Journal of Experimental Child Psychology 159 (2017) 159–174

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Babies know bad dancing when they see it: Older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays

Erin E. Hannon [a,*], Adena Schachner [b], Jessica E. Nave-Blodgett [a]

[a] Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA
[b] Department of Psychology, University of California, San Diego, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 10 July 2016
Revised 17 January 2017

Keywords:
Infant perception
Audiovisual synchrony
Intersensory perception
Music
Music development
Rhythm

ABSTRACT

Movement to music is a universal human behavior, yet little is known about how observers perceive audiovisual synchrony in complex musical displays such as a person dancing to music, particularly during infancy and childhood. In the current study, we investigated how perception of musical audiovisual synchrony develops over the first year of life. We habituated infants to a video of a person dancing to music and subsequently presented videos in which the visual track was matched (synchronous) or mismatched (asynchronous) with the audio track. In a visual-only control condition, we presented the same visual stimuli with no sound. In Experiment 1, we found that older infants (8–12 months) exhibited a novelty preference for the mismatched movie when both auditory information and visual information were available and showed no preference when only visual information was available. By contrast, younger infants (5–8 months) in Experiment 2 did not discriminate matching stimuli from mismatching stimuli. This suggests that the ability to perceive musical audiovisual synchrony may develop during the second half of the first year of infancy.

* Corresponding author.
E-mail address: erin.hannon@unlv.edu (E.E. Hannon).

http://dx.doi.org/10.1016/j.jecp.2017.01.006
0022-0965/

Introduction

The capacity to dance to music is a human universal. In all cultures, people move spontaneously to music, entraining their body movements to the timing of a musical pulse or beat (Large, 2000; Nettl, 1983; Repp, 2005; Snyder & Krumhansl, 2001). Although commonplace, this capacity is not trivial but rather depends on complex and multisensory cognitive processes that develop with age and experience. Dancing to music requires listeners to actively infer a beat from a rich and dynamic musical stimulus, modulate attention toward regularly occurring time points within auditory, visual, and tactile sensory input, form expectations about future events that guide self-generated movements, and continuously monitor these movements for error (Jones & Boltz, 1989; Large & Jones, 1999; Repp & Su, 2013). The current study took an initial step toward understanding the development of these multisensory musical capacities by investigating whether or not young infants can tell when a seen dancer is in or out of synchrony with the beat of heard music.

The term musical "beat" refers to a regularly occurring salient moment in time, often equally spaced or quasi-isochronous, when human listeners are most likely to tap their fingers or feet during music listening (Honing, Bouwer, & Háden, 2014; Lerdahl & Jackendoff, 1983). The musical beat can be considered just one (the most salient) level of the musical "meter," which is made up of multiple hierarchically nested faster and slower levels of pulsation, typically related to the beat by integer ratios (Lerdahl & Jackendoff, 1983; London, 2002). For example, a waltz and a tango might both have 100 beats per minute, but a waltz has a higher level slower pulse every three beats, whereas a tango has higher level pulses every two and four beats. The beat is often highlighted by acoustic features such as louder or longer notes, but a beat can also be perceived in the absence of loudness or duration changes (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Iversen, Repp, & Patel, 2009) and even when there is no acoustic event (Longuet-Higgins & Lee, 1984; Snyder & Krumhansl, 2001). Thus, beat perception is to a surprising extent a top-down, subjective, and listener-driven process; we infer and predict the location of musical beats, and the percept of a beat is not solely the result of bottom-up perceptual input (Honing et al., 2014; Trainor & Hannon, 2013).

Beat perception and production play a central role in human musicality. Beat perception and production are relatively rare among other species even after extensive training (Cook, Rouse, Wilson, & Reichmuth, 2013; Hattori, Tomonaga, & Matsuzawa, 2013; Honing, Merchant, Háden, Prado, & Bartolo, 2012; Schachner, Brady, Pepperberg, & Hauser, 2009; Zarco, Merchant, Prado, & Mendez, 2009). By contrast, this ability emerges spontaneously in most humans during early childhood without explicit musical training. Within days of birth, infants show sensitivity to the beat in simple musical stimuli; while listening to drum patterns, newborns exhibit larger event-related potential (ERP) responses (mismatch negativity) when events are omitted on strong versus weak beat positions (Winkler, Háden, Ladinig, Sziller, & Honing, 2009), and violations of temporal intervals or tempo in metronome-like rhythmic stimuli give rise to behavioral and neural detection responses in 2-month-olds (Baruch & Drake, 1997; Otte et al., 2013). By the middle of the first year, infants categorize rhythms by their underlying beat; for example, 7-month-olds habituated to a set of varied rhythmic sequences that all shared the same underlying beat subsequently exhibited larger dishabituation responses (indicating perception of greater novelty) to rhythms that violated the familiar beat versus those that maintained the beat they heard during the prior habituation phase (Hannon & Johnson, 2005).

Infants also begin to integrate their own body movements with their auditory perception sometime during the first year. For example, when bounced on every second or third beat of an ambiguous rhythm, 7-month-olds later prefer listening to a version of the rhythm containing loudness accents that match the prior bouncing pattern, suggesting that they encode the temporal position of the bounces as reflecting the beat (Phillips-Silver & Trainor, 2005). Some evidence suggests that 3- to 5-month-olds engage in repetitive rhythmic body movements, such as kicking and arm waving, more often in the presence of music than during silence or other non-musical stimuli, such as speech (Fujii et al., 2014; Ilari, 2015; Zentner & Eerola, 2010); however, these movements are relatively infrequent (8% of the trial at most, observed in only some infants).

As in other domains such as language, production appears to lag behind perception during development of beat-based musical behavior. Even though infants can perceive the beat and sometimes move in response to music, they do not move in precise synchrony with music (Ilari, 2015; Zentner & Eerola, 2010). It is not until later childhood that children are capable of consistently moving in precise synchrony with a musical beat (Kirschner & Tomasello, 2009; McAuley, Jones, Holub, Johnston, & Miller, 2006; Provasi & Bobin-Bègue, 2003). Between the ages of 2 and 8 years, children become better at adjusting their regular rhythmic movements to the tempo of an external stimulus (McAuley et al., 2006; Provasi & Bobin-Bègue, 2003) or another drummer (Kirschner & Ilari, 2014), but even 10-year-olds are more variable than adults when tapping to simple rhythms and to music (Drake, Jones, & Baruch, 2000). Thus, the capacity for adult-like entrainment to music does not appear to be fully developed until later childhood.

