Cognition 164 (2017) 1–7

Contents lists available at ScienceDirect

Cognition

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

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
Available online 24 March 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).

http://dx.doi.org/10.1016/j.cognition.2017.01.016
0010-0277/© 2017 Elsevier B.V. All rights reserved.

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, & Trainer, 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 frontocentral negativity in the event-related potential that is present when two auditory objects are perceived, but not when one is perceived, 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 Bahrick, 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, & Vatikiotitis-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, Bahrick, & 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 nonmatching 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-

Download English Version:

https://daneshyari.com/en/article/5041596

Download Persian Version:

https://daneshyari.com/article/5041596

Daneshyari.com