochronous or simple rhythm is lengthened as compared with the other events in the rhythm (Flaunacco et al., 2014; Goswami, Huss, Mead, Fosker, & Verney, 2013; Huss, Verney, Fosker, Mead, & Goswami, 2011). The severity of individuals' deficits in these rhythm perception tasks relates to their deficits in phonological processing and language abilities. One theory suggests that language processing deficits observed especially in individuals with dyslexia may be based on temporal sampling deficits, especially around the range of 2Hz (Goswami, 2011). The popular music with which most individuals engage tends to have a beat frequency around 2Hz (500 ms; Van Noorden & Moelants, 1999), and it has thus been suggested that musical training could be used for rehabilitation and improvement of temporal perception skills in these populations (Goswami, 2011). However, little is known about how individuals with dyslexia, specific language impairments, or poor readers perceive the beat in rich, complex pieces of music. Our task—which is purely perceptual, and adaptable to many different examples of rich, musical stimuli—may be useful for examining beat perception in populations with known timing deficits, and in populations for whom getting a motor (tapping) response is difficult. Thus, future work using our metronome paradigm to study beat and meter perception could examine the relationship between reading readiness, language abilities, and beat perception in rich, complex pieces of music that incorporate multiple levels of temporal structure—much like natural speech.

The use of rich, culturally familiar musical stimuli is a strength of the present work and may explain why we found that even adults with little to no formal music training can perceive multiple levels of meter simultaneously, unlike in prior work (e.g., Geiser et al., 2009, 2010; Jongsma et al., 2004; Palmer & Krumhansl, 1990). In particular, it may be that such results occurred because prior studies used simple rhythms or stimuli that do not contain as many cues to meter as human-performed music. Adult listeners attend to a broad range of cues to beat and meter, such as loudness, melodic contour, melodic repetition, grouping structure, small- and large-scale rhythmic figures, and harmonic progressions and rhythmic structure (Hannon, Snyder, et al., 2004; London, Polak, & Jacoby, 2017; Windsor, 1993). We therefore propose that use of rich, culturally familiar stimuli maximizes the possibility of observing sensitivity to multiple metrical levels in untrained listeners. A limitation of our study is that some participants may have undiagnosed subclinical hearing loss or frequency-specific hearing loss. By using rich musical stimuli presented well above threshold intensity levels even for those with moderate hearing loss and contained myriad cues to meter—both local and global auditory characteristics—we hope that even individuals who may have undiagnosed hearing deficits were able to perceive the majority of cues to musical meter, something that would be more difficult with simple percussive or computer-generated tone rhythms.

Despite using ballroom dance music, we found no evidence that formal dance training improved beat- or measure-level perception in our children, adolescent, or adult participants. This was unexpected, considering individuals move their body in synchrony with multiple levels of metrical structure in music (Burger et al., 2018; Burger, Thompson, et al., 2014; Toiviainen et al., 2010), and this effect is especially pronounced in individuals with formal dance training (Naveda & Leman, 2010). Two possible situations could have hampered our ability to find a relationship between formal dance training and beat and meter perception in the current studies. First, as we did not specifically recruit individuals with moderate-to-high levels of dance training, our low numbers of individuals with formal dance training (both in the child and adult participant groups) may have prevented us from seeing an effect. Second, familiarity with dance styles does change how an individual synchronizes their movements with the metrical structure of music (Naveda et al., 2015). Very few of our participants with formal dance training endorsed having had ballroom dance lessons, suggesting that even those with dance training in our study may have been less familiar with ballroom dance music. Future studies could therefore test individuals with experience dancing to ballroom music, or alternative musical stimuli could be used to better match the experience of the participants.

