



Original Articles

Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object



Nicholas A. Smith^a, Nicole A. Folland^b, Diana M. Martinez^b, Laurel J. Trainor^{b,c,d,*}

^a Perceptual Development Laboratory, Boys Town National Research Hospital, 555 N. 30th Street, Omaha, NE 68131, United States

^b Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

^c McMaster Institute for Music and the Mind, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

^d Rotman Research Institute, Baycrest, University of Toronto, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada

ARTICLE INFO

Article history:

Received 5 October 2015

Revised 17 January 2017

Accepted 24 January 2017

Keywords:

Auditory scene analysis

Auditory development

Multisensory object perception

Simultaneous integration

ABSTRACT

Infants learn to use auditory and visual information to organize the sensory world into identifiable objects with particular locations. Here we use a behavioural method to examine infants' use of harmonicity cues to auditory object perception in a multisensory context. Sounds emitted by different objects sum in the air and the auditory system must figure out which parts of the complex waveform belong to different sources (auditory objects). One important cue to this source separation is that complex tones with pitch typically contain a fundamental frequency and harmonics at integer multiples of the fundamental. Consequently, adults hear a mistuned harmonic in a complex sound as a distinct auditory object (Alain, Theunissen, Chevalier, Batty, & Taylor, 2003). Previous work by our group demonstrated that 4-month-old infants are also sensitive to this cue. They behaviourally discriminate a complex tone with a mistuned harmonic from the same complex with in-tune harmonics, and show an object-related event-related potential (ERP) electrophysiological (EEG) response to the stimulus with mistuned harmonics. In the present study we use an audiovisual procedure to investigate whether infants perceive a complex tone with an 8% mistuned harmonic as emanating from two objects, rather than merely detecting the mistuned cue. We paired in-tune and mistuned complex tones with visual displays that contained either one or two bouncing balls. Four-month-old infants showed surprise at the incongruous pairings, looking longer at the display of two balls when paired with the in-tune complex and at the display of one ball when paired with the mistuned harmonic complex. We conclude that infants use harmonicity as a cue for source separation when integrating auditory and visual information in object perception.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The young infant's ability to organize and process the sensory world is fundamental to virtually all aspects of development. Most environments consist of complex multisensory scenes containing objects with both audible and visible properties. Infants must learn to encode and represent the relevant information from the sensory input in each modality in order to make sense of and interact with people and things in their environment. Here we examine infants' ability to tell whether there are one or two auditory objects present on the basis of auditory harmonicity cues, by capitalizing on their abilities to understand small numbers and to match the number of auditory and visual objects in the stimulus.

Previous research indicates that from a very young age, infants are able to segregate a complex visual scene into representations of the objects in the scene (for a review see: Atkinson, 1998). Within the first few months after birth, infants can make use of features such as texture, shape and size, they can segregate objects based on their relative motion against a background, and they can use physical and subjective contours to segregate and/or discriminate one visual object from another (Atkinson & Braddick, 1992; Curran, Braddick, Atkinson, Wattam-Bell, & Andrew, 1999; Ghim, 1990; Kaufmann-Hayoz, Kaufmann, & Stucki, 1986; Kavšek & Yonas, 2006; Otsuka & Yamaguchi, 2003; Sireteanu & Rieth, 1992; Yonas, Gentile, & Condry, 1991). Between 2 and 4 months, infants are also able to maintain a representation of a visual object across time and space, expect objects to be solid with a coherent structure, and recognize familiar and unfamiliar objects (for reviews see: Shuwairi, Albert, & Johnson, 2007; Wilcox, 1999). While researchers continue to answer important questions about

* Corresponding author at: Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON L8S 4K1, Canada.

E-mail address: LJT@mcmaster.ca (L.J. Trainor).

the development of object perception in the visual domain, far less research has addressed how and when the ability to identify and locate auditory objects develops.

The perception of auditory objects is a challenging process because the sound waves produced by different sources in the environment combine before they arrive at the listener's ear. Auditory scene analysis refers to the auditory system's ability to organize incoming acoustic information by unmixing or segregating the complex signal into streams or auditory objects that are likely to correspond to their multiple corresponding sound sources (Bregman, 1990). Natural sounds that induce a sensation of pitch, such as the human voice, many other animal vocalizations, or musical instruments, typically contain energy at multiple frequencies or harmonics, the lowest of which is referred to as the fundamental (f_0) and corresponds to the perceived pitch. The frequencies of upper harmonics are located at integer multiples of that fundamental. For example, a complex tone with a perceived pitch of 200 Hz typically contains energy at 200, 400, 600, 800 Hz and so on. Although the complex tone contains a number of frequency components, phenomenologically it is experienced as a single sound whose timbre or sound quality is affected by the amount of energy at each harmonic.

