PERCEIVING SIMULTANEOUS SOUNDS IN INFANCY

INFANTS' ABILITY TO PERCEIVE MULTIPLE SIMULTANEOUS AUDITORY OBJECTS

By NICOLE FOLLAND, B.Sc. H

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University

McMaster University PHD PSYCHOLOGY, NEUROSCIENCE & BEHAVIOUR (2015)
Hamilton, Ontario (Psychology, Neuroscience & Behaviour)

TITLE: Infants' ability to perceive multiple simultaneous auditory objects

AUTHOR: Nicole A. Folland

SUPERVISOR: Dr. Laurel J. Trainor

SUPERVISORY COMMITTEE MEMBERS: Dr. Louis A. Schmidt, Dr. Steven Brown

NUMBER OF PAGES: xiii, 119

Abstract

In order to make sense of the world, infants must be able to segregate incoming acoustic information emitted by simultaneous sources. Auditory scene analysis refers to the ability to break down the complex sound signal arriving at the ear into the discrete objects giving rise to it (Bregman, 1990). Chapter 2 uses a conditioned head-turn paradigm to investigate whether 6-month-old infants are able to discriminate a mistuned harmonic within a complex tone, a stimulus that adults perceive as two separate auditory objects. Results suggest that infants are able to perform such a discrimination, but that their threshold is higher than adults. In adults, the perception of two auditory objects is associated with a neural correlate derived from event-related potential (ERP) recordings referred to as the object-related negativity (ORN). Chapter 3 investigates whether the ORN is elicited from infants in response to these mistuned stimuli, indicating that they hear the mistuned harmonic as a distinct auditory object. By 4 months, infants showed a significant frontally-positive, object-related response which transitioned to an adult-like ORN by 8 months. Chapter 4 uses a visual preference paradigm to determine whether 4-month-old infants have an expectation that the number of auditory objects they hear should match the number of objects they see. These infants looked significantly longer at trials where the number of audio and visual objects did not match, suggesting that they integrate information about auditory and visual sources from a young age. Collectively, these findings indicate that by 4 months, infants are able to discriminate a mistuned harmonic, and that this mistuning is perceived as a distinct auditory object that elicits a measurable response in the infant ERP. However, thresholds for discriminating mistuned

harmonics and infant object-related waveforms continue to mature across the first year of birth.

Acknowledgments

First and foremost I would like to thank my parents, William and Mary. You have loved me unconditionally and inspired me to build a dream to chase after. You have instilled in me that I am smart enough, brave enough and passionate enough to take on any challenge in front of me. Words cannot express how much your patience and unwavering faith in me means, especially throughout the past two years. Mom and dad, it is because of you that I am here today.

I would like to extend my sincere gratitude to Dr. Laurel Trainor, my Ph.D. supervisor for your constant support and guidance. Your excellence as a researcher in the infant development community is truly inspiring. A special thank you to Dr. Louis Schmidt who has been a mentor for 13 years. Your support, kindness and encouragement have shaped the researcher and person I am today. Thank you to Dr. Steven Louis, this thesis and many presentations were greatly improved by your thoughtful feedback. To Dr. Blake Butler, this experience would have been a lot more challenging had it not been for your friendship, leadership and sense of humour. I am so grateful to have had you as a lab member and collaborator during my time at McMaster. To Elaine and Nancy, thank you for your part-time counselling services and always making me smile. I also want to acknowledge Kate Einarson and Dave Thompson for your support and help throughout the years.

To my friends and family who have supported and encouraged me throughout this process Michael, Meredith, Lianne, Stephanie and Lindsay R., I love you and thank you!

Andrew, your positive attitude, love and support has meant the world to me and helped me reach this goal. Last but not least, my oldest and dearest friend Lindsay F., your strength, passion and sense of adventure continue to inspire me. Thank you for always being there for me, you have shaped my life in countless ways and I am forever grateful.

TABLE OF CONTENTS

Preface

This thesis consists of three manuscripts of which I am the primary author. Chapters 2 and 3 represent manuscripts that have been published in peer-reviewed journals. Chapter 4 represents a manuscript that has been submitted and is currently under consideration in a peer-reviewed journal.

For each manuscript I was responsible for the experimental design, running of adult and/or infant participants, data collection, data analysis under the supervision of Dr. Laurel Trainor. I was also the primary writer, again under the supervision of Dr. Trainor. Chapter 2 is a reprint of Folland, N.A., Butler, B.E., Smith, N.A., and Trainor, L.J. (2012) Processing simultaneous auditory objects in infancy: Music and mistuned harmonics. *Journal of the Acoustical Society of America*, *131*, 993-997. Chapter 3 is a reprint of Folland, N.A., Butler, B.E., and 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. Chapter 4 corresponds to the submitted manuscript Folland, N.A., Smith, N.A., Martinez, D.M., and Trainor, L.J. (submitted). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition*.

As each Chapter takes a different approach to investigating simultaneous integration in infancy and each is intended to be a stand-alone manuscript there is some repetition within the Chapters.

List of Figures

CHAPTER 1

The ability to identify and locate objects in the environment is fundamental to human survival. Whether identifying a potential threat, communicating with others, or seeking out a potential food source, humans are continually determining the *who*, *what* and *where* of objects and events around them. Because organizing sensory input into representations of objects is such a common daily occurrence, the complexity of the processes involved is not always appreciated.

For example, typically developing adults can look around a complex visual scene such as a library and instantaneously identify the boundaries between each aisle and the book stacks as well as the boundaries between hundreds of books on the shelf. Also, even in a quiet library adults can separate the sound of footsteps walking down the aisle, the clicks of a computer keyboard, and the rustling of pages turning. For decades researchers have been answering questions about how adults take incoming visual and acoustic information from the environment and translate it into a coherent representation of reality. However it is of particular interest to study the development of this ability, as it forms the building block for social, cognitive, language and musical development.

Infant Research Methodology

Many different methods have been employed to study the development of visual and auditory object perception. Most of these techniques fall into one of two broad categories: behavioural measures requiring a motor response from the infant, and electrophysiological measures in which electrical potentials reflecting the depolarizing and firing of neurons are recorded at the surface of the head (Atkinson and Braddick, 2013; Trainor and He, 2013). Behaviourally, visual paradigms relying on eye or head movements are ideal for infants younger than 6 months of age. In experiments such as these, preference to look longer at one stimulus compared to another or dishabituation (increased looking time) to a new stimulus after a priming period with a standard stimulus reflect infants' ability to discriminate between stimuli (Berlyne, 1958; Fantz, 1964). Infants older than 6 months are often tested using a head turn preference procedure or a conditioned head turn procedure. For example, infant auditory perception can be tested in a preference procedure in which looking at one visual display causes one sound (or category of sounds) to be heard while looking at a different visual display causes a different sound (or category of sounds) to be heard. Differential looking times, and therefore differential listening times, to the two stimuli indicate that infants can discriminate them. In a conditioned head turn procedure, infants are taught that upon turning their head to the location of a target stimulus they will be rewarded with a dancing toy or flashing light. Head turns to distracter or background stimuli are not reinforced. If the infant spends more time looking at the target stimulus, this is an indication of behavioural discrimination (e.g., Trainor and Trehub, 1992; Werker, Polka, and Pegg, 1997; Gout, Christophe, and Morgan, 2004; Hollich, 2006; Kooijman, Johnson, and Cutler, 2008).

Electroencephalography (EEG) can be employed to measure event-related potentials (ERPs), which are small voltage changes generated by the depolarization of neurons in the brain that are time-locked to the onset of a stimulus or event (Blackwood and Muir, 1990). Unlike imaging methods such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), ERPs derived using EEG provide an infant-friendly means to study the neural correlates of object perception (Picton, Alain, Woods, John, Scherg, Valdes-Sosa, et al., 1999). Although research involving infant ERPs is still relatively new, cortical ERP components in adults have been well studied. Two exogenous components of interest include the P1, a positive deflection peaking around 50 ms post-stimulus-onset, believed to be generated in primary auditory cortex and the N1, a negative deflection peaking around 100 ms post-stimulus-onset, believed to be generated in secondary auditory cortex (Godey, Schwartz, de Graaf, Chauvel, and Liegeois-Chauvel, 2001; Luck, 2005; Yvert, Fischer, Bertrand, and Montysalo, 2005). Two commonly referenced endogenous ERP components include the mismatch negativity (MMN; Picton, Alain, Otten, Ritter, and Achim, 2000; Näätänen, Paavilainen, Rinne, and Alho, 2007) and the object-related negativity (ORN; Alain et al., 2001). The MMN, thought to reflect a change-detection process, is elicited to the presentation of an occasional deviant or oddball sound stimulus nested in a sequence of frequent standard stimuli. The MMN is seen at fronto-central sites on the scalp as a negative deflection in response to deviant stimuli, peaking between about 150 to 250 ms after stimulus onset depending on stimulus properties (Näätänen et al., 2007). MMN polarity reverses at the back of the scalp, consistent with a primary generator in secondary auditory cortex (Picton et al., 1999). Unlike the N1, which does not appear to be robust until after 4 years of age (Shibasaki and Miyazaki, 1992; Pang and Taylor, 2000; Ponton, Eggermont, Kwong, and Don, 2000; Shanin, Roberts, Chau, Trainor, and Miller, 2008;

Shahin, Trainor, Roberts, Backer, and Miller, 2010) adult-like MMN morphology is seen as young as 2 or 3 months of age for pitch deviants (Picton, Alain, Otten, Ritter, and Achim, 2000; He et al., 2007; 2009; Näätänen et al., 2007; Trainor, 2008; He and Trainor, 2009; Trainor, 2013; Trainor and He, 2013) and by 4 to 6 months for small silent gaps in a tone (Trainor, McFadden, Hodgson, Darragh, Barlow, Matsos, and Sonnadara, 2003). Young infants also show a slow frontally-positive response to occasional deviants, a response whose amplitude decreases with age and is not seen in adults (Moore and Linithicum, 2007, Näätänen et al., 2007; Trainor, 2008).

The ORN, an ERP component first documented by Alain (2001), has been extensively studied in adults, the elderly and school-aged children. However, to date no studies have investigated if and when this component appears in infancy (Alain 2001; Alain 2003). As this component is of particular interest to the development of auditory object perception, I will discuss the ORN in greater detail in the following sections.

Each of these behavioural and EEG methods comes with its own set of limitations. When used in conjunction with one another they are better able to assess functional hearing, which is critical in research involving preverbal infants.

Auditory Input

Despite how crucial auditory object perception is for social, linguistic and musical development, the majority of the infant perceptual and cognitive literature has focused on the visual domain. Thus our understanding of visual object development is much more advanced than our understanding of the development of auditory object perception. In order to understand the problem facing the auditory system, it is necessary to understand how sounds are translated into neural impulses. Sound is created when an object moves or vibrates; these disturbances cause air particles to compress and retract, and the movement of these pressure changes through the air is called a sound wave. When a sound wave reaches the outer ear it cause the eardrum to vibrate. These vibrations travel through the middle ear (ossicles) to reach the inner ear, containing the basilar membrane, which vibrates in a liquid medium. Hair cells from the auditory nerve attach to the basilar membrane, such that the mechanical energy of basilar membrane movement gets translated into neural impulses (Rauschecker, and Shannon, 2002; Yost, 2007).

Pure tones (sine waves) are composed of energy of only a single frequency. However, most natural sounds including the human voice or the sound a musical instrument makes are complex in that they are composed of energy at multiple frequencies. The basilar membrane is organized tonotopically with higher frequency vibrations causing maximal vibration at the stiff and narrow basal end, and lower frequencies causing maximal vibration at the wide, apical end. Therefore, a complex sound will stimulate multiple places along the basilar membrane, corresponding to the different frequency components present in the sound. In other words, the basilar membrane performs a sort of Fourier analysis, decomposing the stimulus into its frequency content. Furthermore, in a typical environment, there are several sound sources. The sound waves created by each source sum in the air and reach the ear as one complex wave. Because each sound source typically contributes energy at frequencies across the spectrum in the frequency decomposition accomplished by the basilar membrane, the various sound sources innervate overlapping areas of the membrane. Thus, unlike in the visual system, where a two-dimensional map of receptors in the retina gives information about objects and their relative locations, further stages of processing in the auditory pathway are necessary in order to construct streams or groupings that reflect each individual sound source from the incoming spectrotemporal information (Bregman, 1990).

Auditory Scene Analysis

The auditory systems' ability to organize incoming acoustic information according to the multiple sources that gave rise to it is referred to as Auditory Scene Analysis (ASA). Much of our understanding of auditory scene analysis comes from the pioneering work of Bregman (1990). Sounds rarely occur in isolation in the real world. One of the greatest feats of the auditory system is taking these co-occurring sounds and separating them into representations of distinct perceptual sources. Bregman hypothesized that there were both bottom-up and top-down influences on ASA. He hypothesized that the bottom up processes are innate and pre-attentive and are based on basic acoustic cues. For example, successive sounds that are similar in frequency, pitch or timbre tend to be perceived as coming from the same source, whereas sounds that are dissimilar tend to be perceived as coming from different sources. Similarly, sounds or sound sequences that are said to have *good continuation*, in that they remain constant or change smoothly across time, are also likely to be perceived as coming from one source, as are sounds that occur in rapid succession (Bregman, 1990).

Top down or schema-based segregation and integration, relies on learned properties of the auditory world (Alain, McDonald, Ostroff, and Schneider, 2001). For example, if two classical pieces of music are interleaved with each other we are better at segregating out one of these pieces if we are familiar with it (Dowling, Lung, and Herrbold, 1987; Bey and McAdams, 2003). Similarly, familiarity with a person's name can aid speech understanding in noise (Newman, 2005).

Bottom-up pre-attentive ASA involves two basic types of processes, *sequential* grouping or segregating of elements across time and *simultaneous* integrating or segregating of elements at the same time (Bregman, 1990). For example, with respect to sequential grouping, a person's voice may change in frequency over time, but the successive speech sounds are grouped as one auditory object. With respect to simultaneous grouping, complex tones have an array of frequency components and, thus, if two complex tones are played simultaneously, the auditory system has to determine which frequency components should be integrated into one auditory object, and which are integrated into a separate auditory object. This thesis focuses on bottom-up integration and segregation.

Bregman and Pinker (1978) conducted an experiment that highlights how these two processes function together to analyze incoming acoustic information (see Figure 1). During this experiment three pure tones were played in a continuous cycle, where tone A was the highest in frequency and was presented first, and tone A alternated with tones B and C. Tone B was higher in frequency than tone C but both were presented simultaneously. Depending on the relative frequencies of tone A and tone C, Bregman and Pinker were able to manipulate how participants perceived the sequence of tones. If, for example, they lowered the frequency of tone A to match that of tone B, participants reported hearing one stream consisting of A and B combined, and a separate tone C. Relative onsets also affect ASA. If they adjusted tone C so that its onset was synchronous with tone B, participants reported hearing a pure high tone, tone A, followed by a complex tone B and C integrated into one tone.

This experiment points to the low-level, bottom-up nature of stream segregation and simultaneous integration whereby the auditory system automatically groups acoustic information based on cues such as frequency, intensity, phase, temporal relations and harmonicity (Bregman and Pinker, 1978; Alain, Reinke, He, Wang, and Lobaugh, 2005; Lipp, Kitterick, Summerfield, Bailey, Paul-Jordanov, 2010).

Since the publication of Bregman's (1990) influential book, *Auditory Scene Analysis: The Perceptual Organization of Sound*, there have been a multitude of studies investigating stream segregation and simultaneous integration in both human and non-human species. However, prior to this thesis, there were no studies on the development of the ability of infants to use cues to segregate simultaneous sound sources.

A. *Stream Segregation in Non-human Animals*

Given the evolutionary importance of distinguishing between different sounds in the environment, it is no surprise that many hearing vertebrates demonstrate some ability to segregate and integrate varying acoustic components (Bass and McKibbon, 2003). For survival, animals must be able to hear a predator approaching or identify their prey amongst an array of environmental sounds. As predicted, researchers have found auditory stream segregation abilities in many different species, including bats, fish, bullfrogs, and songbirds and monkeys (Hulse, and MacDougall-Shackelton, 1997; Moss, and Surlykke, 2001; Fishman, Reser, and Steinschneider, 2004). For example, Hulse et al. (1997) examined the ability of European starlings (*Sturnes vulgaris*), a songbird, to discriminate and identify their own song amongst other species' songs that were presented simultaneously. The starlings were initially conditioned to peck for a food reward for the correct identification of a 10 s recording of their own song. This song was then combined with the songs of several other species to determine whether songbirds have the ability to segregate the stream of their own song in order to obtain the food reward. Hulse et al. (1997) found that this avian species was able to segregate its own song from other species' songs after conditioning for food rewards. In addition to this, Macdougall-Shackleton et al. (1998) have shown that songbirds are also able to segregate unfamiliar stimuli such as pure tone sequences differing in frequency after conditioning.

Fishman et al. (2001, 2004) were the first to study the neural correlates of stream segregation in animals through the use of neurophysiological recordings from the primary auditory cortex (A1). This group investigated whether the adult macaque monkey's (*Macaca fascicularis)* primary auditory cortex was activated when presented with an "ABAB" stream segregation condition similar to that studied in humans (Bregman, 1990). In the "ABAB" condition, two tones that alternate (high, low, high, low…) are continuously presented with various frequency separations that elicit either the perception of one or of two streams. Researchers examined the neural impulses evoked by altering tone duration, presentation rate and frequency separation. They hypothesized that the pattern of recordings would reflect differentiation between instances when one stream is perceived versus instances when two streams are perceived. In two separate studies they found that the neural responses paralleled the psychoacoustic data on adult stream segregation. Several other studies have also found similar results when studying the neural correlates of stream segregation in other species such as monkeys and bats (Micheyl, Tian, Carlyon, and Rauschecker, 2005; Kanwal, Medvedev, and Micheyl, 2003).