The later arrival of precise synchronization with music suggests that despite infants' early perceptual sensitivity to the beat in some tasks, many aspects of beat processing develop gradually. Much evidence suggests that this slow development includes acquisition of abstract, higher-level aspects of musical meter and beat and their multisensory correlates. For example, representations of musical meter and beat undergo substantial developmental changes during early childhood; young infants are initially able to discriminate rhythmic structures from any culture (Hannon & Trehub, 2005b), but between 5 and 12 months they begin to preferentially listen to and better discriminate metrical structures from their own culture (Hannon, Soley, & Levine, 2011; Hannon & Trehub, 2005a; Soley & Hannon, 2010). Listening experience during infancy appears to increase perceptual biases for the more common meters within a given culture, leading to enhanced encoding of duple meters (groups of two or four beats) compared with less common triple meters (groups of three beats) in North America (Bergeson & Trehub, 2006; Gerry, Faux, & Trainor, 2010). The malleability of culture-specific metrical representations may extend into later childhood because exposure to foreign music can reverse or reduce own-culture metrical biases among children up to age 7 years but not among 9-year-olds and adults (Hannon & Trehub, 2005a; Hannon, Bosch, der Nederlanden, & Tichko, 2012). Although only a handful of studies provide data on the development of beat perception and production during childhood (e.g., Drake et al., 2000; Hannon et al., 2012; McAuley et al., 2006; Provasi & Bobin-Bègue, 2003), findings generally support the notion that nuanced aspects of beat perception develop slowly as listeners become more familiar with the music around them and form stable cognitive representations of musical meter.

The development of audiovisual synchrony perception

Our earliest encounters with music include coordinated auditory, visual, tactile, and sensorimotor input; as infants, we are often rocked and bounced during face-to-face caregiver singing and vocal play (Phillips-Silver, 2009; Trainor & Hannon, 2013). Nevertheless, relatively little is known about the development of intermodal music perception. When presented with simple non-musical displays of audible and visible events, infants between 1 and 10 months expect to hear a sound when they see an impact between a bouncing ball, toy, hammer, or object and a surface or between two clapping hands (Bahrick, 1983, 1987, 1988, 2001; Kopp, 2014; Kopp & Dietrich, 2013; Lewkowicz, 1992a, 1992b, 1994, 1996; Spelke, 1979). In these studies of simple non-musical events, infants behaviorally and neurally detect disruption of audiovisual synchrony relations, and they prefer synchronous over asynchronous audiovisual displays. Low-level audiovisual synchrony may also boost young infants' sensitivity to the patterning of multisensory input. For example, 4-month-olds can classify audiovisual patterns on the basis of rhythm, tempo, or emotional content, but they cannot perform the same classification when stimuli are presented unimodally (e.g., using auditory or visual information alone) until they are a few months older (Bahrick & Lickliter, 2000; Flom & Bahrick, 2007; Pickens & Bahrick, 1995, 1997; Walker-Andrews, 1986).

Perception of audiovisual synchrony nevertheless changes with age, particularly for complex multimodal structures such as language. This suggests that experience and acquired knowledge may be prerequisites for perceiving some types of audiovisual synchrony. For example, infants readily detect synchrony between the audio and visual components of isolated spoken syllables (Lewkowicz, 2003, 2010), but they have greater difficulty when syllables form a rhythmic pattern (Lewkowicz, 2003) and only inconsistently respond to audiovisual asynchrony in fluent speech, having particular trouble with foreign speech or non-infant-directed speech (Dodd, 1979; Kubicek et al., 2014; Lewkowicz & Pons, 2013).

Evidence from studies of language suggest that audiovisual perception may undergo a progression wherein young infants initially rely on low-level perceptual aspects of audiovisual synchrony but increasingly learn about and rely on higher-level structures such as language identity (including phonetic and prosodic structure), person identity, emotion, and gender (Lewkowicz, 2014). Whereas very young infants (under 5 months) can use phonetic information to match vowels with facial movements (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003) and know that human faces go with human voices (and not with monkey vocalizations or tones) (Kuhl, Williams, & Meltzoff, 1991; Spelke, 1976; Vouloumanos, Druhen, Hauser, & Huizink, 2009), mounting evidence suggests that experience-driven language- and species-specific knowledge influences how infants match visual and auditory components of vocal stimuli (Lewkowicz, 2014). For example, younger infants perform audiovisual matching equally well whether vocalizations are human or non-human and whether syllables are from a native or non-native language. However, by 8 to 10 months, infants show an advantage for audiovisual matching of species- and language-specific stimuli (Grossmann, Missana, Friederici, & Ghazanfar, 2012; Lewkowicz & Ghazanfar, 2006; Lewkowicz, Leo, & Simion, 2010; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). Similarly, adults are more likely to detect asynchrony in audiovisual speech when facial movements are accompanied by natural speech than by sine wave speech (a non-speech analogue), whereas infants show no such advantage for natural speech (Baart, Vroomen, Shaw, & Bortfeld, 2014).

Recent work has emphasized the notion that similar trajectories may characterize perceptual development of both language and music abilities during infancy and early childhood, particularly those related to rhythm and timing (Brandt, Gebrian, & Slevc, 2012; Hannon, Leveque, Nave, & Trehub, 2016). Just as audiovisual perception of speech may depend on gradual acquisition of higher-level knowledge of language, so too might accurate perception of audiovisual synchrony between dance and music depend on acquired musical knowledge. Music is comparable to speech in both its complexity and its reliance on abstract, culture-specific knowledge that is acquired gradually over the course of development. In addition, although speech is often considered the most ecologically relevant instance of complex audiovisual synchrony perception, perceiving the audiovisual synchrony of a dancer moving to music is equally ecologically valid; this experience is a similarly universal and ancient part of human culture (Large, 2000; Nettl, 1983; Repp, 2005; Snyder & Krumhansl, 2001).

Perceiving audiovisual synchrony of a dancer moving to music may also pose unique challenges distinct from those involved in speech perception. Perceiving dance involves coordinating complex, dynamic inferred structures from both auditory and visual input. The perceiver must infer the beat within a rich and dynamic musical surface while also determining whether the movement/gesture patterns are consistent with that inferred beat. Because there are many potential points of synchrony between heard auditory events and seen movements, this determination is itself complex. For example, some metrical levels are reflected in movement of extremities, others are expressed by the trunk, and still others are expressed by movements toward the boundaries versus center of the personal space (Burger, Saarikallio, Luck, Thompson, & Toiviainen, 2013; Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2014; Naveda & Leman, 2010; Su, 2016; Toiviainen, Luck, & Thompson, 2010). Thus, perceiving audiovisual synchrony in dance displays may depend on musical knowledge and expectations that are acquired during development.

