

PITCH PERCEPTION WITH NO RESOLVABLE SPECTRAL CUES

Adults' and Infants' Perception of Pitch-Evoking Stimuli with
No Resolvable Spectral Cues

By

Blake Edward Butler, B.Sc., M.Sc.

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

McMaster University

© *Copyright Blake E. Butler, 2012*

DOCTOR OF PHILOSOPHY (2012)
Psychology, Neuroscience & Behaviour

McMaster University

TITLE: Adults' and Infants' Perception of Pitch-Evoking Stimuli with No Resolvable
Spectral Cues

AUTHOR: Blake E. Butler, B.MSc., M.Sc. (University of Western Ontario)

SUPERVISOR: Dr. Laurel J. Trainor

NUMBER OF PAGES: xv, 135

Abstract

Pitch perception depends on two types of cues provided by acoustic stimuli: spectral cues arise from the tonotopic organization of the basilar membrane, while temporal cues are contained within the pattern of action potentials generated in auditory nerve fibers. Filtered iterated rippled noise (IRN) can be used to examine how adults and infants perceive stimuli without resolvable spectral cues. Chapter 2 uses event-related potentials (ERPs) to compare adult perception of IRN and complex harmonic stimuli. Functional imaging studies have revealed a common pitch centre along lateral Heschl's gyrus, but lack the temporal resolution to index different stages of processing. Chapter 2 reveals differences between feature extraction-related ERP components elicited in response to different pitch-evoking stimuli. However, no differences were observed in the mismatch negativities (MMN), suggesting pitch percepts formed at this stage of processing are independent of stimulus type. Moreover, source estimates for the MMN were consistent with fMRI studies. Chapter 3 demonstrates that infants can perceive the pitch of stimuli that lack resolvable spectral cues. Eight-month-old infants discriminated between 167 Hz and 200 Hz IRN stimuli. However, performance was poor relative to when resolvable spectral cues were present, and required a period of pitch-priming (Experiment 2). Chapter 4 uses ERPs to further examine infants' perception of IRN. Using an oddball paradigm, 4- and 8-month-old infants were shown to discriminate between 167 Hz and 200 Hz stimuli. However, successful discriminations again required pitch-priming (Experiment 2). Together, these findings suggest that infant pitch percepts are dominated by resolved spectral cues. Infants discriminate between stimuli without resolvable

spectral cues, but appear to require a period of priming to do so, at least until 8 months of age. Thus, infant pitch perception may be impaired in challenging listening situations, such as extracting vocal pitch cues in the presence of background noise.

Acknowledgements

A well-known proverb suggests that it takes a village to raise a child. A developmental psychologist would likely agree with this idea (while, of course, acknowledging the role of genetic predisposition). I would like to extend this idea to graduate school, as I believe it takes a community to produce a doctoral thesis. In that spirit, I would like to acknowledge all of the amazing individuals who helped me reach this point.

First and foremost, I would like to thank my supervisor, Laurel Trainor. Laurel, your dedication to the field, and the impact of your work are inspirational. Thank you for all of the guidance and support you have shown me over the past four years. You have been a wonderful mentor, and I will surely measure my future success, at least in part, by my ability to emulate the traits I admire so much in you.

I would also like to thank each of the individuals who invested their time and energy as members of my supervisory committee. Daphne, your wisdom and thoughtful criticism have been incredibly valuable throughout this process. Thanks for lending your, often unique, perspectives to the issues at hand. Dan, your wealth of knowledge has been a wonderful asset to me. Thanks for your willingness to lend a hand when needed.

This thesis was greatly improved through the assistance and thoughtful feedback provided by the members of the Auditory Development Lab. In particular, I would like to thank former lab member Chao He for the training he provided during our brief overlap in the EEG lab, lab-mom Katie Corrigan for her services as part-time counselor, and Nicole Folland for being a great collaborator. I must also sincerely thank our research assistants

without whom the papers in this thesis, and the work that comprises them would never have come together. Elaine “stroke-in-the-night” Whiskin, thanks so much for all of your hard work training undergrads, running participants, and generally making the lab a more pleasant place to be. Andrea Unrau, thank you for all of the proof-reading and poster tweaking you’ve done over the years. Although not actually a full-time member of the Trainor Lab, I owe a debt of gratitude to Dave “Chips” Thomson for all of the technical assistance he provided during the course of my Ph.D. Thanks so much for all of the times you turned hours of work into minutes, rescued precious testing sessions, and for all of the lessons in the art of brewing. Finally, a huge thanks to all of the junior scientists and their wonderful parents; without you we would not have the freedom to explore the amazing questions that motivate this research.

I would also like to acknowledge some friends who have been an integral part of my time at McMaster. Chris McAllister, your friendship has meant a great deal to me over the past few years. Whether helping me learn Illustrator, planning surprise birthdays/vacations/engagements, or mopping the court with me over at DBAC, you’ve always been there for me. Chris Teeter, Matt Pachai, Adam Reddon, and Dan Case, thanks so much for making Mac such a wonderful place to be for the past 4 years. Whether playing ball, lunching at the Phoenix, or confirming our place in the Winking Judge history books, some of my best times have been spent with you fine gentlemen. Finally, to my long-time friends Jeff Bridgman and Kevin Reeve, thanks for bearing with me while I disappear for weeks or months at a time, only to resurface with nothing new or exciting to report.

I would like to thank my amazing family for the unconditional love and support they have always shown for me; I would truly not be where I am today if not for their kindness. To my parents Brian and Diane, and my sister Kaitlin, I will never be able to adequately articulate my appreciation for everything you have done for me throughout the years. I am so grateful to have you in my corner now and always. Finally, I would like to express my deepest thanks and unending appreciation to my amazing wife, and biggest fan, Lyndsey. Thank you for always supporting me when I need it the most, for being there to listen, regardless of the topic, and for putting away my soapbox when I've finished with it. Your love, patience, and support have guided me through this process, and I couldn't have done it without you. You are my pebble.

Table of Contents

Abstract.....	iii
Acknowledgements.....	v
Table of Contents.....	viii
Preface.....	x
List of Figures.....	xii
CHAPTER 1: Introduction.....	1
Models of Pitch Processing.....	2
Development of Frequency and Pitch Processing.....	9
Methodologies Used in this Thesis.....	12
Conditioned Head-Turn Procedure.....	12
Event-Related Potentials.....	13
Components of Interest: P1, N1 & MMN.....	15
Effects of Attention on the MMN.....	19
Motivation for the Current Thesis.....	22
CHAPTER 2: Sequencing the cortical processing of pitch-evoking stimuli using EEG and source estimation.....	24
Preface.....	24
Abstract.....	27
Introduction.....	28
Methods.....	37
Results.....	46
Discussion.....	53
References.....	59
Afterward.....	67
CHAPTER 3: Development of pitch processing: Infants' discrimination of iterated rippled noise stimuli with unresolved spectral content.....	69
Preface.....	69
Abstract.....	72
Introduction.....	73
Experiment 1 Methods.....	77
Experiment 1 Results.....	80
Experiment 2 Methods.....	81
Experiment 2 Results.....	82
Discussion.....	83
References.....	87
CHAPTER 4: Brief pitch-priming facilitates infants' discrimination of pitch-evoking noise: Evidence from event-related potentials.....	91
Preface.....	91

Abstract	94
Introduction	95
Experiment 1 Methods	100
Experiment 1 Results	104
Experiment 2 Methods	106
Experiment 2 Results	107
Discussion	110
References	115
CHAPTER 5: General Discussion	122
Unique Contributions of this Thesis to the Literature	122
Limitations	126
Future Directions	129
Final Conclusions	130
References for Chapters 1 & 5	132

Preface

This thesis is comprised of three manuscripts. The first empirical chapter (Chapter 2) has been published in a peer-reviewed journal. The empirical work in Chapters 3 and 4 are manuscripts that currently under consideration for publication in peer-reviewed journals. Each empirical chapter represents a single manuscript for which I am the primary author. The remainder of this preface is designed to clarify my contributions to the manuscripts that comprise the empirical chapters of this thesis.

The first empirical chapter is a reprint of Butler, B. E., and Trainor, L. J. (2012). Sequencing the cortical processing of pitch-evoking stimuli using EEG and source estimation. *Frontiers in Psychology*, 3(180), 1-13. My role in this manuscript included experimental design and programming, data collection (running human participants), and data analysis. I was also the primary writer of the manuscript.

The second empirical chapter is the following manuscript: Butler, B. E., Folland, N.A., and Trainor, L. J. (submitted). Development of pitch processing: Infants' discrimination of iterated rippled noise stimuli with unresolved spectral content. *Hearing Research*, Manuscript ID HEARES-D-12-00290. My role in this manuscript included experimental design and programming, data collection, and analysis. I was also the primary writer of the manuscript.

The third empirical chapter is the following manuscript: Butler, B. E., and Trainor, L. J. (submitted). Brief pitch-priming facilitates infants' discrimination of pitch-evoking noise: Evidence from event-related potentials. *Brain and Cognition*, Manuscript ID BRCG-12-271. My role in this manuscript included experimental design and

programming, data collection, and analysis. I was also the primary writer of the manuscript.

As a final note to the reader, each of chapters 2 through 4 represents a manuscript intended to stand along as a published work. As a result, there is a fair amount of redundancy present within the introductory sections to each of these chapters that the reader should be aware of. That being said, the experiments presented within each manuscript are distinct works intended to address separate, yet complementary objectives.

List of Figures

CHAPTER 2

- Figure 1.** Schematic diagram of the delay (D), gain (G) and add networks used to generate the type of IRN stimuli used in the present study (adapted from Yost, 1996). A sample of white noise is added ($G=1$) to itself following some delay. This process is repeated a number of times, and the resulting stimulus has a perceivable pitch component at the inverse of the delay ($1/D$)..... 38
- Figure 2.** Panel **A** shows the power spectra calculated using Welch's method for the IRN (blue) and complex harmonic (red) 200 Hz stimuli used in the experiment. The spectra were calculated with a 11025- point Hamming window applied with a 10925-point overlap. Panel **B** shows a zoomed-in view of three of the upper harmonics of the stimuli, illustrating the difference between the 'rippled noise' spectrum and a complex waveform made up of discrete harmonics. In particular, the harmonics of the complex waveform are higher in amplitude and more defined in that they show sharper peaks whereas those of the IRN stimuli are more spread out in frequency..... 41
- Figure 3.** Electrode groupings. Seventy-six of 128 electrodes were divided into 4 groups (frontal, central, parietal, and occipital) for each hemisphere. Each group contained between 16 and 20 electrodes that were averaged together 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.....	45
Figure 4. Grand average waveforms in response IRN and complex harmonic stimuli for standard trials (panel A) and deviant trials (panel B) for all 8 scalp regions.....	47
Figure 5. Grand average waveforms (standard response, deviant response, and difference wave [deviant-standard]) for IRN stimuli (panel A) and complex harmonic stimuli (panel B). The bars under the waveforms show latencies at which the amplitude of the difference wave was significantly different from zero.....	49
Figure 6. Three-dimensional locations of symmetric regional sources fit using inverse solutions generated by BESA. Panel A shows the location of P1 in both stimulus conditions. Panel B shows the location of N1 in both stimulus conditions. Panels C and D show the location of P1, N1 and MMN in the IRN and complex harmonic conditions, respectively.....	51
Figure 7. Three-dimensional scatterplot showing the Cartesian coordinates of the P1 and N1 components elicited by the IRN and complex harmonic stimuli, as well as the common MMN elicited by both stimulus types. In the BESA Cartesian coordinate system, the X-axis passes through the LPA (left pre-auricular area) and RPA (right pre-auricular area) with right being the positive direction. The Y-axis passes through the nasion, with rostral being positive. The Z-axis is orthogonal to the X- and Y-axes, with dorsal being the positive direction. Components shown are all located in the right hemisphere, while symmetric sources in the left hemisphere have been omitted for clarity.....	52

CHAPTER 3

Figure 1. Panel a) shows a box plot of the d' sensitivity scores observed for the discrimination of a change in the perceived pitch of an IRN stimulus from 167 Hz to 200 Hz. Panel b) shows the individual d' sensitivity scores for each infant in Experiment 2..... 82

Figure 2. Hit rates (dark bars) and false alarm rates (light bars) for each infant in the current study, arranged in order of decreasing sensitivity. Thirteen of sixteen infants showed hit rates that exceeded their rate of false-alarms..... 83

CHAPTER 4

Figure 1. Electrode groupings (see Materials and Methods section for details). Ninety of 124 electrodes were divided into 5 groups (frontal, central, parietal, occipital and temporal) for each hemisphere. Each group contained between 16 and 20 electrodes that were averaged together 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..... 103

Figure 2. Grand average difference waveforms (deviant-standard) for pitch changes recorded in Experiment 1. Responses from each of the 10 electrode regions are presented for both 4-month-old (solid lines) and 8-month-olds (dashed lines)... 105

Figure 3. Grand average difference waveforms (deviant-standard) for pitch changes recorded in Experiment 2. Responses from each of the 10 electrode regions are presented for both 4-month-old (solids lines) and 8-month-olds (dashed lines). The bars under the waveforms show latencies at which the amplitude of the difference wave differed significantly from zero for each age group.....109

CHAPTER 1

There are four basic perceptual features of sound: pitch, loudness, duration, and timbre. Pitch is the perceptual correlate of stimulus frequency, and is important for processing stimuli across a number of domains. For example, pitch is essential to music perception; the melody of a song is determined by the way in which pitch changes over time. Indeed, individuals with an inability to accurately perceive these pitch contours (termed ‘amusia’) report being indifferent to music (Ayotte, Peretz, and Hyde, 2002). The extraction of vocal pitch is also integral to speech processing. Vocal pitch helps identify speakers (e.g. van Dommelen, 1990), and aids in the understanding of speech in the presence of background noise (e.g. Song, Skoe, Banai, and Kraus, 2011). In addition, prosodic pitch contours signal lexical and syntactic information, as well as emotional expression (e.g. Frick, 1985). Finally, the harmonic structure giving rise to the pitch of complex stimuli is a salient cue for the successful separation of overlapping sounds in the environment, and for the attribution of those sounds to their correct sources (Bregman, 1990).

Despite the importance of pitch perception, a number of questions concerning how pitch percepts are formed from frequency cues remain, both in the developing auditory system and in mature cortex. One question in particular relates to how spectral cues and temporal cues (arising from the tonotopic organization of the basilar membrane and the pattern of action potentials generated by pitch-evoking stimuli, respectively [see Background: Models of Pitch Processing for more detail]) are combined to create a

singular pitch representation. In this thesis, I present the first study using EEG to index pitch discrimination at different levels of the auditory cortex, for stimuli with different contributions from spectral and temporal cues to pitch. The precise way in which frequency cues are combined to create a pitch percept remains unknown, and while a number of functional imaging studies have suggested a cortical location that may serve as a common pitch centre, these studies have typically involved fMRI, which has temporal resolution on the order of seconds. The current thesis examines the process of pitch perception using EEG, which provides much finer temporal resolution (on the order of milliseconds). Additionally, a number of studies have demonstrated that infants perceive pitch in an adult-like manner, given salient spectral cues. The infant work presented in this thesis seeks to understand whether infants are capable of extracting pitch from stimuli that rely primarily on temporal cues. Moreover, this thesis examines whether priming infants to listen for temporal cues can affect the cortical processing of pitch.

Background

Models of Pitch Processing

At the level of the cochlea, the mechanical energy contained in an acoustic waveform is translated into a pattern of action potentials that contains two representations of stimulus frequency. The spectral, or place representation, arises from the tonotopic organization of the basilar membrane, which is lined by hair cells. The physical properties of this membrane (e.g. width, rigidity) differ along its length such that high frequencies cause maximal displacement of the membrane basally, and low frequencies displace the

membrane maximally at its apex (Von Békésy, 1960). Because action potentials are only generated by hair cells attached to areas of the basilar membrane that have exceeded some threshold level of displacement, the frequency of harmonic components can be determined based on which auditory nerve fibers have been depolarized. This tonotopic organization is thought to be maintained in the auditory nerve, through subcortical nuclei, and into primary auditory cortex (A1; e.g., see Formisano et al., 2003; Humphries, Liebenthal, and Binder, 2010 for reviews). However, the nature of the tonotopic organization of human A1 remains an issue of much debate. While some fMRI studies suggest that A1 contains multiple frequency gradients (e.g. Striem-Amit, Hertz, and Amedi [2011] report 6 gradients extending as far as the superior temporal sulcus and middle temporal gyrus), Schönweisner, von Cramen, and Rübsamen (2002) have suggested that these gradients might instead arise from neurons processing acoustic features associated with different frequency bands. Still other fMRI studies suggest that frequency representations in A1 simply code high and low frequencies, and are not graded at all (e.g. Langers and van Dijk, 2012).