B. *Simultaneous Integration in Non-human Animals*

Studies on the bullfrog and the midshipman fish have suggested that a range of vertebrates are able to decompose simultaneous sources. Simmons and Bean (2000) studied the bullfrog (*Rana catesbeiana)* in its natural environment. Bullfrogs are known for attracting females with their mating vocalizations, which are very complex and can contain up to 22 harmonics. During breeding periods, males use these multiple-harmonic calls to notify other males of their location and to attract female bullfrogs. It is for this reason that Simmons and Bean hypothesized that bullfrogs would use auditory scene analysis. When tested in their natural environment, male bullfrogs emitted significantly different mating calls to synthetic signals that had one mistuned harmonic compared to synthetic calls that had all harmonics in tune. This provides evidence of sensitivity to harmonic structure in non-human vertebrates.

Several species of fish have communication and mating systems that rely on females recognizing male acoustic calls (Bass and McKibben, 2000). In the midshipman fish (*Porichtys notatus*), males often cohabit in nest-like caves while they wait for females to come and lay their eggs. In order to attract females the males evoke a buzz-like hum that is approximately 100Hz. The assortment of hums from different males at slightly different frequencies in the same area combine to create a collective 'beat' sensation. The female must then decompose the components of the collective auditory signal in order to identify and direct herself to a single male with the hum she desires.

Single-cell recordings suggest that the auditory nerve and the inferior colliculus (midbrain of auditory system) are also sensitive to auditory stimuli that contain one versus two sounds. Both the guinea pig and the chinchilla show evidence of the ability to decompose complex sounds reflected in the firing patterns of neurons involved in the auditory pathway (Palmer, Rees, and Caird, 1990; Sinex, Sabes, and Li, 2002).

Collectively, research investigating stream segregation and simultaneous integration points to the ancient roots and evolutionary importance of the ability to organize incoming acoustic information according to the auditory sources in the environment. For this reason, we expect to see evidence of these pre-attentive, automatic processes of ASA very early in human development.

C. *Stream Segregation in Adults and Children*

Studies of stream segregation in adults date back as far as 1976 (Dannenbring, 1976). Since that time, Bregman and numerous other researchers have focused on what conditions favour stream segregation versus integration. There have also been multiple studies on segregating streams of voices and melodies (Moore, Glasberg, and Peters, 1986; Bey, McAdams, 2002; Ueda, Nakajima, and Akahane-Yamada, 2005). Presentation rate, frequency, and tone duration can all have an effect on stream segregation. For example, larger frequency separation, faster presentation rate, and increasing tone length all favour segregation (Beauvois 1998; Bregman, Ahad, Crum, and O'Reilly 2000).

One very important aspect of stream segregation involves grouping auditory components across time in order to perceive speech. Being able to attend to an individual's voice in an environment where multiple auditory events are present is crucial for learning language and communication, and has been termed the "cocktail party effect" as a party situation often requires picking out and attending to one voice among many. Many studies have investigated the manner in which adult humans separate a stream of speech from noise and, similarly, a melodic line from a piece of music (e.g., Bregman, 1990; Ueda et al., 2005; Marie and Trainor, 2014; Ragert, Fairhurst, and Keller, 2014).

As the current thesis focuses on development, I will focus on previous research involving sequential stream segregation in children and infants. For example, Sussman et al. (2001) used an oddball paradigm to compare behavioural and EEG measures of streaming in adults and children (7-10 years of age) in order to assess whether the pre-attentive processes for stream segregation act in a similar fashion across this age range. They recorded ERPs to infrequent deviant tones using EEG. In the one-stream condition, all tones were played in the same frequency range leading to stream integration. In the two-stream condition, a sequence of repeating tones high in frequency were followed by different (oddball) tones, also high in frequency, while a another sequence of tones played simultaneously at a lower frequency. Participants watched a silent video and were instructed to ignore the sounds. MMNs from the oddball sequence alone were compared to both the one stream and two stream conditions. MMNs were only found in the two-stream condition, indicating participants were not segregating the two streams when the interference tones were present (see Figure 2). In the second experiment, the experimenter asked participants whether they heard two streams or one. Whether or not the specific oddballs were heard was dependent on whether the individual perceived one stream or two. Using an EEG MMN oddball-based paradigm as an index of stream segregation, they found that children and adults demonstrated similar pre-attentive processes for stream segregation on the basis on frequency (see Figure 2; Sussman, Čeponienė, Shestakova, Näätänen, and Winkler 2001).

In a follow-up experiment, Sussman et al. (2007) tested whether the threshold between hearing one versus two streams differed between children aged 5-8 and children aged 9-11 years using the same oddball paradigm as above. The results indicated better performance in older children, suggesting that stream segregation becomes fine-tuned with cortical maturation and/or experience (Sussman, Wong, Horváth, Winkler, and Wang 2007). Using the same frequency-proximity stream segregation experiment as Sussman et al. (2001), Winkler et al. (2003) used EEG to examine the ability to segregate streams in 2-5-day-old infants (see Figure 2). They found that infants showed a mismatch response to the complex oddball stimuli, even when another sequence of tones was presented at a lower frequency. However, the infants did not show a mismatch response when the oddball sequence was interleaved with a sequence of tones in the same frequency range as the oddball. This suggests that the infants were segregating the stream in the two-stream condition but not in the one-stream condition (Winkler, Kushnerenko, Horváth, Čeponienė, Fellmen, Huotilainen et al. 2003). In sum, the array of animal and human research clearly indicates that there is a pre-attentive mechanism for grouping sequential auditory events into coherent streams and objects that is highly conserved across species, although it becomes fine tuned through maturation and/or experience.

D. *Simultaneous Integration in Adults and Children*

The auditory system relies on several low-level cues when initially grouping acoustic information into objects or auditory sources. In humans, the most commonly studied of these is harmonicity, or the relation between the frequencies present in the sound wave. As mentioned previously, natural sounds that give rise to a sensation of pitch, such as those generated by animal vocalizations or musical instruments (Fletcher, 1992), typically contain energy at a fundamental frequency and integer multiples of that frequency ($n*f_0$, where f_0= the fundamental frequency, and n=1, 2, 3..., Fletcher, 1992). For example, a complex sound with a pitch of 200 Hz would typically contain energy at 200 Hz (f_0) and integer multiples of that fundamental (400, 600, 800... Hz). Each of these components is referred to as a harmonic (Yost, 2007). The auditory system uses this fact as a cue for grouping, that is, frequency components in integer multiple relations to a common fundamental frequency are assumed to arise from the same sound source. When one harmonic in a complex sound is mistuned, the auditory system recognizes that it does not fit the harmonic template and it is not grouped with the other frequency components, resulting in the perception of two auditory objects. At very low levels of mistuning, listeners report increased audibility of that harmonic (Rasch, 1978) and the perception of roughness or beating (Hartmann, 1988; Hartmann, McAdams, and Smith, 1990). When a harmonic is sufficiently mistuned, the resulting perception is that of two sounds, one representing the integration of the in-tune harmonics (perceived as a complex tone with pitch at the fundamental frequency), and the second, representing the mistuned harmonic (perceived to have a pitch at the frequency of the mistuned harmonic; Moore, Glasberg, and Peters, 1986; Hartmann, 1988).

Early behavioural studies required participants to match the pitch of the pure sine wave or mistuned harmonic to ensure they were hearing a second auditory object. Researchers noted that the degree of mistuning required to hear a second auditory object was dependent on factors such as the degree of mistuning, the fundamental frequency, the duration of the sound, which harmonic is mistuned, and how many harmonics were in the mistuned complex. For example, there is an increased ability to segregate a mistuned harmonic as the mistuned harmonic number rises, sound duration lengthens, and the degree of mistuning increases (Moore et al., 1985; Moore et al., 1986; Hartmann et al., 1988; Alain et al., 2001a).

Alain and colleagues were the first to combine behavioural and ERP measures to determine the brain's direct response to hearing a complex tone with one harmonic mistuned (Alain et al., 2001a). ERP recordings showed evidence of two components associated with the perception of a mistuned harmonic. The first component, referred to as the object-related negativity or ORN, is seen in both passive and active listening conditions, indicating that it is pre-attentive. This negative deflection in response to the presence of a mistuned harmonic peaks at approximately 160 ms post-stimulus-onset in the fronto-central regions with a reversal in polarity at posterior sites, suggesting a source in auditory cortex. The amplitude of the ORN tends to increase as the salience of the mistuned harmonic increases (Alain et al., 2001; Alain and McDonald, 2007). Although similar in presentation to the MMN, the ORN is not an MMN response as it is present irrespective of stimulus probability, whereas MMN is only present in response to an infrequent deviant auditory event (Lipp et al., 2010). One study found a decrease in the amplitude of the ORN as the probability of hearing a mistuned complex decreased (Alain and Izenburg, 2003), but the ORN is nonetheless present and robust when equal numbers of in-tune (one auditory object) complex tones and complex tones with a mistuned harmonic (two auditory objects) are present.

Although mistuning has been the primary method used to study simultaneous integration, an ORN has been noted in several studies using stimuli that rely on other cues to produce a perception of two auditory objects. Offsetting one harmonic in a complex tone by a small delay results in the perception of two auditory objects and is also associated with an ORN (Lipp et al., 2010). Dichotic sound presentation can also give rise to the perception of two auditory objects. If for example the degree of mistuning is small enough that it does not result in the perception of a second auditory object, but the harmonic is presented in a different location or a small burst of noise is presented to one ear and a stimulus with pitch to the other, the resulting perception will be that of two auditory objects and an ORN will be present (McDonald and Alain, 2005; Johnson et al., 2007). In the same manner, when the onset or offset of a harmonic presented to one ear varies compared to the onset of a complex tone presented to the other ear, an ORN will be elicited (Sanders, Joh, Keen, and Freyman, 2008).

The second ERP component of interest with respect to perceiving two auditory objects is the P400, an attention-dependent component that is only present when participants are required to attend to the task and indicate with a button press whether they heard one sound or two (e.g., an in-tune complex tone or a complex tone with one harmonic mistuned). This positive deflection is seen at approximately 400 ms post-stimulus-onset in adults, and is likely related to top-down influences around making perceptual judgements about the number of auditory objects present (Alain et al., 2001; Alain et al., 2003).

Alain and colleagues have also studied changes with aging associated with discriminating mistuned harmonics (Alain et al., 2001). Previous research noted that the elderly have a particularly difficult time with some aspects of ASA. Congruent with this idea, several studies have that found elderly participants have higher thresholds and a significantly harder time discriminating mistuned harmonics compared to young adults even when controlling for hearing sensitivity (Alain et al., 2002). Additionally, in both EEG and magnetoencephalographic (MEG) studies, the amplitude of the ORN is significantly smaller in the elderly population compared to young and middle-aged adults (Alain et al., 2001; Alain and McDonald, 2007).

Alain et al. (2003) also sought to understand how this process operates in pre-adolescent children aged 8.5-13 years. Researchers employed both behavioural and ERP measures using a 400ms, 200Hz tone with the third harmonic mistuned (Alain, Theunissen, Chevalier, Batty, and Taylor, 2003). Before EEG recordings began, participants were familiarized with the task of pressing a button when they heard two auditory objects. Results suggested that both adults and children were better at identifying the two-object trials when the amount of mistuning was greater. However, children were less sensitive to the mistuning compared to adults, reporting fewer instances of hearing two objects. ERP recordings showed an ORN waveform around 160 ms (peak amplitude range 140-190 ms) post-stimulus-onset, which was slightly later than the adult peak (range 125-175 ms). The ORN amplitude was larger in children than adults, despite the fact that children were behaviourally less sensitive at identifying the inharmonic complexes (Alain et al. 2003).

Motivation for the Current Thesis

In order to segregate a voice or musical melody from background noise, the auditory system must decompose the incoming sound into its spectral components across time, and be able to integrate those components emanating from a common source into a coherent whole. This means that all aspects of auditory scene analysis rely on segregation and integration of simultaneous components. While research investigating streaming of successive sounds has been performed in non-human animals, infants, children and adults, research investigating simultaneous aspects has only been carried out in non-human animals, school-aged children and adults. Despite the significant role that auditory scene analysis plays in development, no research has previously investigated simultaneous integration in children younger than 8.5 years of age (Alain et al., 2003). This thesis provides the first three studies to investigate simultaneous integration in infancy and uses converging evidence from behavioural and ERP measures.

In Chapter 2, we use a conditioned-head turn paradigm to determine if 6-month-old infants are able to discriminate between an in-tune complex tone and a complex tone with one harmonic mistuned. The aim of this chapter is to determine if infants are able to discriminate mistuned harmonics in a similar fashion to older children and adults. In order to discriminate the pitch of complex sounds the auditory system combines the harmonic components to determine the fundamental frequency (Yost, 2007). One piece of evidence suggesting that infants can integrate components of complex tones to create one auditory object comes from research on missing fundamentals. The missing fundamental illusion refers to the fact that adults and infants are able to perceive the pitch of a complex tone even when the fundamental or lowest harmonic is not present. This is because the auditory system derives pitch by compiling or integrating harmonics. Using behavioural measures, Clarkson and Clifton (1985) showed that infants as young as 7 months are able to categorize complex tones even when the fundamental is not physically present. Given that we know that infants can discriminate pitch changes between complex tones we can assume that they, like adults, integrate harmonics to derive pitch (Clarkson, and Clifton, 1985). Using EEG, He and Trainor (2009) showed that 4-month-olds also hear the pitch of the missing fundamental. In Chapter 2, a conditioned head turn method is used to determine infants' thresholds for detecting a mistuned harmonic.

Chapter 3 uses ERP measures to index the neural correlates of perceiving one versus two auditory objects in infancy. Using an experimental design similar to that of the previously mentioned studies by Alain and colleagues (Alain et al., 2001, 2003), infants were presented with in-tune and mistuned complex tones, each on 50% of trials. Given that the ORN is elicited regardless of attention, this experimental paradigm is ideal for use with preverbal infants. A cross-sectional study documented the development of the object-related negativity from 2 to 12 months after birth. Although developmental trajectories of the MMN can vary with task type (i.e., pitch discrimination versus gap discrimination) there is a commonly observed slow positive ERP response to deviant stimuli early in infancy that decreases with age and the development of an earlier adult-like negativity that appears slightly later in development and that increases in amplitude with age (Shibasaki and Miyazaki, 1992; Pang and Taylor, 2000; Ponton et al., 2000; Trainor et al., 2001; 2003; Moore and Linithicum, 2007, Näätänen et al., 2007; Trainor, 2008; He, et al., 2009; Shahin et al., 2010; 2008). We therefore expected that the ORN might look qualitatively different from the adult-like negativity in young infants.

Chapter 4 uses a carefully designed visual preference procedure with infants 4-months of age to further characterize how infants perceive mistuned harmonics as relating to auditory objects. By pairing in-tune and mistuned complex tones with visual stimuli of one or two objects we address whether infants hear a mistuned harmonic as a distinct, second auditory object. We expected infants to look longer at incongruent audiovisual pairings (e.g., one object, two sounds or two objects, on sound) than congruent pairings. This study also investigates whether young infants have expectations linking auditory and visual information about objects in the world, that is, whether when infants hear two objects they also expect to see two objects.

Lastly, Chapter 5 summarizes the results of these three studies and discusses the implications of the work and possible future directions.

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., Izenberg A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15*(7), 1063-1073.

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., Reinke, K., He, Y., Wang, C., and Lobaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience, 17*(5), 811-818.

Alain, C., Shuler, B.M., and McDonald, K.L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America, 111*(2), 990-995.

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., and Braddick, O. (2013). Auditory and musical development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Body and mind* (Volume 1). Oxford: Oxford University Press.

Bass, A., McKibben, J. (2003). Neural mechanisms and behaviors for acoustic communication in teleost fish. *Progress in Neurobiology, 69*(1), 1-26.

Beauvois, M. W. (1998). The effect of tone duration on auditory stream formation. *Perception and Psychophysics, 60*(5), 852-861.

Berlyne, D.E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology, 55*(3), 289-296.

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception and Psychophysics, 64*(5), 844-854.

Bey, C., and McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 267-279.

Blackwood, D.H., Muir, W.J. (1990). Cognitive brain potentials and their application. *The British Journal of Psychiatry, 157*(9), 96-101.

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

Bregman, A. S., Ahad, P. A., Crum, P. A., and O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception and Psychophysics, 62*(3), 626-636.

Bregman, A. S., and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32*(1), 19-31.

Clarkson. M. G., and Clifton, R. K. (1985) Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *Journal of the Acoustical Society of America, 77*(4), 1521-1528.

Dannenbring, G.L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology, 30*(2), 99-114.

Dowling, W.J., Lung, K.M., and Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics, 41*(6), 642-656.

Fantz, R.L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668-670.

Fishman, Y. I., Arezzo, J. C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *Journal of the Acoustical Society of America, 116*(3), 1656-1670.

Fishman, Y. I., Reser, D. H., and Arezzo, J. C. and Steinschneider, M. (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Resolution, 151*, 167-187.

Fletcher, N.H. (1992). *Acoustic systems in biology*. New York: Oxford University Press.

Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P., and Liegeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology, 112*(10), 1850-1859.

Gout, A., Christophe, A., and Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548-567.

Hartmann, W.M. (1988). Pitch perception and the segregation and integration of auditory entities. *Auditory Function, Neurobiological Bases of Hearing,* 623-645.

Hartmann, W.M., McAdams, S., Smith, B.K. (1990). Hearing a mistuned harmonic in and otherwise periodic complex tone. *Journal of the Acoustical Society of America, 88*, 1712-1724.

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

He, C., Hotson, L., Trainor, L.J. (2009). Maturation of cortical mismatch mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia 47,* 218-229.