The current study

Previous work suggests that infants can detect asynchronies in the simplest audiovisual manifestations of musical beat such as a drummer hitting a drum (Bahrick & Lickliter, 2000; Gerson, Schiavio, Timmers, & Hunnius, 2015; Kopp, 2014). However, to date no studies have investigated infants' perception of audiovisual synchrony in complex musical displays even though music is an inherently rich and multimodal component of infants' experiences. As an ecologically valid, complex, and non-speech audiovisual stimulus, music has unique potential to shed light on general aspects of intersensory development. In addition, because musical beat perception is essential for perceiving audiovisual synchrony of dance, a universal human behavior, infants' perception of audiovisual synchrony in music can inform basic questions about the developmental origins of beat-based musical behaviors.

To examine infants' perception of audiovisual synchrony in musical displays, we employed an infant-controlled habituation paradigm, which is used widely in research on infants' intermodal perception (Bahrick & Lickliter, 2000; Bremner et al., 2011; Flom & Bahrick, 2007; Lewkowicz, 1992b, 1996) and may be more sensitive than other infant methods such as paired preference (Lewkowicz, 1992a). In two experiments, we habituated infants to a movie of a person dancing in synchrony with one of two songs. After habituation, we presented infants with two novel movies whose audio tracks were identical and taken from the previously heard song but whose video tracks (of the same dancer) either matched or mismatched the audio track. To determine whether infants could discriminate movies solely on the basis of visual information, we also had a visual-only condition in which infants were presented only with the visual component of movies. If infants are able to extract the beat from the music and match it with regular movements of a dancer, they should look longer when the visual and auditory components of the movies are asynchronous. Moreover, this effect should occur only when both auditory and visual components are available. To determine whether or not musical audiovisual synchrony perception changes over the course of infancy, we examined performance among older infants (Experiment 1) and younger infants (Experiment 2).

Experiment 1

Experiment 1 investigated audiovisual synchrony perception of rich musical stimuli among infants between 8 and 12 months. We made the a priori decision to first target this age range because previous research suggests that during the latter part of the first year after birth, infants begin to exhibit more specialized perceptual abilities for both musical rhythm (Hannon & Trehub, 2005a; Hannon et al., 2011; Soley & Hannon, 2010) and audiovisual synchrony in vocal stimuli (Grossmann et al., 2012; Lewkowicz & Ghazanfar, 2006; Lewkowicz et al., 2010). We reasoned that if infants are capable of perceiving audiovisual synchrony in musical stimuli, we would be most likely to observe this ability during the latter part of the first year.

Method

Participants

A total of 32 infants between 8 and 12 months (M_{age} = 10 months 25 days, range = 8 months 1 day to 11 months 28 days) were recruited from the communities of Cambridge, Massachusetts (n = 20) and Las Vegas, Nevada (n = 12) in the northeastern and western United States, respectively, using birth lists from hospitals and marketing companies. Infants were born full-term and had no known hearing difficulties at the time of testing. An additional 3 infants were excluded from the final sample due to fussing preventing completion of the experiment. Half of the infants were randomly assigned to the audiovisual condition and half to the visual-only condition.

Stimuli

Stimuli consisted of four videos of a smiling adult woman dancing to two prerecorded songs. The dancer was instructed to dance in an exaggerated manner to reflect the musical meter. At the level of the beat she used broad repetitive body movements and gestures (e.g., arm waves), and at the measure level (a higher metrical level) she moved her full body with additional emphasis. In separate takes, the dancer performed two unique dance patterns to each song, with no movements present in both sequences. The dancer wore dark clothing and stood in front of a white backdrop. The videos were recorded using a Sony camcorder (DCR-HC32) in well-lit conditions. Two audiovisual excerpts (denoted A and B) were selected from each song, each one 60 s in length (using Apple iMovie HD 6). The audio tracks of the two excerpts differed minimally, with only subtle acoustic differences in tempo and dynamics across the A and B audio excerpts for each song. The dancer used slightly different gestures across the A and B excerpts of each song, minimizing the likelihood that specific dance movements would cue participants to the matching stimulus during the test phase.

To minimize the influence of a familiar language/lyrics, we selected songs that would not be highly familiar to participants: two primarily instrumental songs from a CD collection of children's world music having musical meters familiar to Western listeners (*Hendry* by Tarika Sammy, tempo of 136 beats per minute, 4/4 meter, vocals in Malagasy, and *Five* by Bobby McFerrin, average tempo of 130 beats per minute, 3/4 meter, vocals as nonsense syllables and vocal patterns). Post-experiment parental report confirmed that songs were novel to all infants.

For the visual-only condition, audio tracks were removed from the four movies. For the audiovisual condition, a clean recording of each song excerpt was added back to the audio track of the movie to create matching and mismatching audiovisual versions of each of the four excerpts. For the matching stimuli, the clean recording of the song precisely matched the original musical excerpt. Thus, repetitive dance movements, such as waving arms and kicks lined up with the beats heard in the music, and larger body movements, such as shifting body orientation and changing of movement from arms to legs or vice versa, emphasized the measure-level beat in the song. For the four mismatching stimuli, the video track of one movie was paired with the audio excerpt from the other song. For example, one mismatching stimulus contained the video track of the dancer moving to the song *Hendry* during Excerpt A, but the audio track was from Excerpt A of *Five*. Thus, due to the different metrical grouping of the beats (three or four per measure/group) and slightly different rates of speed (130 vs. 136 beats per minute), the dancer's movements were out of synchrony with the beats in the music (see Videos 1–4 in the online supplementary material).

Apparatus and procedure

Each infant sat on a caregiver's lap in a dim, sound-attenuated room. The experiment was conducted in a soundproof booth (Industrial Acoustics) in Cambridge and in a similar sound-treated room in Las Vegas. The participant saw the visual component of each movie on a centrally placed 43.2-cm (17-inch) color monitor (Acer AL715) located 173 cm from the infant. Sound in the experiment came from two hidden speakers (Genelec 8020A).

An experimenter monitored infant behavior over closed-circuit television (TruTech DW27TT) with an infrared digital video camera (Sony DCR-HC32) located above the video monitor, focused on the infant. The experimenter was blind to condition (audiovisual or visual-only) and observed the infant while wearing headphones playing music or noise to mask any available sounds and while sitting outside the booth (Cambridge) or behind a curtain (Las Vegas). The experiment was run on a PowerMac Dual 2-GHz PowerPC G5 computer using Habit X software (Cohen, Atkinson, & Chaput, 2000) to control the presentation of stimuli and to code infant visual fixations during the experiment. A second experimenter, also blind to condition, later performed offline frame-by-frame coding of infant looking time using Supercoder Version 1.5 (Holich, 2005), with a third coder performing additional offline coding of a subset of the data to check for reliability. To "blind" infants' caregivers to condition, caregivers holding the infants listened to classical music presented over noise-cancelling headphones (Sony MDR-NC6) for the duration of the experimental session. We instructed caregivers to face the monitor with a neutral expression and to not interact with or cue their infants at any time during the experiment.