When analyzing an auditory scene in which there are two or more simultaneous sound sources (e.g., multiple talkers, musical instruments, animal vocalizations), the brain must integrate the frequency components generated by one source, integrate those generated by a second source, and so on, while segregating the frequency components generated by different objects. The end result is a representation of each sound source in the environment as an auditory object. The auditory system begins by performing a spectrotemporal decomposition of the frequency content over time of the incoming complex sound wave, starting in the cochlea in the inner ear, using both spectral and temporal codes (Eggermont, 2001; McDermott & Oxenham, 2008; Plomp, 1976). *Harmonicity* is a major cue for simultaneous integration of frequency components into the percept of an auditory object (Bregman, 1990). Because the harmonics of natural sounds with pitch are typically at integer ratios of the fundamental, frequencies standing in this relationship are likely produced by the same sound source and thus are more readily integrated into a percept of a single auditory object. When a harmonic is sufficiently mistuned (i.e., deviant from being an integer multiple of the fundamental), it will pop out perceptually from the rest of the frequency components and be perceived as a second auditory object. The cue of harmonicity has been studied in adults, the elderly and school-aged children using complex tones with mistuned harmonics (Alain & McDonald, 2007; Alain, McDonald, Ostroff, & Schneider, 2001; Alain, Theunissen, Chevalier, Batty, & Taylor, 2003).

The question remains as to whether infants are able to use harmonicity cues to group harmonics into auditory objects. In two previous studies, we examined infants' perception of mistuned harmonics. In the first, we used a conditioned head-turn method to show that 6-month-old infants are able to discriminate between an in-tune complex tone and a complex tone that has one harmonic mistuned (Folland, Butler, Smith, & Trainor, 2012). In particular, we found that 6-month-olds detected mistunings as small as 2% of the 3rd harmonic in a complex tone with a 240 Hz fundamental.

In the second study (Folland, Butler, Payne, & Trainor, 2015), we used electroencephalography (EEG) to study this question, measuring a pre-attentive neural correlate of the perception of two auditory objects previously identified in adults (Alain, Arnott, & Picton, 2001). This event-related potential response, referred to as the object-related negativity or ORN, is characterized by a fronto-central negativity in the event-related potential that is present when two auditory objects are perceived, but not when one is per-

ceived, irrespective of stimulus probability. In an effort to map the development of this EEG correlate across the first year, we tested infants between 2 and 12 months using an in-tune complex tone and a complex tone with the third harmonic mistuned by 8%. The two stimuli were played in pseudo-random order, such that each occurred on approximately 50% of trials. This developmental study found that infants aged 2 months showed no evidence of an object-related response, but by 4 months there was a significant frontal object-related response, although it had a longer latency and opposite polarity compared to the adult ORN. By 8–12 months there was evidence of an adult-like ORN response. Event-related responses to stimulus change often manifest with opposite polarity in young infants (He, Hotson, & Trainor, 2007), so this study suggests that by 4 months of age, infants, like adults, process a mistuned harmonic as a separate auditory object. However, because the adult-like response did not emerge until 8 months of age, it would be prudent to find converging evidence before concluding that 4-month-olds use harmonicity cues to determine how many auditory objects are present.

Here we use the fact that infants are adept at auditory-visual correspondences (for a review see Bahrack, 2010) to test our hypothesis, specifically asking whether infants associate an auditory stimulus containing a mistuned harmonic with two visual objects and an auditory stimulus containing in-tune harmonics with one visual object. Much of the evidence that infants show cross-modal matching comes from speech, which is generated through movements that produce correlated visual and auditory information (Yehia, Kuratate, & Vatikiotis-Bateson, 2002). Interestingly, these correlations are enhanced in speech to infants (Smith & Strader, 2014). Infants as young as two months are able to match faces and voices (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003; Walton & Bower, 1993), although cross-modal matching continues to improve with more challenging stimuli (Lewkowicz, Minar, Tift, & Brandon, 2015). At 4 months infants can match shapes to vowel-consonant pairs (Ozturk, Krehm, & Vouloumanos, 2013) and at 5 months can match affect between voices and facial expressions (Vaillant-Molina, Bahrack, & Flom, 2013). Studies involving nonlinguistic stimuli have also found evidence of cross-modal matching. For example, 6-month-old infants are able to match pitch and object size (Prieto-Fernandez, Navarra, & Pons, 2015), 10-month-old infants match higher frequencies with bright objects and lower frequencies with dark objects (Haryu & Kajikawa, 2012), and infants as young as 3–4 months match congruent ascending or descending auditory stimuli and spatial elevation and object width and pitch (Dolscheid, Hunnius, Casasanto, & Majid, 2014).