Cochlear processing also gives rise to a temporal representation of stimulus frequency. Cochlear hair cells are most likely to depolarize during the peak or valley of an acoustic waveform, when the basilar membrane is maximally displaced in one direction or the other. At low frequencies, the response properties of a single neuron can become time-locked to an acoustic stimulus, such that the frequency of action potential generation is equal to the frequency of the stimulus to which it is responding. Beyond the limits of neural firing, a single neuron can no longer accurately track the pattern of peaks

and valleys contained in an acoustic stimulus. However, frequency can be derived from the output of a small population of neurons that innervate a particular area of the basilar membrane. Because stimulus frequency and period are related reciprocally, frequency can be extracted temporally, based on the amount of time that passes between action potentials in the cochlear output (e.g. Cariani and Delgutte, 1996a; 1996b). Thus, the frequency of a sound can be determined from the location of action potential generation (spectral code), from the pattern of action potentials contained within auditory nerve fibers (temporal code), or from some combination of the two.

Naturally-occurring pitch-evoking stimuli typically consist of energy at a fundamental frequency, and at integer multiples of that frequency, referred to as harmonics (e.g. a complex stimulus with a perceived pitch of 200 Hz contains energy at 200 Hz, 400 Hz, 600 Hz, 800 Hz, etc.). While listeners can be trained to hear individual harmonics within a complex stimulus, spectrotemporal processing typically gives rise to a single pitch that corresponds to the fundamental frequency of the stimulus. In fact, the auditory system is capable of perceiving a pitch corresponding to the fundamental frequency of a complex harmonic stimulus, even when the stimulus contains no energy at that frequency. This phenomenon, known as the pitch of the missing fundamental, was first reported more than 150 years ago, and explains how a listener is able to accurately perceive low-pitch stimuli through band-limited channels such as a telephone or inexpensive speakers integrated into some laptop computers and smartphones.

Exactly how pitch cues are combined to form a singular representation of pitch is not fully understood. Historically, a number of models have been proposed that consider

different relative contributions from the spectral and temporal representations of frequency information. In his residue theory, Schouten (1940) suggested that the limited resolution of cochlear frequency analysis precluded spectral cues from contributing meaningfully to pitch, instead proposing that pitch was derived from temporal information arising from the frequency components of a complex sound that cannot be resolved by the cochlea. While the harmonics of a complex sound are linearly spaced, the auditory filters along the basilar membrane are spaced logarithmically. At low frequencies, the bandwidth of these filters is sufficiently small, such that each harmonic of a complex sound is contained within an individual filter. However, at higher frequencies, multiple harmonics fall within the bandwidth of a single auditory filter and thus, cannot be resolved by the cochlea. These unresolved harmonics produce a signal with a periodic, or quasiperiodic envelope at the cochlear output that is preserved by neural phase-locking in auditory nerve fibers (Brugge, Anderson, Hind, and Rose, 1969), and in the cochlear nucleus (Møller, 1970). Schouten (1940) suggested that analysis of this phase-locked cochlear output gave rise to pitch, irrespective of the presence of resolved spectral content. However, a number of subsequent studies provided evidence inconsistent with the residue model. For example, low-order harmonics, particularly the third through fifth, were found to make the greatest contribution to the pitch of the missing fundamental (Plomp, 1967; Ritsma, 1967), despite being resolved in the cochlea. Moreover the pitch percept was shown to be relatively insensitive to harmonic phase (Patterson, 1973; Wightman, 1973), contrary to what was predicted by the residue model. Finally, and most strikingly, dichotic presentation of harmonics (the presentation of

alternating harmonics to the left and right ears) was shown to yield the same pitch as diotic or monotic presentation (Houtsma and Goldstein, 1972); a singular representation of pitch was created despite the fact that the stimuli presented to the left and right ears produced two different cochlear outputs, neither of which would produce the pitch of the fundamental in isolation.

In light of the apparent importance of spectral cues, subsequent pitch models focused on central mechanisms based on pattern recognition for harmonics resolved in the auditory periphery (e.g. Goldstein, 1973; Terhardt, 1979). For example, Goldstein (1973) proposed an optimum processor model in which the pitch of a complex stimulus is determined by the closest match between the resolved harmonics of that stimulus and a central series of templates representing the harmonic spacing of complex tones of different pitches. Alternatively, Terhardt (1979) provided a model in which the best-fitting subharmonic common to the resolved components present in the stimulus could approximate the fundamental frequency of a harmonic stimulus. While these models may sufficiently explain how pitch can be determined from resolved harmonic content, they fail to account for evidence which suggests that pitch can be determined in the absence of resolved spectral content. For example, there is physiological evidence that even when all of the resolved components of a harmonic tone are removed, correlates of the missing fundamental can still be observed in firing patterns of auditory nerve fibers (Brugge et al. 1969) and in the cochlear nucleus (Evans, 1977). Moreover, there are a number of situations in which pitch can be perceived from an acoustic stimulus in the absence of any resolved spectral content; examples include high-pass filtered harmonic tones (Houtsma

and Smurzynski, 1990), Huggins pitch (white noise which is diotic at all frequencies except for a narrow band over which the interaural phase shifts progressively through 360 degrees, eliciting a pitch at the frequency of transition; Cramer and Huggins, 1958), amplitude-modulated white noise (Burns and Viemeister, 1976), and iterated rippled noise (a sample of frozen, white noise that is repeatedly added to itself following some delay, eliciting a pitch equal to the inverse of the delay; Yost, Patterson, and Sheft, 1996).

In order to account both for the dominance of spectrally resolved content in the formation of pitch percepts, and the persistence of these percepts in the absence of spectral cues, modern pitch theories typically include two complementary pitch mechanisms: one that makes use of cues from spectrally resolved components and a second that is based on temporal information extracted in their absence. Such dual-mechanism hypotheses are supported by behavioural studies which show that performance on pitch-related tasks using complex harmonic stimuli degrades in a biphasic manner as the lowest harmonic present increases from the region of resolved to the region of unresolved spectral content (Houtsma and Smurzynski, 1990). It should be noted, however, that this pattern of performance could also be explained by a single pitch mechanism that performs differently depending on the type of pitch cues provided by the auditory signal.

It is now commonly accepted that, where available, both spectral and temporal pitch cues are involved in the formation of a pitch percept. However, how and where these cues might be combined into a singular representation of pitch remains unclear. To date, attempts to find a representation of pitch in the frequency following response (FFR)

elicited from subcortical nuclei have been unsuccessful (Gockel, Carlyon, Mehta, and Plack, 2011). The FFR is elicited by the repeated presentation of auditory stimuli, and reflects phase-locked activity across a population of subcortical neurons. Gockel and colleagues (2011) noted that while the FFR demonstrates that subcortical structures preserve temporal information that may be important to pitch, there is no evidence of subcortical pitch processing beyond that present in the auditory periphery. Pantev, Elbert, Ross, Eulitz, and Terhardt (1996) reported MEG evidence of a neural population in primary auditory cortex that responded to the pitch of a missing fundamental stimulus, suggesting a representation of pitch in primary auditory cortex (A1). However, it has been suggested that the response they recorded may have been the result of a cochlear distortion, or a response to a band of masking noise centred at this frequency (Walker, Bizley, King, and Schnupp, 2011). In fact, electrophysiological recordings in the macaque have failed to find pitch-sensitive neurons in primary auditory cortex (Schwarz and Tomlinson, 1990). Rather, a pitch-responsive area has been identified just beyond primary auditory cortex in the marmoset (Bendor and Wang, 2005). Functional imaging studies in human listeners provide support for the existence of a common pitch centre beyond primary auditory cortex, along the lateral aspect of Heschl's gyrus (Griffiths, Buchel, Frackowski, and Patterson, 1998; Patterson, Uppenkamp, Johnsrude, and Griffiths, 2002; Penagos, Melcher, and Oxenham, 2004; Hall, Barrett, Akeroyd, and Summerfield, 2005; Puschmann, Uppenkamp, Kollmeier, and Thiel, 2010) or in planum temporale (Hall and Plack 2009). However, these studies lack the temporal resolution to identify when this percept is formed, relative to other aspects of auditory processing.

This thesis takes advantage of the high temporal resolution of electroencephalography (EEG) to index the formation of pitch percepts derived from stimuli that differ in the saliency of their spectral and temporal pitch cues. This allowed an extension of the understanding of cortical pitch processing beyond locating the cortical area of pitch percept formation; it allows for the sequencing of events involved in the extraction of spectral and temporal features from the stimulus, and ultimately, the binding of these features into a single auditory object.

Development of Frequency and Pitch Processing

In the human fetus, movement recorded by ultrasound in response to auditory stimuli is taken as behavioural evidence of the onset of hearing, and is typically first observed around 19 weeks gestational age for pure tones at 500 Hz (Hepper and Shahidullah, 1994). Lower tones (250 Hz) appear to elicit a behavioural response by 27 weeks gestational age, while responses to higher tones (1-3 kHz) occur between 33 and 35 weeks gestational age (Hepper and Shahidullah, 1994). The properties of acoustic signals present in the mother's environment are greatly affected by distortions that occur as the signal passes through the mother's abdomen and the fluid of the amniotic sac. Despite this distortion, there is some evidence that the fetus can extract and encode basic stimulus features, such as frequency, from external acoustic signals. For example, Shahidullah and Hepper (1994) demonstrated that the fetus can discriminate between 250 Hz and 500 Hz pure tone stimuli at 35 weeks, but shows no evidence of doing so at 27 weeks gestational age. This was determined using a habituation-dishabituation technique

in which the fetus is presented with a stimulus of one frequency until movement responses cease, at which point a second frequency is presented; if movement occurs in response to this new frequency, it is taken as evidence that the fetus can discriminate it from the original stimulus.

Frequency discrimination thresholds improve dramatically following birth, but subsequently show a slow progression toward adult-like performance. Three- to six-month-old infants demonstrate difference limens as low as 20 Hz for a 1000Hz stimulus (a difference of 2%; Olsho, Schoon, Sakai, Terpin, and Sperduto, 1982). This is a marked improvement over prenatal discrimination, however, thresholds are still elevated when compared to adult difference limens (0.3 – 0.6% at 1000 Hz; Olsho, Koch, and Halpin, 1987), and adult levels of frequency discrimination are not achieved until about 10 years of age (Jensen and Neff, 1993; Thompson, Cranford and Hoyer, 1999). The reasons for this slow rate of development are not fully understood; peripheral frequency resolution does not appear to be the limiting factor. The bandwidths of cochlear filters, as measured in masking experiments, show that they are mature for low frequencies at birth and for high frequencies by 6 months of age (Spetner and Olsho, 1990). The function of outer hair cells, which serve to further refine place cues by manipulating basilar membrane movement, also do not appear to be the limiting factor. Research in human preterm neonates suggests that functional maturation of outer hair cells is nearly complete by 33 weeks gestational age (Morlet, Collet, Salle, and Morgon, 1993), with full structural maturity typically achieved by 35 weeks gestational age (Pujol and Uziel, 1988). Thus, any proposed place cue-related delay in frequency discrimination must be related to the

processing of these cues, rather than their peripheral encoding. Alternatively, delayed maturation of frequency discrimination may reflect delayed development of the mechanism responsible for temporal encoding (e.g. precise phase-locking of auditory neurons) and/or processing (i.e. the ability of auditory areas to make use of these temporal cues).

While the limits of pitch perception in infancy have not been measured, it has been demonstrated that infants as young as eight months of age are capable of discriminating complex stimuli that differ by 20 % (e.g. 160 and 200 Hz; Clarkson and Clifton, 1985). Eight-month-old infants have also been shown to discriminate complex stimuli that differ in pitch by 20% in the absence of energy at the fundamental frequency (Clarkson and Clifton, 1985). Additionally, their perception of pitch is unaffected by a low-pass noise masker. This suggests that for infants, as adults, the pitch of the missing fundamental depends on harmonic structure rather than combination tones resulting from cochlear non-linearities, as any combination tones produced would be rendered inaudible by the masker. Electrophysiological measures suggest that a cortical representation of the pitch of the missing fundamental emerges between 3 and 4 months of age (He and Trainor, 2009). Collectively, these results suggest that by 4 months of age, harmonic structure contributes to pitch percept formation in a manner similar to adult listeners. In support of this idea, infant perception of inharmonic complexes (those in which harmonic spacing is not at integer multiples of a common fundamental frequency; e.g. a complex tone with harmonics at 220, 320, 420 and 520 Hz) has been shown to be qualitatively similar to that of adults, degrading with the degree of inharmonicity (Clarkson and

Clifton, 1995). In sum, these results suggest that some frequency cues are available to the infant auditory system by birth, and that by 4 months of age, infants are processing pitch-evoking stimuli in a manner that is qualitatively adult-like. However, these studies have largely employed salient spectral pitch cues; it remains unknown to what degree temporal cues contribute to pitch perception in infancy. In this thesis, I examine infants' behavioural responses to IRN stimuli, which contain primarily temporal pitch cues, in an effort to determine whether infants can employ these cues in a manner similar to adults. Furthermore, I use EEG responses to examine cortical representations of temporal pitch discrimination in infancy. Finally, I examine the effect of demonstrating the pitch represented by these temporal cues to infants on their cortical responses to IRN stimuli. These studies contribute to the small body of work addressing the role of temporal cues in infant pitch perception, and are the first to use stimuli other than complex harmonic tones. IRN stimuli are created from a sample of white noise, and while the resultant spectra contain some peaks at harmonics of the perceived pitch, this spectral information is much less defined than complex harmonic tones. Thus, these stimuli rely to a greater extent on temporal pitch cues than spectral cues, and provide a means of examining pitch perception in the absence of resolved spectral cues.

Methodologies Used in this Thesis

Conditioned Head-Turn Paradigm

In adult listeners, behavioural discrimination can be easily measured using a wide variety of experimental paradigms that involve active evaluation of stimuli (e.g. 3-

alternative, forced-choice or same/different tasks). However, evaluating behavioural discrimination in infants is more difficult. In Chapter 3, I employed a conditioned head-turn paradigm (see Werker, Polka, and Pegg, 1997 for review) to determine whether infants could discriminate between auditory stimuli of two different pitches. In this method, turning toward a sound source following the presentation of a deviant, target stimulus is reinforced by the appearance of a toy that moves and makes noise. Subsequent turns toward the deviant stimulus in the presence of standard stimuli demonstrate that the infant can discriminate between the two. These ‘hits’ are coded along with ‘misses’ (failures to turn toward the deviant stimulus), and ‘false alarms’ (turns toward the sound source following the presentation of a standard stimulus). The signal detection measure d' takes each of these response types into account, providing a measure of sensitivity that is relatively free of response biases. Thus, the conditioned head-turn procedure produces an unbiased estimate of infant behavioural discrimination. However, as with any infant method, this value likely underestimates discrimination because infants may be unlikely to respond to near-threshold stimuli due to non-sensory factors (e.g. see Nozza, 1995 for review).

Event-Related Potentials

Imaging methods, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), offer impressive spatial resolution, and have been widely used in the study of pitch processing, particularly in adult listeners. However, there are a number of shortcomings to these methods that limit their use. Firstly, these

measures (e.g. blood-oxygenation level dependant [BOLD] measures in fMRI) offer poor temporal resolution, such that the sequence of events involved in a particular response may be uninterpretable. Secondly, they measure metabolic responses that are only presumed to reflect activity in nearby populations of neurons. Finally, these techniques offer a number of challenges specific to developmental work, including: parental apprehension toward participation, design challenges related to reducing movement-related noise, and the lack of useful structural models of infant brains. Fortunately, event-related potentials (ERPs) collected using EEG provide a method by which many of these challenges can be overcome. ERPs represent brain responses time-locked to the onset of an event, and provide millisecond-range temporal resolution. Moreover, when a dense electrode array is used (e.g. the 128- and 124-electrode nets used in chapters 2 and 4, respectively), source estimation can be performed with spatial resolution that is adequate to discriminate between auditory areas, and indeed between areas within a single auditory area, as demonstrated in Chapter 2. In this thesis, EEG is used to index the formation of a pitch percept from stimuli that contain differing contributions from spectral and temporal pitch cues. The fine temporal resolution of EEG enabled the examination of components related to feature extraction (occurring automatically following the presentation of a stimulus) as well as endogenous processing-related components (occurring when an incoming stimulus differs from a previously established auditory memory) to determine at what point a representation of pitch is formed.