The experiment used an infant-controlled habituation paradigm. Prior to the beginning of each trial, the monitor flashed red to orient the infant's attention toward the monitor and loudspeaker. When the infant's gaze was directed to the video screen, the experimenter began the trial. The experimenter monitored the infant's looking behavior in real time, pressing a button to record the onset and offset of looks toward the stimulus. The trial automatically ended when the infant looked away from the monitor for more than 2 s or more. The monitor would then again flash red to recapture the infant's attention and begin the subsequent trial.

During the habituation phase of the audiovisual condition, we repeatedly presented the infant with one synchronous 60-s movie of the dancer moving to Excerpt A or B of either *Hendry* or *Five*. During the habituation phase of the visual-only condition, the same movies were presented but with no sound. On each habituation trial, the movie played for 60 s or until the infant looked away for more than 2 s, whichever occurred first. The same movie was played on every trial until the infant reached the criterion for habituation (average fixation decrement of 50% over 3 trials relative to average fixation of the first 3 trials using a sliding window) or until 12 trials had elapsed.

During the test phase of the audiovisual condition, 3 matching trials and 3 mismatching trials were presented in alternation for up to 60 s per trial. As described above, the audio track for both matching and mismatching trials contained a new excerpt from the same song heard during habituation (e.g., if the infant heard Excerpt A of *Hendry* during habituation, the infant heard Excerpt B of *Hendry* during test). For matching trials the video track matched the original video-recording, so the dancer was in synchrony with the music, whereas for mismatching trials infants saw a video track from a different song, so the dancer was out of synchrony with the music (Fig. 1). Thus, both stimuli used during the test phase had identical audio tracks but contrasting video tracks, one synchronous and the other asynchronous with the audio.

In the visual-only condition, infants saw the same movies in the same alternating trial orders, but they saw video tracks only without auditory accompaniment. Because no audio was present, there was no possibility of synchrony or asynchrony and, thus, no "matching" or "mismatching" between audio and visual. In principle, silent videos could contain dynamic tempo cues or movement patterns that might allow them to be matched across Excerpts A and B, so the same match–mismatch terminology was also applied to trials in the visual-only condition. Thus, on "matching" trials infants in the visual-only condition saw a novel video excerpt of the dancer moving to the same song she had been moving to during habituation (although the music was never present), and on "mismatching" trials infants saw a novel video clip of the dancer moving to a different song (although again the music was not present).

Equal numbers of infants were tested in the audiovisual ($n = 16$) and visual-only ($n = 16$) conditions. The order in which movies were presented (A or B during habituation) and the order of test stimuli (matching or mismatching first) were counterbalanced between participants. Thus, each video was used for both matching and mismatching trial types. Infants were randomly assigned to excerpts from each song. Slightly more infants were presented with *Hendry* ($n = 18$) than with *Five* ($n = 14$). To ensure that this asymmetry did not affect results, song was included as a factor in the analysis below.

Results and discussion

Inter-rater reliability for frame-by-frame video coding was high ($r = .99$). Post-habituation looking times (in seconds) were averaged across the three matching and mismatching trials and were submitted to a 2 × 2 × 2 mixed-design analysis of variance (ANOVA) with a within-participants factor of trial type (matching or mismatching) and between-participants factors of condition (audiovisual or visual-only) and song (*Hendry* or *Five*). This analysis revealed a main effect of trial type, $F(1, 28) = 7.74$, $p = .01$, $\eta_p^2 = .22$. Overall, infants looked longer during mismatching trials ($M = 20.4$, $SEM = 2.7$) than during matching trials ($M = 16.3$, $SEM = 2.2$). Although there was a trend toward longer overall looking in the audiovisual condition ($M = 22.2$, $SEM = 3.9$) than in the visual-only condition ($M = 14.5$, $SEM = 2.8$), this main effect of condition was not significant, $F(1, 28) = 2.41$, $p = .13$. As predicted, there was an interaction between trial type and condition, $F(1, 28) = 6.14$, $p = .02$, $\eta_p^2 = .18$. As illustrated in Fig. 2, after habituation to audiovisual synchrony, infants in the audiovisual condition looked significantly longer during mismatching trials ($M = 26.02$, $SEM = 4.01$) than during matching trials ($M = 18.4$, $SEM = 3.8$), $t(15) = 3.33$, $p = .005$. By contrast, infants in the visual-only condition looked for comparable durations during matching trials ($M = 14.2$, $SEM = 2.4$) and mismatching trials ($M = 14.8$, $SEM = 3.2$), $t(15) = 0.34$, $p = .74$. There were no other significant main effects or interactions.

Of the 32 infants tested, 30 reached the criterion for habituation (average number of trials to habituation = 7.3, $SD = 2.8$), with 2 infants failing to reach the criterion (1 in the audiovisual condition and 1 in the visual-only condition). When the same analyses as above were repeated excluding these 2 infants, the same results were obtained.

Infants discriminated matching from mismatching movies only when the audio track was available, providing evidence that audiovisual synchrony between the inferred musical beat and the dance movements was critical to successful discrimination of test stimuli. The results of Experiment 1, thus,

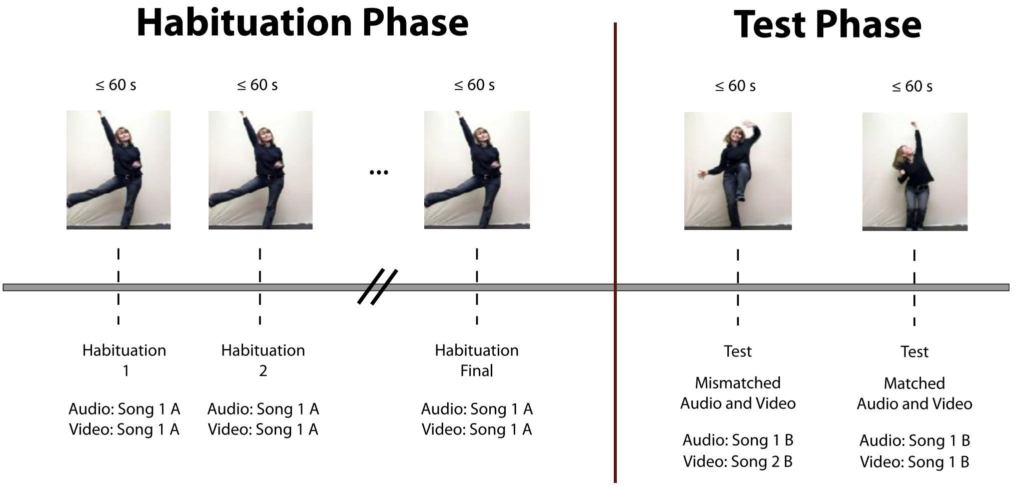

Fig. 1. Trial structure of audiovisual condition. A flashing red screen attracted the infant's gaze to the screen to begin each trial, and trials ended when the infant looked away for 2 s or more. During the habituation phase, an audiovisual movie of a person dancing in synchrony with music was repeatedly presented until looking time (in seconds) declined to the criterion. During the test phase, each of two videos was played in alternation. On matching test trials, infants viewed a novel video excerpt of the same person dancing in synchrony to the same song. On mismatching test trials, infants heard the same audio track as on matching trials but viewed a mismatched video track in which the person danced at a tempo and meter that was not synchronized with the music. We compared infants' looking time with the mismatched versus matched test trials.

suggest that by 8–12 months infants are able to extract the beat from heard music and match it with regular movements of a seen dancer.