The ability to parse incoming sensory information into individual objects is fundamental to the understanding of number. A number of studies show that infants are sensitive to the congruence between the number of objects presented through different modalities (Coubart, Izard, Spelke, Marie, & Streri, 2013; Féron, Gentaz, & Streri, 2006; Izard, Sann, Spelke, & Streri, 2009; Starkey, Spelke, & Gelman, 1983; Wilcox, Woods, Tuggy, & Napoli, 2006). Some of these studies show infant preferences for numerically matching stimuli, and some for numerically non-matching stimuli (see Cantrell & Smith, 2013, for a review). For example, whereas Jordan and Brannon (2006) found that infants preferred visual displays with the number of faces corresponding to the number of voices heard, other work with different sounds and objects has shown that infants prefer visual displays in which the number of objects does not match the number of sounds heard (Feigenson, 2011; Kobayashi, Hiraki, & Hasegawa, 2005; Kobayashi, Hiraki, Mugitani, & Hasegawa, 2004; Moore, Benenson, Reznick, Peterson, & Kagan, 1987).

In much of this previous work, the individuation of objects was more or less taken for granted, with the auditory portion of the dis-

play typically consisting of a series of discrete sounds presented in sequence, in conjunction with a number of unconnected visual objects, creating little ambiguity about the number of auditory or visual objects presented. With the exception of [Jordan and Brannon's \(2006\)](#) study, in which they presented infants with simultaneous voices, the role of auditory stream segregation and auditory object perception in the generation of number percepts have been largely unexplored. In the present study, we connect infants' ability to use harmonicity cues to determine how many auditory objects are present at the same time with their ability to match the number of auditory and visual objects present.

The primary goal of the current study was to extend our previous work ([Folland et al., 2012](#); [Folland et al., 2015](#)) to younger infants, using a behavioural visual preference measure to determine if 4-month-olds are able to use harmonicity as a cue to auditory stream segregation.

Visual preference techniques have a long history in the developmental object perception literature and provide an ideal method to study pre-verbal 4-month-olds ([Bower, 1974](#); [Spelke, 1985](#)). The present study examined infants' visual preferences for displays of either one or two bouncing balls paired with one (in-tune complex tone) or two (complex tone with one mistuned harmonic) auditory objects, which were synchronous with the ball (or balls) hitting the floor. This procedure therefore tests infants' sensitivity to the numerical correspondences between the auditory and visual information.

Some studies have shown that infants prefer congruent audiovisual stimuli in which information in the auditory and visual modalities match, whereas others have shown that infants are more interested in stimuli where there is a mismatch between auditory and visual stimuli. For example, a number of studies have shown that infants prefer visual displays that match auditory displays in terms of synchrony ([Dodd, 1979](#)), emotion ([Walker-Andrews, 1986](#)), and sex of talker ([Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991](#)). In contrast, previous work has demonstrated that infants preferentially attend to displays in which the numerosity of auditory and visual objects do not match, particularly if one audiovisual stimulus is presented at a time ([Kobayashi et al., 2004](#); [Kobayashi et al., 2005](#); [Moore et al., 1987](#); [Wilcox et al., 2006](#)). Given that our method requires matching the number of auditory and visual objects present, we expected that if infants use harmonic cues to the number of auditory objects present, they would most likely prefer to look at visual displays incongruent with the number of auditory objects, that is, look longer at one ball when two auditory objects (i.e., mistuned harmonic) were present and longer at two balls when one auditory object (i.e., in-tune harmonics) was present.

2. Methods

2.1. Participants

78 full term infants aged 4 months (± 2 weeks) participated in the study (30 males, mean age 4.1 ± 0.20). Upon arrival at the lab, caregivers provided written consent and completed questionnaires regarding musical training and exposure. All infants were healthy at the time of testing and caregivers reported no history of frequent ear infections, pressure-equalizing tubes, or familial hearing impairment. Infants who were very fussy ($n = 8$), who completed fewer than 4 trials ($n = 4$), or for whom we experienced technical difficulties ($n = 2$) were excluded from analyses. The final sample consisted of 64 infants with a mean age of 4.1 months. Infants were randomly assigned to either the In-tune ($n = 32$) or Mistuned ($n = 32$) condition in a between-subjects design. After

the experiment, infants were given a certificate of participation and a bath toy as a token of appreciation.

2.2. Stimuli

The properties of the mistuned and in-tune complex tones were chosen so as to maximize saliency for infant listeners. The pitches of these stimuli fall within the range of infant directed speech, they contain resolvable harmonics, and the frequency differences used fall well within infant discrimination limits ([Olsho, Schoon, Sakai, Turpin, & Sperduto, 1982](#)). The harmonics also feature a 6 dB/octave roll off to parallel natural music and speech sounds which typically range between -4 and -12 dB/octave ([Hall, 1980, p. 206](#); [Sundberg, 1991, p. 118](#)). Two complex tones were created using Adobe Audition 6.0, each with a duration of 500 ms, including 50 ms rise and fall times, and a 6 dB/octave roll off. The in-tune complex tone had a fundamental frequency of 240 Hz and included the first 6 harmonics (240, 480, 720, 960, 1200 and 1440 Hz) in random phase. This tone is perceived by adults as one sound (one auditory object). The mistuned complex tone was identical to the in-tune complex tone except that the 3rd harmonic was mistuned upwards by 8% resulting in a frequency of 777.6 Hz. The mistuned complex tone is perceived by adults as two sounds (two auditory objects), one with a pitch of 240 Hz, consisting of a perceptual integration of the five in-tune harmonics, and the other with a pitch of 777.6 Hz, consisting of the mistuned harmonic. The sounds were presented through two audiological GSI speakers connected to a NADC352 stereo integrated amplifier in an Industrial Acoustics Company booth using a Macintosh G4 computer located outside the booth.