ERPs also provide a non-invasive means of collecting data from infant listeners. Furthermore, the conditioned head turn is limited to infants older than 5 months of age,

but ERPs can be measured in newborns. The GSI nets used in Chapters 2 and 4 are baby-friendly, employing electrodes that are contained within sponges soaked in a saline solution. Moreover, EEG systems are more robust to movement-related noise than are fMRI or PET systems. If an individual moves during an EEG recording, that portion of the data can be discarded and recording continued once the individual is still again because the electrodes are on the head. This is not possible with fMRI where the equipment is immobile and the head must remain still with respect to the equipment. In this thesis, I use EEG to observe the cortical representation of temporal pitch discrimination in infants, and to compare those representations to those from adults.

Components of interest: P1, N1, & MMN

An acoustic stimulus elicits a sequence of exogenous ERP components (those related to the mere presence of an auditory event) that reflect the stages of processing of that stimulus. The cortical ERP components include the P1, a positive-going deflection that peaks approximately 50 ms post-stimulus-onset in adults, thought to originate from primary auditory cortex (Godey, Schwartz, de Graaf, Chauvel, and Liegeois-Chauvel, 2001; Yvert, Fischer, Bertrand, and Montysalo, 2005), and the N1, a negative-going deflection that peaks approximately 100 ms post-stimulus-onset in adults, that is thought to have multiple sources in secondary auditory cortex (Godey et al., 2001; Yvert et al., 2005). Importantly, these components are thought to arise following extraction of the basic features of an auditory signal (e.g. onset/offset, amplitude, frequency, etc), but before those features are bound into a perceivable auditory object (Näätänen and Winkler,

1999). In this thesis, P1 and N1 were measured in adults as indices of pitch encoding in auditory cortex to temporal cues.

In addition to feature-related components, ERP experiments can be designed to elicit components related to endogenous processing of auditory objects. For example, an infrequent auditory deviant presented among a repeated train of standard auditory stimuli will elicit a mismatch negativity (MMN) component thought to reflect an updating of auditory trace memory (see Näätänen, Paavilainen, Rinne, and Alho, 2007 for review). In adults, the MMN presents as a negative-going deflection in the difference waveform (deviant – standard) that peaks approximately 150 ms post-stimulus onset. This latency overlaps with the N1 component and as a result, early critics suggested that the MMN simply reflected an amplification of the N1. However, a number of subsequent findings confirm that the MMN component is distinct from the N1. For example, while N1 latency is independent of stimulus type and closely time-locked to stimulus onset (see Näätänen and Picton, 1987 for review), MMN latency has been shown to be highly variable, and related to the degree of difference between standard and deviant stimuli (Tiitinen, May, Reinikainen, and Näätänen, 1994; Lang et al., 1990; Winkler, Tervaniemi, and Näätänen, 1997). In addition, it is possible to elicit an MMN in the absence of an N1 response. For example, while the elicitation of an N1 component only occurs following the presentation of a sound, the omission of a tone from a series presented at a short, constant SOA (Yabe, Tervaniemi, Reinikainen, and Näätänen, 1997; Yabe et al., 1998) or the omission of the second tone in a regularly presented tone pair (Tervaniemi, Saarinen, Paavilainen, Danilova, and Näätänen, 1994) will elicit an MMN response. Finally,

pharmacological evidence suggests that the neurotransmitters involved in producing the MMN differ from those involved in the production of exogenous components. For example, intracortical recordings in awake monkeys have demonstrated that the MMN response is eliminated following NMDA receptor antagonist administration, while the N1 response remains intact (Javitt, Steinschneider, Schroeder, and Arezzo, 1996). In sum, multiple lines of evidence suggest that the MMN is dissociable from the N1 component, and likely represents a response to deviation within a stream of stimuli. MMN can be used as a cortical measure of stimulus discrimination, as the component is only elicited if the deviant stimulus is perceptibly different than the standard stimulus. Furthermore, it can be recorded from both adult and infant listeners. In this thesis, MMN is used as a measure of the ability to discriminate stimuli with different pitch in infants compared to adults.

Two complementary cortical sources have been proposed for the adult MMN component. ERPs recorded by implanted electrodes in the cat auditory cortex suggest that one source is located in the rostroventral part of secondary auditory cortex (Pincze, Lakatos, Rajkai, Ulbert, and Karmos, 2001), the exact location of which has been shown in humans to be dependant on the acoustic feature eliciting the mismatch response (Paavilainen, Alho, Reinikainen, Sams, and Näätänen, 1991). An additional source has been localized to frontal cortex in humans using PET (Dittmann-Balcar, Jüptner, Jentzen, and Schall, 2001; Müller, Jüptner, Jentzen, and Müller, 2002) and has been confirmed by lesion studies (Alho, Woods, Algazi, Knight, and Näätänen, 1994; Alain, Woods, and Knight, 1998). It has been proposed that the involvement of this frontal source in the

process generating the MMN provides evidence for a central role of this process in involuntary orienting, or attention switching, to a change in the acoustic environment (Näätänen 1979; Näätänen and Picton, 1987). A change in an incoming signal could represent a cue to meaningful changes in one's environment; the elicitation of an MMN to a deviant stimulus appears to be an integral step in the redirection of attention toward the source of such stimuli. This of particular interest in the current thesis: because the MMN results from the perception of auditory objects, it must be elicited at or beyond the point at which stimulus features, such as spectral or temporal pitch cues, have been bound into a singular representation of a sound stimulus (Ritter, Sussman, and Molholm, 2000). In chapter 2 of this thesis, I compare the MMN component across different stimulus types to determine whether the source estimates are consistent with functional imaging studies of a common pitch-processing centre.

While much is now known about the generators and response characteristics of the adult MMN, analogues in infant listeners are still poorly understood. The first observable deviance-related component recorded in very young infants typically takes the form of a slow, frontally-distributed positivity (e.g. Trainor, Samuel, Desjardins, and Sonnadara, 2001; Friederici, Friedrich, and Weber, 2002; Winkler et al., 2003; Friedrich, Weber, and Friederici, 2004; Novitski, Huotilainen, Tervaniemi, Näätänen, and Fellman, 2007; He, Hotson, and Trainor, 2009b). This component is later replaced by a negative-going deflection of shorter latency (e.g. Alho et al., 1990; Cheour et al., 1998; Trainor et al., 2001, 2003; Kushnerenko et al., 2002; He, Hotson, and Trainor, 2007; 2009a; 2009b), which resembles the adult MMN. The age at which these components emerge differs

depending on the stimulus feature eliciting the response. For example, detection of a change in the pitch of complex stimuli elicits a slow positivity at 2 months of age, while a negative MMN-like response emerges at 3 months of age (He et al., 2007). However, in complexes with a missing fundamental, no response is recorded in 3-month-olds, while 4-month-olds show a robust, short-latency negativity (He and Trainor, 2009). These differences in developmental timeline are in accordance with MEG data from adult listeners that suggest MMN responses elicited by deviations in different stimulus features have at least partially different neural sources (Giard et al., 1995). In this thesis, I used the development of the MMN response elicited in response to IRN stimuli as a cortical measure of the development of temporal pitch discrimination development.

Effects of Attention on the MMN component

The MMN response is thought to reflect endogenous, pre-attentive change-detection, and as such, should not be modulated by directed attention. Early descriptions of the MMN noted no significant differences in the amplitude of adult MMN responses elicited by frequency-deviants between active and passive listening conditions (Näätänen, Gaillard, and Mäntysalo, 1978; Näätänen, Gaillard, and Mäntysalo, 1980; Sams, Alho, and Näätänen, 1984). However, paradigms involving the active direction of attention toward a particular area in auditory space have demonstrated that the MMN response to changes in stimuli occurring outside of the focus of attention are attenuated or eliminated depending on their distance from the target area (Arnott & Alain, 2002). Moreover, the idea that MMN is unaffected by attention was challenged by dichotic listening tasks in

which adult participants were instructed to ignore one auditory stream in order to detect intensity-deviants in the other. The MMN amplitudes elicited by deviant stimuli in the attended stream were significantly larger than the MMN elicited by deviant stimuli in the ignored stream, even when target-related and response-related potentials were accounted for (Woldorff, Hackley, and Hillyard, 1991).

Woldorff and colleagues (1991) suggested that directed attention to one auditory stream initiated a gating function that prevented the extraction of features from stimuli in the ignored stream. Näätänen, Paavilainen, Tiitinen, Jiang, and Alho (1993) replicated the effect of attention on the adult MMN in response to intensity-deviants, but failed to observe an effect of attention on the MMN elicited by frequency-deviants using the same dichotic listening paradigm. It is unparsimonious to suggest that attention might gate the perception of one stimulus parameter (intensity) in the ignored channel, while having no effect on another (frequency). Moreover, the fact that an MMN was observed in response to stimuli in the ignored channel suggests those stimuli have been discriminated, regardless of relative response amplitude. Thus, Näätänen and colleagues (1993) suggested that the presence of a MMN to changes in the ignored stream was sufficient evidence that the features of ignored stimuli had been extracted. They speculated that, rather than preventing the generation of a response, the diversion of attention away from intensity-deviants in the ignored stream prevented amplification of the MMN.

It has been noted that the presentation of deviants of similar intensity to each ear in these dichotic listening paradigms (Woldorff et al., 1991; Näätänen et al., 1993) sets up a competition for MMN generation that is biased by the goals of the subject (i.e. to ignore

one channel in order to detect deviants in the other; Sussman, Winkler, and Wang, 2003). Conversely, the largely different frequency deviants presented to each ear by Näätänen and colleagues (1993) provide no such basis for competition. Thus, Sussman and colleagues (2003) suggest an alternate explanation, in which direct competition between auditory streams, rather than featural differences, explain the differential effects of attention on intensity- and frequency-deviants. In support of this hypothesis, Sussman and colleagues (2003) have shown that presenting dichotic auditory streams containing similar frequency-deviants and dissimilar intensity-deviants can reverse the attentional effects observed by Näätänen et al. (1993).

While these models are of great importance to experimental designs with competing channels, the fundamentally important finding for the current thesis is that the MMN component can be robustly elicited in passive listening conditions. In traditional oddball paradigms for adult listeners, such as the one employed in Chapter 2 of this thesis, passive listening is employed to avoid eliciting attention-dependant components such as the Nd-P300 complex, which is thought to be related to maintenance and rehearsal of an attentional trace, further processing of attended stimuli, or an arousal response (Alho, Woods, and Algazi, 1994). The Nd component overlaps with the MMN both in latency and distribution and can make interpretation of results more difficult in conditions where attention is directed toward auditory stimuli. In electrophysiological studies of infant discrimination, such as the one in Chapter 4, the ability to elicit responses in passive listening conditions may be even more important, as attention is difficult to maintain. Moreover, behavioural evidence of subtle discriminations between

stimuli is difficult to obtain using age-appropriate methods such as the conditioned head-turn procedure. The insensitivity of the MMN component to attention in diotic listening tasks where attention is not deliberately focused elsewhere suggests that modulation of the MMN response following a brief priming period, as seen in Chapter 4, cannot be attributed to attentional effects; it more likely reflects changes in the way that temporal pitch cues are being perceived.

Motivation for the Current Thesis

This thesis focuses on pitch perception in the case where spectral cues are unresolved, and temporal cues provide the main determinants of stimulus pitch. The methods used to date to identify a common pitch processing centre in auditory cortex have not had the temporal resolution to examine the sequence of events involved in the formation of the pitch percept. In Chapter 2, ERPs and source estimation are used to index the feature extraction and pitch discrimination processes involved in perceiving the pitch of stimuli that rely on temporal and spectral cues to different degrees. The aim of this chapter was to test the predictions that arise from functional imaging in humans suggesting that pitch is first represented beyond primary auditory cortex. If the first representation of pitch is indeed located along lateral HG, we would expect that the MMN component elicited by a pitch change would have a source at or beyond this point in the hierarchy of processing. Moreover, if this pitch representation is common to all pitch-evoking stimuli, we would not expect to find significant differences in its latency or source location between stimulus types. Conversely, the properties of those components

that arise from the analysis of basic stimulus features (e.g. P1, N1) may differ significantly between different pitch-evoking stimuli.

Although infants appear to process pitch-evoking stimuli in a manner qualitatively similar to adults, very little is known about the relative contribution of temporal pitch cues. Chapter 3 explores whether 8-month-old infants can behaviourally discriminate between stimuli that contain primarily temporal pitch cues. If d' sensitivity exceeds chance levels in a conditioned head-turn procedure, we can conclude that 8-month-old infants are capable of extracting pitch from the temporal cues present in IRN stimuli. However, previous studies have demonstrated that infant pitch perception relies on salient spectral cues to a greater extent than does adult pitch perception (Clarkson and Clifton, 1995; Clarkson and Rogers, 1995). Thus, if 8-month-olds can indeed perceive temporal pitch cues, it is expected that their performance will be degraded relative to the discrimination of stimuli that contain resolved spectral cues. Experiment 1 of Chapter 4 uses ERPs to examine the cortical representation of temporal cue-based pitch discrimination in 4-month-old and 8-month-old infants. If an infrequent change in the pitch of IRN stimuli elicits a mismatch response, we can conclude that young infants are able to discriminate pitch-evoking stimuli in the absence of spectral cues. Experiment 2 in Chapter 4 examines whether providing a period of priming in which the pitch of IRN stimuli is emphasized through the addition of pure tones affects the infants' ability to discriminate IRN stimuli when these spectral cues are once again removed. If this pitch-priming serves to accentuate the basis on which IRN stimuli can be discriminated, it is hypothesized that the MMN component elicited by a pitch change may be strengthened.

Finally, Chapter 5 provides a summary of these results, and suggests possible future directions.

CHAPTER 2: PREFACE

Butler, B. E., and Trainor, L. J. (2012). Sequencing the cortical processing of pitch-evoking stimuli using EEG and source estimation. *Frontiers in Psychology*, 3(180), 1-13, doi: 10.3389/fpsyg.2012.00180

As discussed in Chapter 1, it is widely accepted that temporal cues arising from the firing pattern of action potentials across a population of auditory nerve fibers make a meaningful contribution to pitch perception. This is particularly evident in instances where spectral pitch cues cannot be resolved by the cochlea. Both spectral and temporal pitch cues appear to be preserved in the auditory nerve, through subcortical structures, and into primary auditory cortex (A1): however, electrophysiological recordings in animal models (Bendor and Wang, 2005) and the majority of functional imaging studies in humans (Griffiths et al., 1998; Patterson et al., 2002; Penagos et al., 2004; Hall et al., 2005; Puschmann et al., 2010) have suggested that pitch is first represented beyond primary auditory cortex, along the anterolateral ridge of Heschl's gyrus (HG).

Collectively, these studies have demonstrated, with substantial spatial resolution, that this portion of lateral HG is responsive to a wide variety of pitch-evoking stimuli. Thus, a convincing case can be made for the presence of a common pitch-processing centre located just beyond A1. However, to date, the processing of pitch-evoking stimuli

has not been undertaken using methods with sufficient temporal resolution to examine the time course of events involved in extracting pitch from stimuli with varied contributions from spectral and temporal cues. This chapter will use event-related potentials (ERPs) and source estimation to examine the feature extraction and object formation processes involved in adult pitch perception, highlighting differences between different pitch-evoking stimuli, and ultimately supporting the notion of a common pitch centre in lateral HG.

Sequencing the cortical processing of pitch-evoking stimuli using EEG analysis and source estimation

Blake E. Butler¹ and Laurel J. Trainor¹

¹ Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada.