Experiment 2

Experiment 2 examined whether the audiovisual synchrony perception observed among older infants would also be observed among younger infants. If basic sensitivity to the beat is sufficient for perceiving audiovisual synchrony in complex music displays, young infants should also succeed on the current task given previous evidence of beat sensitivity even among newborns (Winkler et al., 2009). However, if higher-level knowledge plays a role in audiovisual synchrony perception, younger infants should have difficulty because they are less likely to have acquired nuanced and culture-specific knowledge of musical beat and meter.

Method

Participants

A total of 32 infants participated in Experiment 2 (M_{age} = 6 months 10 days, range = 4 months 25 days to 7 months 29 days). Infants were recruited from the Las Vegas community using family marketing lists. All infants were born full-term and had no known hearing difficulties at the time of testing. An additional 3 infants were excluded from the final sample due to technical error ($n = 1$), premature birth ($n = 1$), or fussiness preventing completion of the experiment ($n = 1$). As in Experiment 1, half of the infants were randomly assigned to the audiovisual condition and half to the visual-only condition.

Stimuli, apparatus, and procedure

We used the same stimuli as in Experiment 1: four audiovisual matching, four audiovisual mismatching, and four visual-only dance stimuli. Infants were tested while sitting on their caregivers' laps in a sound-treated room. All aspects of the apparatus and procedure were identical to those in Exper-

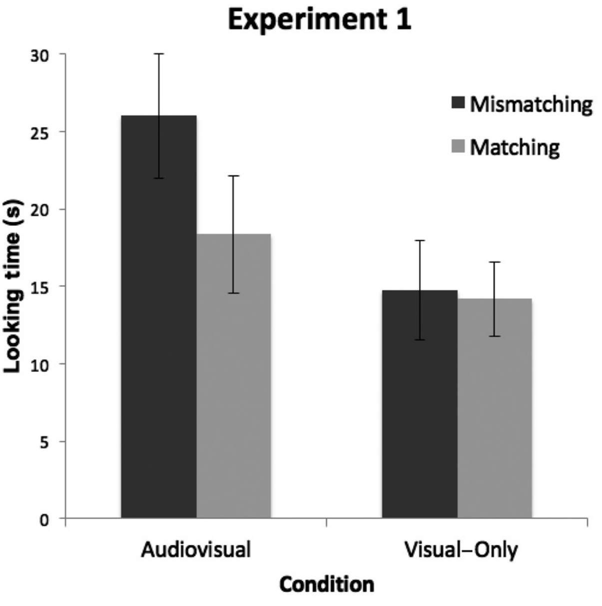

Fig. 2. Looking time (in seconds) of 8- to 11-month-olds in Experiment 1. Error bars represent standard errors.

iment 1. Equal numbers of infants heard each of the two songs during habituation (*Hendry:* $n = 16$; *Five:* $n = 16$).

Results and discussion

Inter-rater reliability for frame-by-frame video coding was high ($r = .99$). Post-habituation looking times (in seconds) were averaged across the three matching and mismatching trials and submitted to a $2 \times 2 \times 2$ mixed-design ANOVA with a within-participants factor of trial type (matching or mismatching) and between-participants factors of condition (audiovisual or visual-only) and song (*Hendry* or *Five*). This analysis revealed a main effect of condition, $F(1, 28) = 4.89$, $p = .035$, $\eta_p^2 = .15$, but no other significant main effects or interactions. Infants looked longer during the audiovisual condition ($M = 15.83$, $SEM = 2.70$) than during the visual-only condition ($M = 9.75$, $SEM = 1.52$) (Fig. 3), a finding consistent with previous infant research (Bahrick & Lickliter, 2000). However, unlike older infants, younger infants did not discriminate matching from mismatching stimuli in either condition.

Of the 32 infants tested, 23 reached the criterion for habituation (average number of trials to habituation = 7.5, $SD = 2.4$). Of the 9 infants who did not reach the criterion, only 1 was in the audiovisual condition. The same analyses as above were repeated without these infants, and the same results were found, again suggesting that younger infants did not discriminate matching from mismatching stimuli even in the audiovisual condition.

The current experiments also aimed to address the question of whether or not musical audiovisual synchrony develops during infancy. If it does, we would expect to see stronger novelty preferences in older infants than in younger infants. To examine novelty preference, we converted raw looking times from Experiments 1 and 2 to proportion of total looking time (PTLT) toward the mismatched stimulus by dividing the average duration of looking during the mismatching trials by the combined average looking during all test trials. Thus, scores higher than .50 indicate a preference for the novel mismatching stimulus. Because PTLT expresses novelty preferences as a proportion of total looking time, it removes other sources of variability such as age-driven changes in overall looking behavior, allowing us to directly compare novelty preference among infants of different ages.

A larger number of infants failed to habituate in Experiment 2 ($n = 9$) than in Experiment 1 ($n = 2$). When comparing the two age groups, this difference in prevalence of habituation could pose a problem because dishabituation (i.e., the novelty preference) should be more systematic for infants who have fully habituated (Wetherford & Cohen, 1973). In accord with this idea, PTLT scores for infants who met the habituation criterion were higher ($M = .54$, $SEM = .02$) than for infants who did not meet the criterion ($M = .43$, $SEM = .03$), $t(62) = 2.61$, $p = .01$. Therefore, we included only those infants who successfully habituated ($n = 15$ younger and $n = 15$ older in the audiovisual condition; $n = 8$ younger and $n = 15$ older in the visual-only condition).