He, C., Trainor, L.J. (2009). Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience, 29,* 7718-7722.

Hollich, G. (2006). Combining techniques to reveal emergent effects in infants' segmentation, word learning, and grammar. *Language and Speech, 49*(1), 3-19.

Hulse, S. H., MacDougall-Shackleton, S. A. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (Sturnus vulgaris). *Journal of Compartative Pyschology, 111*(1), 3-13.

Johnson, B.W., Hautus, M.J., Duff, D.J., and Clapp, W.C. (2007). Sequential processing of interaural timing differences for sound source segregation and spatial localization: Evidence from event-related cortical potentials. *Psychophysiology, 44*(4), 541-551.

Kanwal, J.S., Medvedev, A.V., and Micheyl, C. (2003). Neurodynamics for auditory stream segregation: Tracking sounds in the mustached bat's natural environment. *Network Computation in Neural Systems, 14*(2003), 413-435.

Kooijman, V., Johnson, E. K., and Cutler, A. (2008). Reflections on reflections of infant word recognition. *Early language development: Bridging brain and behaviour*, 91-114.

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P.J., and Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia, 48*(5), 1417-1425.

Luck, S.J. (2005). *An introduction to the event-related potential technique*. MIT Press, Cambridge MA, 1st Edition.

MacDougall-Shackelton, S. A., Hulse, S. H., Gentner, T. Q., and White, W. (1998) Auditory scene analysis by European starlings (Sturnus vulgaris): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America, 103*(6), 3581-3587.

Marie, C., and Trainor, L. J. (2012). Development of simultaneous pitch encoding: Infants show a high voice superiority effect. *Cerebral Cortex*, bhs050.

McDonald, K.L., and Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related potentials. *Journal of the Acoustical Society of America, 118*(3), 1593-1604.

Micheyl, C., Tian, B., Carlyon, R.P., and Rauschecker, J.P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron, 48*(1), 139-148.

Moore, B.C., Glasberg, B.R., and Peters, R.W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80*, 479.

Moore, B.C., Peters, R.W., and Glasberg, B.R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America, 77*(5), 1861-1867.

Moore, J.K., and Linithicum, F.H. Jr. (2007). The human auditory system: A timeline of development. *International Journal of Audiology, 46*(9), 2007.

Moss, C. F., Surlykke, A. (2001). Auditory scene analysis in bats. *Journal of the Acoustical Society of America, 110*(4), 2207-2226.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544-2590.

Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, *41*(2), 352.

Palmer, A. R., Rees, A., and Caird, D. (1990). Interaural delay sensitivity to tones and broad band signals in the guinea-pig inferior colliculus. *Hearing research*, *50*(1), 71-86.

Pang, E.W., and Taylor, M.J. (2000). The development of the N1 from age 3 to adulthood: an examination of speech and non-speech stimuli. *Clinical Neurophysiology, 111*(3), 388-397.

Picton, T.W., Alain, C., Otten, L., and Ritter, W. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neurotology*, *5*(3-4), 111-139.

Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., … & Trujillo, N.J. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-Otology, 4*, 64-79.

Ponton, C.W., Eggermont, J.J., Kwong, B., and Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Clinical Neurophysiology, 111*(2), 220-236.

Ragert, M., Fairhurst, M.T., and Keller, P.E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS ONE, 9*(1), e84085.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acta Acustica united with Acustica*, *40*(1), 21-33.

Rauschecker, J. P., and Shannon, R. V. (2002). Sending sound to the brain. *Science, 295*(5557), 1025-1029.

Sanders, L.D., Joh, A.S., Keen, R.E., and Freyman, R.L. (2008). One sound or two? Object-related negativity indexes echo perception. *Perception and Psychophysics, 70*(8), 1558-1570.

Shanin, A.J., Roberts, L.E., Chau, W., Trainor, L.J., and Miller, L.M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage, 41*(1), 113-122.

Shahin, A.J., Trainor, L.J., Roberts, L.E., Backer, K.C., and Miller, L.M. (2010). Development of auditory phase-locked activity for music sounds. *Journal of Neurophysiology, 103*(1), 218-229.

Shibasaki, H., and Miyazaki, M. (1992). Event-related potential studies in infants and children. *Journal of Clinical Neurophysiology, 9*(3), 408-418.

Simmons, A. M., and Bean, M. E., (2000). Perception of mistuned harmonics in complex sounds by the bullfrog (Rana catesbeiana). *Journal of Comparative Psychology, 114*(2), 167-173.

Sinex, D.G., Sabes, J.H., and Li, H. (2002). Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hearing Research, 168*(1-2), 150-162.

Sussman, E., Čeponienė, R., Shestakova, A., Näätänen, R., Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hearing Research, 153,* 108-114.

Sussman, E., Wong, R., Horváth, J., Winkler, I., and Wang, W. (2007). The development of perceptual organization of sound by frequency separation in 5-11-year-old children. *Hearing Research, 225*, 117-127.

Trainor, L. J. (2008). Event related potential measures in auditory developmental research. In L. Schmidt and S. Segalowitz (Eds.), *Developmental psychophysiology: Theory, systems and methods* (pp. 69-102). New York: Cambridge University Press.

Trainor, L. J. (2013). Development of pitch perception and the processing of simultaneous sounds in infancy. *Canadian Hearing Report, 8*(4)*,* 35-36.

Trainor, L. J., and He, C. (2013). Auditory and musical development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Body and mind* (Volume 1) (pp. 310-337). Oxford: Oxford University Press.

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., and Sonnadara, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology, 51,* 5-15.

Trainor, L. J., Samuel, S. S., Galay, L., Hevenor, S. J., Desjardins, R. N., and Sonnadara, R. (2001). Measuring temporal resolution in infants using mismatch negativity. *NeuroReport, 12,* 2443-2448.

Ueda, K., Nakajima, Y., and Akahane-Yamada, R. (2005). An artificial environment is often a noisy environment: Auditory scene analysis and speech perception in noise. *Journal of physiological anthropology and applied human science, 24,* 129-133.

Werker, J.F., Polka, L., and Pegg, J.E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting, 6*, 171-178.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., ... and Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, *100*(20), 11812-11815.

Yost, W.A. (2007). Perceiving sounds in the real world: An introduction to human complex sound perception. *Frontiers in Bioscience, 12,* 3461-3467.

Yvert, B., Fischer, C., Bertrand, O., and Montysalo, J. (2005). Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage, 28,* 140-153.

FIGURES

Figure 1: Ambiguous pure tone stimuli used by Bregman and Pinker (1978). Listeners heard a repeating sequence of pure tone A followed by a complex tone containing two pure tone components B & C. Whether A & B were integrated into a single stream separate from C, or whether B & C were integrated into a single stream separate from A depended on the proximity of tones A & B and the synchronization of B & C onset and offsets. From Bregman and Pinker (1978) Figure 1, page 22.

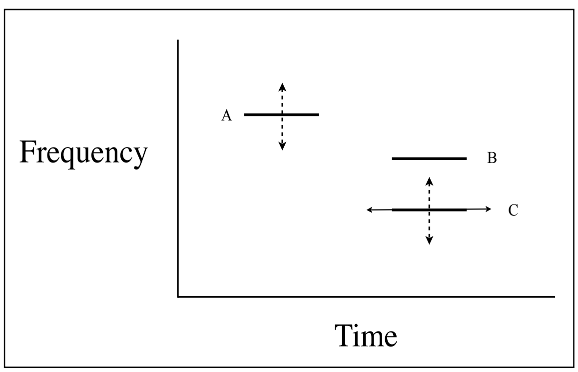
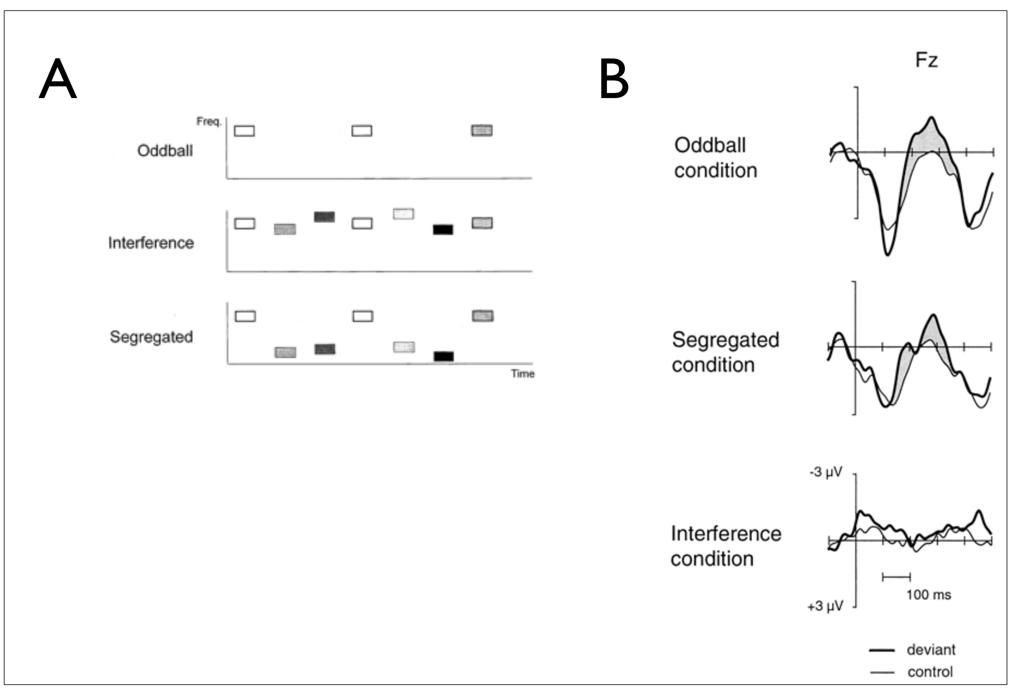


Figure 2: Panel A shows the stimulus paradigm used by Sussman et al. (2001). Children experienced three different auditory conditions. The first *oddball* sequence consisted of a sequence of tones, with 90% being standard tones and 10% oddball tones. The *interference* condition consisted of the same oddball sequence with two additional tones interleaved between each tone close in frequency. The *segregated* condition consisted of the oddball sequence with two additional tones interleaved between each tone, but further away in frequency. Panel B shows ERP results from the three conditions. In the *oddball* and *segregated* conditions an MMN to the oddball stimulus was present indicating children heard two streams in the segregated condition. In the *interference* condition no MMN was present indicating children integrated the tones into one stream and no longer heard the oddball tone. From Sussman et al. (2001), Figure 1 & 2, page 111 & 112.



CHAPTER 2

PREFACE:

Auditory scene analysis, the ability to organize incoming acoustic information according to the multiple auditory sources that gave rise to them, is necessary for language and music development (e.g., Bregman, 1990; Trainor 2015). Harmonic structure or harmonicity is one cue for simultaneous integration (separating two simultaneous auditory objects), but it had not been studied previously in infants. The current chapter investigates whether or not infants are able to discriminate between in-tune (perceived as one auditory object by adults) harmonic complexes and complexes in which one harmonic is mistuned (perceived as two auditory objects) and how their thresholds for detecting a mistuning in one harmonic compare to adults. Adults and 6-month-old infants participated in a behavioural study in which they were required to discriminate in-tune and mistuned complex tones with varying degrees of mistuning.

Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes

Nicole A. Folland*, Blake E. Butler*, Nicholas A. Smith+ & Laurel J. Trainor*

*Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4L8

+Perceptual Development Laboratory, Boys Town National Research Hospital, 555 N. 30th Street, Omaha, Nebraska, 68131

Abbreviated title: Mistuning discrimination in infants

Correspondence:
Laurel Trainor
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
L8S 4B2
LJT@mcmaster.ca

Abstract

The ability to separate simultaneous auditory objects is crucial to infant auditory development. Music in particular relies on the ability to separate musical notes, chords, and melodic lines. Little research addresses how infants process simultaneous sounds. The present study used a conditioned head-turn procedure to examine whether 6-month-old infants are able to discriminate a complex tone (240 Hz, 500 ms, 6 harmonics in random phase with a 6 db rolloff per octave) from a version with the 3rd harmonic mistuned. Adults perceive such stimuli as containing two auditory objects, one with the pitch of the mistuned harmonic and the other with pitch corresponding to the fundamental of the complex tone. Adult thresholds were between 1% and 2% mistuning. Infants performed above chance levels for 8%, 6% and 4% mistunings, with no significant difference between conditions. However, performance was not significantly different from chance for 2% mistuning and significantly worse for 2% compared to all larger mistunings. These results indicate that 6-month-old infants are sensitive to violations of harmonic structure and suggest that they are able to separate two simultaneously sounding objects.

I. INTRODUCTION

Much music contains multiple simultaneous notes whether from different instruments playing simultaneously or one instrument playing several notes at the same time. However, the sound wave that reaches the ear is a composite of the sound energy from each source. Therefore the auditory system must analyze the spectrotemporal properties of the input in order to determine the number and identity of the auditory objects that gave rise to it (Bregman, 1990). This process of parsing out distinct auditory events and sources to create meaningful percepts is known as *auditory scene analysis* (Bregman, 1990). One cue to whether a set of frequency components originated from one source is the harmonic relations between them. Sounds with pitch typically have energy at harmonics that are integer multiples of a fundamental. Reflecting this, the auditory system tends to integrate frequency components standing in integer relations into a single object or percept (Hartmann, 1996; Lin and Hartmann, 1998). Conversely, when a frequency component does not fit with a set of such components it will tend to be heard as a separate auditory object originating from a different source.

In music, analysis of frequency relations is critical for perceiving pitch, as well as for segregating the different musical sounds in polyphonic music, for the perception of harmony, and for perceiving consonance and dissonance. With respect to the last point, recent research indicates that perception of consonance is more highly correlated with whether the harmonics are integer multiples of a fundamental (McDermott, Lehr, and Oxenham, 2010) than whether there is beating or roughness present, as suggested by Plomp and Levelt (1965). Thus, analysis of harmonicity appears to be crucial for pitch perception, for identification of simultaneous auditory objects and for the perception of consonance and dissonance.

Frequency processing in the cochlea appears to be mature at birth (Teas, Klein, and Kramer, 1982; Abdala and Chatterjee, 2003). Psychophysical and auditory brainstem studies of the frequency resolution of spatial encoding mechanisms concur that 3-month-old infants show adult-like processing for frequencies below 4000 Hz, and that by 6 months of age infants show mature processing for all frequencies (Abdala and Folsom, 1995; Folsom, 1985; Folsom and Wynne, 1987; Olsho, 1985; Spetner and Olsho, 1990). Pitch discrimination thresholds for high frequency pure tones (over 4000 Hz) also rely predominantly on spatial encoding mechanisms, and follow a similar developmental time as for frequency resolution (Olsho, Koch, and Halpin, 1987). However, pitch discrimination at lower frequencies relies additionally on temporal mechanisms and remains immature until around 10 or 11 years of age (Maxon and Hochberg, 1982; Werner, 2007). Nonetheless, pitch discrimination at low frequencies is mature enough in early infancy to support the fine discrimination needed for music perception (Trainor and Corrigall, 2010; Trainor and He, 2011). At 2 months of age infants are able to discriminate vowel sounds that differ in frequency by 3% (Swoboda, Morse, and Leavitt, 1976) and show cortical EEG responses to a 6% change in the pitch of piano tones (He, Hotson, and Trainor, 2007). Unpublished work in our lab indicates that by 4 months infants show EEG responses to a 3% change in the pitch of guitar tones and by 5 months can behaviorally discriminate a 2% change in the frequency of pure tones (Olsho, Schoon, Sakai, Turpin, and Sperduto, 1982).

Whether infants integrate frequency components into a single pitch percept is a more difficult question. In adults, when the auditory system is presented with several harmonics of a complex tone in the absence of the fundamental frequency, the pitch of the complex sound can still be perceived even though there is no energy at the fundamental frequency. Using event-related potentials derived from EEG recordings, He, Hotson and Trainor (2009) showed that by 4 months of age infants also perceive the pitch of the missing fundamental.

In the present study we examined infants' sensitivity to violations in harmonicity that lead to the perception of two auditory objects in adults. Previous behavioral studies with adults indicate an improved ability to segregate a mistuned harmonic with increased length of stimulus, increased amount of mistuning, and decreased harmonic number. Young adults (aged 22-24 years) have thresholds that vary between 0.5 and 8% mistuning depending on the stimuli used, while children (aged 8-13 years) and the elderly are less sensitive and report fewer instances of hearing two objects across all amounts of mistuning (Hartmann, McAdams and Smith, 1990; Alain, McDonald, Ostroff, and Schneider, 2001; Alain, Theunissen, Chevalier, Batty, and Taylor, 2003; Alain and McDonald, 2007). Electrophysiological (EEG) and neuromagnetic (MEG) studies in adults and children (aged 8-13) have revealed a neural component that is associated with the perception of two auditory objects and is thus referred to as the object-related negativity (ORN). It has a frontally negative and posteriorly positive topography consistent with activation in auditory areas and is present regardless of stimulus probability (Alain and Schuler, 2002). The ORN occurs in the absence of attention or awareness so it is thought to reflect a bottom-up or low-level process that is largely unconscious, such that the mistuning acts as a preattentive cue that facilitates segregation of harmonically unrelated frequency components (Alain, Arnott, and Picton, 2001). If the segregation of a mistuned harmonic involves low-level processing it might be expected to be present early in development. No research to date has used the mistuning paradigm to address infants' sensitivity to harmonic structure. Here we use a conditioned head-turn procedure (Werker, Polka, and Pegg, 1997) to measure infants' ability to detect a mistuning in the third harmonic of a 6-component tone in comparison to that of adults. The conditioned head-turn procedure is ideal for testing auditory perception in infants around 6 months of age because they possess adequate control over head movements and are entertained by the reinforcers (Werker et al., 1997).