The visual orienting and visual test stimuli were created in Apple QuickTime format using Adobe Director 11, and presented using software developed in Max/MSP/Jitter 5 on a 23-inch Apple Cinema HD display via a Macintosh G4 computer. The visual orienting stimulus was a 3.8 cm black-and-white spotted looming ball in the center of the screen, subtending a maximum visual angle of 4.4° . There were two visual test stimuli: a one-ball video and a two-ball video. The one-ball video depicted a single 3.8 cm (visual angle 4.4°) dark grey bouncing ball. The two-ball video depicted both a 3.8 cm dark grey bouncing ball and a 1.3 cm (visual angle 1.5°) white bouncing ball (see [Supplementary Material](#)). Both bouncing balls were shaded to appear 3-dimensional in shape and were coordinated with the sounds such that they fell with a realistic acceleration trajectory and the sound began when they hit a black bar, representing the ground, near the bottom of the screen.

As shown in [Fig. 1](#), each video was 2000 ms in duration and looped continuously for the duration of infant looking on a given trial. Each video began with the ball (or balls) at the apex of its bounce, with each ball falling and hitting the ground at the 1000 ms mark. This impact coincided with the onset of the tone complex. Because the white ball fell from a higher initial position than the black ball, it travelled at a slightly faster average speed ($9.4^\circ/s$) than the dark ball ($7.7^\circ/s$) in order to hit the ground at the same time. This difference in speed was intended to reinforce the percept of the two balls as separate objects, and to counteract potential grouping effects that may have been produced had the two balls moved at the same speed (i.e., through the Gestalt principle of common fate).

To adults, the large ball was perceived to produce the complex tone with a low fundamental frequency and the smaller white ball was perceived to produce the high-pitched mistuned harmonic. Although this percept is consistent with previous research on cross-modal correspondences ([Marks, 1975](#); [Mondloch & Maurer, 2004](#)), in the present study it is not necessary to make assumptions about how infants might be matching particular elements of the in-tune or mistuned tone complexes with elements of the one-

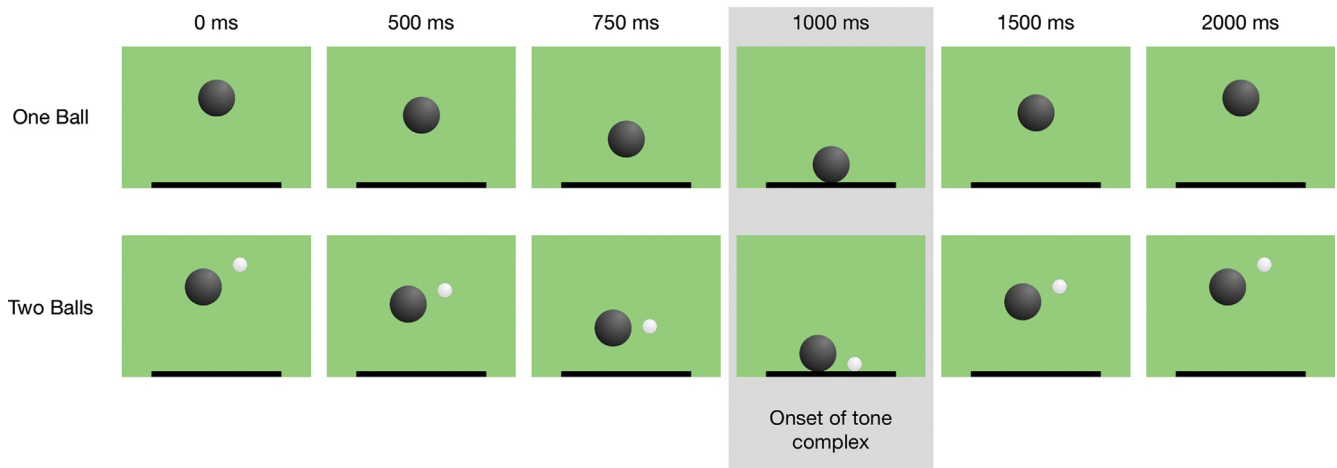


Fig. 1. Audiovisual stimuli. An illustration of the one-ball (top row) and two-ball (bottom row) video stimuli, showing the positions of the balls at a series of time points. In both conditions the balls hit the ground at 1000 ms, which coincided with the onset of the tone complex (in-tune or mistuned, depending on the condition). The videos looped, with the balls returning to their initial position at the 2000-ms point.

and two-ball videos, because our paradigm tests infants' sensitivity to the congruence between the auditory and visual stimuli as a whole. The visual test stimulus on a trial contained either one or two balls and was presented on either the left or right side of the screen over a neutral green background. Whether a particular infant saw stimuli with one ball on the left side (and two balls on the right side), or vice versa, remained constant through the experiment, but this factor was counterbalanced between infants.