We submitted PTLT scores of habituated infants to a 2 × 2 univariate ANOVA with between-participants factors of condition (audiovisual or visual-only) and age group (younger or older). This analysis revealed an interaction between condition and age group, $F(1,\ 49) = 4.56$, $p = .038$, $\eta_p^2 = .085$, and no other significant effects. As shown in Fig. 4, older infants' PTLT scores were significantly higher in the audiovisual condition ($M = .61$, $SEM = .03$) than in the visual-only condition ($M = .47$, $SEM = .03$), $t(28) = 3.10$, $p < .01$, whereas younger infants' PTLT scores did not differ between the audiovisual condition ($M = .53$, $SEM = .03$) and the visual-only condition ($M = .56$, $SEM = .07$), $t(21) = 0.42$, $p = .68$. One-tailed Bonferroni-corrected t tests confirmed that only older infants' PTLT scores in the audiovisual condition were significantly different from chance (.50), $t(14) = 3.75$, $p < .01$. Thus, older infants, but not younger infants, showed a novelty preference for the asynchronous audiovisual test stimulus. This result is unlikely to be an artifact of diminished attention or engagement on the part of younger infants during habituation because the same number of infants successfully habituated in both age groups ($n = 15$) and looking times during the habituation phase (averaged over the final 6 trials) were comparable for younger infants ($M = 26.63$, $SEM = 2.43$) and older infants ($M = 25.20$, $SEM = 2.01$), $t(30) = 0.455$, $p = .65$. This suggests that the differences observed across the two age groups reflect a true developmental change in sensitivity to musical audiovisual synchrony.

Across both experiments, infants failed to differentiate silent test videos even when we considered only those infants who successfully habituated. This indicates that successful discrimination depended on the presence of both auditory and visual information. Alternatively, it is possible that infants—regardless of age—were simply more attentive in the audiovisual condition than in the

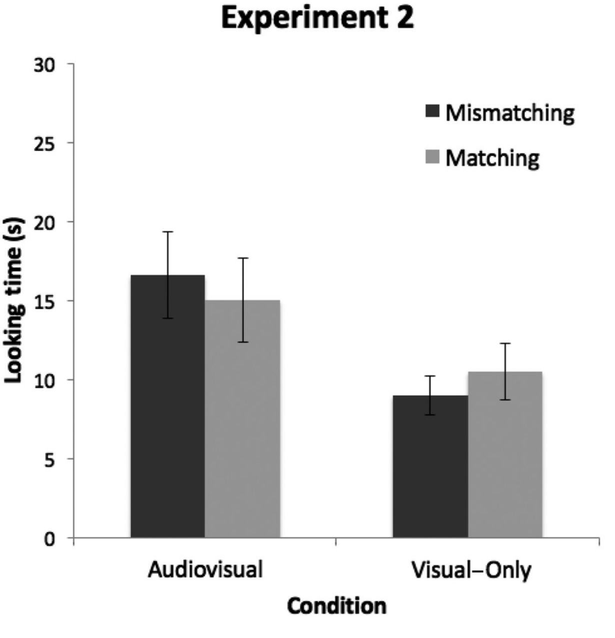

Fig. 3. Looking time (in seconds) of all 6-month-olds in Experiment 2. Error bars represent standard errors.

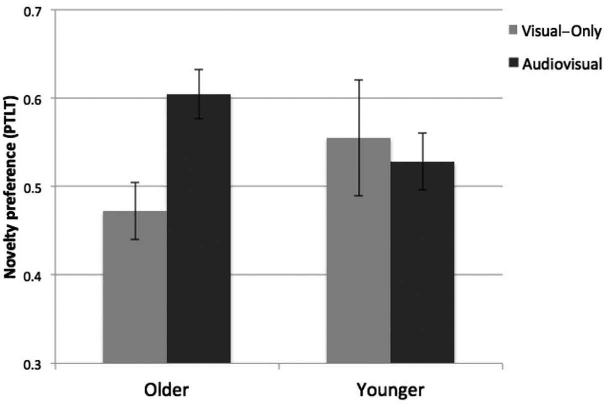

Fig. 4. Percentage total looking time (PTLT) toward the mismatching stimulus versus the matching stimulus for infants who successfully habituated in Experiments 1 and 2. Error bars represent standard errors.

visual-only condition because of the arousing properties of the music, and their inattention in the visual-only condition led to their failure. By this account, if infants paid closer attention to the stimuli in the visual-only condition, they would have discriminated those test stimuli just as did older infants in the audiovisual condition. To test this alternative account, we collected additional data from adults, reasoning that adults should have greater ability to control their attention during the task than infants. If even adults have difficulty in discriminating the visual-only stimuli, this would strongly suggest that infants' failure was not solely the result of inattention. We presented adults ($N = 24$) with the same habituation and test videos that were shown to infants, and we asked them to provide similarity ratings for each of two test videos compared with one prior habituation video.[1] We found that even adults had difficulty in discriminating the matching video from the mismatching video in the visual-only condition. Adults robustly differentiated matching from mismatching audiovisual test stimuli, giving higher similarity ratings to the matching test video ($M = 4.04$, $SEM = 0.26$) than to the mismatching test video ($M = 3.16$, $SEM = 0.22$), $t(23) = 2.64$, $p = .015$, but they did not differentiate matching from mismatching stimuli in the visual-only condition (matching $M = 3.36$, $SEM = 0.21$; mismatching $M = 3.13$, $SEM = 0.18$), $t(23) = 1.13$, $p = .27$, Bonferroni-corrected t tests. A repeated-measures ANOVA showed a significant interaction between trial type (matching or mismatching) and condition (audiovisual or visual-only), $F(1, 23) = 5.81$, $p = .024$, $\eta_p^2 = .20$, as well as main effects of trial type, $F(1, 23) = 5.27$, $p = .03$, $\eta_p^2 = .19$, and condition, $F(1, 23) = 8.28$, $p = .008$, $\eta_p^2 = .27$. Adults' difficulty in discriminating stimuli in the visual-only condition strongly suggests that infants' similar failure in the visual-only condition was not simply due to inattention. Instead, this failure appears to reflect a true inability to discriminate the stimuli when audiovisual synchrony/asynchrony is not present.

General discussion

Across two experiments, older infants (8–12 months), but not younger infants (5–8 months), were able to perceive audiovisual synchrony of a person dancing to music. When both visual information and auditory information (and, thus, information about audiovisual synchrony) were available, only older infants showed a post-habituation preference for a mismatched (asynchronous) video relative to a matched (synchronous) video. By contrast, when the same visual stimuli were presented without sound, neither younger nor older infants showed a preference for either test stimulus. This failure was not simply a result of lesser attention; even adults performed poorly at discriminating the same visual-only videos. Thus, these data provide evidence that older infants, but not younger infants, were

[1] Ratings were on a scale of 1 to 7. Each adult participated in both conditions (audiovisual and visual-only), counterbalanced for condition order and test video order.

sensitive to the audiovisual synchrony of a dancer moving to music, suggesting that perception of complex musical audiovisual synchrony develops over the course of the first year of infancy.