II. METHOD:

A. Participants

Ten adult listeners (4 males, mean age = 24.4 yrs +/- 1.879) with no reported hearing impairments and 24 6-month-old infants (11 males, mean age = 6.28 months +/- 0.198) participated. Five infants were excluded because they did not pass initial training and 1 because of equipment failure. An additional 21 infants were excluded because they were too tired and fussy to complete the second experimental block. All infants were screened at birth to ensure no sensorineural hearing loss according to the Ontario Infant Hearing Program (Hyde, 2005). After giving informed consent to participate, parents of the infants were asked to fill out a brief questionnaire for auditory screening purposes. According to the questionnaire, no infants had a history of frequent ear infections, pressure-equalizing tubes, or hearing impairment in the family and all were healthy at the time of testing.

B. Stimuli

Stimuli were created using Adobe Audition 6.0. The in-tune complex tone was 500 ms in duration including 50 ms rise and fall times, had a pitch of 240 Hz and contained the first 6 harmonics (240, 480, 720, 960, 1200, and 1440 Hz) in random phase with a 6 dB/octave roll off. The mistuned sounds were created by mistuning the 3rd harmonic (720 Hz) of the in-tune complex upward by 1, 2, 4, 6, or 8% (e.g. in the 8% mistuned condition, the 3rd harmonic was equal to 720Hz x 1.08 = 777.6 Hz). Because previous literature has not addressed whether young infants are able to discriminate any amount of mistuning of any harmonic, we chose stimulus parameters that gave rise to a clear, salient perception of two auditory objects in adults (Hartman et al., 1990; Lin and Hartman, 1998; Moore, Glasberg and Peters; 1986).

C. Infant Procedure

Parents sat in the sound attenuated chamber (Industrial Acoustics Company) with their infant sitting on their lap facing the experimenter. Sounds were presented by a

Macintosh G4 computer through an NAD C352 amplifier to a GSI Audiological speaker located to the left of the infant. The experimenter sat behind a small desk that housed several stuffed toys and a button box out of the infant's view. The button box was connected to the computer through a custom-built interface to a National Instruments PCI-DIO96 I/O card. Under the speaker to the left of the infant was a cabinet with four compartments that each housed a mechanical toy and lights used to reward the infant for correct responses. These toys and lights were also controlled by the computer through the custom-built interface. The cabinet had a smoked Plexiglas front such that each mechanical toy was not visible unless the lights in that compartment were illuminated. The parent and experimenter listened to continuous music through headphones that masked the stimuli the infant was hearing in order to ensure that they could not bias the infant's behavior.

The in-tune stimulus repeated continuously with an SOA of 1500ms at approximately 70 dB(A), measured at the location of the infant's head, over a noise floor of 26 dB(A). When the infant was attentive and facing forward (the experimenter attracted the infant's attention with stuffed toys if necessary) the experimenter pressed a button that signalled to the computer to present a trial. On half of the trials, the mistuned complex replaced one instance of the in-tune complex (change trials) and on the other half, the in-tune complex was presented (control trials). The experimenter was not aware of whether a change or control trial was presented or exactly when a trial was presented. The experimenter pressed a second button on the button box whenever the infant turned his or her head at least 45^{o} toward the speaker following the presentation of a trial. The computer kept track of head turns. If the infant turned his or her head within a 2000 ms window post onset of a change trial, the computer rewarded the infant by turning on the lights and a mechanical toy in one of the four compartments of the toy cabinet. Each experimental phase (condition) consisted of 24 trials, 12 control and 12 change trials, presented in quasi-random order such that no more than 3 control trials were presented in a row. Each experimental phase was preceded by a training phase in which the intensity of the mistuned complexes was 6 dB higher [76 dB(A)] than the in-tune complexes in order to teach infants that head turns to a change in stimulus were rewarded with animated toys. During training there were no control trials, and thus only hits and misses were recorded. In order to proceed to the experimental phase, infants were required to successfully turn toward the speaker on four consecutive change trials within 20 trials. If this criterion was not met, testing ended, and the infant's data were not used. During testing, all stimuli were presented at 70 dB(A).

Because young infants can only remain attentive and cooperative for a short time, we were only able to test each infant in two mistuning conditions. All infants were run on the 8% mistuning condition first. If the infant completed the 8% condition and was not fussy, then they began the training for *one* of the 6%, 4% or 2% conditions. Eight infants completed testing in each of the groups (A: 8% and 6%; B: 8% and 4%; and C: 8% and 2%). For each condition, the data of interest were the number of hits (correct turns to a mistuned complex) and false alarms (incorrect turns to an in-tune complex) during the testing phase.

D. Adult Procedure

Adults were tested individually in as similar a manner to infants as possible. Adults were seated in the sound attenuated chamber directly across from the experimenter. The stimulus presentation and procedure were the same as with infants except that adults raised a hand instead of turning to indicate when they heard two objects. Unlike infants, however, each adult was tested in all mistuning conditions (8%, 6%, 4%, 2%, and 1%).

III. RESULTS

Infants.

A *d′* score was calculated for each infant for each condition. Individual d' values were calculated to avoid the underestimation of sensitivity that can result from calculating d' based on average hit and false alarm rates from individuals with similar d' scores, but different response biases (Macmillan and Creelman, 2005, pp. 8-9). A correction was applied to adjust for perfect proportions in which proportions of 1 and 0 (representing infinite d' values) were converted to 1-1/(2N) and 1/(2N), respectively where N is the number of trials on which the proportion is based (Macmillan and Creelman, 2005, pp. 332). All three groups of infants performed similarly in the initial 8% mistuning condition and a one-way ANOVA showed no significant differences across the groups, $F(2,21) = 0.24$, $p = 0.79$. A second one-way ANOVA comparing performance in the 6%, 4%, and 2% mistuning conditions showed a significant effect of condition, $F(2, 21) = 17.34$, $p < 0.001$. Post hoc tests indicated that performance in the 2% mistuning condition was significantly worse than in both the 4% ($p < 0.001$) and 6% ($p < 0.001$) conditions, which did not differ significantly ($p = 0.64$). T-tests on each condition revealed that performance was significantly above chance levels at 8% ($p < 0.001$), 6% ($p = 0.001$) and 4% ($p < 0.001$) mistunings, but not at 2% ($p = 0.603$; Figure 1).

Adults.

Using d' scores as the dependent measure, adults performed significantly above chance at all mistuning levels (all $p < 0.001$). A one-way ANOVA revealed a significant effect of mistuning condition, $F(4, 22) = 10.421$, $p < 0.05$. Tukey's post-hoc tests revealed that performance in the 1% mistuning condition was significantly worse compared to all other conditions (Figure 2).

IV. DISCUSSION

The results indicate that infants are sensitive to violations in harmonic structure at 6 months of age. Specifically, infants are able to discriminate when the 3rd harmonic in a 6-harmonic complex tone is mistuned, with a threshold between 2% and 4%. This threshold is in line with infant pitch discrimination abilities as measured for pure tones (Werner, 2007) and guitar tones (Trainor, Lee, and Bosnyak, in press). Adults tested with a similar procedure showed thresholds below 1% (also in line with reported pitch discrimination abilities; Olsho, 1984), suggesting that sensitivity to harmonic structure improves with age. Previous work indicates that 4-month-old infants can integrate harmonics in order to perceive the pitch of a complex tone even when the fundamental is missing (He et al., 2009). The current study suggests that infants are also able to successfully segregate components that do not fit the harmonic structure of a complex sound, provided that the deviation is large enough. This suggests that infants can use harmonicity cues to detect multiple simultaneous auditory sound sources or auditory objects, although further experiments are needed to verify whether infants actually hear separate objects.

The ability to process the harmonic structure of complex sounds is important for reasons that extend beyond the musical domain. Infants' ability to segregate the auditory scene into auditory objects (or streams) corresponding to sound sources in their auditory worlds is critical for the remarkable language learning that occurs within the first years of life. Although several studies have demonstrated infants' ability to perceptually organize sequential, non-overlapping stimuli (Demany, 1982; McAdams and Bertoncini, 1997; Smith and Trainor, 2011), virtually all real-world auditory environments consist of simultaneous, overlapping sounds (e.g., multiple talkers, or a single talker amid extraneous environmental sounds). For this reason, the present examination of the perception of simultaneous auditory objects, provides a point of departure for increased understanding of how infants make sense of the noisy world into which they are born.

V. REFERENCES

Abdala, C., and Chatterjee, M. (2003). Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans. *The Journal of the Acoustical Society of America*, *114*(2), 932-943.

Abdala, C., and Folsom, R. C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *The Journal of the Acoustical Society of America*, *98*(2), 921-930.

Alain, C., Arnott, S. R., and 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*(5), 1072.

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

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

Alain, C., Schuler, B. M., and McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *The Journal of the Acoustical Society of America*, *111*(2), 990-995.

Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., and Taylor, M. J. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research*, *16*(2), 210-218.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development*, *5*(2), 261-276.

Folsom, R. C. (1985). Auditory brain stem responses from human infants: Pure-tone masking profiles for clicks and filtered clicks. *The Journal of the Acoustical Society of America*, *78*(2), 555-562.

Folsom, R. C., and Wynne, M. K. (1987). Auditory brain stem responses from human adults and infants: Wave V tuning curves. *The Journal of the Acoustical Society of America*, *81*(2), 412-417.

Hartmann, W. M. (1996). Pitch, periodicity, and auditory organization. *The Journal of the Acoustical Society of America*, *100*(6), 3491-3502.

Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *The Journal of the Acoustical Society of America*, *88*(4), 1712-1724.

He, C., Hotson, L., and Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of cognitive neuroscience*, *19*(5), 878-892.

He, C., Hotson, L., and Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: effects of presentation rate and magnitude of change. *Neuropsychologia*, *47*(1), 218-229.

Hyde, M. L. (2005). Newborn hearing screening programs: overview. *Journal of Otolaryngology*, *34*(2), S70.

Lin, J. Y., and Hartmann, W. M. (1998). The pitch of a mistuned harmonic: evidence for a template model. *The Journal of the Acoustical Society of America*, *103*(5), 2608-2617.

Macmillan, N.A., Creelman, C.D. (2005). Detection theory: A user's guide (2nd edition). Mahwah (NJ): Erlbaum.

Maxon, A. B., and Hochberg, I. (1982). Development of psychoacoustic behavior: sensitivity and discrimination. *Ear and Hearing*, *3*(6), 301-308.

McAdams, S., and Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *The Journal of the Acoustical Society of America*, *102*(5), 2945-2953.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*(11), 1035-1041.

Moore, B. C., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, *80*(2), 479-483.

Olsho, L. W. (1984). Infant frequency discrimination. *Infant behavior and development*, *7*(1), 27-35.Olsho, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior and Development*, *8*(4), 371-384.

Olsho, L. W., Koch, E. G., and Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *The Journal of the Acoustical Society of America*, *82*(2), 454-464.

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., and Sperduto, V. (1982). Preliminary data on frequency discrimination in infancy. *The Journal of the Acoustical Society of America*, *71*(2), 509-511.

Plomp, R., and Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *The journal of the Acoustical Society of America*, *38*(4), 548-560.

Smith, N. A., and Trainor, L. J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distracters. *Infancy*, *16*(6), 655-668.

Spetner, N. B., and Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Development*, *61*(3), 632-652.

Swoboda, P. J., Morse, P. A., and Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development*, 459-465.

Teas, D. C., Klein, A. J., and Kramer, S. J. (1982). An analysis of auditory brainstem responses in infants. *Hearing research*, *7*(1), 19-54.

Trainor, L. J., Lee, K., and Bosnyak, D. J. (2011). Cortical plasticity in 4-month-old infants: specific effects of experience with musical timbres. *Brain topography*, *24*(3-4), 192-203.

Trainor, L. J., and Corrigall, K. A. (2010). Music acquisition and effects of musical experience. In *Music perception* (pp. 89-127). Springer New York.

Trainor, L.J., and He, C. (2011). Auditory and musical development. In P. D. Zelazo (Ed.), The Oxford handbook of developmental psychology. Vol. 1: Body and mind. New York: Oxford University Press.

Trainor, L. J., Tsang, C. D., and Cheung, V. H. (2002). Preference for sensory consonance in 2-and 4-month-old infants. *Music Perception*, *20*(2), 187-194.

Werker, J. F., Polka, L., and Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and parenting*, *6*(34), 171-178.

Werner, L.A. (2007). Human Auditory Development. In R. Hoy, P. Dallos, and D. Oertel (Eds.), The Senses: A comprehensive Reference. Volume 3: Audition (pp. 871-874). New York, NY: Academic Press.

FIGURES

Figure 1: Infant discrimination of the mistuned 3rd harmonic. Group A performed equally well at 8% and 6% mistunings. Group B preformed equally well at 8% and 4% mistunings. Group C performed significantly better at 8% than 2% mistunings. Infants were significantly above chance levels at discriminating the 8, 6, and 4% mistunings but not the 2% mistuning. Error bars reflect standard error of the mean.

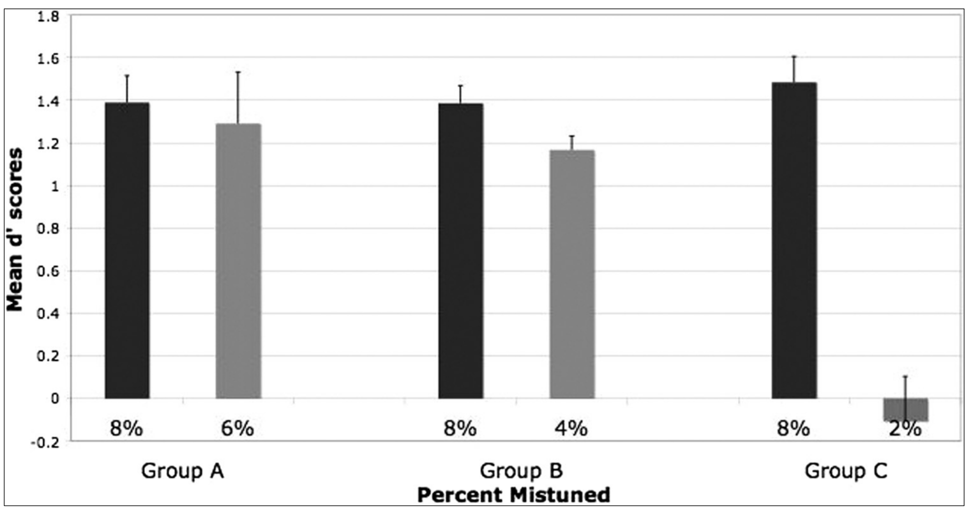
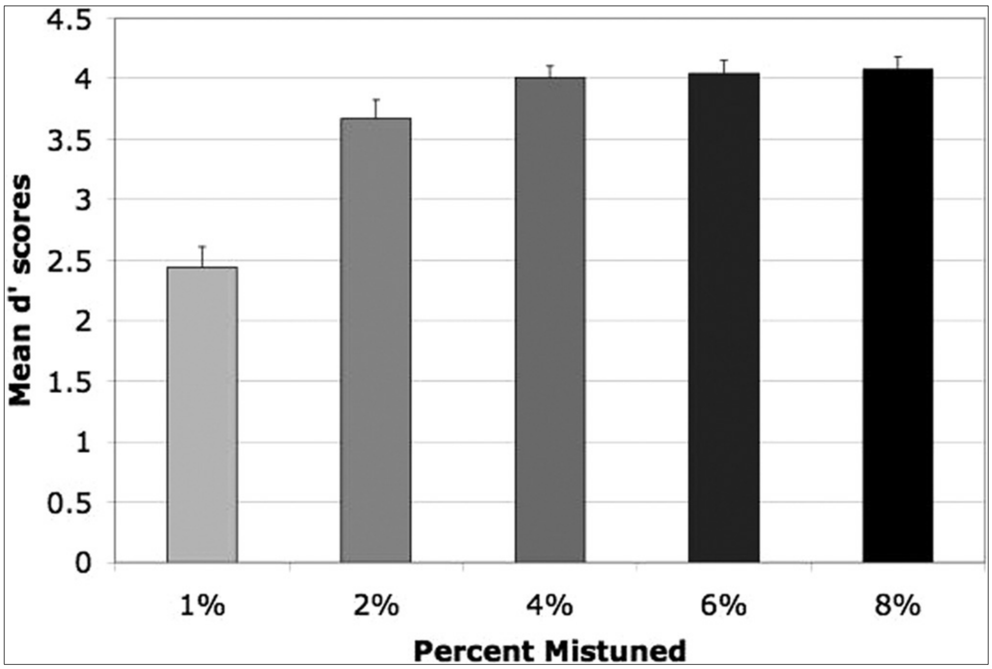


Figure 2: Adult discrimination of the mistuned 3rd harmonic. Adults performed well at all levels of mistuning, although significantly worse at 1% compared to all other levels of mistuning tested. Error bars represent standard error of the mean.



CHAPTER 3

Reproduced with permission from Folland, N.A., Butler, B.E., and 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.Copyright 2015, MIT Press Journals.

PREFACE:

In the previous chapter I demonstrated that 6-month-old infants are able to discriminate between an in-tune complex tone one in which the third harmonic was mistuned by 8, 6, and 4% but not by 2%. This suggests that infants are able to detect mistuned harmonics but their thresholds for hearing the mistunings are higher than those of adult listeners, who easily detected mistunings of 1%. Because no previous studies had used these stimuli with infants it was unknown if they would be able to make simple discriminations between in-tune and mistuned complexes. Thus, chapter 2 was a necessary building block for studies investigating simultaneous integration in infancy using mistuned harmonics.