From these two audio and two visual stimuli (see Fig. 1), four audiovisual test stimuli were created. Two of the audiovisual stimuli were congruent, that is, the audio and visual information matched (two bouncing balls with the mistuned complex tone; one bouncing ball with the in-tune complex tone), and two were incongruent, that is, the auditory and visual information did not match (two bouncing balls with the in-tune complex tone; one bouncing ball with the mistuned complex tone).

2.3. Procedure

After obtaining informed consent, the infant and caregiver(s) were brought into the sound booth and the infant was placed in a car seat 50 cm in front of the monitor. To ensure that infants were not distracted, floor to ceiling black curtains surrounded the car seat and computer screen. Each caregiver was asked to remain seated behind the infant and to remain quiet for the duration of the experiment.

Infant looking times were recorded by two independent observers located outside the sound booth. Both observers were blind to which condition the infant was being tested in. The observers viewed the infant's eye movements on a monitor outside the sound booth, on which a live feed of the infant's head was shown from a camera positioned beneath the computer screen. The observers controlled the experiment using independent, silent keypads that were each connected to the computer that presented the stimuli. When the infant was attentive, each observer indicated with a button press that the infant was ready for a trial. When both observers indicated that the infant was ready, the orienting flashing ball appeared in the middle of the screen. Each observer indicated with a second button press when the infant's attention was on the middle of the screen. Once both observers had pressed this second button, the orienting stimulus disappeared and a test stimulus was presented. During the test stimulus, one of the two audiovisual stimuli was presented on either the right or the left side of the monitor. When the infant looked at the visual display (presented

on either the right or left side of the screen), each observer pressed a third button, which they held down for as long as the infant looked at the stimulus. When the infant looked away from the stimulus, observers released their button. The looking time counter for the trial began when both observers had pressed their buttons and it ended when both observers had released their buttons for at least 2 s. In this way, across trials, the amount of time infants spent looking at congruous and incongruous auditory/visual stimuli was recorded. On a particular trial, if an observer released their button, but repressed it within 2 s, the trial continued. The next trial began when both observers indicated that the infant was ready for the next orienting stimulus. The experiment ended when the infant completed 16 trials or became too fussy to continue.

Infants were randomly assigned to either the In-Tune Condition, in which they heard only the in-tune complex tone, or to the Mistuned Condition, in which they heard only the complex tone with the mistuned third harmonic. Within each of these conditions, half of the infants were first presented with the visual stimulus containing two balls, followed by the visual stimulus containing one ball, in alternating fashion. The other half were presented with the visual stimulus containing one ball, followed by two balls. Crossed with this factor, half of the infants were first presented with the stimulus on the left and half with the stimulus on the right. This design ensured that any observed looking time differences were related to audiovisual congruence, and not to side or primacy biases.

If infants perceived the mistuned complex tone as two separate auditory objects, we expected them to look longer on incongruent trials in which there was one bouncing ball. Similarly, if infants perceived the in-tune tone as a single auditory object, we expected them to look longer on incongruent trials in which there were two bouncing balls.

3. Results

The total number of trials completed ranged from 6 to 16 trials (mean = 10.9, SD = 3.3). An initial mixed design ANOVA with factors visual presentation side (one ball left, two balls right; two balls left, one ball right) and first trial (one ball; two balls) revealed no significant effects, meaning that infants showed no evidence of left/right side bias, or primacy bias in their responses. Consequently, the data were collapsed across these factors. The raw data can be found in [Supplementary Material](#).

Next, a two-way repeated measures ANOVA, with the between-subjects factor tone complex (in-tune vs. mistuned) and within-subjects factor visual stimulus (one ball vs. two balls) was performed. No significant main effect was found for tone complex, $F(1,62) = 0.057$, *ns*, or for visual stimulus, $F(1,62) = 0.82$, *ns*. This demonstrates infants did not look longer overall when the audio component consisted of an in-tune or mistuned tone complex. Similarly, they did not show an overall visual preference for 1 or 2 balls. However, as predicted, a highly significant tone complex by visual stimulus interaction was found, $F(1,62) = 34.29$, $p < 0.0001$, demonstrating that infants' visual preferences were driven by which tone complex the visual stimulus was paired with. As shown in Fig. 2, infants looked longer at the two-ball video than the one-ball video when presented with an in-tune tone complex (indicating a single auditory object), and they looked longer at the one-ball video in the context of a mistuned tone complex (indicating two auditory objects). In other words, infants showed a preference for audiovisual mismatch or incongruence in terms of object numerosity.