Running Title: EEG measures of pitch-evoking stimuli

Keywords: Pitch, EEG, Mismatch Negativity, Iterated Rippled Noise

Correspondence:

Dr. Laurel J. Trainor
McMaster University
Department of Psychology, Neuroscience, & Behaviour
Auditory Development Lab
1280 Main Street West,
Hamilton, ON, L8S4L8, Canada
ljt@mcmaster.ca

Abstract

Cues to pitch include spectral cues that arise from tonotopic organization and temporal cues that arise from firing patterns of auditory neurons. fMRI studies suggest a common pitch center is located just beyond primary auditory cortex along the lateral aspect of Heschl's gyrus, but little work has examined the stages of processing for the integration of pitch cues. Using EEG, we recorded cortical responses to high-pass filtered iterated rippled noise (IRN) and high-pass filtered complex harmonic stimuli, which differ in temporal and spectral content. The two stimulus types were matched for pitch saliency, and a mismatch negativity (MMN) response was elicited by infrequent pitch changes. The P1 and N1 components of event-related potentials (ERPs) are thought to arise from primary and secondary auditory areas, respectively, and to result from simple feature extraction. MMN is generated in secondary auditory cortex and is thought to act on feature-integrated auditory objects. We found that peak latencies of both P1 and N1 occur later in response to IRN stimuli than to complex harmonic stimuli, but no latency differences between stimulus types for MMN. The location of each ERP component was estimated based on iterative fitting of regional sources in the auditory cortices. The sources of both the P1 and N1 components elicited by IRN stimuli were located dorsal to those elicited by complex harmonic stimuli, whereas no differences were observed for MMN sources across stimuli. Furthermore, the MMN component was located between the P1 and N1 components, consistent with fMRI studies indicating a common pitch region in lateral Heschl's gyrus. These results suggest that while the spectral and temporal processing of different pitch-evoking stimuli involves different cortical areas during early

processing, by the time the object-related MMN response is formed, these cues have been integrated into a common representation of pitch.

1. Introduction

Pitch is the perceptual correlate of stimulus frequency, and is important across a number of domains. Pitch can be used to convey prosodic and semantic information in speech (e.g., Frick, 1985; see Moore, 2008 for a review), and is central to music perception. In addition, pitch information provides one of the primary cues for separating overlapping sounds and attributing them to their correct sources in a complex soundscape (Bregman, 1990). Sounds with pitch typically contain energy at a fundamental frequency and at harmonics at integer multiples of the fundamental frequency. Normally, the different frequency components are integrated into a single percept whose pitch corresponds to the fundamental frequency. In fact, even if the energy at the fundamental frequency is not present in the stimulus, the pitch percept corresponding to that frequency remains. This phenomenon is known as the pitch of the missing fundamental, and it emphasizes that pitch extraction is a complex process that depends on spectrotemporal processing of information contained in the sound stimulus. The mechanisms of pitch extraction can be studied by using a variety of stimuli that evoke pitch sensations.

At the level of the cochlea, frequency information is represented in two ways. Physical characteristics of the basilar membrane, such as the variation in stiffness along its length, give rise to a place-based representation (Von Békésy, 1960). Energy at different frequencies causes maximal displacement of the membrane at different locations

along its length, generating a tonotopic organization such that high frequencies are represented near the base of the cochlea, while low frequencies are represented more apically. The mechanical energy contained in the sound wave is converted to an electrical signal via the depolarization of inner hair cells such that tonotopic organization is maintained in the auditory nerve, through subcortical nuclei, and into primary auditory cortex (e.g., see Formisano et al., 2003; Humphries et al., 2010 for reviews). The second frequency representation is based on the periodicity of action potentials in auditory nerve fibers. Because inner hair cells depolarize when the basilar membrane is maximally displaced, firing across a population of auditory nerve fibers occurs at time intervals that represent the inverse of the frequency of the acoustic signal (e.g., Delgutte and Cariani, 1992; Cedolin and Delgutte, 2007).

The neural mechanisms that underlie pitch perception are not yet entirely understood. One class of models is based on place or tonotopic information. For example, Goldstein (1973) described a place-based model in which the harmonic structure of an acoustic stimulus could be matched to a harmonic template to extract pitch. However, such models have at least two inherent limitations. The first is that such a mechanism would require resolution of individual harmonics, but studies have shown that listeners can identify the pitch of complex harmonic stimuli even when the spectral content of those stimuli is confined to the region where harmonics are unresolved (de Boer, 1976). The second limitation of models that rely solely on place cues is that, although the tonotopicity of the basilar membrane may allow for the extraction of place cues at low intensities, these tonotopic maps have been shown to degrade progressively

with increasing stimulus intensity beginning at auditory thresholds (Moore, 2003). A second class of models is based on the temporal firing code, suggesting that the tonotopic organization of the auditory system may not be required for pitch extraction. Rather, these theories suggest that pitch extraction depends on a mechanism that takes advantage of the distribution of inter-spike intervals in the auditory nerve to calculate the fundamental frequency of a complex auditory stimulus (e.g., Cariani and Delgutte, 1996). These theories account for the limitations of spectral models, as the temporal code could represent the fundamental frequency of a harmonic complex sound even if only unresolvable harmonics were present in the stimulus, and such models are more robust to changes in sound level. However, purely temporal models of pitch perception cannot explain why the salience of harmonic stimuli containing resolvable harmonics exceeds the salience of stimuli containing exclusively unresolved harmonics (Houstma and Smurzynski, 1990; Shackleton and Carlyon, 1994). Thus, current working models of pitch extraction typically take into account both the cues arising from the tonotopy of the auditory system and the temporal fine structure in the signal, signal envelope cues, or both (e.g., Moore and Gockel, 2011; Oxenham et al., 2009; Santurette and Dau, 2011).

Although a substantial literature has developed examining these models, how and where spectral and temporal frequency information is combined to give rise to the percept of pitch is still largely unknown. Although frequency is analyzed subcortically, pitch does not appear to be represented until at least the level of primary auditory cortex. For example, Gockel, Carlyon, Mehta, and Plack (2011) failed to find a representation of pitch in the frequency following response (FFR) elicited from subcortical nuclei.

Electrophysiological recordings in animal models (Merzenich and Brugge, 1973; Merzenich et al., 1974) and functional imaging studies in humans (Lauter et al., 1985; Pantev et al., 1988) have shown that primary auditory cortex contains a tonotopic frequency organization, with neurons selectively responsive to high frequency sounds located medially on the superior temporal plane, and those responsive to low frequency sounds located more laterally. Using magnetoencephalography (MEG), Pantev et al. (1996) found that neural populations in A1 were responsive to the fundamental frequency of a complex stimulus containing harmonics of that fundamental, but no energy at f_0 itself. They concluded that the tonotopic organization of A1 is based on periodicity rather than frequency, suggesting that primary auditory cortex contains the first representation of stimulus pitch. However, it has since been suggested that the response recorded by Pantev and colleagues (1996) may have been in response to the distortion product commonly found at f_0 in missing fundamental stimuli, or to a band of masking noise that was centered at this frequency (Walker et al., 2011). Moreover, electrophysiological recordings in awake macaques have failed to find any individual neurons in A1 that are responsive to the pitch of the missing fundamental (Schwarz and Tomlinson, 1990). Instead, it appears that pitch is first represented beyond primary auditory cortex.

Physiological recordings from a region just beyond primary auditory cortex at the anterolateral low-frequency border of A1 in the marmoset have identified a restricted region of cells responsive to pure tones and complex tones of the same pitch (Bendor and Wang, 2005). The search for an equivalent neural population in human cortex has predominantly employed iterated rippled noise (IRN), a class of stimuli created by

delaying a copy of a noise waveform and adding it back to the original noise. When this process is repeated a number of times, the resultant stimulus exhibits temporal regularity at the period of the delay, which is heard as a pitch equal to the inverse of that delay (e.g. a delay of 5 ms results in a perceived pitch of 200 Hz; Patterson et al., 1996). IRN stimuli can be high-pass filtered to remove all of the energy in the region of the resolvable harmonics, leaving a high-frequency spectrum that resembles white noise, although with an increasing number of iterations it will contain some unresolved spectral information. These filtered IRN stimuli still elicit a pitch percept, suggesting that the introduction of temporal regularity by the iterated delay-and-add process provides cues to pitch that are predominantly represented by a temporal code. Thus, IRN stimuli primarily engage the temporal pitch mechanism, and provide a useful comparison for stimuli with more salient spectral cues such as pure tones or complex harmonic sounds.

Collectively, PET and fMRI studies using IRN stimuli to investigate pitch processing in humans have suggested that the pitch center is located along the lateral aspect of Heschl's gyrus (Griffiths et al., 1998; Patterson et al., 2002; Hall et al., 2005). However, it has been rightly suggested that in order for a single area of cortex to be considered a true pitch center, it must respond to all pitch-evoking stimuli. Using a wider variety of stimulus types, Hall and Plack (2009) found that although Heschl's gyrus was indeed responsive to IRN stimuli, responses from this region to other pitch-evoking stimuli were largely variable both between conditions and between listeners. They found that the most overlap in regions of activation across different types of pitch-evoking stimuli was not located along Heschl's gyrus, but rather along the planum temporale (PT).

This led to the suggestion that lateral Heschl's gyrus is selectively responsive to the physical characteristics of IRN rather than the pitch percept it creates, and that the PT is the most likely candidate for a common pitch center. However, recent studies have implicated lateral Heschl's gyrus in response to a variety of pitch-evoking stimuli, including tones in noise, Huggins pitch, and binaural band-pitch stimuli (Puschmann et al., 2010), harmonic complexes (Penagos et al., 2004), and click trains (Gutschalk et al., 2004). Thus, the question of where a common pitch-processing center might exist remains an issue of much debate.

Studies aiming to examine a common pitch representation in human auditory cortex have overwhelmingly relied on the high spatial resolution of functional imaging techniques like fMRI (Patterson et al., 2002; Penagos et al., 2004; Hyde et al., 2008; Hall and Plack, 2009; Puschmann et al., 2010; Barker et al., 2011a). These methods are well-suited to localizing neural populations generating specific responses, but they are unable to reveal detail with respect to the temporal sequence of events involved in the processing of pitch-evoking stimuli. One might expect that early processing of spectral and temporal cues is accomplished in separate areas, but that an integrated pitch percept is later generated in a common area. The finer temporal resolution of electroencephalography (EEG) is better suited to address this question, as the relative timing of components from event-related potentials (ERPs) can help index stages of processing.

Acoustic stimuli evoke a characteristic sequence of ERPs including long-latency, cortically-generated components such as the P1 and N1. The P1 response is observed as a frontally positive-going deflection with a peak occurring roughly 50 ms after the onset of

the acoustic stimulus. The generators of this component are thought to reside in primary auditory cortex, located predominantly in intermediate sections of Heschl's gyrus (Godey et al., 2001; Yvert et al., 2005). Conversely, the N1 component is seen as a frontally negative-going deflection peaking roughly 100 ms after stimulus onset, and is thought to have multiple generators in secondary auditory areas including lateral Heschl's gyrus and PT (Godey et al., 2001; Yvert et al., 2005). In addition to these potentials, there are also components that arise from violations of expectations for sound features, such as the mismatch negativity (MMN). The MMN is elicited in response to an infrequent deviant auditory stimulus occurring within a sequence of repeating standard stimuli, such as when a stimulus with one pitch is repeated from trial to trial, but is occasionally replaced by a stimulus with different pitch (e.g., Näätänen et al., 1978; Picton et al., 2000; Näätänen, et al., 2007). This response is thought to reflect the updating of auditory memory caused by a mismatch between the deviant auditory signal and the sensory-memory of the standard stimulus. Based on a review of the literature, Näätänen and Winkler (1999) have suggested that while N1 likely reflects simple feature detection, the MMN likely operates on perceptible auditory objects (see also Winkler et al., 2005; Takegata et al. 2001; Takegata et al., 1999; Sussman et al., 1998). In the case of pitch perception, this means that while the N1 and MMN components may share overlapping generators (Picton et al., 2000), N1 is likely generated before the formation of a single pitch percept whereas MMN is generated after the formation of a single pitch percept based on all the available cues. Furthermore, the location of MMN activity in the brain is slightly different depending on whether the feature being modified is pitch, duration or sound location

(Giard et al., 1995). Thus, MMN is sensitive enough to employ for examining the location of an integrated pitch percept. By comparing the P1, N1, and MMN elicited by pitch-evoking stimuli containing either primarily spectral or primarily temporal cues, we can determine at what stage of auditory processing these different codes for pitch are combined into a common pitch representation. Moreover, by contrasting the N1 and MMN responses, we can compare the representation of pitch-evoking stimuli in secondary auditory areas before and after the stimulus features are combined into a pitch percept.

An important consideration that is often overlooked in the literature on pitch processing is the effect of pitch saliency on the measurement of neural signatures of the percept. It has been suggested rightly that in order for an area of cortex to be considered a true pitch center, it must show a graded response to acoustic stimuli such that the response increases with increased pitch saliency (Hall and Plack, 2009). A number of researchers have measured fMRI responses to stimuli of varying pitch saliency in an effort to isolate areas of auditory cortex that would satisfy this criterion (Hall et al., 2005; Hall and Plack, 2009; Puschmann et al., 2010; Barker et al., 2011a). Furthermore, using MEG, Krumboltz and colleagues found that pitch onset responses are correlated with pitch saliency even when measured independently from sound onset by transitioning from white noise to iterated noise with a very similar spectrum (Krumbholz et al. 2003). However, studies comparing stimuli with different cues to pitch (spectral, temporal, binaural) typically fail to equate the pitch saliency across the different stimulus types. By failing to equate saliency across different pitch-evoking stimuli, these studies introduce an

additional confound in that different areas of auditory cortex may make different contributions to pitch representation depending on pitch saliency. Indeed, in an attempt to observe the effect of pitch saliency on fMRI recordings, Penagos and colleagues (2004) elicited responses to complex stimuli that differed in their harmonic structure such that some contained harmonics in low spectral regions (340-1100 Hz: the strong pitch condition) while others contained harmonics in high spectral regions (1200-2000 Hz: the weak pitch condition). These contrasts differed in both pitch saliency and spectral content, and control contrasts revealed that activity in anterolateral auditory cortex was sensitive to saliency and not to other differences between these stimuli. This result indicates that the responses of pitch-sensitive regions depend on pitch saliency. Barker and colleagues (2011a) measured discrimination thresholds for each of the stimulus types used in their experiment (pulse trains and complex tones comprised of unresolved harmonics) and determined that they were linearly related to the degree of regularity in the stimulus. However, they did not report whether saliency differed between stimulus types, nor did they control for saliency during fMRI recording.

Thus, to accurately locate areas of overlapping activity across stimuli containing different cues to pitch using functional imaging, the saliency of the pitch percept must be controlled. This consideration is equally important for ERP analysis of pitch perception. Processing-related components such as the MMN have been shown to differ in amplitude and latency depending on the complexity of the stimulus and the difficulty of the discrimination (e.g., see Picton et al., 2000). For example, the latency of the MMN elicited by stimuli that differ in periodicity pitch is longer than the MMN evoked by an

equivalent frequency change in pure tone stimuli, suggesting that the extraction of periodicity-based pitch requires more complex processing than extraction of spectral pitch (Winkler et al., 1997). Picton and colleagues (2000) have suggested, further, that in order to determine the relative timing for discrimination of different types of auditory features, it is essential to measure MMN latencies when the difficulties of the discriminations are equivalent.

In the present study, we compare responses to two different stimulus types that evoke pitch sensations, IRN and complex tone stimuli containing unresolved harmonic components. Furthermore, we matched them for the saliency of their perceived pitch (see methods for details). We compare the amplitude and latency of cortically-generated ERPs including components thought to be generated before (P1, N1) and after (MMN) a pitch percept is extracted. We use dipole-fitting methods to determine the locations of the neural sources of these components, and make comparisons across stimuli to investigate when and where a common representation of pitch might exist in auditory cortex.

2. Materials and Methods

2.1. Participants

Twelve normal hearing adults participated (3 males, mean age = 21.8 years, SD = 1.67). All of the subjects participated in each condition of the experiment. After the nature of the study was described, each subject gave informed consent to participate, and filled in a short survey outlining their hearing and musical history. No subject reported any history of hearing impairment. No participants were highly musically trained (mean

= 3.1 ± 2.8 years of formal instruction) and no significant correlations were found between amount of musical training and any of the measures reported below ($p > 0.2$ for all correlations). The study procedures were approved the McMaster Research Ethics Board.

2.2. Stimuli

Two types of stimuli were created for this experiment: high-pass filtered IRN stimuli and high-pass filtered harmonic complexes. All stimuli were created using Adobe Audition, were 450 ms in duration, employed 10 ms linear onset and offset ramps, and were presented at 70 dBA.

Iterated rippled noise stimuli were created by generating a sample of frozen white noise, and adding it to itself following a delay equal to the inverse of the frequency of the desired pitch percept (for a schematic illustration, see Figure 1).