Having identified that infants are able to discriminate between in-tune and mistuned complexes allowed us to be able to use electrophysiological measures to determine if there is a neural correlate associated with hearing two simultaneous auditory objects. The object-related negativity is a component seen in young children and adults irrespective of stimulus probability that correlates with the perception of a mistuned harmonic as a separate auditory object (Alain et al., 2003). The current chapter investigated whether infants aged 2 to 12 months show the same object-related negativity when presented with sequences of equiprobable in-tune and mistuned complex tones. If infants show an object-related negativity to the mistuned complex tones, this would provide evidence that not only can they discriminate mistuned harmonic complexes, but also that they perceive them in the same manner as children and adults, that is, as two auditory objects, one a higher-pitched pure tone and the other a lower-pitched complex tone).

Cortical representations sensitive to the number of perceived auditory objects emerge between 2 and 4 months of age: Electrophysiological evidence

Nicole Folland[1], Blake E. Butler[1], Jennifer E. Payne[1], & Laurel J. Trainor[1]

[1]Department of Psychology, Neuroscience and Behaviour, McMaster Institute for Music and the Mind, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4L8

Abbreviated title: Harmonic cues for auditory object formation in infancy

Correspondence:
Laurel Trainor
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
L8S 4B2
LJT@mcmaster.ca

Abstract

Sound waves emitted by two or more simultaneous sources reach the ear as one complex waveform. Auditory scene analysis involves parsing a complex waveform into separate perceptual representations of the sound sources (Bregman, 1990). Harmonicity provides an important cue for auditory scene analysis. Normally, harmonics at integer multiples of a fundamental frequency are perceived as one sound with a pitch corresponding to the fundamental frequency. However, when one harmonic in such a complex, pitch-evoking sound is sufficiently mistuned, that harmonic emerges from the complex tone and is perceived as a separate auditory object. Previous work has shown that the percept of two objects is indexed in both children and adults by the object-related negativity (ORN) component of the event-related potential (ERP) derived from EEG recordings (Alain et al., 2001). Here we examine the emergence of object-related responses to an 8% harmonic mistuning in infants between 2 and 12 months of age. Two-month-old infants showed no significant object-related response. However, in 4- 12-month-olds a significant frontally-positive component was present and by 8-12 months, a significant fronto-central object-related negativity, similar to that seen in older children and adults. This is in accordance with previous research demonstrating that infants younger than 4 months of age do not integrate harmonic information to perceive pitch when the fundamental is missing (He et al., 2009). The results indicate that the ability to use harmonic information to segregate simultaneous sounds emerges at the cortical level between 2 and 4 months of age.

I. INTRODUCTION

Many listening environments contain multiple sound-producing objects. These sound waves arrive at the cochlea as a single, complex waveform. The auditory system is tasked with separating this complex waveform into individual representations that correspond to the sound sources. This is referred to as auditory scene analysis (Bregman, 1990), and relies on complex spectrotemporal processing beginning with the pattern of action potentials in auditory nerve fibers (e.g., Ibrahim and Bruce, 2010) and involving processing in multiple brainstem nuclei as well as auditory cortex (e.g., Krishnan, Gandour, and Bidelman, 2012).

Pitch-evoking stimuli are typically comprised of energy at a fundamental frequency, and integer multiples of that frequency, called harmonics. Four-month-old infants integrate harmonics into a single pitch percept in a manner that is qualitatively adult-like (He and Trainor, 2009). Harmonic structure provides one cue for identifying components of a complex signal that should be grouped together into a single sound (i.e., those that fit a harmonic template). Indeed, when one harmonic of a pitch-evoking sound is mistuned, adult listeners report hearing the mistuned harmonic as a second object (Goldstein, 1978; Hartman, 1996; Lin and Hartman, 1998). Behaviorally, young adults detect mistunings of less than 1% in a single harmonic (Alain, Arnott, and Picton, 2001; Folland, Butler, Smith, and Trainor, 2012). However, children (aged 8-13 years; Alain, Theunissen, Chevalier, Batty, and Taylor, 2003) and the elderly (Alain, Mcdonald, Ostroff, and Schneider, 2001; Alain and McDonald, 2007) are less able to use mistuning to determine the number of sound sources, reporting fewer instances of hearing out a mistuned harmonic across all degrees of mistuning.

While auditory scene analysis is essential for the development of auditory function, language acquisition, and musical processing, little research has addressed its development. Moreover, the majority of existing literature focuses on the segregation of sequentially presented stimuli (e.g., Demany, 1982; McAdams and Bertoncini 1997; Sussman, Čeponienė, Shestakova, Näätänen, and Winkler, 2001), with less research addressing the important issue of segregating simultaneously presented sounds. Folland et al. (2012) demonstrated behaviourally that 6-month-old infants are able to discriminate 4% but not 2% mistunings in the 3rd harmonic of a 6-harmonic complex sound. However, it remains unclear whether infants perceive the mistuned harmonic as a separate auditory object in the way that adults do, or whether their discrimination is based on other perceptual features (e.g., differences in timbre).

The present study uses the object-related negativity (ORN) component of the event-related potential (ERP) derived from EEG recordings to determine whether infants' behavioural discrimination of mistuned harmonics reflects the formation of separate auditory objects, and seeks to determine at what age harmonicity is employed as a cue for auditory scene analysis. In adults, the ORN component is observed when two simultaneous objects are presented, but not when one auditory object is presented (Alain et al., 2001). If, like adults, infants perceive the mistuned harmonic to be a second auditory object, an ORN should be elicited in response to a mistuning exceeding their perceptual threshold. If, on the other hand, infants' discrimination of mistunings does not lead to object formation, no ORN response should be produced.

II. METHOD:

A. Participants

Full-term infants between 2 and 12 months of age (n = 153) participated in the study. Five age groups, each consisting of 16 infants and one age group consisting of 32 infants formed the final sample (see Table 1 for details). An additional 41 infants were not included in the analyses (13 because they fell asleep, 10 because they had an insufficient number of artifact-free trials, and 18 due to fussiness). No parent reported that his or her infant had a history of frequent ear infections, pressure-equalizing tubes, or familial hearing impairment. All infants weighed more than 2500g at birth and were born between 38 and 42 weeks gestation. All infants were screened at birth to ensure no sensorinerual hearing loss according to the Ontario Infant Hearing Program (Hyde, 2005). Prior to testing, parents gave informed written consent, and were asked to fill out a brief questionnaire outlining music and language experience. All experimental procedures were approved by the McMaster Research Ethics Board.

B. Stimuli

Stimuli were synthesized using Adobe Audition (6.0), with a sample rate of 44,100 Hz and 16-bit resolution. The in-tune complex tone had a pitch of 240 Hz and contained the first 6 harmonics (240, 480, 720, 960, 1200, and 1440 Hz) added in random phase with a 6 dB/octave roll off. The mistuned complex was identical to the in-tune complex except that the 3rd harmonic was mistuned upward by 8% (777.6 Hz as opposed to 720 Hz). Both stimuli were 500 ms in duration, including 50 ms rise and fall times.

In adults, several factors affect the ease of segregation, including the frequency of the complex tone, which harmonic is mistuned, the sound duration, and the degree of inharmonicity. Because no previous study has elicited an ORN in response to a mistuned harmonic in infants, we chose stimulus parameters that gave rise to a clear, salient perception of two auditory objects in adults (Hartman, McAdams, and Smith, 1990; Lin and Hartman, 1998; Moore, Glasberg, and Peters, 1986), and which 6-month-old infants have been shown to clearly discriminate behaviorally (Folland et al., 2012).

C. Procedure

The infant was seated on his or her parent's lap in a sound attenuating room. Stimuli were presented at 70dB(A) with a background noise level of 30 dB(A) through a WestSun loudspeaker (WestSun Jason Sound, JSIP63, Mississauga, ON, Canada) located 1 meter in front of the infant. Stimuli were presented using E-Prime 1.1 software (Psychology Software Tools) from a Dell OptiPlex280 computer with an Audigy 2 platinum sound card (Creative Labs, Singapore). Infants were tested in a passive listening condition in which they watched a silent movie (Baby Einstein) and were entertained by the experimenter using silent toys and puppets in order to minimize movement. Each infant heard a total of 500 trials (250 in-tune complex tones, and 250 mistuned complex tones) lasting 13 minutes, presented in quasi-random order, with the constraint that no more than 8 of the same tone were played in a row. The interstimulus interval was randomized between 800 ms and 1000 ms.

D. Recording and Analysis

All infants were fitted with a 124-electrode Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene, OR, USA) and electrode impedances were maintained below 50 kΩ. EEG data were recorded continuously using NetStation 4.3.1 software. Electrodes were referenced to the vertex (Cz), and a 0.1-400 Hz hardware filter was used during recording. The data were filtered offline between 0.5 Hz and 20 Hz and were resampled at 200 Hz. Artifacts such as those arising from head movements, eye blinks and ocular movement were blocked using a procedure described previously (Mourad, Reilly, de Bruin, Hasey and MacCrimmon, 2007; Fujioka, Mourad, He, and Trainor, 2011). Data were then segmented into 700 ms epochs including a 100 ms pre-stimulus baseline. Responses to the in-tune and mistuned complexes were averaged separately, and re-referenced using an average reference. Finally, difference waves for each infant were created by subtracting the response to the in-tune complex tone from the response to the mistuned complex tone. Grand averages of the in-tune, mistuned, and difference waveforms were created by averaging across all of the infants in each age group. For statistical analysis, 90 electrodes were selected and divided into 5 regions for each hemisphere, representing frontal, central, parietal, occipital and temporal scalp regions (see Figure 1). Within each region the waveforms were averaged across electrodes. Peripheral electrodes were excluded in order to reduce the impact of muscle artifacts from the eyes, face and neck. Midline electrodes were excluded to allow for comparisons between hemispheres. If the mistuning of the third harmonic evoked an object-related response, a significant difference would be expected between the mistuned (two objects) and in-tune (one object) waveforms.

III. RESULTS

The morphology of the ERP waveforms changed dramatically between 2 and 12 months of age. To reflect this, infants were divided into three groups: 2-month-olds; 4- to 6-month-olds; and 8- to 12-month-olds. This produced groups with similar waveform morphology across participants (Figures 2 to 4).

In-tune (one auditory object), mistuned (two auditory objects) and difference waveforms for 2-month-old infants are shown in Figure 2. In-tune and mistuned waveforms are dominated by a frontally-positive slow wave, which reverses at posterior sites, consistent with a source in auditory cortex. To assess whether in-tune and mistuned responses were significantly different in the absence of a clear peak, the waveform for each 2-month-old was divided into 100 ms segments, and the mean amplitude across each segment was measured. One-sample t-tests on the mean amplitude in each electrode region revealed that the response did not differ significantly between in-tune and out-of-tune responses for any 100 ms segment. After correcting for multiple comparisons, all p values were <0.05.

For the older two age groups, two components were seen in the difference waves, one a frontally-negative component peaking around 230 ms and the other a frontally-positive component peaking around 350 ms, both accompanied by reversals at occipital sites (Figures 3 and 4). The first of these corresponds to the ORN component seen in adults in response to the presence of two auditory objects. The second likely reflects an immature response to the presence of two objects, as discussed below. The amplitude of each component was measured for each infant by taking the average amplitude across a 50 ms time window centered at the time of each peak in that age group's grand average difference waveform. Separate one-sample t-tests were performed for each component in each age group to determine whether the mean amplitude was significantly different from baseline in any electrode region. These components invert in polarity between frontal and occipital electrode sites, resulting in near zero recordings at parietal electrodes. Thus, the parietal electrode region was not included in statistical analyses.

The early negativity did not approach significance in any electrode region in the 4- to 6-month-old infants (all $ps>0.5$), but the late positive component was significant in the left frontal ($t[31]=5.77$, $p<0.001$), right frontal ($t[31]=3.30$, $p=0.008$), and left central ($t[31]=3.59$, $p=0.004$) electrode regions, after correcting for multiple comparisons.

The 8- to 12-month-old group showed a larger early negative deflection than the 4- to 6-month-old group, particularly in the frontal and central regions of the right hemisphere (Figure 4). The average amplitude in the right frontal area failed to reach significance after adjusting for multiple comparisons (adjusted $p=0.19$) but there was a trend for significance at the right central region ($t[47]=2.49$, adjusted $p=0.06$). The late positive component reached significance at the frontal (right: $t[47]=5.49$, adjusted $p<0.001$; left: $t[47]=5.94$, adjusted $p<0.001$), central (right: $t[47]=3.59$, adjusted $p<0.001$; left: $t[47]=5.83$, adjusted $p<0.001$), and occipital (right: $t[47]=4.41$, adjusted $p<0.001$; left: $t[47]=4.87$, adjusted $p<0.001$) electrode regions bilaterally, and in the right temporal region ($t[47]=3.42$, adjusted $p<0.001$).

Repeated-measures ANOVAs were run separately for the early negative and late positive waveform deflections with electrode region (frontal, central, occipital, temporal) and hemisphere as within-subject factors, and age group (4-6 months, 8-12 months) as a between-subjects factor. The polarities of recordings from temporal and occipital electrode regions were inverted prior to statistical analysis so that region effects would reflect differences in waveform amplitude rather than differences arising from the inversion of polarity common to responses from auditory cortex.

The early negative deflection showed a significant region by age-group interaction ($F[3,234]=4.69$, $p=0.003$). Post-hoc independent sample t-tests on each region revealed that the amplitude of the early negativity was greater in 8- to 12-month olds than 4- to 6-month-olds at frontal ($t[78]=2.41$, $p=0.02$) and occipital ($t[78]=2.47$, $p=0.02$) electrode regions. No other significant effects were observed for this deflection.

The late positive deflection showed no main effect of group. However there was a significant effect of electrode region ($F[3,234]=34.67$, $p=<0.001$) with the amplitude in frontal regions exceeding the amplitude in the central regions ($t[79]=4.00$, $p<0.001$), while the frontal, central, and occipital regions each exceeded the amplitude in temporal regions (all ps <0.001). Additionally, response amplitudes in the left hemisphere exceeded those recorded in the right hemisphere ($F[1,78]=5.11$, $p=0.03$).

IV. DISCUSSION

The present study infers a cortical representation of concurrent auditory stream segregation in infant listeners. Using an experimental paradigm designed to elicit an ORN component in response to the perception of two simultaneous auditory objects, we demonstrated that by 4 months of age, infants, like adults, are able to detect a mistuned harmonic within a complex tone based on its emergence as a separate auditory object. The ORN is thought to reflect a bottom-up process that is largely preconscious, with the segregation of frequency components that do not fit the harmonic structure of a complex sound occurring preattentively (Alain, Arnott, et al., 2001; Alain and Izenberg, 2003; Synder, Alain, and Picton, 2006). Finding an object-related response in young infants is consistent with the idea that perceptual organization occurs preattentively.

Two-month-old infants showed no object-related response to the mistuned harmonic stimulus. It is possible that a larger mistuning might elicit a response from these listeners or that their neural generators are oriented in such a way that we cannot see their activity. However, this is unlikely, as previous studies have shown that infants are easily able to perceive changes in frequency of this magnitude both with behavioural (e.g., Olsho, Schoon, Sakai, Turpin, and Sperduto, 1982) and EEG (e.g., He, Hotson, and Trainor, 2009) methods. The failure of 2-month-olds to perceive two auditory objects on the basis of harmonic cues at the cortical level is particularly interesting in that a previous study examining infants' responses to the pitch of the missing fundamental concluded that infants are first able integrate harmonic information into a single pitch percept at 4 months of age (He and Trainor, 2009). Thus, integrating harmonics into a single object with pitch, and using mistuning to separate harmonics into different objects, appear to emerge at the same time in development.

The object-related responses recorded from 4- to 6-month-old infants were dominated by a late, frontally-positive deflection. The object-related responses recorded from 8- to 12-month-old infants also contained this positive component, but contained a shorter latency frontally-negative deflection as well, resembling the ORN recorded in older children and adults. This age-dependent pattern of frontally-positive responses in younger infants and the later emergence of shorter latency, adult-like frontally negative responses has been demonstrated previously for other neural components, including the mismatch negativity (MMN) response (e.g., He, Hotson, and Trainor, 2007, 2009; Tew, Fujioka, He, and Trainor, 2009). This pattern of responses across age is likely related to maturational changes affecting synaptic connections occurring at this time (Trainor, 2008; 2012).

To maximize the number of artifact-free infant trials we chose to use one in-tune stimulus and one mistuned stimulus. Consequently, we cannot say with certainty whether the object-related infant response seen here reflects the detection of mistuning independent of concurrent sound perception. While this study is an important first step towards understanding simultaneous sound perception in infancy, we encourage future studies to manipulate stimulus properties previously shown to affect thresholds for detecting mistuning in adults, such as sound duration and the degree of mistuning (Alain et al., 2001).

It should also be noted that because the in-tune and mistuned stimuli were presented with equal probability in the present study, neither a mismatch negativity (MMN) component, nor any responses related to activation of non-refractory neurons would have been elicited. The frontally negative object-related response in the 8- to 12-month old infants resembled the ORN recorded previously in older children and adults (Alain et al., 2003) although it was at a longer latency in the infants (peak latency of 270 ms) than in 8.5- to 12.5-year-old children (160 ms; Alain et al., 2003). This latency difference is likely due to maturational factors, although it is possible that the fact that our stimuli had longer rise times (50 ms) than those used by Alain and colleagues (10 ms) may have contributed to the latency difference (Alain et al., 2001; 2003). Previous research has demonstrated that although the mechanisms involved in complex pitch perception first appear functional around 4 months of age (He and Trainor, 2009), even simple frequency discrimination does not reach adult levels until 10 years of age (Jensen and Neff, 1993; Thompson, Cranford, and Hoyer, 1999). Thus, it is likely that as the perception of pitch improves, so too does the ability to separate auditory objects based on how well they fit harmonic structures, and that this improvement is indexed by a decrease in the latency of the ORN component.