Musical training impacted beat- and measure-level perception in adults, but did not impact children as strongly as adults, and generally only affected beat-level perception. The impact of music training on children and adolescents was not as large as the impact of age, after controlling for all other factors. This may seem surprising given that some child musicians possess extraordinary and rhythmically precise musical performance skills (Ruthsatz & Detterman, 2003; Young & Nettelbeck, 1995), and also given evidence from prior work that even brief music training can improve children's rhythmic production (Slater, Tierney, & Kraus, 2013). One possibility is that our child sample did not have enough variation in music training: many children in our sample had no

music or dance training, and mean years of music and dance training ranged between 1 and 3 years even in the oldest groups. Perhaps having a larger sample of musically trained children would reveal a link between music training and sensitivity to measure-level information. However, it is not clear that improved finger tapping, the ability to perform a temporally precise rendition of a piece of music, or even sensitivity to beat necessarily predict sensitivity to multiple levels of the metrical hierarchy.

To our knowledge, no studies have explicitly and systematically examined perception or production of multiple metrical levels among young listeners with and without music training. Thus, a goal for future work is to directly examine this question by using tasks such as the present one with highly musically trained children. As the current study used a cross-sectional design, future studies could use a longitudinal design to attempt to disentangle the impact of age and musical training on the development of beat and meter perception in children, adolescents, and adults. Following younger and older adults receiving musical training for the first time could also give further insight to the effects of musical training on the perception of multiple levels of musical meter after adulthood has been reached.

In summary, we provide novel evidence that adult listeners can perceive multiple levels of metrical structure in a metronome-matching task, and music training enhances this ability, although it is not required. We also provide evidence that perceptual sensitivity to multiple levels of musical beat and meter emerges gradually over the course of childhood and adolescence. This is presumably a result of developmental changes in general timing capacities along with the gradual accumulation of culture-specific experience.

Context

The work we describe in this article was motivated by strong claims by music theorists that perception of multiple beat levels occurs while listening to most pieces of western music. Until the current study, this assumption had not been tested by many empirical studies, and none that we know of used ecologically valid musical materials or tested a large group of young listeners with a wide range of age and musical experience. This article also builds on prior work by the second and third authors (Erin E. Hannon, and Joel S. Snyder), who have published numerous studies on beat and meter perception in infants, children, and adults, using measures of perceptual judgment, finger tapping, and electrical brain activity. To rule out the influence of motor limitations in children, we chose to develop a new perceptual paradigm that could reveal sensitivity to whether a fast and a slow beat level match real musical excerpts. This task bears some similarity to a beat-matching task used by Hannon and Snyder in the first study they published together (Hannon et al., 2004) as well as the Harvard Beat Assessment Test and the Beat-Alignment Test (Fujii & Schlaug, 2013; Iversen & Patel, 2008).