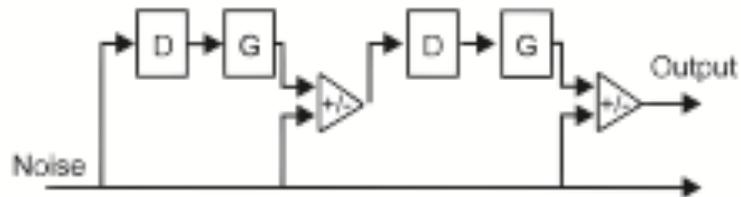


Figure 1. Schematic diagram of the delay (D), gain (G) and add networks used to generate the type of IRN stimuli used in the present study (adapted from Yost, 1996). A sample of white noise is added ($G=1$) to itself following some delay. This process is repeated a number of times, and the resulting stimulus has a perceivable pitch component at the inverse of the delay ($1/D$).

This process was repeated a total of 16 times, as pitch perception for IRN stimuli has been shown to plateau at 16 iterations; further iterations do not increase pitch salience (Patterson et al., 1996). The delay time was set to either 6 ms, in order to create a signal with a perceived pitch of 167 Hz, or 5 ms, in order to create a signal with a perceived pitch of 200 Hz. To ensure equal power across the length of the stimuli, the first and last 100 ms (which contain a gradual increase and decrease in power, respectively, resulting from the iterative delay-and-add process) were removed. The IRN stimuli were then high-pass filtered at 2600 Hz (high-ordered Butterworth filter), representing the 13th harmonic of the 200 Hz stimulus, to remove spectral content in the range of the resolvable harmonics. IRN stimuli have characteristic ripples in their power spectra at low frequencies; high-pass filtering in this way removes these spectral pitch cues.

Complex harmonic stimuli with fundamental frequencies of 167 and 200 Hz were generated by adding harmonics of equal intensity from the fundamental frequency to 20 kHz. These harmonic stimuli were high-pass filtered at 2600 Hz such that the pitch-evoking information was constrained to the same spectral area as in the IRN stimuli. Despite having spectral information limited to the region typically considered to contain primarily unresolved harmonics, these complexes evoked a stronger pitch sensation than the IRN stimuli, likely because their spectral peaks were more defined (see Figure 2). To equate saliency between the stimulus types, these complexes were presented in white noise at an intensity that was determined as follows. Six listeners with musical experience (mean = 10.8 years of formal instruction) were recruited to participate in a pitch-saliency matching exercise. Participants with musical training were chosen as they

readily understood the concept of pitch strength. Given that both stimulus types were novel, it is unlikely that relative pitch saliency would differ between participants with and without musical training. The participants listened to the IRN stimulus and compared it to the complex harmonic stimulus in a white noise masker, each presented over Sennheiser HDA 200 headphones. They were asked to “adjust the intensity of the noise masker until the strength of the perceived pitch was equal between their newly created combination and the pitch-matched IRN stimulus”. They were able to hear each stimulus as many times as needed, and the matching exercise concluded when they determined that they had created the best possible match. The signal-to-noise ratio (SNR) was recorded for each of the two stimuli (30.3 ± 2.1 dB for the 167 Hz stimulus and 28.0 ± 1.5 dB for the 200 Hz stimulus). These SNRs were applied to the level of white noise added to the complex harmonic stimuli. Inter-subject reliability was high for this exercise (standard error = 0.92 dB) suggesting that this type of perceptual matching exercise was reliable for matching pitch saliency between stimulus types.

Power spectra for the IRN and complex tone stimuli are shown in Figure 2, panel A. Following 16 iterations of the process used to generate IRN sounds, some spectral information emerges in their power spectra. However, close examination of the information contained at these spectral peaks reveals that while the harmonic stimuli contain distinct peaks at harmonics of the fundamental frequency, the spectral cues contained in the IRN stimuli are much less clearly defined (Figure 2, panel B).

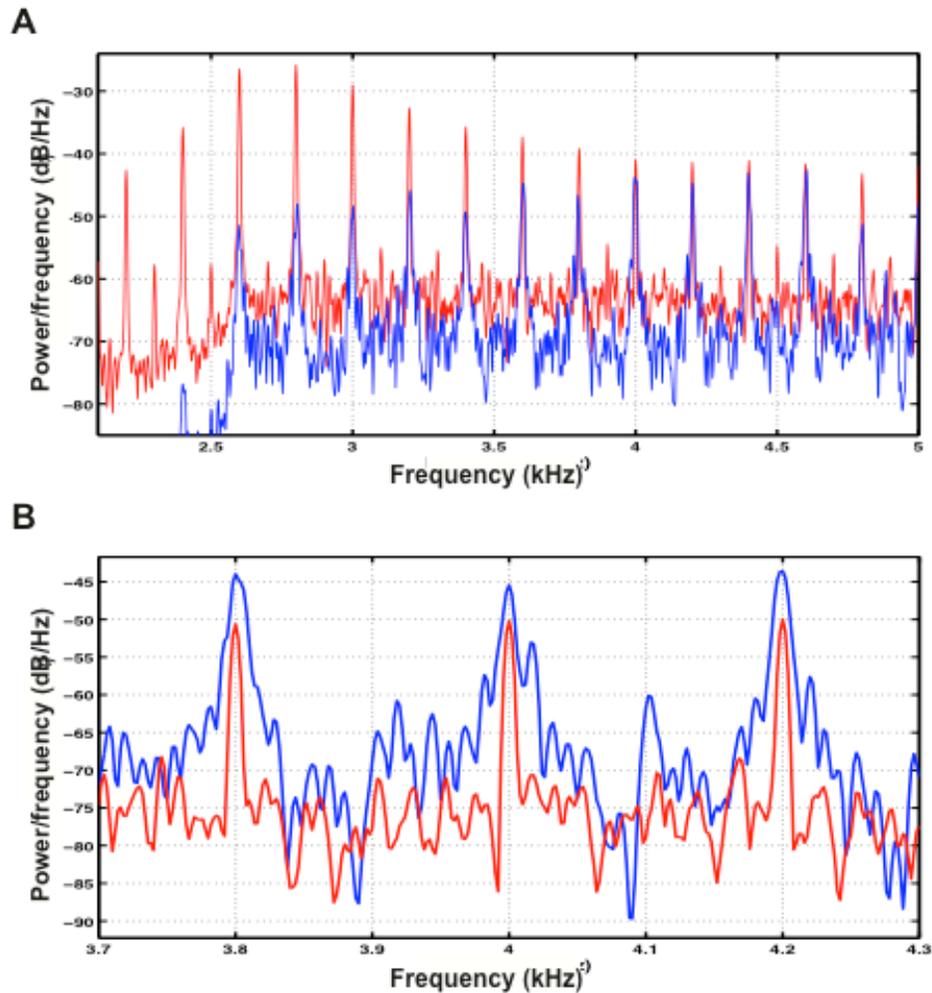


Figure 2. Panel **A** shows the power spectra calculated using Welch's method for the IRN (blue) and complex harmonic (red) 200 Hz stimuli used in the experiment. The spectra were calculated with a 11025- point Hamming window applied with a 10925-point overlap. Panel **B** shows a zoomed-in view of three of the upper harmonics of the stimuli, illustrating the difference between the 'rippled noise' spectrum and a complex waveform made up of discrete harmonics. In particular, the harmonics of the complex waveform are higher in amplitude and more defined in that they show sharper peaks whereas those of the IRN stimuli are more spread out in frequency.

It is also possible that cochlear interactions may introduce some spectral cues to IRN stimuli at the position on the basilar membrane corresponding to the fundamental

frequency. This would not be the case for the harmonic complexes as they were presented in broadband noise to equate for pitch saliency. However, potential difference tones corresponding to the fundamental frequency in the case of the IRN stimuli would be of low amplitude. Furthermore, it has been demonstrated using a low-pass noise masker that the effects of IRN in lateral HG cannot be explained by a response to distortion products (Hall and Plack, 2009).

2.3. Procedure

Participants were seated comfortably in a sound attenuating room, and EEG signals were collected using a 128-channel EGI HydroCel GSN electrode net connected to NetStation 4.2 software. Data from each of the 128 electrodes were digitized at 1000 Hz with a vertex reference and bandpass filter of 0.1-400 Hz, while electrode impedance was maintained below 50 k Ω (the EGI system is high impedance). In each condition, a traditional oddball paradigm was employed whereby the standard stimulus (perceived pitch of 167 Hz) was presented on 85% of trials, and a deviant (oddball) stimulus (perceived pitch of 200 Hz) was presented on the remaining 15% of trials. All stimuli were delivered to a speaker positioned 1 m in front of the listener using a presentation program written in E-prime. Each participant heard both the IRN and complex harmonic stimulus conditions in different blocks of trials, and the order of presentation was randomized. In each of the two blocks, participants heard a total of 750 stimuli (595 standard and 105 deviant) at an SOA of 800 ms. The whole recording lasted roughly 20

minutes. Participants watched a silent movie for the duration of the experiment and were asked to minimize movements and eye blinks during testing phases.

2.4. Analysis

Data were resampled offline at 200 Hz, converted to an average reference, and bandpass filtered between 0.5 and 20 Hz with a roll-off of 24 dB/octave. For each condition, event-related potentials (ERPs) to the deviant and standard stimuli (excluding those standards that immediately followed a deviant) were averaged separately for each individual, using EEProbe software. Filtered continuous data were segmented into 500 ms epochs containing a 100 ms baseline and 400 ms post-stimulus onset. This epoch contained all of the components of interest (P1, N1, and MMN), and eliminated any possible stimulus-offset responses that were not of interest in the present study. EEG responses exceeding $\pm 120 \mu\text{V}$ were considered artifact, thus any epoch in which the response from any electrode exceeded this limit was removed before averaging.

Difference waves were created for each subject by subtracting the standard waveform from the deviant waveform. Finally, data from all subjects were averaged to create grand average waveforms. To visualize the waveforms, 76 electrodes were selected and divided into 4 groupings in each hemisphere, and averaged within each grouping to represent brain responses recorded at the frontal (20 electrodes), central (20 electrodes), parietal (20 electrodes), and occipital (16 electrodes) regions (Figure 3). This virtual electrode montage has been used successfully in previous EEG studies to illustrate the average

responses observed across scalp regions (e.g., Marie and Trainor, 2012; Trainor et al. 2011; He & Trainor, 2009).

The peak latency and amplitude of the P1, N1 and MMN components were measured in each subject at each of the scalp regions defined in Figure 2. P1 and N1 were defined as the first positive and negative-going deflections in the frontal and central regions, respectively (with polarity reversed in the parietal and occipital regions), and were measured in each individual in response to the standard stimulus in each stimulus condition (IRN and complex harmonic). MMN was defined as the largest negative peak between 100 and 300 ms at the frontal and central regions (with polarity again reversed in the parietal and occipital regions), and was measured from the difference wave (deviant-standard) for each individual for each stimulus condition. P1, N1 and MMN were largest at the frontal regions, so six separate analyses of variance (ANOVA) were conducted on the peak amplitude and latency at frontal regions for each component to test whether they varied across condition and hemisphere.

Source analyses were performed to compare the locations of the generators of the P1, N1 and mismatch negativity (MMN) components across stimulus conditions as follows. Regional sources, each containing three orthogonal dipoles, were fit for each component in each subject using the 4-shell ellipsoid model included in the Brain Electrical Source Analysis (BESA) software package. Two sources were fixed to the eyes to account for any residual muscle-related activity that was not eliminated during artifact rejection.

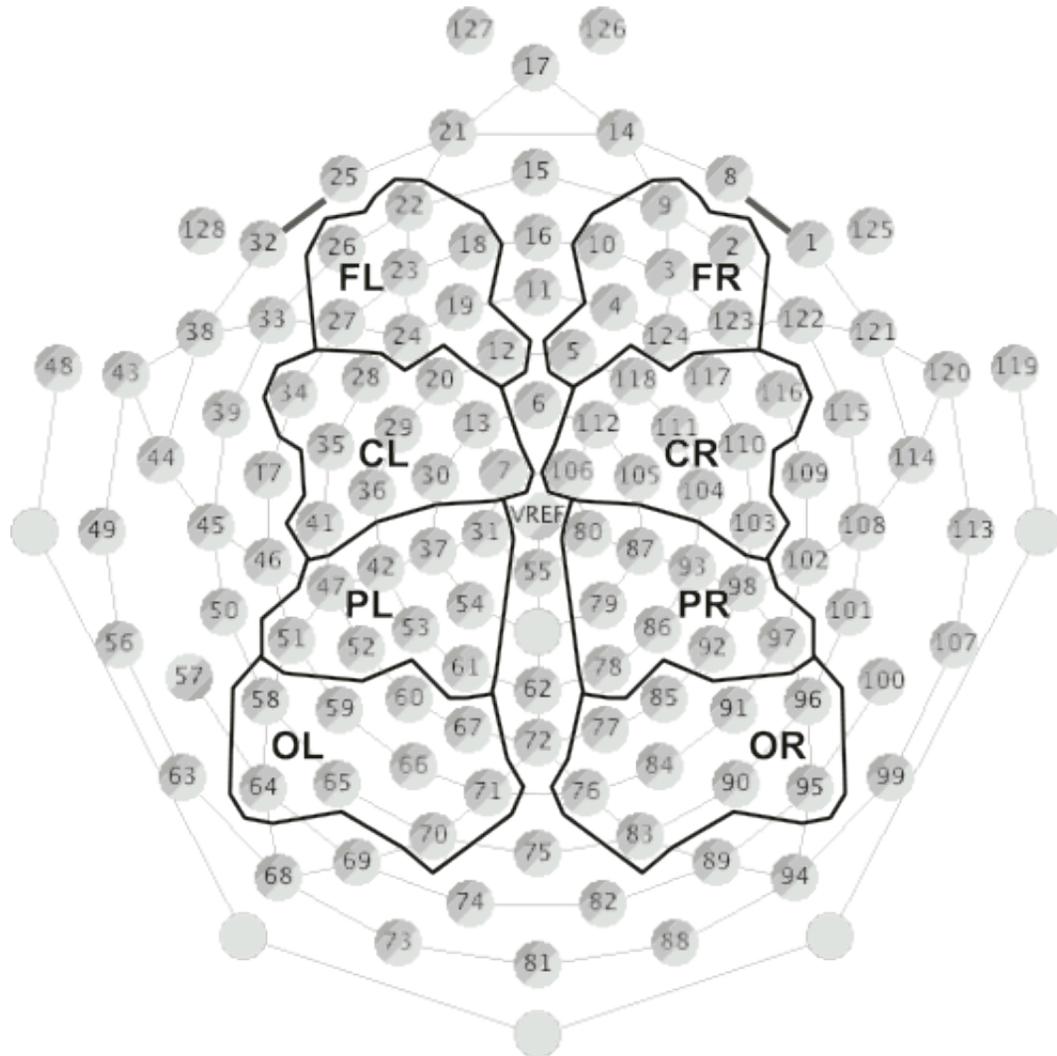


Figure 3. Electrode groupings (see Methods section for details). Seventy-six of 128 electrodes were divided into 4 groups (frontal, central, parietal, and occipital) for each hemisphere. Each group contained between 16 and 20 electrodes that were averaged together to represent EEG responses from that 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.

An additional regional source was then fit in the auditory cortex of each hemisphere, and the two were constrained to be symmetric between hemispheres. The

3D location of the symmetrical sources was determined by the inverse solutions generated by BESA. Participants for whom these regional sources accounted for less than 85% of the variance in the field patterns for one or more components were excluded from further analysis (in the remaining 8 of 12 subjects, this fitting paradigm accounted for a mean variance of 95%, 92%, and 92% in the P1, N1, and MMN components, respectively). It has been suggested that in addition to generators in auditory cortex, frontal areas may contribute to the MMN response, however the fitting of an additional pair of regional sources did not significantly reduce the residual variance.

3. Results

Group average ERP waveforms for each of the electrode groups across stimulus conditions are presented in Figure 4. P1 and N1 reverse polarity between frontal and occipital regions, indicative of activity originating in auditory cortex. Separate repeated-measures ANOVAs examined P1 and N1 peak amplitude and latency in response to the standard stimuli, as measured in the frontal regions of each listener. For each analysis, within-subject factors included stimulus type (IRN, complex) and hemisphere (left, right). The order of stimulus blocks was treated as a between-subjects factor, and was shown to have no effect on the amplitude or latency of the P1 and N1 components. There were no significant effects of hemisphere or stimulus type on P1 or N1 amplitude. For latency, there was also no significant effect of hemisphere for either component ($p=0.95$ and $p=0.37$ for P1 and N1 respectively), but there was a significant effect of stimulus type for both P1 latency [$F(1,11)=19.83$, $p=0.001$] and N1 latency [$F(1,11)=73.04$, $p<0.001$].