The present study indicates that the ability to separate simultaneous sound sources on the basis of harmonic cues emerges between 2 and 4 months of age, the same time frame during which cortical responses emerge that reflect integration of harmonics into a pitch percept. These abilities relate to auditory scene analysis and likely contribute to linguistic, musical and social development during this time period.

V. REFERENCES

Alain, C., Arnott, S.T., and 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., and Izenberg A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15*(7), 1063-1073.

Alain, C., McDonald, K.L., Ostroff, J.M., and 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., and Taylor, M. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research, 16,* 210-218.

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

Bregman, A.S. (1990). Auditory scene analysis: the perceptual organization of sounds. London, UK: MIT Press.

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

Folland, N.A., Butler, B.E., Smith, N.A., and 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.

Fujioka, T., Mourad, N., He, C., and Trainor, L.J. (2011). Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clinical Neurophysiology, 122,* 43-51.

Goldstein, J.L. (1978). Mechanisms of signal analysis and pattern perception in periodicity pitch. *Audiology 17,* 421-445.

Hartmann, W.M. (1996). Pitch periodicity, and auditory organization. *Journal of the Acoustical Society of America, 103,* 2608-2617.

Hartmann, W.M., McAdams, S., and Smith, B.K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America, 88,* 1712-1724.

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

He, C., Hotson, L., and Trainor, L.J. (2009). Maturation of cortical mismatch mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia 47,* 218-229.

He, C., and Trainor, L.J. (2009). Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience, 29,* 7718-7722.

Hyde, M.L. (2005). Newborn hearing screening programs: Overview. *Journal of Otolaryngology, 34,* S70-S78.

Ibrahim, R.A., and Bruce, I.C. (2010). Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In: Lopez-

Poveda, E.A., Palmer, A.R., Meddis, R., (Eds.), The Neurophysiological Bases of Auditory Perception. Springer, New York. pp. 429–438.

Jensen, J.K., and Neff, D.L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science 4*(2), 104-107.

Krishnan, A., Gandour, J.T., and Bidelman, G.M. (2012). Experience-dependent plasticity in pitch encoding: From brainstem to auditory cortex. *Neuroreport, 23,* 498-502.

Lin, J-Y., and Hartmann, W.M. (1998). The pitch of a mistuned harmonic: Evidence for a template model. *Journal of the Acoustical Society of America, 103,* 2608-2617.

McAdams, S., and Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America, 102*(5), 2945-2953.

Moore, B.C.J., Glasberg, B.R., and Peters, R.W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80,* 479-483.

Mourad, N., Reilly, J.P., de Bruin, H., Hasey, G., and MacCrimmon, D. (2007). A simple and fast algorithm for automatic suppression of high-amplitude artifacts in EEG data. *Acoustics, Speech and Signal Processing, 1,* I393-I396.

Olsho, L.W., Schoon, C., Sakai, R., Turpin, R., and Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology, 18*(5), 721-726.

Sussman, E., Čeponienė, R., Shestakova, A., and Näätänen, R., Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hearing Research, 153,* 108-114.

Synder, J.S., Alain, C., and Picton, T.W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18*(1), 1-13.

Tew, S., Fujioka, T., He, C., and Trainor, L. (2009). Neural representation of transposed melody in infants at 6 months of age. *Annals of the New York Academy of Sciences, 1169,* 287-290.

Thompson, N.C., Cranford, J.L., and Hoyer, E. (1999). Brief-tone frequency discrimination by children. *Journal of Speech, Language, and Hearing Research, 42,* 1061-1068.

Trainor, L.J. (2008). Event related potential measures in auditory developmental research. In: Developmental psychophysiology: theory, systems and methods (Schmidt L & Segalowitz S, eds), pp69-102. New York: Cambridge UP.

Trainor, L.J. (2012). Musical experience, plasticity, and maturation: Issues in measuring developmental change using EEG and MEG. *Annals of the New York Academy of Sciences 1252,* 25-36.

FIGURES

Figure 1: Electrode groupings (see Methods section for details). Ninety of 124 electrodes were divided into 5 regions (frontal, central, parietal, occipital and temporal) for each hemisphere. Each region contained between 16 and 20 electrodes that were averaged to represent EEG responses from that scalp region. The remaining channels around the perimeter of the net were excluded from analysis to avoid artifacts resulting from muscle activity in the face and neck, and channels along the midline were removed to allow for comparison between hemispheres.

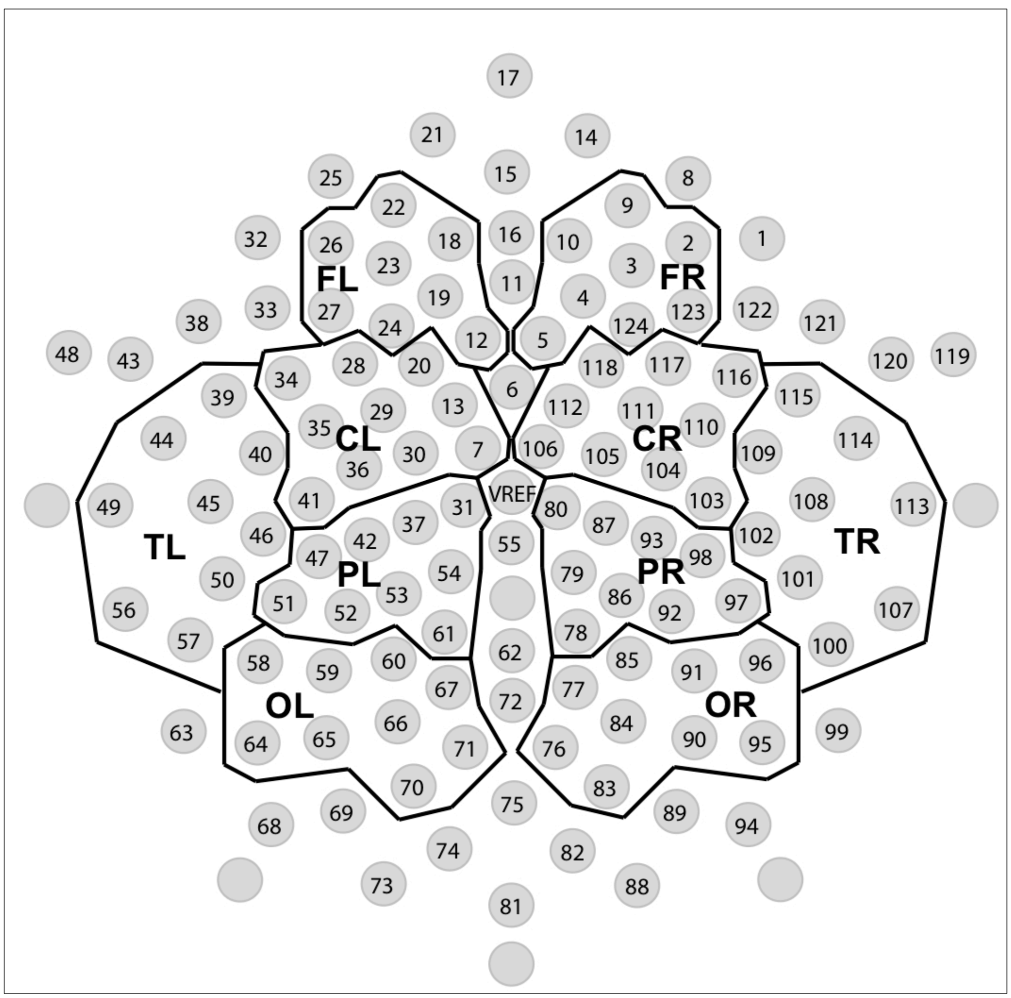


Figure 2: Two-month-olds – Panel A shows the grand average responses of 2-month-old infants to the in-tune (black) and mistuned (red) tones. Panel B shows the grand average difference waveforms (mistuned – in-tune). In each case, responses from each of the 10 electrode regions are presented.

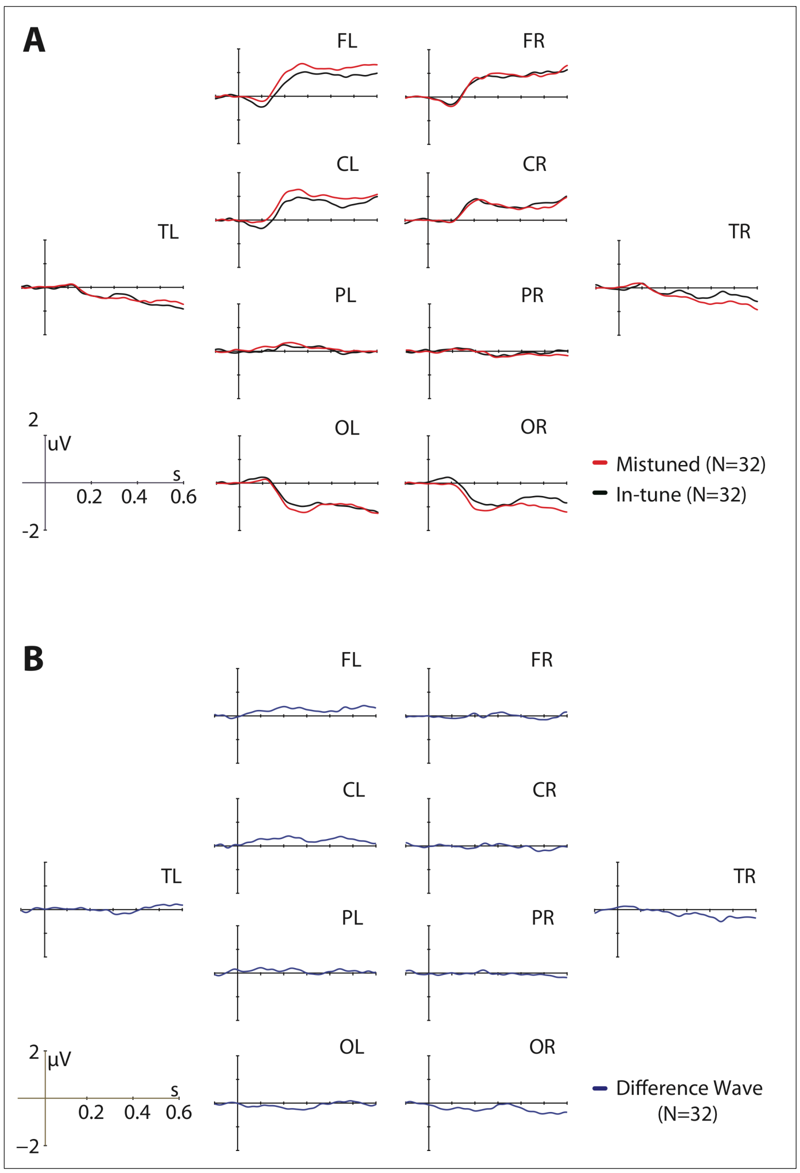


Figure 3: Four- to six-month-olds – Panel A shows the grand average responses of 4- to 6-month-old infants to the in-tune (black) and mistuned (red) tones. Panel B shows the grand average difference waveforms (mistuned – in-tune). In each case, responses from each of the 10 electrode regions are presented.

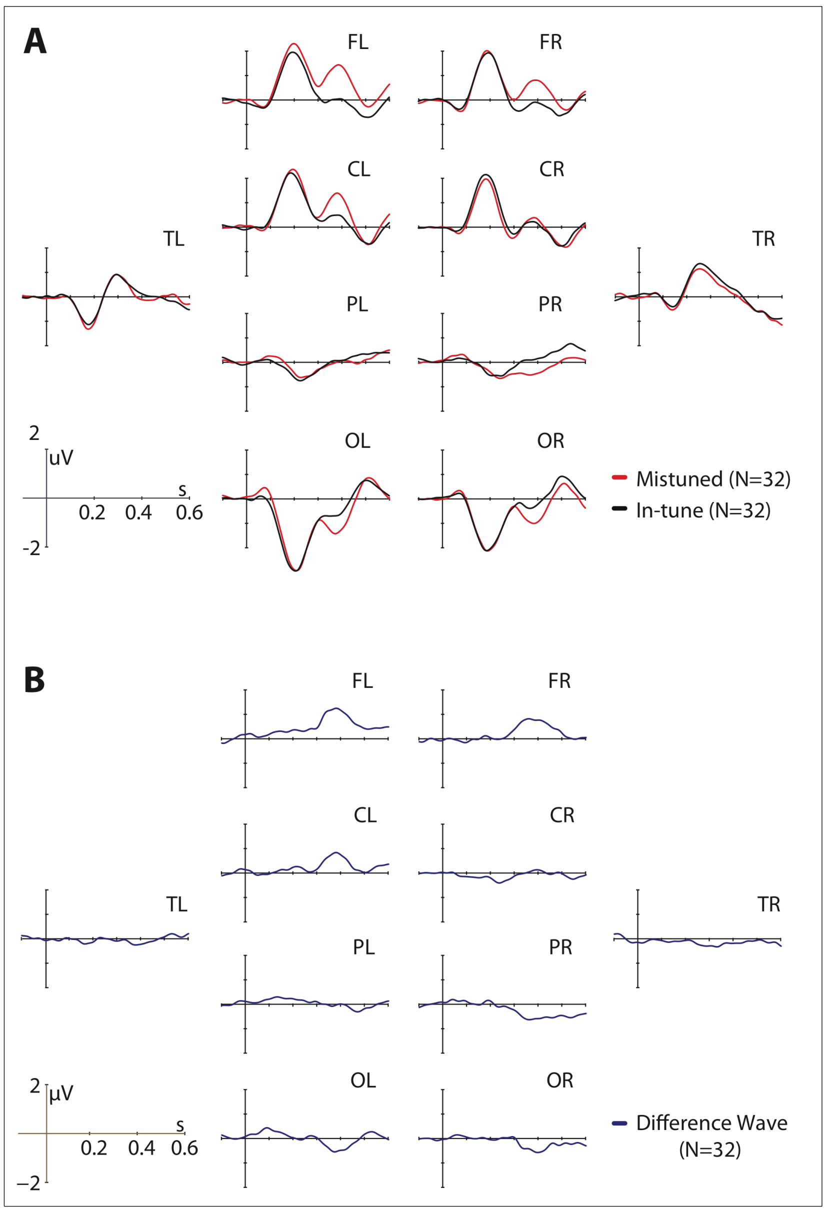
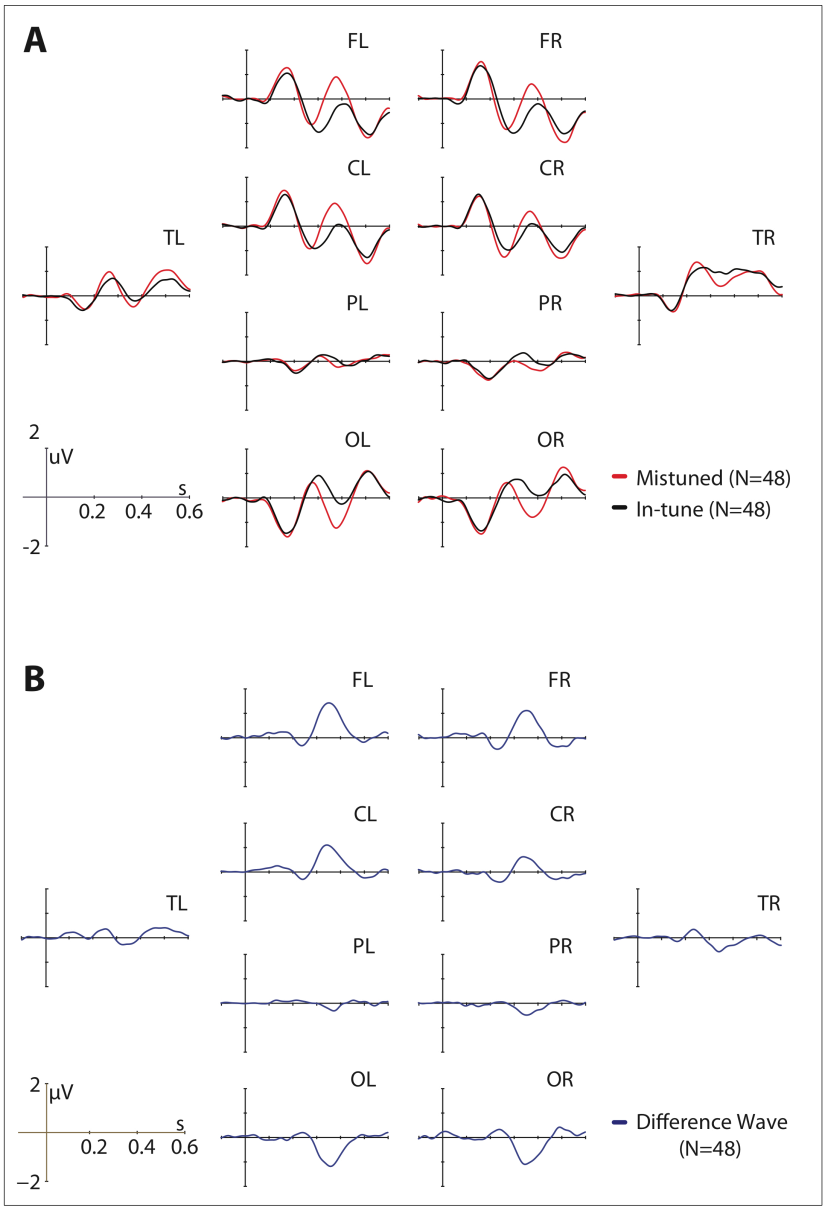


Figure 4: Eight- to twelve-month-olds – Panel A shows the grand average responses of 8- to 12-month-old infants to the in-tune (black) and mistuned (red) tones. Panel B shows the grand average difference waveforms (mistuned – in-tune). In each case, responses from each of the 10 electrode regions are presented.