4. Discussion

The current study found that 4-month-old infants use harmonicity as a cue to determine the number of auditory objects present in the environment and that they expect the number of visual objects making sound to match the number of auditory objects. Specifically, infants looked longer at incongruent audiovisual displays, containing one ball and two auditory objects (i.e., tone complex with a mistuned harmonic) or two balls and one auditory object (i.e., in-tune harmonic), compared to congruent displays with one ball and one auditory object or two balls and two auditory objects.

Adults use a number of cues to determine the number of sound sources in their auditory environments (Bregman, 1990), but harmonicity is particularly important for sounds with pitch, which prominently include communication sounds such as vocalizations and musical tones. During the early months after birth, infants are attracted to speech and music (Corbeil, Trehub, & Peretz, 2013), and over the first year, infants' brains become specialized through experience-driven plasticity for the particular language, voices and musical structures in their environment (e.g., Friendly, Rendall, & Trainor, 2013; Hannon & Trainor, 2007; Johnson, Westrek, Nazzi, & Cutler, 2011; Kuhl, Williams, Lacerda, Stevens, & Lindblom,

2006; Werker & Tees, 1999). In order for this to happen, infants must be able to separate voices and musical tones from other concurrent sounds in the environment. Thus, infants' sensitivity to harmonicity is critical for auditory scene analysis, as well as for early speech and musical development.

Adults are very sensitive to harmonic mistunings. Previous studies suggest young adults (aged 22–24) are able to discriminate mistunings as small as 0.5% depending on the properties of the mistuned complex (Alain et al., 2001). In an earlier behavioural study using a conditioned head-turn response, we found evidence that 6-month-old infants were able to detect mistunings as small as 2% in the third harmonic of a 6-harmonic complex, but we were not able to show sensitivity to smaller mistunings (Folland et al., 2012). For this reason, in the present study, we employed an easily detectable 8% mistuning. It would be valuable for future studies to map infants' increasing sensitivity to smaller mistunings over the first year after birth and to determine how these relate to mistuning thresholds for the perceptual separation of auditory objects. The ability to resolve and encode a sensory cue and the ability of processes downstream to use these cues in the formation of percepts are separate but related. While this previous work demonstrates infants' detection of mistunings and sensitivity to harmonicity cues, the present study provides new evidence that infants make use of these cues in their formation of auditory objects.

The results of the present study are also consistent with our previous EEG study, in which we found that 4-month-olds show a frontally-positive object-related ERP response to stimuli with mistuned harmonics when presented in the context of in-tune stimuli (Folland et al., 2015). The results of the present study indicate that 4-month-olds perceive the mistuned harmonic stimulus as two auditory objects, which provides corroboration that the object-related response in 4-month-old infants (Folland et al., 2015), though different in morphology from that of adults, is nevertheless a neural correlate of the perception of two auditory objects. Interestingly, Folland et al. (2015) did not find evidence of an object-related ERP response at 2 months of age. It would therefore be interesting for future studies to investigate behavioural manifestations of object-related processing related to harmonicity cues in infants younger than 4 months of age, as infants are already learning about voices and faces and other objects in their environment at this time.

In addition to showing that 4-month-olds use harmonicity as a cue for the number of auditory objects in the environment, to our knowledge, this is the first study to show that infants link their perception of multiple auditory objects to expectations about how many visual objects should be present at the same time based on harmonicity cues. Previous work by Wilcox et al. (2006) found that 4.5 month-olds are also able to individuate one versus two objects across time. Specifically, they found that infants looked significantly longer at a display with one object after previously hearing two different sounding rattles compared to after hearing two rattles that made the same noise. However, they did not find a significant audio-visual matching when using computer generated musical notes. The current study shows that when the sounds and visual objects are presented concurrently and aspects of the sounds match aspects of the visual objects, 4-month-old infants link what they are hearing to visual objects and use harmonicity cues to create expectations about the number of objects.

As our understanding of auditory scene analysis has grown in typically developing children and adults, researchers have begun to investigate auditory scene analysis in special populations. For example, a failure to efficiently organize incoming acoustic information into auditory objects may be one of the reasons some individuals with autism find loud, busy environments overwhelming or aversive. Lodhia, Brock, Johnson, and Hautus (2014) found that

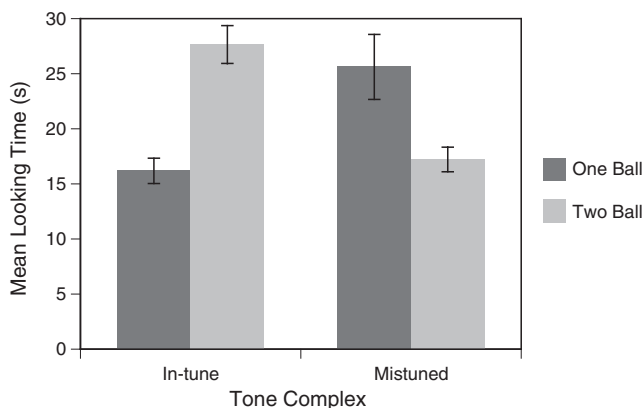


Fig. 2. Mean infant looking time. Mean looking time to one-ball and two-ball visual stimuli as a function of auditory tone complex. Standard error bars are shown. Infants in the In-tune Condition (one auditory object) looked significantly longer at the incongruent visual pairing (two balls). Infants in the Mistuned Condition (two auditory objects) looked significantly longer at the incongruent visual pairing (one ball).

compared to a control group of verbal-, IQ- and age-matched controls, adults with autism showed a significantly smaller ORN. Given the finding from the current study that 4-month-old infants can perceive multiple auditory objects, combined with the previous study indicating that the neural correlates of auditory object representation can be measured at this age (Folland et al., 2015), it is possible that object-related responses in young infants could be used as a test for risk for autism.