Both the P1 and N1 components occurred later in response to the IRN stimuli than to the complex harmonic stimuli, suggesting that the extraction of stimulus features is takes more time in the IRN condition.

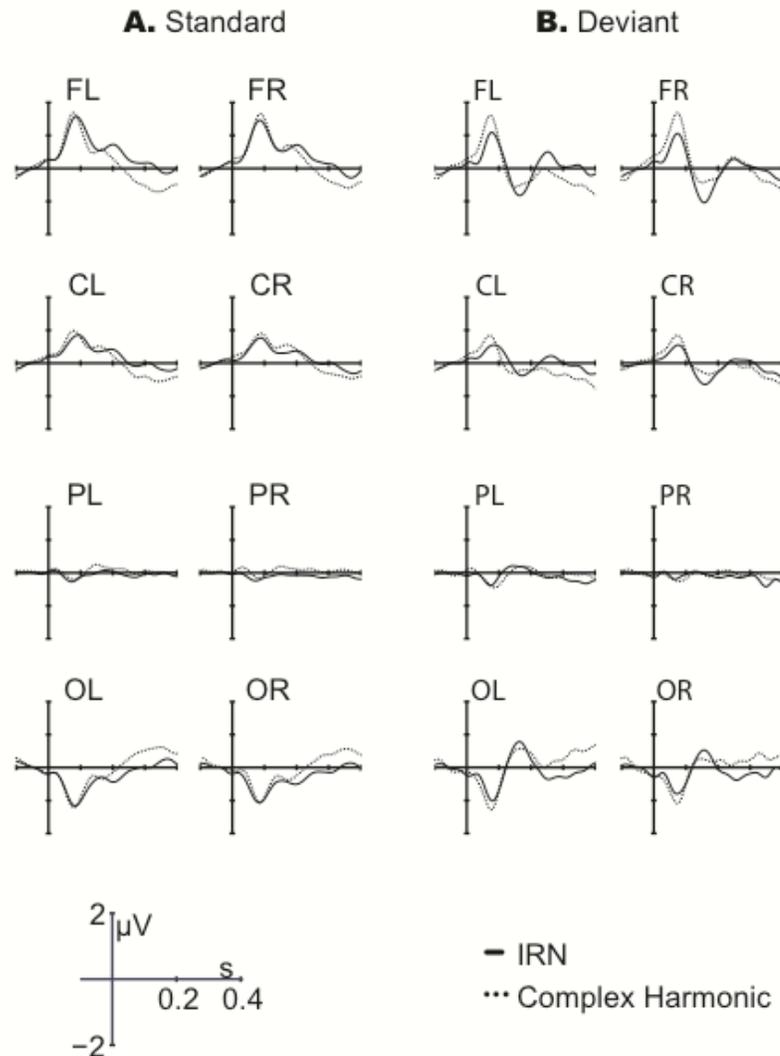


Figure 4. Grand average waveforms in response IRN and complex harmonic stimuli for standard trials (panel **A**) and deviant trials (panel **B**) for all 8 scalp regions.

Analyses were also conducted to examine peak amplitudes and latencies of the P1 and N1 components in response to deviant stimuli, and results were similar. N1 latency again demonstrated a significant effect of stimulus type ($F(1,11) = 24.55, p < 0.001$). For P1 latency, the trend was in the same direction as for standards, but failed to reach significance ($p = 0.28$), most likely because there were few deviant stimuli presented.

Figure 5 compares the grand average difference waves (deviant-standard) showing the MMN elicited by pitch changes in the IRN and complex harmonic stimuli. The front-to-back reversal is consistent with a generator in auditory cortex. Within-subject t-tests revealed latencies at which the difference waveforms were significantly different from zero (i.e., latencies at which the response to the deviant stimulus was significantly different from that to the standard, as shown by the bars beneath the waveforms in Figure 5). Separate repeated-measures ANOVAs examined MMN peak amplitude and latency, as measured in the frontal regions of each listener. For each analysis, within-subject factors were stimulus type (IRN, complex) and hemisphere (left, right). The order of stimulus blocks was again treated as a between-subjects factor, and was shown to have no effect on MMN amplitude or latency. For MMN latency, unlike P1 and N1 components, stimulus type had no significant main effect [$F(1,11) = 0.33, p = 0.58$]. However, the amplitude of the MMN was larger for IRN stimuli than for complex harmonic stimuli [$F(1,11) = 13.89, p = 0.003$].

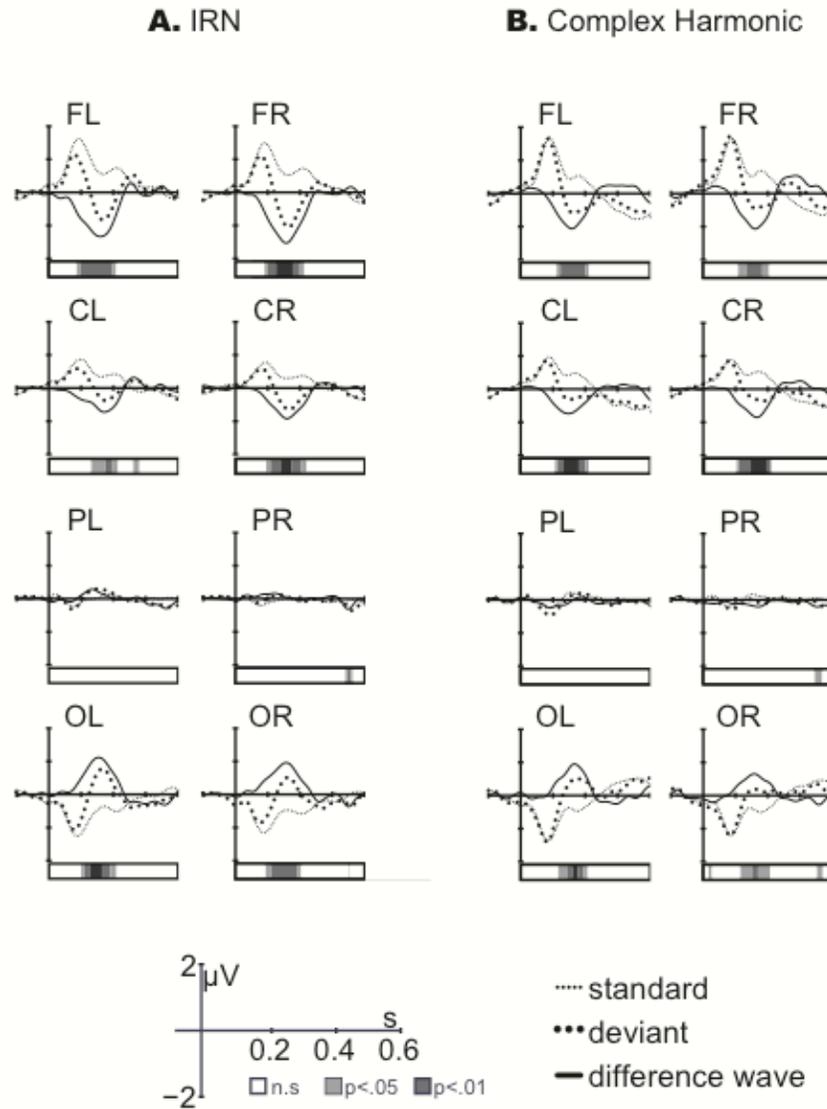


Figure 5. Grand average waveforms (standard response, deviant response, and difference wave [deviant-standard]) for IRN stimuli (panel **A**) and complex harmonic stimuli (panel **B**). The bars under the waveforms show latencies at which the amplitude of the difference wave was significantly different from zero.

The increased MMN amplitude in the IRN condition suggests that more neurons were recruited for analysis of pitch in the IRN case than in the case of the complex harmonic stimuli, but the time taken to form a pitch percept was similar.

Figure 6 depicts the regional sources fit for the P1, N1, and MMN components of the grand average waveforms for each stimulus condition using inverse solutions generated by BESA. The locations of symmetric dipoles fit using this method were recorded for each component in each participant for each stimulus condition. Within-subject t-tests were performed to determine whether the location of these source estimates differed between stimulus types.

While no significant differences were observed in the rostrocaudal ($p=0.37$ and $p=0.31$ for P1 and N1, respectively) or mediolateral ($p=0.80$ and $p=0.77$ for P1 and N1, respectively) planes, both the P1 and N1 components generated in response to the complex harmonic stimuli were determined to have originated from sources dorsal to those generated in response to the IRN stimuli [$t(7)=5.90$, $p=0.001$ for P1, $t(7)=5.56$, $p=0.001$ for N1].

Conversely, source estimates of the MMN component for IRN and complex harmonic stimuli were not shown to differ significantly along the dorsoventral [$t(7)=1.71$, $p=0.13$], rostrocaudal [$t(7)=0.03$, $p=0.97$], or mediolateral planes [$t(7)=0.49$, $p=0.64$]. Thus these source models reveal differences between the P1 and N1 components elicited in response to IRN versus complex harmonic stimuli, while failing to find a difference between stimulus types for MMN.

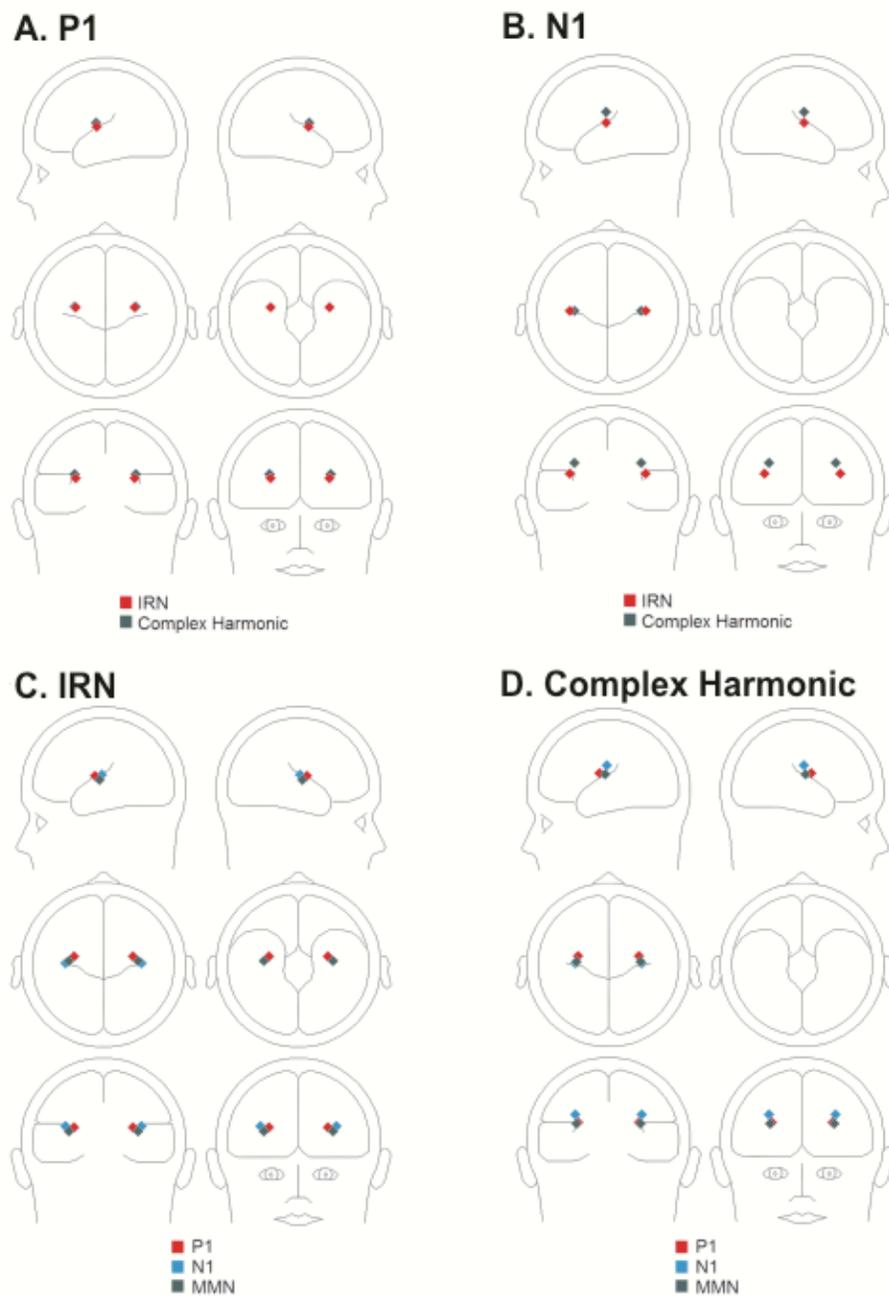


Figure 6. Three-dimensional locations of symmetric regional sources fit using inverse solutions generated by BESA. Panel **A** shows the location of P1 in both stimulus conditions. Panel **B** shows the location of N1 in both stimulus conditions. Panels **C** and **D** show the location of P1, N1 and MMN in the IRN and complex harmonic conditions, respectively.

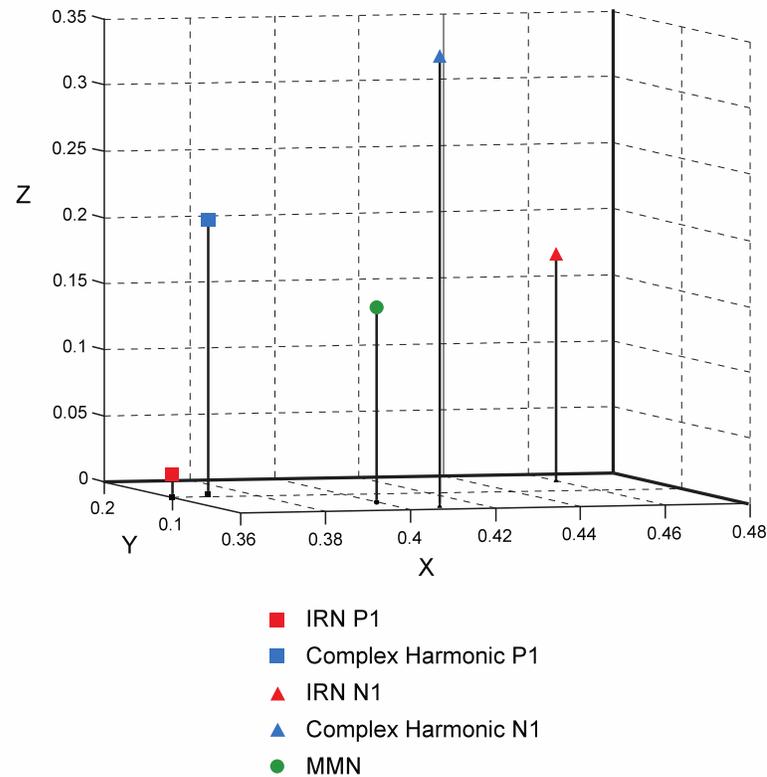


Figure 7. Three-dimensional scatterplot showing the Cartesian coordinates of the P1 and N1 components elicited by the IRN and complex harmonic stimuli, as well as the common MMN elicited by both stimulus types. In the BESA Cartesian coordinate system, the X-axis passes through the LPA and RPA with right being the positive direction. The Y-axis passes through the nasion, with rostral being positive. The Z-axis is orthogonal to the X- and Y-axes, with dorsal being the positive direction. Components shown are all located in the right hemisphere, while symmetric sources in the left hemisphere have been omitted for clarity.

Collectively, these data suggest that the pitch representations elicited by IRN and complex harmonic stimuli are processed in different cortical areas in primary auditory cortex (indexed by P1) and early processing in secondary auditory areas (indexed by N1

response). However, the MMN components elicited in response to each stimulus type had a common source in secondary auditory cortex. While our source analysis does not allow us to determine the precise anatomical location of the MMN response to pitch, the data suggest it is located between the regions generating P1 and N1 (see Figure 7), consistent with a source in the area in lateral Heschl's gyrus that has been revealed by fMRI studies using stimuli with a variety of cues to pitch (Griffiths et al., 1998; Patterson et al., 2002; Puschmann et al., 2004; Gutschalk et al., 2004; Penagos et al., 2004; Hall et al., 2005).