References

- Audacity Team. (2020). Audacity(R): Free Audio Editor and Recorder (version 2.1.2) [Computer software]. Retrieved from <https://www.audacityteam.org/>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>
- Baruch, C., & Drake, C. (1997). Tempo discrimination in infants. *Infant Behavior & Development*, 20, 573–577. [http://dx.doi.org/10.1016/S0163-6383\(97\)90049-7](http://dx.doi.org/10.1016/S0163-6383(97)90049-7)
- Béglé, V., Benoit, C. E., Correa, A., Cutanda, D., Kotz, S. A., & Dalla Bella, S. (2017). “Lost in time” but still moving to the beat. *Neuropsychologia*, 94, 129–138. <http://dx.doi.org/10.1016/j.neuropsychologia.2016.11.022>
- Bergeson, T. R., & Trehub, S. E. (2006). Infants’ perception of rhythmic patterns. *Music Perception*, 23, 345–360. <http://dx.doi.org/10.1525/mp.2006.23.4.345>
- Bialystok, E., & Depape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 565–574. <http://dx.doi.org/10.1037/a0012735>
- Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, 6, 145–238.
- Braun Janzen, T., Thompson, W. F., & Ranvaud, R. (2014). A developmental study of the effect of music training on timed movements. *Frontiers in Human Neuroscience*, 8, 801. <http://dx.doi.org/10.3389/fnhum.2014.00801>
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The “ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14, 362–366. <http://dx.doi.org/10.1111/1467-9280.24441>
- Brown, S., & Jordania, J. (2013). Universals in the world’s musics. *Psychology of Music*, 41, 229–248. <http://dx.doi.org/10.1177/0305735611425896>
- Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research*, 82, 1195–1211.
- Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., & Toiviainen, P. (2014). Hunting for the beat in the body: On period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, 8, 903. <http://dx.doi.org/10.3389/fnhum.2014.00903>
- Cameron, D. J., Bentley, J., & Grahn, J. A. (2015). Cross-cultural influences on rhythm processing: Reproduction, discrimination, and beat tapping. *Frontiers in Psychology*, 6, 366. <http://dx.doi.org/10.3389/fpsyg.2015.00366>
- Celma-Miralles, A., de Menezes, R. F., & Toro, J. M. (2016). Look at the beat, feel the meter: Top-down effects of meter induction on auditory and visual modalities. *Frontiers in Human Neuroscience*, 10, 108. <http://dx.doi.org/10.3389/fnhum.2016.00108>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18, 2844–2854. <http://dx.doi.org/10.1093/cercor/bhn042>
- Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: Effects of music background. *Frontiers in Neuroscience*, 10, 229. <http://dx.doi.org/10.3389/fnins.2016.00229>
- Clayton, M. (2000). *Time in Indian music*. New York, NY: Oxford University Press.
- 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, 42–45. <http://dx.doi.org/10.20982/tqmp.01.1.p042>
- Davies, M. E. P., & Plumbley, M. D. (2007). Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 15, 1009–1020.
- Dawson, J. F., & Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope

- difference test. *Journal of Applied Psychology*, 91, 917–926. <http://dx.doi.org/10.1037/0021-9010.91.4.917>
- Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32, 341–365. <http://dx.doi.org/10.1068/p3370>
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, 81, 181–187. <http://dx.doi.org/10.1016/j.neubiorev.2017.02.011>
- Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, 53, 25–33. <http://dx.doi.org/10.3758/BF03211712>
- Drake, C., & Ben El Heni, J. (2003). Synchronizing with music: Intercultural differences. *Annals of the New York Academy of Sciences*, 999, 429–437. <http://dx.doi.org/10.1196/annals.1284.053>
- Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54, 277–286. <http://dx.doi.org/10.3758/BF03205262>
- Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77, 251–288. [http://dx.doi.org/10.1016/S0010-0277\(00\)00106-2](http://dx.doi.org/10.1016/S0010-0277(00)00106-2)
- Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception*, 18, 1–23. <http://dx.doi.org/10.2307/40285899>
- Eerola, T., Luck, G., & Toiviainen, P. (2006). An investigation of preschoolers' corporeal synchronization with music. *Proceedings of the 9th International Conference on Music Perception and Cognition* (pp. 472–476). Bologna, Italy.
- Einarson, K. M., & Trainor, L. J. (2015). The effect of visual information on young children's perceptual sensitivity to musical beat alignment. *Timing & Time Perception*, 3, 88–101. <http://dx.doi.org/10.1163/22134468-03002039>
- Einarson, K. M., & Trainor, L. J. (2016). Hearing the beat: Young children's perceptual sensitivity to beat alignment varies according to metric structure. *Music Perception*, 34, 56–70. <http://dx.doi.org/10.1525/mp.2016.34.1.56>
- Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37, 1–7. <http://dx.doi.org/10.3758/BF03207132>
- Fitzroy, A. B., & Sanders, L. D. (2015). Musical meter modulates the allocation of attention across time. *Journal of Cognitive Neuroscience*, 27, 2339–2351.
- Flaunagacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., . . . Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, 8, 392. <http://dx.doi.org/10.3389/fnhum.2014.00392>
- Fujii, S., & Schlaug, G. (2013). The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Frontiers in Human Neuroscience*, 7, 771.
- Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS ONE*, 9, e97680. <http://dx.doi.org/10.1371/journal.pone.0097680>
- Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *The Journal of Neuroscience*, 35, 15187–15198. <http://dx.doi.org/10.1523/JNEUROSCI.2397-15.2015>
- Fujioka, T., Zendel, B. R., & Ross, B. (2010). Endogenous neuromagnetic activity for mental hierarchy of timing. *The Journal of Neuroscience*, 30, 3458–3466. <http://dx.doi.org/10.1523/JNEUROSCI.3086-09.2010>
- Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception—Training increases hierarchical metre processing. *European Journal of Neuroscience*, 32, 1979–1985. <http://dx.doi.org/10.1111/j.1460-9568.2010.07462.x>
- Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex*, 45, 93–102. <http://dx.doi.org/10.1016/j.cortex.2007.09.010>
- Gerry, D. W., Faux, A. L., & Trainor, L. J. (2010). Effects of Kindermusik training on infants' rhythmic enculturation. *Developmental Science*, 13, 545–551. <http://dx.doi.org/10.1111/j.1467-7687.2009.00912.x>
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15, 3–10. <http://dx.doi.org/10.1016/j.tics.2010.10.001>
- Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, 49, 1363–1376. <http://dx.doi.org/10.1016/j.cortex.2012.05.005>
- Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research*, 220, 51–61. <http://dx.doi.org/10.1007/s00221-012-3114-8>
- Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, 19, 893–906. <http://dx.doi.org/10.1162/jocn.2007.19.5.893>
- Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, 45, 54–61. <http://dx.doi.org/10.1016/j.cortex.2008.01.005>
- Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, 50, 354–377. <http://dx.doi.org/10.1016/j.cogpsych.2004.09.003>
- Hannon, E. E., Nave-Blodgett, J. E., & Nave, K. M. (2018). The developmental origins of the perception and production of musical rhythm. *Child Development Perspectives*, 12, 194–198. <http://dx.doi.org/10.1111/cdep.12285>
- Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 956–974. <http://dx.doi.org/10.1037/0096-1523.30.5.956>
- Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science*, 14, 865–872. <http://dx.doi.org/10.1111/j.1467-7687.2011.01036.x>
- Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 543–548. <http://dx.doi.org/10.1037/a0027225>
- Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science*, 16, 48–55. <http://dx.doi.org/10.1111/j.0956-7976.2005.00779.x>
- Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 12639–12643. <http://dx.doi.org/10.1073/pnas.0504254102>
- Hannon, E. E., Vanden Bosch der Nederlanden, C. M., & Tichko, P. (2012). Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Annals of the New York Academy of Sciences*, 1252, 92–99. <http://dx.doi.org/10.1111/j.1749-6632.2012.06466.x>
- Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, 9, 155–183. <http://dx.doi.org/10.2307/40285527>
- Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive

- abilities across the life span. *Psychological Science*, 26, 433–443. <http://dx.doi.org/10.1177/0956797614567339>
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277. <http://dx.doi.org/10.1016/j.cognition.2018.04.007>
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam, the Netherlands: TT-Publikaties.
- Hox, J. J. (2010). *Multilevel analysis techniques and applications* (2nd ed.). New York, NY: Routledge. <http://dx.doi.org/10.4324/9780203852279>
- Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, 47, 674–689. <http://dx.doi.org/10.1016/j.cortex.2010.07.010>
- Ilari, B. (2015). Rhythmic engagement with music in early childhood: A replication and extension. *Journal of Research in Music Education*, 62, 332–343. <http://dx.doi.org/10.1177/0022429414555984>
- Iversen, J. R., & Patel, A. D. (2008). The beat alignment test (BAT): Surveying beat processing abilities in the general population. *Proceedings of the 10th International Conference on Music Perception and Cognition*, Sapporo, Japan.
- Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, 1169, 58–73. <http://dx.doi.org/10.1111/j.1749-6632.2009.04579.x>
- Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27, 359–370. <http://dx.doi.org/10.1016/j.cub.2016.12.031>
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323–355. <http://dx.doi.org/10.1037/0033-295X.83.5.323>
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459–491. <http://dx.doi.org/10.1037/0033-295X.96.3.459>
- Jones, M. R., & Pfordresher, P. Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51, 271–291. <http://dx.doi.org/10.1037/1196-1961.51.4.271>
- Jongsma, M. L., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and non-musicians. *Biological Psychology*, 66, 129–152. <http://dx.doi.org/10.1016/j.biopsycho.2003.10.002>
- Kraus, N., Strait, D., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, 1252, 100.
- Kirschner, S., & Ilari, B. (2014). Joint drumming in Brazilian and German preschool children: Cultural differences in rhythmic entrainment, but no prosocial effects. *Journal of Cross-Cultural Psychology*, 45, 137–166. <http://dx.doi.org/10.1177/0022022113493139>
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102, 299–314. <http://dx.doi.org/10.1016/j.jecp.2008.07.005>
- Ladinig, O., Honing, H., Haden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception*, 26, 377–386. <http://dx.doi.org/10.1525/mp.2009.26.4.377>
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119–159. <http://dx.doi.org/10.1037/0033-295X.106.1.119>
- Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37. http://dx.doi.org/10.1207/s15516709cog2601_1
- Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*, 1169, 46–57. <http://dx.doi.org/10.1111/j.1749-6632.2009.04550.x>
- Leow, L.-A., Parrott, T., & Grahn, J. A. (2014). Individual differences in beat perception affect gait responses to low- and high-groove music. *Frontiers in Human Neuroscience*, 8, 811. <http://dx.doi.org/10.3389/fnhum.2014.00811>
- Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, 1, 229–252. <http://dx.doi.org/10.2307/40285257>
- Lerdahl, F., & Jackendoff, R. (1985). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Lester, J. (1986). *The rhythms of tonal music*. Carbondale, IL: Southern Illinois University Press.
- London, J. (2002). Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception*, 19, 529–550. <http://dx.doi.org/10.1525/mp.2002.19.4.529>
- London, J. (2012). *Hearing in time*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199744374.001.0001>
- London, J., Polak, R., & Jacoby, N. (2017). Rhythm histograms and musical meter: A corpus study of Malian percussion music. *Psychonomic Bulletin & Review*, 24, 474–480. <http://dx.doi.org/10.3758/s13423-016-1093-7>
- Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition*, 77, 298–303. <http://dx.doi.org/10.1016/j.bandc.2011.07.007>
- Marchant, J. L., & Driver, J. (2013). Visual and audiovisual effects of isochronous timing on visual perception and brain activity. *Cerebral Cortex*, 23, 1290–1298. <http://dx.doi.org/10.1093/cercor/bhs095>
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135, 348–367. <http://dx.doi.org/10.1037/0096-3445.135.3.348>
- McCormack, T., Brown, G. D. A., Maylor, E. A., Darby, R. J., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology*, 35, 1143–1155. <http://dx.doi.org/10.1037/0012-1649.35.4.1143>
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22, 1425–1433. <http://dx.doi.org/10.1177/0956797611416999>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9, e89642. <http://dx.doi.org/10.1371/journal.pone.0089642>
- Naveda, L., & Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Perception*, 28, 93–111. <http://dx.doi.org/10.1525/mp.2010.28.1.93>
- Naveda, L., Martínez, I. C., Damesón, J., Ghiena, A. P., Herrera, R., & Ordás, M. A. (2015, June). Cross-cultural comparisons of unconstrained body responses to Argentinian and Afro-Brazilian music. In *International Symposium on Computer Music Multidisciplinary Research* (pp. 464–482). Dordrecht, the Netherlands: Springer International Publishing.
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, 31, 10234–10240. <http://dx.doi.org/10.1523/JNEUROSCI.0411-11.2011>
- Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, 32, 17572–17581. <http://dx.doi.org/10.1523/JNEUROSCI.3203-12.2012>