To make sense of their environments, infants and adults alike must solve a general perceptual problem that is common across sensory modalities, namely to organize the complex arrays of incoming sensory information to correspond to individual things in the world. Despite general principles that operate across domains (e.g., Bregman's, 1990 *Auditory Scene Analysis* draws much inspiration from Gestalt principles of visual grouping), previous research has primarily focused on this problem within individual modalities, more or less independently of each other. Previous research on how infants make sense of and organize their perception of multiple objects in their environment has largely focused on the visual domain (e.g., Johnson, Slemmer, & Amso, 2004; Kaufmann-Hayoz et al., 1986; Peterhans & von der Heydt, 1991). Existing studies of auditory scene analysis in infancy have largely focused on sequential cues (e.g., Demany, 1982; McAdams & Bertoncini, 1997; Smith & Trainor, 2011; Winkler et al., 2003). The present results extend previous studies by showing that infants can use harmonicity cues to separate simultaneous auditory objects and relate the number of auditory objects to the number of visual objects present.

Given the multisensory nature of our environments, containing things that are simultaneously seen and heard, it is particularly important to understand how solving this perceptual problem within one modality might relate to solving it in another. Future work using bistable or multistable stimuli (Cook & Van Valkenburg, 2009; O'Leary & Rhodes, 1984; Sterzer, Kleinschmidt, & Rees, 2009), whose perceptual organization is ambiguous, would provide a way of examining how bidirectional and reciprocal audiovisual influences on perceptual organization may develop in infancy.

Acknowledgments

This research was supported by grants from the Natural Science and Engineering Research Council of Canada (RGPIN-2014-0470) and the Canadian Institutes of Health Research (MOP 42554) to LJT, the National Institutes of Health (1P20GM109023) to NAS, and a scholarship to NF from the Natural Science and Engineering Research Council CREATE in Auditory Cognitive Neuroscience. The authors declare no competing financial interests. The authors would like to thank the families that participated in the study as well as Dave Thompson and Blake Butler for their assistance and technical support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2017.01.016>.

References

Alain, C., Arnott, S. T., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 1072–1089.

Alain, C., & McDonald, K. L. (2007). Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. *Journal of Neuroscience*, 27(6), 1308–1314.

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2001). Age-related changes in detecting a mistuned harmonic. *Journal of the Acoustical Society of America*, 109(5), 2211–2216.

Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., & Taylor, M. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research*, 16, 210–218.

Atkinson, J., & Braddick, O. (1992). Visual segmentation of oriented textures by infants. *Behavioural Brain Research*, 49(1), 123–131.

Atkinson, J. (1998). The 'where and what' or 'who and how' of visual development. In F. Simion & G. Butterworth (Eds.), *The development of sensory, motor and cognitive capacities in early infancy: From perception to cognition* (pp. 3–24). Sussex: Psychology Press.

Bahrick, L. E. (2010). Intermodal perception and selective attention to intersensory redundancy: Implications for typical social development and autism. In G. Bremner & T. D. Wachs (Eds.), *Blackwell handbook of infant development* (2nd ed., pp. 120–166). Oxford, England: Wiley/Blackwell.

Bower, T. G. R. (1974). *Development in infancy*. San Francisco: Freeman.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. Cambridge, MA: MIT Press.

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, 128, 331–352.

Cook, L. A., & Van Valkenburg, D. L. (2009). Audio-visual organisation and the temporal ventriloquism effect between grouped sequences: Evidence that unimodal grouping precedes cross-modal integration. *Perception*, 38, 1220–1233.

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology*, 4(372), 1–11.

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2013). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, 17, 11–22.

Curran, W., Braddick, O. J., Atkinson, J., Wattam-Bell, J., & Andrew, R. (1999). Development of illusory-contour perception in infants. *Perception*, 28, 527–538.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development*, 5, 261–276.

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

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6), 1256–1261.

Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, 157, 1–42.

Feigenson, L. (2011). Predicting sights from sounds: 6-month-olds' intermodal numerical abilities. *Journal of Experimental Child Psychology*, 110(3), 347–361.

Féron, J., Gentaz, E., & Streri, A. (2006). Evidence of amodal representation of small numbers across visuo-tactile modalities in 5-month-old infants. *Cognitive Development*, 21, 81–92.