To our knowledge, this article presents the first evidence to date that infants can perceive audiovisual synchrony in a complex musical display such as a dancer moving to music. Music is comparable to speech in its richness and ecological validity, and both music and language are acquired in parallel during early childhood. Yet, despite a rich literature on development of audiovisual speech perception, relatively little work has explored the development of audiovisual music perception. Whereas previous work has shown that infants perceive audiovisual synchrony in very simple displays where individual temporally distinct sounds correspond precisely to individual visual trajectory reversals or impact events (Bahrick & Lickliter, 2000;Gerson et al., 2015; Kopp, 2014), the current experiments show that by the end of the first year infants succeed in a much more complex intermodal synchrony perception task. To succeed in this task, the participant must infer a musical beat from sound information and match it to seen movement patterns, parsing and comparing many potential points of synchrony between the complex movements of a dancer and the multilayered event structure in real music (Burger et al., 2013, 2014; Naveda & Leman, 2010; Su, 2016; Toiviainen et al., 2010). Therefore, this article takes an important first step by revealing that the capacity to perceive this type of complex, real-world, musical audiovisual synchrony emerges during infancy.

Interestingly, this is the same developmental time frame during which language experience begins to influence audiovisual synchrony perception of speech (Grossmann et al., 2012; Lewkowicz & Ghazanfar, 2006; Lewkowicz et al., 2010). In addition, culture-specific biases in rhythm and beat processing begin to emerge during this developmental window (Hannon & Trehub, 2005a; Hannon et al., 2011; Soley & Hannon, 2010). Thus, one account of the current data is that the developmental changes seen in the current experiments are the result of acquired knowledge about music, similar to the experience-based nature of the changes seen in language. For instance, older infants may possess more robust representations of musical meter, which allows them to better compare the musical periodicities they hear with body movement periodicities they see.

If infants' culture-specific musical experiences influence their perception of musical audiovisual synchrony, this would predict that at some age infants should better detect audiovisual asynchrony in dance to familiar music than in dance to music from another culture. In this light, it is perhaps surprising that we did not observe any effects of song in either experiment given that the meter of one song (*Hendry*, in 4/4) is purported to be more prevalent in North American music than the meter of the other song (*Five*, in 3/4) (Gerry et al., 2010). However, these asymmetries in prevalence might not reflect all infant-directed music given that prevalence has been examined in only a limited (Kindermusik) repertoire (Gerry et al., 2010). Thus, infants in our sample may have had equal prior experience with both meters. To more directly address the role of experience, future research is needed to compare audiovisual synchrony perception of culturally familiar and unfamiliar music and dance displays with a larger sample. Ideally, this future work would involve a cross-cultural sample of infants with differing musical experience to directly tease apart the effects of cultural familiarity from perceptual characteristics of the musical stimuli.

Perception of musical audiovisual synchrony may also be influenced by individual infants' specific experiences in listening and moving to music. Audiovisual synchrony perception differs for adults with and without specific musical training, with piano players and drummers showing enhanced audiovisual synchrony perception only for displays of their own instrument (piano playing and drumming, respectively) (Lee & Noppeney, 2011; Petrini, Russell, & Pollick, 2009). Adult dancers exhibit differential brain responses while watching videos of movements that are within their repertoire compared with movements from a different type of dance (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005). Exposure to specific dances, such as the Charleston or Samba, may also determine whether or not adults can accurately perceive audiovisual synchrony in displays of those dances (Naveda & Leman, 2010). Recent evidence suggests that 6-month-olds who had just 5 min of experience in attempting to play a drum outperformed 6-month-old control infants without such experience on a simple audiovisual synchrony matching task (Gerson et al., 2015). Future work is needed to investigate whether or not individual differences in experience with drumming, dancing, and perhaps even walking and crawling predict how well infants perceive musical audiovisual synchrony.

We observed success at discrimination of matching from mismatching test stimuli in the audiovisual condition but not in the visual-only condition (for older infants in Experiment 1). In principle, infants might have been able to distinguish matching from mismatching stimuli even without auditory information because the visual dance movements of the matching stimulus more closely matched the movements seen during habituation given that they were drawn from the same song. However, infants' failure to discriminate these test stimuli without sound, even when able to discriminate the audiovisual versions of the same stimuli, suggests that older infants' success in the audiovisual condition was based on detecting audiovisual synchrony and asynchrony and not on visual information alone. In addition, even adults had difficulty in discriminating matching from mismatching videos when the videos were presented without sound, further supporting the idea that visual information alone could not have driven infants' discrimination and, thus, that audiovisual synchrony played a crucial role. Infants' failure, as well as adults' failure, in the visual-only condition is also interesting because it suggests that participants may have been unable to perceive a beat or coherent pattern within the visual-only stimuli. This interpretation is consistent with the notion that intersensory redundancy enhances perception of temporal structures (Bahrick & Lickliter, 2000). It may be that only when auditory musical information was available were older infants and adults able to parse a regular beat and metrical patterning from the visual stimuli.

The current findings also constrain theoretical explanations for the slow gradual development of the ability to move in synchrony with music. Although precocious sensitivity to a musical beat in simple auditory stimuli has been demonstrated among newborns and infants with some tasks and stimuli (Winkler et al., 2009), production (e.g., precise synchrony of movement to music) lags behind perception of rhythm and beat in music, such that young children are often unable to accurately move in time with a musical beat (Ilari, 2015; Zentner & Eerola, 2010). Our findings argue against a motor limitation account of this lag. By this motor account, infants and young children have an early-developing grasp of the beat but are unable to dance in synchrony with music solely because they lack the necessary motor coordination (Cirelli, Wan, & Trainor, 2014; Trainor & Cirelli, 2015; Zentner & Eerola, 2010). In contrast to the predictions of this account, the current findings rather support the notion that the rich multimodal perception of musical beat and its accompanying behaviors emerge slowly over the course of infancy and perhaps during childhood. The protracted development of musical audiovisual synchrony perception may reflect the relatively slow developmental trajectory of intermodal integration in general (Gori, Del Viva, Sandini, & Burr, 2008; Massaro, 1984; Nava & Pavani, 2013; Tremblay et al., 2007), which may arise as children continuously recalibrate information across changing sensory modalities, as they learn to establish correspondence between sensory signals, and as they adjust the weighting of information from earlier versus later developing sensory modalities (Ernst, 2008; Lewkowicz, 1988).

Overall, the current findings provide the first evidence that infants perceive musical audiovisual synchrony in complex naturalistic stimuli by the end of their first year. In contrast, younger infants fail at the same task, suggesting that this capacity emerges during later infancy. These findings provide critical insight into the developmental origins of beat-based musical behavior, a universal and fundamentally human activity, and expand our growing understanding of how infants learn to combine and coordinate complex information to perceive audiovisual events in the world through multiple senses.

Acknowledgment

Preparation of this manuscript was supported by a grant from the National Science Foundation (BCS-1052718) awarded to E.E.H.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2017.01.006.

References

Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition, 130*, 31–43.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development, 6*, 429–451.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development, 10*, 387–416.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development, 59*, 197–209.