- Otte, R. A., Winkler, I., Braeken, M. A. K. A., Stekelenburg, J. J., van der Stelt, O., & Van den Bergh, B. R. H. (2013). Detecting violations of temporal regularities in waking and sleeping two-month-old infants. *Biological Psychology*, 92, 315–322. <http://dx.doi.org/10.1016/j.biopsycho.2012.09.009>
- Overly, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. *Dyslexia*, 9, 18–36. <http://dx.doi.org/10.1002/dys.233>
- Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergartners. *Journal of Experimental Child Psychology*, 167, 354–368. <http://dx.doi.org/10.1016/j.jecp.2017.11.009>
- Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728–741. <http://dx.doi.org/10.1037/0096-1523.16.4.728>
- Paul, B. T., Sederberg, P. B., & Feth, L. L. (2015). Imagined temporal groupings tune oscillatory neural activity for processing rhythmic sounds. *Timing & Time Perception*, 3, 172–188. <http://dx.doi.org/10.1163/22134468-03002042>
- Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3–18. <http://dx.doi.org/10.1037/0096-1523.7.1.3>
- Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411–440. <http://dx.doi.org/10.2307/40285311>
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. <http://dx.doi.org/10.3102/10769986031004437>
- Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronization in 21/2- and 4-year-old-children. *International Journal of Behavioral Development*, 27, 220–231. <http://dx.doi.org/10.1080/01650250244000290>
- Puyjarinet, F., Bégel, V., Lopez, R., Dellacherie, D., & Dalla Bella, S. (2017). Children and adults with Attention-Deficit/Hyperactivity Disorder cannot move to the beat. *Scientific Reports*, 7, 11550. <http://dx.doi.org/10.1038/s41598-017-11295-w>
- Rajendran, V. G., Harper, N. S., Garcia-Lazaro, J. A., Lesica, N. A., & Schnupp, J. W. H. (2017). Midbrain adaptation may set the stage for the perception of musical beat. *Proceedings: Biological Sciences*, 284, 20171455. <http://dx.doi.org/10.1098/rspb.2017.1455>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5, 199–213. <http://dx.doi.org/10.1037/1082-989X.5.2.199>
- Repp, B. H. (2010). Do metrical accents create illusory phenomenal accents? *Attention, Perception, & Psychophysics*, 72, 1390–1403. <http://dx.doi.org/10.3758/APP.72.5.1390>
- Repp, B. H., London, J., & Keller, P. E. (2013). Systematic distortion in musicians' reproduction of cyclic three-interval rhythms. *Music Perception*, 30, 291–305. <http://dx.doi.org/10.1525/mp.2012.30.3.291>
- Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20, 403–452. <http://dx.doi.org/10.3758/s13423-012-0371-2>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley. <http://dx.doi.org/10.1002/9780470316696>
- Ruthsatz, J., & Detterman, D. K. (2003). An extraordinary memory: The case study of a musical prodigy. *Intelligence*, 31, 509–518. [http://dx.doi.org/10.1016/S0160-2896\(03\)00050-3](http://dx.doi.org/10.1016/S0160-2896(03)00050-3)
- Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. *The Journal of Neuroscience*, 19, 10074–10081. <http://dx.doi.org/10.1523/JNEUROSCI.19-22-10074.1999>
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114. <http://dx.doi.org/10.2307/3002019>
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 8987–8992. <http://dx.doi.org/10.1073/pnas.1414495112>
- Schubotz, R. I., Friederici, A. D., & von Cramon, D. Y. (2000). Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *NeuroImage*, 11, 1–12. <http://dx.doi.org/10.1006/nimg.1999.0514>
- Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology*, 4, 622. <http://dx.doi.org/10.3389/fpsyg.2013.00622>
- Slater, J., Tierney, A., & Kraus, N. (2013). At-risk elementary school children with one year of classroom music instruction are better at keeping a beat. *PLoS ONE*, 8, e77250. <http://dx.doi.org/10.1371/journal.pone.0077250>
- Snyder, J. S., Hannon, E. E., Large, E. W., & Christiansen, M. H. (2006). Synchronization and continuation tapping to complex meters. *Music Perception*, 24, 135–146. <http://dx.doi.org/10.1525/mp.2006.24.2.135>
- Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception*, 18, 455–489. <http://dx.doi.org/10.1525/mp.2001.18.4.455>
- Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24, 117–126. <http://dx.doi.org/10.1016/j.cogbrainres.2004.12.014>
- Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: A cross-cultural comparison. *Developmental Psychology*, 46, 286–292. <http://dx.doi.org/10.1037/a0017555>
- Sowiński, J., & Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia*, 51, 1952–1963. <http://dx.doi.org/10.1016/j.neuropsychologia.2013.06.027>
- Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. (2011). *Optimal design plus empirical evidence: Documentation for the "Optimal Design" software*. Retrieved from <http://hlmssoft.net/od/od-manual-20111016-v300.pdf>
- Stark, A. M., Davies, M. P., & Plumbley, M. D. (2009). Real-time beat-synchronous analysis of musical audio. *Proceedings of 12th International Conference on Digital Audio Effects (DAFx)*. Retrieved from http://dafx09.comopolimi.it/proceedings/data/DAFx09_Proceedings.pdf
- Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions*, 7, 44. <http://dx.doi.org/10.1186/1744-9081-7-44>
- Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat synchronization across the lifespan: Intersection of development and musical experience. *PLoS ONE*, 10, e0128839. <http://dx.doi.org/10.1371/journal.pone.0128839>
- Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, 29, 334–348. <http://dx.doi.org/10.1111/j.1467-9817.2006.00312.x>
- Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology*, 102, 120–129. <http://dx.doi.org/10.1016/j.jphysparis.2008.03.007>
- Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *The Journal of Neurosci-*