TABLES

Table 1: Infant age groups

	# of infants	Mean age (SD)	Analysis group
2m	n=32	2.3 (+/- 0.18)	A
4m	n=16	4.5 (+/- 0.26)	B
6m	n=16	6.3 (+/- 0.22)	B
8m	n=16	8.2 (+/- 0.29)	C
10m	n=16	10.2 (+/- 0.15)	C
12m	n=16	12.1 (+/- 0.14)	C

CHAPTER 4

Folland, N.A., Smith, N.A., Martinez, D.M., and Trainor, L.J. (submitted). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition*.

PREFACE:

Chapter 3 provided electrophysiological evidence of the development of an object-related response in infants aged 4-months and older. Given that Chapter 2 investigated 6-month-old infants but Chapter 3 found evidence of an object-related response as early as 4-months, the current Chapter investigates whether 4-month-old infants also discriminate behaviourally between in-tune and mistuned complex tones.

In a carefully designed visual preference study, 4-month-old infants were presented with congruent and incongruent audiovisual pairings of one or two auditory objects (in-tune or mistuned complex tone) and one or two visual objects (one or two bouncing balls). Chapter 4 takes our investigation of simultaneous integration in infancy one-step further and asks whether or not 4-month-old infants have expectations that the number of auditory objects they are hearing should match the number of visual objects they are seeing. If infants hear the mistuned harmonic as a second auditory object, then we would expect to see significant differences in looking time between the congruent and incongruent stimuli because hearing one sound but seeing two objects would be odd or surprising. Chapter 4 is the final study in this series that aimed to identify how early infants use harmonic cues to determine the number of auditory objects in an environment.

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

Nicole, A. Folland[1], Nicholas, A. Smith[2], Diana M. Martinez[1] & Laurel J. Trainor[1,3,4]

[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4L8

[2]Perceptual Development Laboratory, Boys Town National Research Hospital, 555 N. 30th Street, Omaha, Nebraska, 68131

[3]McMaster Institute for Music and the Mind, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 3L8

[4]Rotman Research Institute, Baycrest, University of Toronto, 3560 Bathurst Street, Toronto, Ontario, Canada, M6A 2E1

Abbreviated title: Harmonic cues to object perception

Correspondence:
Laurel Trainor
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
L8S 4B2
LJT@mcmaster.ca

ABSTRACT

Infants learn to use auditory and visual information in order to organize the sensory world into identifiable objects with particular locations. 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 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 et al., 2003). Previous work by our group demonstrated that 4-month-old infants are also sensitive to this cue. They 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 investigated whether infants actually perceive a complex tone with an 8% mistuned harmonic as emanating from two objects, rather than merely detecting the mistuned cue. We paired an in-tune complex tone with a visual display that either contained one or two bouncing balls. As well, we paired a complex tone with a mistuned harmonic with each of these visual displays. 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 integrate auditory and visual information in object perception, and use harmonicity as a cue for source separation.

I. 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 abstract 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. The present study examines the relation between auditory and visual object perception.

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 & Braddick, 2013). 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; Sireteanu & Rieth, 1992; Kaufmann-Hayoz, Kaufmann, & Stucki, 1986; Curran, Braddick, Atkinson, Wattam-Bell, & Andrew, 1999; Ghim, 1990; Kavsek & Yonas, 2006; Otsuka & Yamaguchi, 2003; 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; Wilcox, Woods, Tuggy, & Napoli, 2006). While researchers continue to answer important questions about 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 listeners' ear. Auditory scene analysis refers to the auditory systems' ability to organize incoming acoustic information by unmixing or segregating the signal into streams or auditory objects that likely correspond to the multiple sound sources that gave rise to the complex sound wave reaching the ear (Bregman, 1990). Natural sounds that induce a sensation of pitch, such as the human voice, many other animal vocalizations, or a 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 (i.e., 1:2:3:4…) of that fundamental. For example, a complex tone with a perceived pitch of 200 Hz likely contains energy at 200, 400, 600, 800 Hz and so on. Although the 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, 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 (Plomp, 1976; Eggermont, 2001; McDermott & Oxenham, 2008). *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 easily integrated. 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 et al., 2001, 2003, 2007).

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 complex tone with a 240 Hz fundamental.

In the second study, we used electroencephalography (EEG) to study this question. Alain and colleagues identified a pre-attentive neural correlate of the perception of two auditory objects in adults, referred to as the object-related negativity or ORN. The

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 perceived, irrespective of stimulus probability (Alain et al., 2001). 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.

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 (Folland, Butler, Payne, & Trainor, 2015). Event-related responses to stimulus change are 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.

The current study had two goals. The first was to extend our previous work (Folland et al., 2012) to younger infants to determine if 4-month-olds are able to use harmonicity as a cue to stream segregation. To do this we were required to shift to a visual preference procedure that is more suitable to children of this age. This shift is advantageous because it allowed us to pursue a second goal of examining the relation between infants' auditory and visual object perception in the context of a multisensory audiovisual scene.

Visual preference techniques have a long history in the developmental object perception literature and provide an ideal paradigm to study pre-verbal 4-month-olds (Bower, 1974; Spelke, 1985). The present study used the violation of expectation paradigm (Fantz, 1964) with visual 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 sounded when the balls hit the floor, to determine if infants perceive two auditory objects. We expected that if infants perceived two auditory objects, they would look longer at a visual display with one bouncing ball compared to two, as this would be unexpected and surprising. If, on the other hand, the mistuned harmonic does not lead to the perception of two objects, we do not expect infants to look longer at the visual stimulus with one bouncing ball. This paradigm therefore allows us to examine how object perception in one modality influences object perception in another.

II. METHOD:

A. 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 the 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.

B. Stimuli

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 5 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 and were presented on a 23–inch Apple Cinema HD display. 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^{o}. There were two visual test stimuli. The first consisted of a single 3.8 cm (visual angle 4.4^{o}) dark grey bouncing ball and the second consisted of both the 3.8 inch dark grey bouncing ball and a 1.3 cm (visual angle 1.5^{o}) white bouncing ball (see Supplementary Material for videos of the stimuli). 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. To adults, the large ball was perceived to produce the complex tone with low fundamental frequency and the smaller white ball was perceived to produce the high-pitched mistuned harmonic. The visual test stimuli were presented on the left or right side of the screen over a neutral green background.

From these two audio and two visual stimuli (see Figure 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).

C. 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 infants' 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 same Macintosh G4 computer. 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 spend 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 at incongruent trials on 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 at incongruent trials on which there were two bouncing balls.

III. 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 primary bias in their responses. We therefore collapsed the data across these factors, and created a within-subjects factor, Congruency, with two levels: total time spent looking at the congruent stimuli (either 2 auditory objects/2 balls or 1 auditory object/1 ball) and total looking time at incongruent trials (either 2 auditory object/1 ball or 1 auditory object/2 balls). An ANOVA with Auditory Condition (in-tune; mistuned) as a between-subjects factor and Congruency as a repeated measure revealed only a significant effect of congruency such that infants in both conditions spent more time looking at the incongruent trials, $F(1, 62) = 436.8$, $p < 0.001$. Figure 2 shows that infants who heard the in-tune tone (one auditory object) spent longer looking at the display with two balls, and infants who heard the mistuned tone (two auditory objects) spent longer looking at the display with one ball. Together these results provide evidence that infants do in fact hear a mistuned harmonic as a second auditory object.

IV. DISCUSSION

The current study determined 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., stimulus 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., Werker & Tees, 1999; Kuhl, 1998; Johnson, Westrek, Nazzi, & Cutler, 2011; Friendly, Rendall, & Trainor, 2013 Hannon & Trainor, 2007). 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 infants were able to detect mistunings as small as 4% in the third harmonic of a 6-harmonic complex, but we were not able to show sensitivity to smaller mistunings (Folland et al., 2012). In the present study, we employed an easily detectable 8% mistuning. It would be interesting 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 strong support that the object-related response in 4-month-old infants (Folland et al., 2012), though different in morphology, 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.

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. The importance of the ability to combine information from multiple sensory modalities for communicative purposes has been acknowledged as far back as Darwin (1872). However, how early infants are able to integrate information across sensory domains still remains unclear (for a review see Bahrick, 2010). Research on language development shows matching between visual and auditory modalities in infancy, although various abilities appear to emerge at different ages. There is evidence that infants as young as two months are able to match faces and voices (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003; Walton & Bower, 1993), but cross-modal matching of female speakers articulating a monologue, may not emerge until 12 months of age (Lewkowicz et al., 2014). Infants can match affect between voices and facial expressions at 5 month, but appear not to be able to do so at younger ages (Vaillant-Molina, Bahrick, & Flom, 2013). Some studies have found evidence in 4-month-old infants who were able to spontaneously match shapes to vowel-consonant pairs such as 'kiki' but not to vowels or consonants alone (Ozturk, Krehm, & Vouloumanos, 2013). Studies such as these suggest that experience with speech and faces may play an important role in the development of multisensory coherence (Lewkowicz, Minar, Tift, & Brandon, 2014).

At the same time, 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 look significantly longer at stimuli that pair higher frequencies with bright objects and lower frequencies with dark objects (Haryu & Kajikawa, 2012), and infants as young as 3- to 4-

months show evidence of cross-modal matching of congruent ascending or descending auditory stimuli and spatial elevation and object width and pitch (Dolscheid, Hunnius, Casasanto, & Majid, 2014).

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 effect of audiovisual 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 et al. (2014) found that 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., Peterhans & von der Heydt, 1991; Kaufmann-Hayoz et al., 1986; Johnson 2004). Existing studies of auditory scene analysis in infancy have largely focused on sequential cues (e.g., Demany, 1992; McAdams & Bertoncini, 1997; Winkler, Kushnerenko, Horváth, Čeponienė, Fellman, Huotilainen et al., 2003; Smith & Trainor, 2011).

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. The present study reflects an attempt to understand the development of multisensory object perception, showing that auditory perceptual organization guides infants visual processing of related stimuli. Future work using bistable or multistable stimuli (Cook & Van Valkerberg, 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 infancy.

V. 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.

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.

Bregman, A.S. (1990). Auditory scene analysis: the perceptual organization of sounds. London, UK: MIT Press.

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

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.

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

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

Folland, N.A., Butler, B.E., Payne, J.E. & Trainer, 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.

Fantz, R.L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668-670.

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

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.

Kaufmann-Hayoz, R., Kaufmann, F., & Stucki, M. (1986). Kinectic 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.

Lewkowicz, D.J., Minar, N.J., Tift, A.H., & Brandon, M. (2014). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Pscyhology, 130*, 147-162.

Peterhans, E., & von der Heydt, R. (1991). Subjective contours-bridging the gap between psychophysics and physiology. *Trends in neurosciences*, *14*(3), 112-119.

O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, *35*, 565–569.

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.

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. & 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 infacy. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 323-364). Norwood, NJ: Ablex.

Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*(7), 310–318.

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young Infants Match Facial and Vocal Emotional Expressions of Other Infants. *Infancy*, *18*(s1), E97-E111.

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.

Yonas, A., Gentile, D.A., & Condry, K. (1991). Infant perception of illusory contours in apparent motion displays. *Bulletin of the Psychonomic Society, 28*, 480.

FIGURE CAPTIONS

Figure 1. Audiovisual stimuli: Left: Visual representation of the one-ball audiovisual stimulus. Congruent sound pairing with the in-tune (one ball, one auditory object) 240 Hz complex tone. Incongruent sound pairing with the mistuned (3rd harmonic mistuned; one ball, two auditory objects) 240 Hz Complex tone. Right: Visual representation of the two-ball audiovisual stimulus. Congruent sound pairing with the mistuned (3rd harmonic mistuned; two balls, two auditory objects) 240 Hz complex tone. Incongruent sound pairing with the in-tune (two balls, one auditory objects) 240 Hz complex tone.

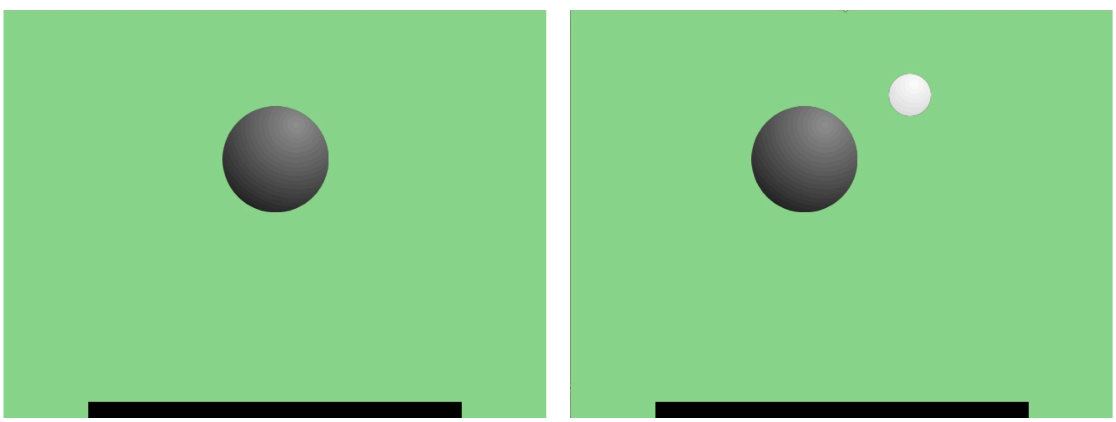
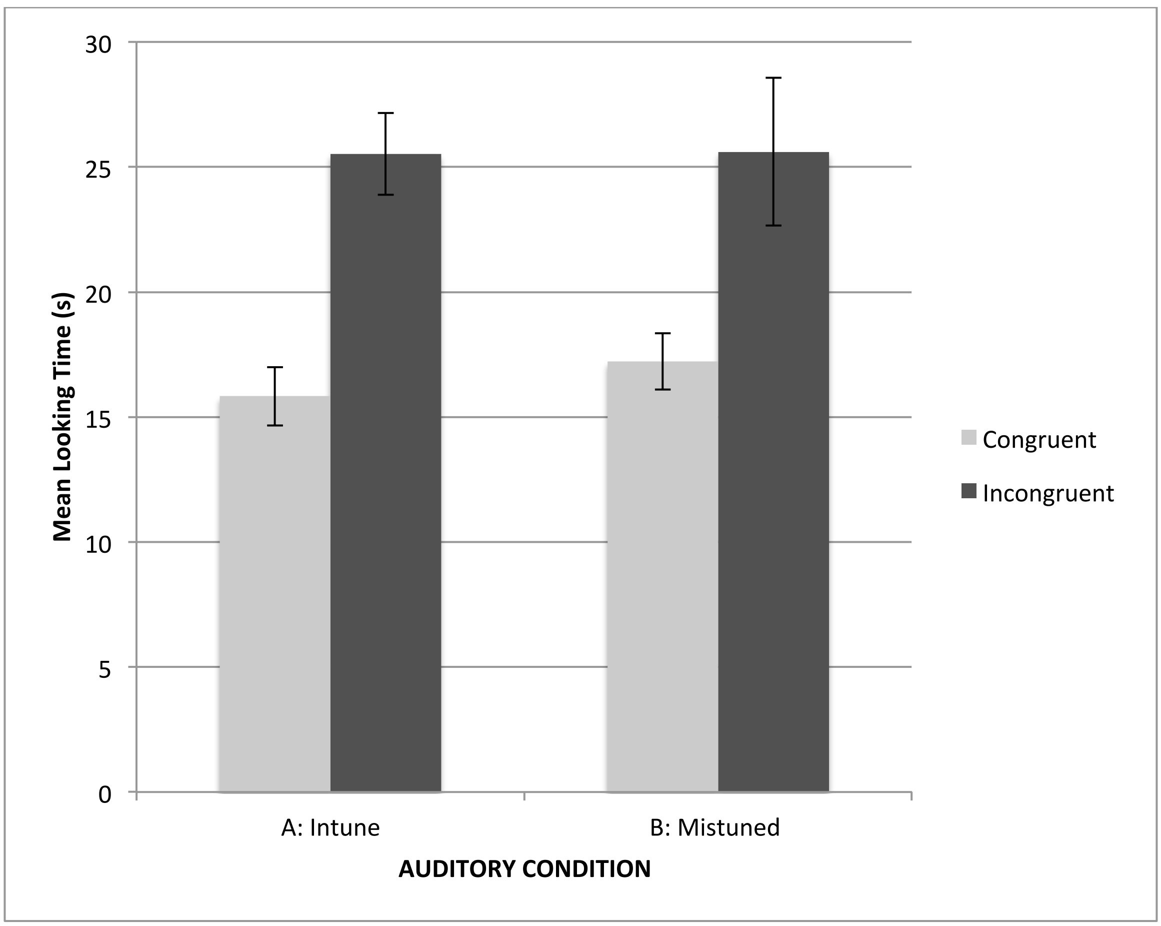


Figure 2. Mean infant looking time: Average looking time to congruent and incongruent trials by auditory condition. Infants in the In-tune Condition (one auditory object) looked significantly longer at the incongruent visual pairing (two balls). Infants in the Mistuned Condition (two auditory objects) looked significantly longer at the incongruent visual pairing (one ball).



CHAPTER 5

Thesis Findings and Unique Contributions

Across Chapters 2, 3 and 4 I used a combination of behavioural and electrophysiological measures to investigate if and when infants are able to use harmonicity cues to separate two concurrent auditory objects. This ability, referred to as simultaneous integration, had previously only been investigated in school-aged children, young adults, and the elderly (Alain et al., 2001a, 2003). Together, these three studies greatly increase our understanding of simultaneous integration in infancy. This ability was demonstrated in two behavioural and one electrophysiological study using mistuned harmonic stimuli. Most natural sounds including the human voice and the sounds of musical instruments are complex sounds composed of energy at integer multiples of the lowest, or fundamental, frequency. When one harmonic in a complex tone is sufficiently mistuned, adults and school-aged children report hearing that harmonic as a separate auditory object. How infants process such stimuli was previously unknown.