Folland, N. A., Butler, B. E., Payne, J. E., & Trainor, L. J. (2015). Cortical representations sensitive to the number of perceived auditory objects emerge between 2 and 4 months of age: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, 27(5), 1060–1067.

Folland, N. A., Butler, B. E., Smith, N. A., & Trainor, L. J. (2012). Processing simultaneous auditory objects: Infants' ability to detect mistunings in harmonic complexes. *Journal of the Acoustical Society of America*, 131, 993–997.

Friendly, R. H., Rendall, D., & Trainor, L. J. (2013). Plasticity after perceptual narrowing for voice perception: Reinstating the ability to discriminate monkeys by their voices at 12 months of age. *Frontiers in Psychology*, 4(718), 1–8.

Ghim, H. (1990). Evidence for perceptual organization in infants: Perception of subjective contours by young infants. *Infant Behavior and Development*, 13, 221–248.

Hall, D. E. (1980). *Musical acoustics: An introduction*. Belmont, CA: Wadsworth Publishing Co.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466–472.

Haryu, E., & Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory-visual correspondences in 10-month-old infants. *Infant Behavior & Development*, 35(4), 727–732.

He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, 19, 878–892.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, 106(25), 10382–10385.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-Month-Olds. *Infancy*, 6, 185–201.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14, 1002–1011.

Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceeding of the National Academy of Sciences*, 103(9), 3486–3489.

Kaufmann-Hayoz, R., Kaufmann, F., & Stucki, M. (1986). Kinetic contours in infants' visual perception. *Child Development*, 57(2), 292–299.

Kavšek, M., & Yonas, A. (2006). The perception of moving subjective contours by 4-month-old infants. *Perception*, 35(2), 215–227.

Kobayashi, T., Hiraki, K., & Hasegawa, T. (2005). Auditory-visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science*, 8(5), 409–419.

- Kobayashi, T., Hiraki, K., Mugitani, R., & Hasegawa, T. (2004). Baby arithmetic: One object plus one tone. *Cognition*, *91*(2), B23–B34.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (2006). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.
- Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162.
- Lodhia, V., Brock, J., Johnson, B. W., & Hautus, M. J. (2014). Reduced object related negativity response indicates impaired auditory scene analysis in adults with autistic spectrum disorder. *PeerJ*, *2*, e261.
- Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*, 303–331.
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America*, *102*, 2945.
- McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*, 452–463.
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 133–136.
- Moore, D., Benenson, J., Reznick, J. S., Peterson, M., & Kagan, J. (1987). Effect of auditory numerical information on infants' looking behavior: Contradictory evidence. *Developmental Psychology*, *23*, 665–670.
- O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, *35*, 565–569.
- Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology*, *18*, 721–726.
- Otsuka, Y., & Yamaguchi, M. K. (2003). Infants' perception of illusory contours in static and moving figures. *Journal of Experimental Child Psychology*, *86*(3), 244–251.
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186.
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196.
- Peterhans, E., & von der Heydt, R. (1991). Subjective contours-bridging the gap between psychophysics and physiology. *Trends in Neurosciences*, *14*(3), 112–119.
- Plomp, R. (1976). *Aspects of tone sensation*. London: Academic.
- Prieto-Fernandez, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior & Development*, *38*, 77–81.
- Shuwairi, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. *Psychological Science*, *18*(4), 303–307.
- Sireteanu, R., & Rieth, C. (1992). Texture segregation in infants and children. *Behavioural Brain Research*, *49*(1), 133–139.
- Smith, N. A., & Strader, H. L. (2014). Infant-directed visual prosody: Mothers' head movements and speech acoustics. *Interaction Studies*, *15*, 38–54.
- Smith, N. A., & Trainor, L. J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distracter. *Infancy*, *16*, 655–668.
- Spelke, E. S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 323–364). Norwood, NJ: Ablex.
- Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*, 179–181.
- Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*(7), 310–318.
- Sundberg, J. (1991). *The science of musical sounds*. San Diego, CA: Academic Press.
- Vaillant-Molina, M., Bahrack, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, *18*(s1), E97–E111.
- Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors. Relation of eye and voice? *Developmental Psychology*, *22*, 373–377.
- Walker-Andrews, A. S., Bahrack, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*, 55–75.
- Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, *16*, 233–243.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535.
- Wilcox, T. (1999). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology*, *37*, 97–155.
- Wilcox, T., Woods, R., Tuggy, L., & Napoli, R. (2006). Shake, rattle, and ... one or two objects? Young infants' use of auditory information to individuate objects. *Infancy*, *9*(1), 97–123.
- Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., ... Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceeding of the National Academy of Sciences*, *100*, 11812–11815.
- Yehia, H., Kuratate, T., & Vatikiotitis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, *30*, 555–568.
- Yonas, A., Gentile, D. A., & Condry, K. (1991). Infant perception of illusory contours in apparent motion displays. *Bulletin of the Psychonomic Society*, *28*, 480.