- ence, 33, 14981–14988. <http://dx.doi.org/10.1523/JNEUROSCI.0612-13.2013>
- Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception*, 28, 59–70. <http://dx.doi.org/10.1525/mp.2010.28.1.59>
- Toiviainen, P., & Snyder, J. S. (2003). Tapping to Bach: Resonance-based modeling of pulse. *Music Perception*, 21, 43–80. <http://dx.doi.org/10.1525/mp.2003.21.1.43>
- Toussaint, G. T. (2013). *The geometry of musical rhythm: What makes a “good” rhythm good?* Boca Raton, FL: CRC Press.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: Effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences*, 1252, 129–138. <http://dx.doi.org/10.1111/j.1749-6632.2012.06462.x>
- Trehub, S. E., & Hannon, E. E. (2009). Conventional rhythms enhance infants’ and adults’ perception of musical patterns. *Cortex*, 45, 110–118. <http://dx.doi.org/10.1016/j.cortex.2008.05.012>
- Ullal-Gupta, S., Hannon, E. E., & Snyder, J. S. (2014). Tapping to a slow tempo in the presence of simple and complex meters reveals experience-specific biases for processing music. *PLoS ONE*, 9, e102962. <http://dx.doi.org/10.1371/journal.pone.0102962>
- van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28, 43–66. <http://dx.doi.org/10.1076/jnmr.28.1.43.3122>
- Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, 24, 560–564. <http://dx.doi.org/10.1016/j.neuroimage.2004.08.039>
- Wechsler, D., & Hsiao-pin, C. (2011). *WASI-II: Wechsler Abbreviated Scale of Intelligence*. Bloomington, MN: Pearson.
- Weiss, M. W., Cirelli, L. K., McDermott, J. H., & Trehub, S. E. (2020). Development of consonance preferences in Western listeners. *Journal of Experimental Psychology: General*, 149, 634–649. <http://dx.doi.org/10.1037/xge0000680>
- Windsor, W. L. (1993). Dynamic accents and the categorical perception of metre. *Psychology of Music*, 21, 127–140. <http://dx.doi.org/10.1177/030573569302100203>
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 2468–2471. <http://dx.doi.org/10.1073/pnas.0809035106>
- Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, 15, 179–206. <http://dx.doi.org/10.1023/A:1013880723925>
- Woodruff Carr, K. W., Tierney, A., White-Schwoch, T., & Kraus, N. (2016). Intertrial auditory neural stability supports beat synchronization in preschoolers. *Developmental Cognitive Neuroscience*, 17, 76–82. <http://dx.doi.org/10.1016/j.dcn.2015.12.003>
- Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 14559–14564. <http://dx.doi.org/10.1073/pnas.1406219111>
- Young, R. L., & Nettelbeck, T. (1995). The abilities of a musical savant and his family. *Journal of Autism and Developmental Disorders*, 25, 231–248. <http://dx.doi.org/10.1007/BF02179286>
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 5768–5773. <http://dx.doi.org/10.1073/pnas.1000121107>
- Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS ONE*, 9, e99868. <http://dx.doi.org/10.1371/journal.pone.0099868>

Received February 4, 2019

Revision received May 7, 2020

Accepted May 14, 2020 ■