4. Discussion

In the current experiment, we recorded participants' ERP responses to occasional pitch changes in high-pass filtered IRN and complex harmonic stimuli matched for perceptual pitch saliency. We used the fact that different ERP components reflect different stages of processing to investigate when and where in auditory cortex different cues to pitch might be integrated into a common representation. Particularly useful for this question is the fact that P1 (around 50 ms after stimulus onset, originating in primary auditory cortex) and N1 (around 100 ms after stimulus onset, originating in secondary auditory cortex) responses are thought to reflect early feature processing whereas MMN (around 150 ms) is thought to operate on perceptible auditory objects (Näätänen and Winkler, 1999). Specifically, we found no difference in the amplitude of P1 and N1 responses across stimuli, but both P1 and N1 responses occurred earlier for complex harmonic than IRN stimuli, suggesting somewhat different processes at work for different

cues to pitch at these stages of processing. The converse was true for MMN, with responses to pitch changes not different in latency, but larger for IRN than complex harmonic stimuli, suggesting that although more neurons might be recruited for the analysis of sounds with less spectral pitch information, the timing of the formation of an integrated pitch percept appears to be similar for sounds regardless of spectral and temporal cues. In any case, the most important findings were in the results of the regional source estimation analyses. As expected, P1 sources were located more rostral and dorsal than sources for N1, consistent with P1 originating in primary auditory cortex and N1 in secondary auditory cortex. Furthermore, for both P1 and N1, source locations were highly significantly different for the IRN and complex harmonic stimuli. This difference suggests that at these pre-object integration stages of processing, different features of the stimuli give rise to processing in somewhat different brain regions. On the other hand, there was no significant difference between the source locations of the MMN for the IRN and complex harmonic stimuli. This suggests that at this stage of processing an integrated pitch perception is formed, regardless of the particular pitch cues present. It is important to note that the P1 and N1 components elicited in response to standard and deviant stimuli of the same stimulus type are expected to be very similar. It is possible that they reflect in part, processing of non-pitch features (Barker et al., 2011b). However, neural activity related to these features should be largely eliminated in the difference wave, and the MMN component should reflect primarily pitch processes. Therefore, the fact that we see no latency or source localization difference for the MMN response between stimulus types, despite different latencies and source localizations for the P1 and

N1 components suggests a common process of object formation. Furthermore, the MMN response to pitch change was located between the locations of the P1 and N1 generators, which is consistent with fMRI studies indicating the existence of a pitch center located adjacent to primary auditory cortex along the lateral aspect of Heschl's gyrus. In sum, different cues to pitch appear to be processed in somewhat different regions during early cortical processing, but by about 155 ms after stimulus onset, an integrated pitch percept is formed regardless of particular cues to pitch, as indexed by the MMN response.

The fine temporal resolution of EEG recording allows examination of individual stages of auditory processing that differ on a millisecond scale. However, the spatial resolution of this method is not as good as that of functional imaging methods such as fMRI, as the locations of cortical activity must be modeled on the basis of electrical field distributions at the scalp. Source location estimation is accomplished through inverse modeling, in which source locations are estimated and the pattern of activity that they would generate at the head surface is calculated. An iterative process is employed whereby the sources are moved in location and orientation until the surface pattern best matches that measured during the experiment (see Baillet et al., 2001; Michey et al., 2004; Hallez et al., 2007 for reviews). With such a process, there are multiple solutions for any particular component and stimulus condition. Consequently, the source estimate represents the solution that accounts for the greatest amount of variance in the data rather than a direct recording of electrical activity. Although this does offer a limitation in terms of localizing component sources, the dipole fitting method used in the present study was sufficient to spatially resolve responses from primary and secondary auditory cortices (as

represented by the P1 and N1 components). Moreover, it was sufficient to show highly significant differences within auditory areas across stimulus types (e.g., differences in both P1 and N1 sources for IRN compared to complex harmonic stimuli). Thus, finding a common source location for the MMN component across stimulus types is unlikely due to insufficient spatial resolution.

The pitch salience of the stimuli in the current experiment was perceptually equated, but there were minor differences in the spectral content as the white noise masker in the complex harmonic condition contained energy at low frequencies that was not present in the IRN stimulus. However, it has been demonstrated previously that pitch-related areas of auditory cortex respond to the salience of the perceived pitch rather than to other stimulus differences. For example, Penagos and colleagues (2004) contrasted fMRI responses to stimuli that differed either in spectral content alone or in both spectral content and pitch saliency and found that spectral inequality did not result in any differential activation in lateral Heschl's gyrus. Moreover, based on the tonotopic organization of auditory cortex, one would expect slight differences in spectral content to shift source estimates of the P1 and N1 components mediolaterally (Humphries et al., 2010), rather than dorsoventrally as observed in the current study. Thus, the differences in P1 and N1 sources revealed in the present study most likely reflect differential processing of sound features, such as the relative salience of different pitch cues, rather than minor differences in spectral content.

There is some evidence that the right auditory cortex is specialized for the processing of pitch. Lesion studies suggest that patients with surgical excisions of the

right, but not left auditory cortex show deficits in perceiving the missing fundamental (Zatorre, 1988), in processing complex spectral structure (Sidtis and Volpe, 1988) and in discriminating melodic pitch patterns (Zatorre, 1985, 1988). These findings have been supported by recent functional imaging studies demonstrating that the auditory areas in the right hemisphere are selectively activated in both pitch perception (Hyde et al., 2008) and production tasks (Perry et al., 1999). Zatorre (2002) has suggested that this functional asymmetry may represent a tradeoff in processing such that the auditory cortical systems in the two hemispheres have evolved complementary specializations. Our study did not demonstrate any appreciable hemispheric differences, which at first glance seems discordant with these findings. However, there are two important experimental parameters that warrant consideration here: the difficulty of the discrimination and the complexity of the task. In an fMRI study of pitch processing, Hyde and colleagues (2008) demonstrated a right hemisphere advantage in discriminating pitch changes between 6.25 and 200 cents. However, they noted that BOLD response in the left planum temporale increased for the larger pitch changes. Hyde et al. suggested that both hemispheres likely contribute to pitch discrimination, with the right hemisphere advantage being most evident for fine frequency discriminations. Thus, the finding in the present study of a bilateral response is consistent with Hyde et al., as we used a relatively large pitch change (167-200 Hz represents a change of more than 300 cents). It is also possible that the right hemisphere advantage in pitch processing is task dependant. For example, Johnsrude et al. (2000) found that patients with excisions that encroached on lateral Heschl's gyrus in the right hemisphere were impaired in detecting the direction of

a pure tone frequency change, but simple frequency discrimination for those same stimuli was unaffected in these patients. The passive pitch discriminations measured in our experiment did not require detection of the direction of pitch change, and therefore likely would have diminished any right hemisphere advantage.

Modern theories of pitch extraction typically include both spectral and temporal pitch cues, but the details of how and when the two pitch codes are combined into a common representation of the percept are still largely unknown. Temporal integration windows widen throughout the ascending auditory pathway (Walker et al., 2011) and, accordingly, it has been demonstrated that the conversion of temporal pitch cues into a more stable code occurs at or before the level of primary auditory cortex (Griffiths et al., 1998), possibly in the inferior colliculus (Langner and Schreiner, 1988). On the other hand, spectral pitch cues are well represented in the tonotopic organization of the auditory system that exists at least into primary auditory cortex (Humphries et al., 2010). While fMRI is well-suited to localize pitch-responsive areas of cortex with good spatial resolution, it can tell us little about the temporal sequence involved in pitch processing. In the present study, we have used EEG to show that early cortical processing of different sound features, such as the relative salience of different pitch cues, is accomplished in somewhat different areas in primary and secondary auditory cortex, while at a later stage (about 150 ms after stimulus onset), an integrated pitch perception appears to emerge in a common processing area. In future studies it would be interesting to use measures with good temporal resolution such as EEG or MEG in order to compare the extraction of pitch percepts across a wide range of stimuli that contain a variety of cues to pitch.

Acknowledgements

The authors thank Elaine Whiskin for help with data collection, Chris McAllister for help preparing figures, and Dan Bosnyak for helpful comments during manuscript preparation.

This research was funded by a grant to LJT from the Canadian Institutes of Health Research.

References

- Baillet, S., Moshier, J.C., and Leahy, R.M. (2001). Electromagnetic brain mapping. *IEEE Signal Proc. Mag.* 18, 14-30.
- Barker, D, Plack, C.J., and Hall, D.A. (2011a). Human auditory cortical responses to pitch and to pitch strength. *Neuroreport* 22, 111-115.
- Barker, D, Plack, C.J., and Hall, D.A. (2011b). Reexamining the evidence for a pitch-sensitive region: A human fMRI study using iterated ripple noise. *Cereb. Cortex*, doi:10.1093/cecor/bhr065.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161-1165.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.
- Cariani, P., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch saliency. *J. Neurophysiol.* 76, 1698-1716.
- Cedolin, L., and Delgutte, B. (2005). Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve. *J. Neurophysiol.* 94, 347-362.

- Cedolin, L., and Delgutte, B. (2007). “Spatio-temporal representation of the pitch of complex tones in the auditory nerve,” in *Hearing – From Sensory Processing to Perception*, ed. B. Kollmeier, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Berlin: Springer-Verlag), 61-70.
- de Boer, E. (1976). “On the “residue”, and auditory pitch perception,” in *Handbook of Auditory Physiology*, ed. W.D. Keidel, and W.D. Neff (Berlin: Springer-Verlag), 479-583.
- Delgutte, B., and Cariani, P. (1992). “Coding of the pitch of harmonic and inharmonic complex tones in the interspike intervals of auditory nerve fibers,” in *The Processing of Speech*, ed. M.E.H. Schouten (Berlin: Mouton-De Gruyer), 37-45.
- Formisano, E., Kim, D., Di Salle, F., van de Moortele, P., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40, 859-869.
- Frick, R.W. (1985). Communicating emotion: The role of prosodic features. *Psychol. Bull.* 97, 412-429.
- Giard, M.H., Lavikainen, J., Reinikainen, K., Bertrand, O., Pernier, J., and Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model study. *J. Cognitive Neurosci.* 7, 133-143.
- Gockel, H. E., Carlyon, R. P., Mehta, A., and Plack, C. J. (2011). The frequency following response for dichotic pitch stimuli: No evidence for pitch encoding. *J. Acoust. Soc. Am.* 129, 2592-2592.

- 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. *Clin. Neurophysiol.* 112, 1850-1859.
- Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54, 1496-1516.
- Griffiths, T.D., Buchel, C., Frackowski, R.S.J., and Patterson, R.D. (1998). Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.* 1, 422-427.
- Gutschalk, A., Patterson, R.D., Scherg, M., Uppenkamp, S., and Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage* 22, 755-766.
- Hall, D.A., Barrett, D.J.K., Akeroyd, M.A., and Summerfield, A.Q. (2005). Cortical representations of temporal structure in sound. *J. Neurophysiol.* 94, 3181-3191.
- Hall, D.A., and Plack, C.J. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576-585.
- Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K.P., Fabri, S.G., Huffel, S.V., and Limahieu, I. (2007). Review on solving the forward problem in EEG source analysis. *J. Neuroeng. Rehabil.* 4.
- He, C., and Trainor, L. J. (2009). Finding the pitch of the missing fundamental in infants. *J. Neurosci.* 29, 7718-7722.
- Houstma, A.J.M. and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304-310.

Humphries, C., Liebenthal, E., and Binder, J.R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage* 50, 1202-1211.

Hyde, K.L., Peretz, I., and Zatorre, R.J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632-638.

Johnsrude, I.S., Penhune, V.B., and Zatorre, R.J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155-163.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., and Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing centre in Heschl's gyrus. *Cereb. Cortex* 13, 765-772.

Langner, G., and Schreiner, C.E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60, 1799-1822.

Lauter, J.L., Herscovitch, P., Formby, C., and Raichle, M.E. (1974). Tonotopic organization in human auditory cortex revealed with positron emission tomography. *Hearing Res.* 20, 199-205.

Marie, C., and Trainor, L. J. (2012). Development of simultaneous pitch encoding: Infants show a high voice superiority effect. *Cereb. Cortex*
doi:10.1093/cercor/bhs050.

Merzenich, M.M., and Brugge, J.F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res.* 50, 275-296.

Merzenich, M.M., Knight, P.L., and Roth, G.L. (1974). Representation of cochlea within auditory cortex in the cat. *J. Neurophysiol.* 38, 231-249.

Micheyl, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., and de Peralta, R.G.

(2004). EEG source imaging. *Clin. Neurophysiol.* 115, 2195-2222.

Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). London:

Emerald Group Publishing

Moore, B.C.J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people.

J. Assoc. Res. Otolaryngol. 9, 399-406.

Moore, B.C.J., and Gockel, H.E. (2011). Resolvability of components in complex tones

and implications for theories of pitch perception. *Hearing Res.* 276, 88-97.

Näätänen, R., Gaillard, A.W.K., and Mantysalo, S. (1978). Early selective-attention effect

on evoked potential reinterpreted. *Acta Psychol.* 42, 313-329.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity

(MMN) in basic research of central auditory processing: A review. *Clin.*

Neurophysiol. 118, 2544-2590.

Näätänen, R., and Winkler, I. (1999). The concept of auditory stimulus representation in

cognitive neuroscience. *Psychol. Bull.* 125, 826-859.

Oxenham, A.J., Micheyl, C., and Keebler, M.V. (2009). Can temporal fine structure

represent the fundamental frequency of unresolved harmonics. *J. Acoust. Soc. Am.*

125, 2189-2199.

Pantev, C., Elbert, T., Ross, B., Eulitz, C., and Terhardt, E. (1996). Binaural fusion and

the representation of virtual pitch in human auditory cortex. *Hearing Res.* 100,

164-170.

- Pantev, C., Hoke, M., Lehnertz, K., Lutkenhoner, B., Anoglanakis, G., and Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroen. Clin. Neuro.* 69, 160-170.
- Patterson, R.D., Handel, S., Yost, W.A., and Datta, A.J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.* 100, 3286-3294.
- Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767-776.
- Penagos, H., Melcher, J.R., and Oxenham, A.J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* 24, 6810-6815.
- Perry, D.W., Zatorre, R.J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A.C. (1999). Localization of cerebral activity during simple singing. *Neuroreport* 10, 3979-3984.
- Picton, T.W., Alain, C., Otten, L., and Ritter, W. (2000). Mismatch negativity: Different water in the same river. *Aud. Neuro-otol.* 5, 111-139.
- Puschmann, S., Uppenkamp, S., Kollmeier, B., and Thiel, C.M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage* 49, 1641-1649.
- Santurette, S., and Dau, T. (2011). The role of temporal fine structure information for the low pitch of high-frequency complex tones. *J. Acoust. Soc. Am.* 129, 282-292.

- Schwarz, D.W., and Tomlinson, R.W. (1990). Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J. Neurophysiol.* 64, 282-298.
- Shackleton, T.M. and Carlyon, R.P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95, 3529-3540.
- Sidtis, J.J., and Volpe, B.T. (1988). Selective loss of complex-pitch of speech discrimination after unilateral lesion. *Brain Lang.* 34, 235-245.
- Sussman, E., Gomes, H., Noursak, J. M. K., Ritter, W, and Vaughan, H. G. Jr. (1998). Feature conjunctions and auditory sensory memory. *Brain Res.*, 793, 95-102.
- Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., and Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neural populations. *Neuroreport*, 12, 525-529.
- Takegata, R., Paavilainen, P., Näätänen, R., and Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neurosci. Letters*, 2, 109-112.
- 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 Topog.* 24, 192-203.
- Von Békésy, G. (1960). *Experiments in Hearing*. New York: McGraw-Hill.
- Walker, K.M.M., Bizley, J.K., King, A.J., Schnupp, J.W.H. (2011). Cortical encoding of pitch: Recent results and open questions. *Hearing Res.* 271, 74-87.

Winkler, I., Czigler, I., Sussman, E., Horváth, J., and Balázs, L. (2005). Preattentive binding of auditory and visual stimulus features. *J. Cog. Neurosci.* 17, 320-339.

Winkler, I., Tervaniemi, M, and Näätänen, R. (1997). Two separate codes for missing-fundamental pitch in the human auditory cortex. *J. Acoust. Soc. Am.* 102, 1072-1082.

Yost, W.A. (1996). Pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* 100, 3329-3335.

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.

Zatorre, R.J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia* 23, 31-41.

Zatorre, R.J. (1988). Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566-572.

Zatorre, R.J. (2002). Neural specializations for tonal processing. *Ann. N.Y. Acad. Sci.* 930, 193-210.