In Chapter 2, a behavioural measure indicated that 6-month-old infants were able to discriminate between a 240 Hz complex tone that was in-tune and a 240 Hz complex tone that had one harmonic mistuned by 8, 6, or 4%, but they were not able to perform the task with a 2% mistuning. This study was an important first step as it demonstrated that 6-month-old infants are sensitive to mistunings of a harmonic in a complex tone, although their thresholds for detecting such mistunings were higher than those of adults. However, although this provided evidence to support the notion that infants are able to separate two concurrent auditory stimuli, it did not directly address whether infants actually perceive the mistuned harmonic as a separate auditory object.

Following on the results from Chapter 2, Chapter 3 used event-related potential measures (ERPs) to track the development of the neural correlate referred to as the object-related negativity, or ORN, across the first year after birth. The ORN is a fronto-central negative deflection that reverses in polarity at occipital and mastoid sites and peaks around 140-180 ms post stimulus onset. It is thought to reflect the perception of two concurrent auditory objects. The ORN is seen in school-aged children and adults on trials associated with the perception of two auditory objects, but not on trials where one object is perceived. The EEG paradigm employed in Chapter 3 was similar in nature to previous work investigating the ORN in adults and school-aged children (Alain et al., 2001, 2003). Infants aged 2-12 months were presented with two equally probable stimuli, one an in-tune 240 Hz complex tone and one a 240 Hz complex tone with the third harmonic mistuned by 8%. If infants were hearing the mistuned harmonic as a separate sound, we expected to see an object-related response to the mistuned stimulus but not the in-tune stimulus. In the youngest participants, aged 2 months, there was no evidence of any object related response. In infants aged 4-6 months, we identified an object-related response associated with the mistuned trials and not with the in-tune trials. Although the morphology of the waveform was frontally-positive and delayed in latency compared to previous ORN studies in adults and children, similar infant-adult difference in waveform morphology are seen for other ERP components such as the MMN. We therefore interpret this result as providing evidence that by 4-months infants perceive a mistuned harmonic as separate from the rest of a complex tone. Infants aged 8-12 months also showed this positive object-related response in addition to a negative deflection resembling the ORN seen in adults. This study was the first to provide evidence of a cortical representation of concurrent auditory stream segregation in infancy. It not only increases our understanding of the development of the ORN response but also increases our understanding of the development of infant cortical EEG responses in general. By testing infants at two-month intervals, we are able to look globally at developmental trends. Other commonly studied infant cortical responses, such as the MMN, show a similar developmental trajectory whereby younger infants show early frontally-positive responses and older infants show a more adult-like, earlier negativity (He et al., 2009; He, Hotson, and Trainor 2007; Tew, Fuijoka, He, and Trainor, 2009). This study of 2- to 12-month-old infant cortical responses provides additional evidence that as the infant brain matures and cortical connections are strengthened, cortical responses change from slower, frontally-positive waveforms to earlier, fronto-central negativities.

Chapter 4 provided new insight into how young infants process mistuned harmonics as well as how they relate this information to the visual world around them. Four-month-old infants participated in an audiovisual preference procedure that paired bouncing ball visual stimuli with the same in-tune and mistuned complex tones from Chapter 3. The experiment was designed in such a way that if infants truly perceived a mistuned harmonic as a separate, discrete auditory object, they would find one of two visual displays incongruent or surprising when paired with a sound stimulus. In line with results from Chapter 3, 4-month-old infants showed behavioural evidence of simultaneous integration. Infants looked longer at audiovisual displays that were incongruent; for example, two sounds (mistuned complex tone) paired with only one bouncing ball or one sound (in-tune complex tone) paired with two bouncing balls. While this paralleled our previous study, it also provided evidence that by 4-months infants have expectations that the number of visual and auditory objects in the sensory world should match.

From birth, infants experience an auditory environment that is dynamic and changes across time. Infants are often confronted with situations where there are more than one auditory object or event occurring simultaneously. Acoustic components or sound waves from each of these auditory events often overlap in frequency and reach the infant ear at the same point in time (Trainor, 2015). In order to make sense of which components belong to which auditory event, infants must be able to segregate acoustic components that emanate from a different auditory object while integrating acoustic components that come from the same auditory object. In adults and children, harmonic structure is one pre-attentive cue used to separate simultaneous auditory objects. Chapters 2 to 4 are the first in the auditory scene analysis literature to investigate whether infants also use harmonic structure as a cue for auditory scene analysis.

In this thesis, I studied whether infants are able to segregate a mistuned harmonic as a separate auditory object while integrating the remaining harmonics in the complex tone into a single auditory object, what this might sound like, and if infants are able to apply information gained from auditory scene analysis to the visual scene. Results from Chapters 2 to 4 indicate that infants are able to organize incoming acoustic information when there are two simultaneous auditory objects. Results from these studies indicate that by 4 months, infants segregate a mistuned harmonic as a distinct auditory object and that they integrate the remaining harmonics into a single complex tone with the perception of pitch. Being able to separate simultaneous complex sounds on the basis of harmonic structure has clear implications for language and music development as most natural sounds such as the human voice and the sound an instrument makes are complex and contain multiple harmonics. Infants must be able to separate acoustic information from multiple auditory events that overlap in frequency in order to hear a voice as distinct from background noise or to perceive the melody in a piece of music. In addition, many perceptual features of sound itself rely on this ability to separate simultaneous sounds and overlapping frequency components such as pitch, loudness, timbre and location (Bregman, 1990).

Bregman referred to two different aspects of auditory scene analysis, bottom-up or preattentive processing and top-down or schema-based processing. The type of auditory scene analysis seen in the studies within this thesis provide evidence for bottom-up preattentive processing, which is automatic and occurs irrespective of familiarity with the sounds. Although this aspect of auditory scene analysis has been commonly assumed to be innate (Bregman, 1990; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Alain et al., 2001). Chapters 2 and 3 are the first to provide evidence of preattentive simultaneous integration in infancy. These studies provide evidence that shortly after birth, infants are able to automatically organize incoming acoustic information into a coherent representation of the auditory objects in the environment as opposed to a collection of unrelated harmonics.

Collectively these three studies make a meaningful contribution to the auditory scene analysis literature. They are the first to demonstrate that infants in the first year after birth are able to process two concurrent auditory stimuli, the first to identify an object-related neural response in infancy, and the first to show that infants not only perceive two concurrent auditory objects, but they are able to link how many auditory sources they are hearing to how many visual objects they are seeing on the basis of harmonicity cues.

Limitations and Future Directions

One difficulty with using mistuned harmonic stimuli is that the perception of one versus two auditory objects is highly dependent on the properties of the mistuned complex tone. For example, larger degrees of mistuning and longer stimulus length lead to an increased ability to detect mistunings (Hartmann et al., 1988; Bregman, 1990). In Chapter 2, we noted that 6-month-old infants were unable to detect 2% mistunings in a 240 Hz complex tone, and in Chapter 3 we noted that 2-month-old infants showed no object-related response to an 8% mistuning in a 240 Hz complex tone. However it is possible that different stimulus properties such as a longer stimulus length or a larger degree of mistuning could have changed infants' sensitivity. To determine whether or not 2-month-olds are sensitive to any amounts of mistuning, one could use a similar EEG paradigm as in Chapter 3 to determine if a larger degree of mistuning could elicit an object-related response in 2-month-olds, for example, a mistuning of 10 or 20%. Additionally, it would be useful to test 2-month-olds in the visual preference procedure featured in Chapter 4 to determine if these young infants show behavioural sensitivity to mistuned harmonics, despite not showing a cortical object related response.

Although rewarding, research involving young preverbal infants comes with it own set of challenges. One key limitation to this work, in particular, is that based on the age of the infants we examined and their attention span, we were not able to test any infants using both our behavioural and electrophysiological measures. Additionally, Chapter 3 investigating the development of the ORN across the first year was a cross-sectional representation of development. To properly study the trajectory of this neural correlate across time it would be desirable to study the same infants from birth to 12 months in a longitudinal research design.

Another limiting factor to this research is the specificity of using mistuned harmonics alone to understand simultaneous integration as a whole. The auditory system makes use of many cues to simultaneous integration and segregation, including but not limited to harmonicity, location, and onset-synchrony (McDonald and Alain, 2005; Lipp et al., 2010). To truly understand the development of auditory object formation it would be necessary to investigate whether the same age-related trends apply to other cues. For example, when one harmonic in a mistuned complex tone is slightly offset in time compared to the other harmonics, this also creates the perception of two auditory objects. Although these offset harmonic stimuli would be problematic in an electrophysiological study that relied on stimulus onset, they do provide an additional behavioural stimulus that could be substituted for the mistuned harmonics in Chapter 2 and 3. Additionally, infants rarely hear two sounds in isolation; in real world environments there is a constant interaction between simultaneous integration (segregating two or more sounds occurring at the same time) and stream segregation (segregating two or more sounds across time) (Bregman, 1990).

While the work in this thesis focuses solely on mistuned harmonic stimuli, these three studies provide a useful framework for future experiments to investigate specific questions such as: what are infant listeners' thresholds for detecting mistuned harmonics in silence and in background noise? does musical experience make infant listeners' more sensitive behaviourally and/or in terms of neural responses? and how does abnormal hearing development affect infants' ability to segregate a mistuned harmonic?

Implications and Final Conclusions

It is difficult to imagine what our daily lives would be like without auditory scene analysis. Our perception of the sensory world relies on the auditory system's ability to organize incoming acoustic information into a meaningful percept. Linguistic, social and musical development could not occur without the ability to segregate the auditory environment. In this thesis, I examined at what age infants are able to separate two simultaneous auditory objects, an ability that had previously not been investigated in children younger than 8 years of age. Not only does this work provide new insights into typical auditory development, it also provides a means to ask new questions about atypical auditory development.

Through looking at cases where auditory scene analysis is not functioning optimally, we can begin to understand just how critical this ability truly is. Autism, developmental dyslexia, and specific language impairments (SLIs) are all associated with problems in auditory processing (Lodhia, Brock, Johnson, and Hautus, 2014; Rabinowicz, Silipo, Goldman, and Javitt, 2000; Sutter, Petkov, Baynes, and O'Connor, 2000; Benasich, and Tallal, 2002). In the first two years after birth there are few signs that indicate whether a child will have future language impairments or autism. This is one reason why studying the ability of infants to organize their perceptual world as well as understanding the developmental time course for auditory processing capabilities is crucial. Approximately 7.4% of children entering kindergarten each year are classified as having a SLI (Tomblin, Records, Buckwalter, Zhang, Smith, and O'Brien, 1997), and Autism Spectrum Disorder (ASD) is reported in approximately 1 out of every 165 children (Benasich, and Tallal, 2002; Autism Society Canada).

Many autistic individuals show difficulties extracting speech or relevant auditory information from background noise (Boatman et al., 2001; Alcantara et al., 2004; Teder-Salejarvi et al., 2005; Groen et al., 2009). The processing required to separate two co-occurring sounds such as a human voice from the background music and noise at a party relies on the auditory system's ability to perform simultaneous integration. In a recent study by Lodhia et al. (2014), as well as an unpublished study by our group, teenagers and young adults with autism were found to have a significantly smaller ORN compared to age- and verbal-IQ-matched controls. Together these two studies suggest that individuals with autism have difficulties with simultaneous segregation. There is increasing interest in determining how early autism can be identified and what signs of autism are present in preverbal infants. The understanding gained from this thesis concerning the normal development of auditory scene analysis in infancy therefore has potentially important implications for early diagnosis of autism.

Individuals with developmental dyslexia are also faced with problems related to auditory scene analysis. One of the problems that many individuals with developmental dyslexia face is the inability to process rapidly changing auditory events or sounds (Christmann, Lachmann, and Steinbrink, 2015; Sussman, SteinSchneider, Lee, and Lawson, 2015; Dole, Meunier, and Hoen, 2014). When tested on their ability to segregate streams on the basis of presentation rate, researchers found that adults with dyslexia segregate streams at a slower rate than controls. This deficit in auditory processing is likely linked to the vast array of difficulties dyslexia can cause, such as discriminating phonemes, determining phoneme order, and the acquisition of literacy (Helenius, Uutela, and Hari, 1999). In a comprehensive longitudinal study, Benasich and Tallal (2002) found that the single best indicator of language abilities at the age of three was the child's ability to process rapidly presented complex tones. A variety of controls and two different language measures were employed in infants with and without a family history of SLI. Results showed that the ability to process rapid complex sounds was a better predictor of language ability then family history. Similar to autism, early identification and remediation are key for developmental dyslexia and specific language impairments. Given that our methods are infant-friendly, provide both behavioural and cortical insight, and can be employed at least as early as 2-months, this could provide an exciting means to investigate early auditory processing abilities in typically developing infants and infants classified as at risk for dyslexia, SLI, or autism.

In summary, by 4-months, infants are able to behaviourally discriminate a mistuned harmonic in a complex tone and show a cortical object-related response to the presentation of two but not one auditory object. Results suggest that at 6 months, infants are able to discriminate mistunings of 8, 6, and 4% compared to adults who show sensitivity to mistunings as small as 1%. This indicates that although young infants are able to process these stimuli their sensitivity likely increases with age. The results of these two behavioural studies are paralleled by the developmental changes seen in our electrophysiological data, in which a progression was evident between 2 to 12 months from no object-related response, to a delayed, positive object-related response to a more adult-like, negative response. Over the past two decades researchers have used neurophysiological and behavioural measures to study auditory scene analysis and simultaneous integration with harmonic complexes. The work of this thesis uses both neurophysiological and behavioural techniques to investigate simultaneous integration in infancy. Collectively, this work fills a gap in the auditory scene analysis literature addressing when, and in what manner, infants are able to process two simultaneous auditory objects.

Afterward

An important study by Bendixen et al. (2015) investigating simultaneous integration in newborn infants was published after submitting this thesis to my committee for consideration. This study points to the timely nature of this line of research and the importance of investigating how infants process harmonicity cues in infancy. Bendixen et al., measured ERP responses in newborns to the presentation of two auditory conditions. The first condition, a single-cue condition used the identical stimuli from Chapter 3 of this thesis. The second condition, a double-cue condition, employed a standard in-tune complex and a mistuned complex with the second harmonic in the complex tone mistuned by 8% and offset by 100 ms. In both conditions, researchers noted that the mistuned waveforms were negatively displaced in comparison to the in-tune waveforms between 360 and 400 ms, suggesting the possible presence of an object-related negativity. While this object-related negativity was significant when newborns were presented with two cues, it did not reach significance in the one cue condition using the stimuli from our cross-sectional study. This is in line with our work, which did not find an object-related response until 4 months of age. As the morphology of this response was negative and delayed in comparison to the response we found in 2-month-olds, it is possible that these differences could be due to maturational differences between neonates and 2-month-olds. It is also possible that they are due to methodological differences such as reference electrodes, artifact reduction techniques or filter settings, which differed between the two studies.

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

Alcantara, J.I., Weisblatt, E.J.L., Moore, B.C.J., and Bolton, P.F. (2004). Speech-in-noise perception in high functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 45*(6),1107-1114.

Benasich, A. A., and Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural Brain Research, 136*, 31-49.

Boatman, D., Alidoost, M., Gordan, B., Lipsky, F., and Zimmerman, W. (2001). Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *Journal of Cognitive Neuroscience*, *15*, 226-235.

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

Christmann, C.A., Lachmann, T., and Steinbrink, C. (2015). Evidence for a general auditory processing deficit in development dyslexia from a discrimination paradigm using speech versus nonspeech sounds matched in complexity. *Journal of Speech, Language, and Hearing Research, 58,* 107-121.

Dole, M., Meunier, F., and Hoen, M. (2014). Functional correlates of the speech-in-noise perception impairment in dyslexia: An MRI study. *Neuropsychologia, 60,* 103-114.

Groen, W.B., van Orsouw, L., ter Huurne, N., Swinkels, S., van der Gaag, R.J., Buitelaar, J.K., and Zwiers, M.P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders, 39*(3),742-750.

Hartmann, W.M. (1988). Pitch perception and the segregation and integration of auditory entities. *Auditory Function, Neurobiological Bases of Hearing,* 623-645.

He, C., Hotson, L., Trainor, L.J. (2009). Maturation of cortical mismatch mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia 47,* 218-229.

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

Helenius, P., Uutela, K. and Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain, 122*, 907-913.

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P.J., and Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia, 48*(5), 1417-1425.

Lodhia, V., Brock, J., Johnson, B.W., and Hautus, M.J. (2014). Reduced object related negativity response indicates impaired auditory scene analysis in adults with autistic spectrum disorder. *PeerJ*, *2*, e261.

McDonald, K.L., and Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related potentials. *Journal of the Acoustical Society of America, 118*(3), 1593-1604.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in neurosciences*, *24*(5), 283-288.

Rabinowicz, E. F., Silipo, G., Goldman, R., and Javitt, D. C. (2000). Auditory sensory dysfunction in schizophrenia. *Archives of General Psychiatry, 57*, 1149-1155.

Sussman, E., SteinSchneider, M., Lee, W., and Lawson, K. (2015). Auditory scene analysis in school-aged children with developmental language disorders. *International Journal of Psychophysiology, 95*(2), 113-124.

Sutter, M. L., Petkov C., Baynes K. and O'Connor. (2000). Auditory scene analysis in dyslexics. *Cognitive Neuroscience, 11*, (9), 1967-1971.

Teder-Salejarvi, W.A., Pierce, K.L., Courchesne, E., and Hillyard, S.A. (2005). Auditory spatial localization and attention deficits in autistic adults. *Cognitive Brain Research, 23*,221-234.

Tew, S., Fuijoka, T., He, C., and Trainor, L.J. (2009). Neural representation of transposed melody in infants at 6 months of age. *The Neurosciences and Music III: Disorders and Plasticity, 1169,* 287-290.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260.