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology, 79*, 253–270.

Bahrick, L. E., & Lickliter, R. L. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*, 190–201.

Baruch, C., & Drake, C. (1997). Tempo discrimination in infants. *Infant Behavior and Development, 20*, 573–577.

Bergeson, T. R., & Trehub, S. E. (2006). Infants' perception of rhythmic patterns. *Music Perception, 23*, 345–360.

Brandt, A., Gebrian, M., & Slevc, R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsyg.2012.00327.

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., Spring, J., & Bremner, M. E. (2011). Two- to eight-month-old infants' perception of dynamic auditory–visual spatial colocation. *Child Development, 82*, 1210–1223.

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.

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception, 30*, 517–533.

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*. http://dx.doi.org/10.3389/fnhum.2014.00903.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15*, 1243–1249.

Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2014). Fourteen-month-old infants use interpersonal synchrony as a cue to direct helpfulness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*. http://dx.doi.org/10.1098/rstb.2013.0400.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition [computer software]*. Austin: University of Texas.

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non-vocal mimic. *Journal of Comparative Psychology, 127*, 412–427.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*, 478–484.

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.

Ernst, M. O. (2008). Multisensory integration: A late bloomer. *Current Biology, 18*, R519–R521.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*, 238–252.

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of dancing and singing to music in three- to four-month-old infants. *PLoS ONE, 9*(5), e97680.

Gerry, D. W., Faux, A. L., & Trainor, L. J. (2010). Effects of Kindermusik training on infants' rhythmic enculturation. *Developmental Science, 13*, 545–551.

Gerson, S. A., Schiavio, A., Timmers, R., & Hunnius, S. (2015). Active drumming experience increases infants' sensitivity to audiovisual synchrony during observed drumming actions. *PLoS ONE, 10*(6), e0130960.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*, 694–698.

Grossmann, T., Missana, M., Friederici, A. D., & Ghazanfar, A. A. (2012). Neural correlates of perceptual narrowing in cross-species face–voice matching. *Developmental Science, 15*, 830–839.

Hannon, E. E., Bosch, Vanden., 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.

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.

Hannon, E. E., Leveque, Y., Nave, K., & Trehub, S. E. (2016). Exaggeration of language-specific rhythms in English and French children's songs. *Frontiers in Psychology, 7*. http://dx.doi.org/10.3389/fpsyg.2016.00939.

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.

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science, 16*, 48–55.

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.

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports, 3*. http://dx.doi.org/10.1038/srep01566.

Holich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5) [computer program]*. West Lafayette, IN: Purdue University.

Honing, H., Bouwer, F. L., & Háden, G. P. (2014). Perceiving temporal regularity in music: The role of auditory event-related potentials (ERPs) in probing beat perception. In H. Merchant & V. de Laufente (Eds.), *Neurobiology of interval timing* (pp. 305–323). New York: Springer.
Honing, H., Merchant, H., Háden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE, 7*(12), e51369.
Ilari, B. (2015). Rhythmic engagement with music in early childhood: A replication and extension. *Journal of Research in Music Education, 62*, 332–343.
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.
Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459–491.
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.
Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology, 102*, 299–314.
Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience, 9*, 56–67.
Kopp, F., & Dietrich, C. (2013). Neuronal dynamics of audiovisual synchrony and asynchrony perception in 6-month-old infants. *Frontiers in Psychology, 4*. http://dx.doi.org/10.3389/fpsyg.2013.00002.
Kubicek, C., De Boisferon, A. H., Dupierrix, E., Pascalis, O., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS ONE, 9*(2), e89275.
Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.
Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 829–840.
Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science, 19*, 527–566.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119–159.
Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences of the United States of America, 108*, E1441–E1450.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Lewkowicz, D. J. (1988). Sensory dominance in infants: Ten-month-old infants' response to auditory–visual compounds. *Developmental Psychology, 24*, 155–171.
Lewkowicz, D. J. (1992a). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior and Development, 15*, 297–323.
Lewkowicz, D. J. (1992b). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*, 519–528.
Lewkowicz, D. J. (1994). Limitations on infants' response to rate-based auditory–visual relations. *Developmental Psychology, 30*, 880–892.
Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1094–1106.
Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*, 795–804.
Lewkowicz, D. J. (2010). Infant perception of audio–visual speech synchrony. *Developmental Psychology, 46*, 66–77.
Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*, 292–315.
Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 6771–6774.
Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy, 15*, 46–60.
Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development, 37*, 90–94.
London, J. (2002). Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception, 19*, 529–550.
Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1*, 424–441.
Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development, 55*, 1777–1788.
McAuley, J., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General, 135*, 348–367.
Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: Converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Development, 84*, 604–616.
Naveda, L., & Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Perception, 28*, 93–111.
Nettl, B. (1983). *The study of ethnomusicology: Twenty-nine issues and concepts*. Urbana: University of Illinois Press.
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.
Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development, 22*, 237–247.
Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*, 191–196.
Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition, 110*, 432–439.
Phillips-Silver, J. (2009). On the meaning of movement in music, development, and the brain. *Contemporary Music Review, 28*, 293–314.
Phillips-Silver, J., & Trainor, L. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science, 308*, 1430.

Pickens, J., & Bahrick, L. E. (1995). Infants' discrimination of bimodal events on the basis of rhythm and tempo. *British Journal of Developmental Psychology, 13*, 223–236.
Pickens, J., & Bahrick, L. E. (1997). Do infants perceive invariant tempo and rhythm in auditory–visual events? *Infant Behavior and Development, 20*, 349–357.
Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 10598–10602.
Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2½- and 4-year-old children. *International Journal of Behavioral Development, 27*, 220–231.
Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969–992.
Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*, 403–452.
Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology, 19*, 831–836.
Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception, 18*, 455–489.
Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: A cross-cultural comparison. *Developmental Psychology, 46*, 286–292.
Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology, 8*, 553–560.
Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626–636.
Su, Y. (2016). Visual tuning and metrical perception of realistic point-light dance movements. *Scientific Reports, 6*. http://dx.doi.org/10.1038/srep22774.
Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception, 28*, 59–70.
Trainor, L. J., & Cirelli, L. (2015). Rhythm and interpersonal synchrony in early social development. *Annals of the New York Academy of Sciences, 1337*, 45–52.
Trainor, L. J., & Hannon, E. E. (2013). Musical development. In D. Deutsch (Ed.), *The psychology of music* (pp. 423–497). Tokyo: Academic Press.
Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE, 2*(8), e742.
Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 18867–18872.
Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology, 22*, 373–377.
Wetherford, M. J., & Cohen, L. B. (1973). Developmental changes in infant visual preferences for novelty and familiarity. *Child Development, 44*, 416–424.
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.
Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology, 102*, 3191–3202.
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.