AFTERWARD

My committee provided some interesting comments on the nature of an “auditory object” and the interpretation of the MMN data in this regard. As these comments came subsequent to the publication of the manuscript comprising Chapter 2, I have written this short commentary to address these points. Chapter 2 uses ERP components to demonstrate that different pitch-evoking stimuli are processed in distinct cortical areas during feature extraction, but not following the formation of an auditory object. However, a definition of the term ‘auditory object’ was not provided. Exactly what constitutes an object is debatable. For the purposes of this chapter, Bregman’s (1990) “mental description of a sound source in the environment rather than the source itself or the sounds it emits” provides a useful working definition.

The relationship between the MMN component and auditory objects is complex. While Chapter 2 relies on the observation that an MMN is elicited following object formation (e.g. see Ritter et al. 2000), it is also pointed out that the scalp distribution of the MMN differs depending on the stimulus feature eliciting the response (e.g. see Paavilainen et al. 1991). For the purposes of this chapter, these potentially discordant ideas can be interpreted as follows: an auditory object has distinct features (pitch, duration, etc.) that are constructed from information present in the stimulus (harmonic structure, onset/offset, etc.). The MMN is elicited following object formation, with the feature that differs from the representation of the standard stimulus influencing the location of MMN generation. Thus, the common source for the MMN elicited by

complex harmonic and IRN stimuli suggests that the feature being changed (i.e. pitch) is perceptually the same regardless of stimulus type.

A second question concerns whether there is enough power to detect differences in MMN location, should they exist, between complex harmonic and IRN stimuli. Given that previous studies have detected differences in MMN location between sounds of low and high pitch (Tiitinen et al., 1994), and given that we found robust differences in P1 and N1 locations between harmonic complex and IRN stimuli, the present findings suggest that any potential differences in location that might exist between these stimulus types are likely minimal and that they involve overlapping neural populations.

CHAPTER 3: PREFACE

Butler, B. E., Folland, N.A., and Trainor, L. J. (submitted). Development of pitch processing: Infants' discrimination of iterated rippled noise stimuli with unresolved spectral content. *Hearing Research*, Manuscript ID HEARES-D-12-00290

Chapter 2 provided electrophysiological evidence that adult listeners can discriminate between IRN stimuli of different pitches. Moreover, the mismatch response generated by an occasional change in IRN pitch appeared to originate in the same area of cortex as similar responses to other pitch-evoking stimuli. This suggests that adult listeners are processing stimulus pitch in a similar way, regardless of whether the cues to pitch are spectral, temporal, or a combination of the two.

A variety of studies have examined pitch processing in infants. Many of these have suggested that infants process pitch-evoking stimuli in a qualitatively adult-like manner beginning at a very young age. For example, behavioural experiments have shown that eight-month-old infants can discriminate complex stimuli that differ in pitch by 20% (Clarkson & Clifton, 1985). Additionally, He and Trainor (2009) demonstrated that a cortical representation of pitch processing emerges between 3 and 4 months of age. However, each of these studies employed harmonic stimuli that provided robust spectral pitch cues; the question of how infants perceive pitch-evoking stimuli in the absence of spectral cues has not been addressed.

This chapter describes a visually-reinforced conditioned head-turn procedure used to examine whether 8-month-old infants can perceive a change in the pitch of IRN stimuli that contain no resolved spectral pitch cues. In Experiment 1, a traditional paradigm was used in which the target stimulus used during the training phase was presented at a higher amplitude than the background stimulus. In Experiment 2, a pure tone was added to the target IRN stimulus during the training phase in order to emphasize the pitch component of the IRN. Collectively, the experiments of Chapter 3 demonstrate that, following a period of pitch-priming, 8-month-old infants can discriminate between IRN stimuli of different pitches. However, the low sensitivity values reported in this chapter suggest that their perception of these stimuli is greatly impaired, relative to stimuli that contain salient spectral pitch cues.

Submitted to Hearing Research

September 25, 2012

**Development of pitch processing: Infants' discrimination of iterated rippled noise
stimuli with unresolved spectral content**

Blake E. Butler^a, Nicole A. Folland^a, and Laurel J. Trainor^a

^a Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada.

Correspondence:

Dr. Laurel J. Trainor
McMaster University
Department of Psychology, Neuroscience, & Behaviour
1280 Main Street West,
Hamilton, ON, Canada, L8S 4L8
ljt@mcmaster.ca

Abstract

Sound frequency is extracted at the level of the cochlea, and is represented by two neural codes: a spectral (place) code that is maintained by tonotopic maps extending into primary auditory cortex, and a temporal code based on the periodicity of action potentials in auditory nerve fibres. To date, little work has examined infants' ability to perceive pitch when spectral content cannot be resolved by cochlear filters; the present experiments do so using high-pass filtered iterated rippled noise (IRN) stimuli. Using a conditioned head-turn paradigm, 8-month-old infants were shown to discriminate above chance levels a change in the perceived pitch of such high-passed filtered IRN stimuli from 167 Hz to 200 Hz, but only when first exposed to a training IRN target stimulus that emphasized pitch through the addition of a sine wave tone to the IRN stimulus at the frequency of the perceived pitch. However, even after this period of pitch priming, performance was quite poor relative to previous studies using stimuli with resolved spectral content. These results support the idea that 8-month-olds can perceive pitch when only unresolved spectral content is present in the stimulus, but that such processing is not yet robust.

Keywords: Infant; Pitch; Development; Iterated Rippled Noise; Conditioned Head-Turn

IRN = Iterated rippled noise
3AFC = 3-Alternative, forced-choice

1. Introduction

The accurate formation of a pitch percept is important for identifying sounds in the environment, and for acquiring language and music. Pitch can be used to convey prosodic and semantic information in speech (e.g., Frick, 1985; see Moore, 2008 for a review), and is central to music perception (see Koelsch & Siebel, 2005; McDermott & Oxenham, 2008; Trainor & Corrigan, 2010 for reviews). Pitch information also provides a primary cue for separating overlapping sounds and correctly attributing them to their sources (Bregman, 1990). Sounds with pitch are complex in that they typically contain energy at a fundamental frequency and at harmonics that are at integer multiples of the fundamental frequency. The formation of a single pitch percept from this frequency information is a complex process that depends on spectrotemporal processing of the sound stimulus.

A number of studies examining how infant listeners perceive pitch-evoking stimuli have demonstrated that infants are capable of rather sophisticated pitch discriminations. For example, behavioral evidence indicates that 3- to 6-month-old infants show pure tone difference limens as low as 2% at 1000 Hz (Olsho, Schoon, Sakai, Terpin, & Sperduto, 1982). Eight-month-old infants have been shown to discriminate complex stimuli that differ by 20% (e.g. 160 and 200 Hz; Clarkson & Clifton, 1985), although this value represents a commonly used interval and is likely well above the threshold of discrimination. Similar pitch changes have been used to demonstrate that infants, like adults, are sensitive to the pitch of the missing fundamental (a stimulus in which a pitch percept is formed based on harmonic spacing, despite a lack of any energy

at that fundamental frequency; Clarkson & Clifton, 1985). Montgomery and Clarkson (1997) demonstrated further that the addition of a low-frequency noise masker does not impair the ability of 8-month-old infants to discriminate missing-fundamental stimuli. Thus, as in adults, infants' ability to perceive the pitch of the missing fundamental is not due to low-frequency combination tones resulting from non-linearities in the inner ear. Electrophysiological measures suggest that cortical representations of the pitch of the missing fundamental emerge between 3 and 4 months of age (He & Trainor, 2009). Furthermore, Clarkson and Clifton (1995) demonstrated that 7-month-old infants can discriminate pitch changes in inharmonic complexes, and that, as in adults, performance is related to the degree of inharmonicity in a manner that is qualitatively similar to adult performance. Collectively, these results suggest that infants, like adults, use the harmonic structure of complex tones to determine their pitch.

Although infants appear to process pitch-evoking stimuli in an adult-like manner by 4 months of age, little research has examined how infant listeners perceive stimuli containing limited spectral pitch cues. This question is a relevant one, as extracting pitch given limited spectral content is necessary for communication by telephone (where information is carried by a limited passband), or for the extraction of pitch-based information in the presence of masking noises of various spectra as found in the everyday environment. Complex tones containing harmonics of a fundamental frequency can be high-pass filtered such that spectral content is limited to the region beyond which individual harmonics can be resolved by the basilar membrane. While the spacing of harmonics in a complex tone is linear, the tonotopic organization of the basilar membrane

is logarithmic. As such, the characteristic places along the membrane corresponding to the lower harmonics of a complex sound are sufficiently spaced such that each harmonic falls within its own frequency channel. However, at higher harmonics, the bandwidth of frequency channels on the basilar membrane exceeds the spacing of harmonics such that multiple harmonics fall into the same frequency channel, activating the same cochlear nerve fibers. These harmonics are considered to be beyond the limit of cochlear resolvability (e.g. Moore, 2008).

In adults, pitch salience is greater for harmonic stimuli that contain spectrally resolved components compared to those that contain only high, unresolved components (Ritsma, 1962). Moreover, for stimuli that contain both resolved and unresolved components, the resolved components (in particular harmonics three through five) make the greatest contribution to the pitch percept (Plomp, 1967; Ritsma, 1967). The dominance of resolved harmonics is evident in performance on pitch-related tasks. For example, the performance of adult listeners on pitch interval-matching tasks degrades (Houtsma & Goldstein, 1972; Houtsma & Smurzynski, 1990) and difference limens for pitch increase (Houtsma & Smurzynski, 1990) as the lowest component present in a complex harmonic stimulus is increased. However, although low-frequency, resolved components may dominate pitch perception, high-frequency, unresolved components are sufficient to elicit a pitch percept in adult listeners. For example, performance on a pitch interval-matching task remains well above chance, even for harmonic stimuli that contain no resolvable components (Houtsma & Smurzynski, 1990).

Based on a number of studies reporting qualitatively adult-like pitch perception in infants by 4 months of age (Clarkson & Clifton, 1985, 1995; He & Trainor, 2009; Montgomery & Clarkson, 1997), it is reasonable to predict that infants may also perceive a pitch percept for stimuli containing only unresolvable pitch cues. However, one previous study found that although 7- to 8-month-old infants were able to successfully categorize complex stimuli containing resolvable harmonics according to pitch, there was no evidence that they could do so when only unresolvable harmonics were present (Clarkson & Rogers, 1995). In the present study we examined infants' ability to detect pitch changes in the absence of information in the region of resolvable harmonics using iterated rippled noise (IRN) stimuli. IRN stimuli are created by generating a sample of frozen white noise, and adding it to itself following a delay equal to the inverse of the frequency of the desired pitch percept. Although the resultant stimuli contain spectral ripples, high-pass filtering can remove spectral cues in the region of resolvable harmonics while preserving the sensation of pitch. The strength of this pitch sensation, and resultant pitch discrimination thresholds, are dependent upon a number of stimulus parameters, including: the length of delay used to create the IRN, the number of iterations of the delay-and-add process, and the filter settings employed. For example, using a 3-alternative, forced-choice method designed to target 70.7% accuracy, Barker, Plack, and Hall (2011) demonstrated that adults can discriminate between 100 Hz and 160 Hz IRN stimuli, band-pass filtered between 1 and 2 kHz. Butler and Trainor (2012) presented electrophysiological evidence that adults can discriminate between 167 and 200 Hz IRN stimuli high-pass filtered at 2.6 kHz. The current study used a visually-reinforced,

conditioned head-turn procedure to determine whether 8-month-old infants could discriminate behaviourally between these same stimuli.

2. Experiment 1

2.1. Method

2.1.1. Participants

Five healthy 8-month-old infants (3 males; mean age = 251 +/-3.17 days) participated. An additional four infants failed to complete the training phase of the experiment, and one infant completed the training phase, but was unable to complete the experimental phase due to fussiness. All infants were born within 2 weeks of full term, were healthy at the time of testing, and no parent reported a history of chronic ear infection or hearing impairment. All research protocols were approved by the McMaster Research Ethics Board.

2.1.2. Stimuli

Iterated rippled noise stimuli identical to those of Butler & Trainor (2012) were created. The delay-and-add process was repeated a total of 16 times, as pitch perception for IRN stimuli has been shown in adults to plateau at 16 iterations; further iterations do not increase pitch salience (Patterson, Handel, Yost, & Datta, 1996). The delay time was set to either 6 ms, in order to create a signal with a perceived pitch of 167 Hz, or 5 ms, in order to create a signal with a perceived pitch of 200 Hz. To ensure equal power across the length of the stimuli, the first and last 100 ms (which contain a gradual increase and decrease in power, respectively, resulting from the iterative delay-and-add process) were

removed, resulting in stimuli with a total length of 450 ms. Following 16 iterations, some spectral peaks appear in the power spectra of IRN stimuli, although these peaks have a rippled appearance and thus provide spectral cues that are much less clear than those provided by the clear peaks present in traditional harmonic stimuli. Regardless, the IRN stimuli were then high-pass filtered at 2600 Hz (high-order Butterworth filter), representing the 13th harmonic of the 200 Hz stimulus, to remove spectral content in the range of the resolvable harmonics.

The stimuli were pilot tested on 6 adults using the infant procedure described below (the only difference being that adults raised their hand rather than turned their head to indicate the presence of a change in pitch). All adults were 100% correct (12/12 hits, 0/12 false alarms).

2.1.3. Procedure

After the procedure was explained and the consent form was signed by a parent, each infant was tested individually, seated on his/her parent's lap facing the experimenter inside an Industrial Acoustics Co. sound-attenuating booth with a flat-spectrum GSI speaker to the infants' left. The speaker was located above a box containing four compartments, each of which housed a mechanical toy and lights. The box had a smoked Plexiglas front such that the toys were not visible unless the lights in that compartment were illuminated. Infants were tested using the go/no-go conditioned head-turn response procedure in which head turns toward the speaker are reinforced with an illuminated, moving toy only if the turn occurs within 2 s of the onset of a change in sound. Sound stimuli were presented by an Apple G4 computer, through an NAD C352 stereo

integrated amplifier, while both the parent and the experimenter wore headphones and heard continuous music that masked the stimuli.

The background IRN stimulus had a pitch of 167 Hz, and was played continuously throughout both the training and experimental phases with a stimulus onset asynchrony of 2 seconds and an amplitude of 58 dB(A) over a background noise level of 26 dB(A). When the child's attention was focused on the experimenter, she called for a trial using a button box connected to a computer via a custom-built interface to a NI PCI-DIO96 I/O card. Head turns by the infant toward the speaker were recorded to computer by the experimenter pressing another button on the button box.

During the *training phase*, all trials were change trials in which the 167 Hz IRN background stimulus was replaced by the 200 Hz IRN. If an infant made a turn toward the speaker of at least 45° within 2 s of the onset of a change stimulus, the computer illuminated and moved one of the toys located beneath the speaker for 2 s. Once the reinforcement had ended and the experimenter had regained the infant's attention, the experimenter called for the next trial. During the *training phase*, target stimulus amplitude was 6 dB higher (64 dB[A]) to help the infant learn the contingency between a head-turn response to a 200 Hz pitch sensation and the visual reinforcement (the illuminated, moving toy). In order to pass training, the infant needed to make 4 consecutive correct head-turns to the change in pitch within 20 trials.

During the *testing phase*, twenty-four trials (12 change trials and 12 no-change trials) were presented in quasi-random order for each subject with the constraint that no more than two no-change trials were presented consecutively. For change trials, the 167

Hz IRN background stimulus was replaced by a 200 Hz IRN stimulus of equal amplitude. For no-change trials, the background 167 Hz IRN stimulus continued. Turns recorded during change trials were coded as hits, while turns made within 2 seconds of the onset of no-change trials were considered false alarms. Hit and false alarm rates were converted to individual d' sensitivity measures for each infant. In addition, the hit and false alarm rates across all infants were combined to create a measure of group sensitivity.

2.2. Results and Discussion

Half of the ten infants tested were unable to complete this experiment, suggesting that the discrimination was difficult. Of the remaining five infants, only two had d' -prime values greater than chance ($d'=0$). Macmillan and Kaplan (1985) have suggested that averaging hits and false alarms across subjects can yield a reliable, unbiased estimate of average d' when the number of trials available for each listener is limited, such as in the present study. In this case, combining the data across listeners yielded a group sensitivity of $d' = 0.24$. This poor performance is in contrast with perfect performance in the pilot adult subjects as described in the stimulus section.

These results suggest that infants are unable to perceive the pitch of these high-pass filtered IRN stimuli. However, it is possible that infants' perception is naturally drawn to the very salient noisy timbral quality of these stimuli, and that they would show perception of their pitch if their attention could be drawn to their pitch. Thus, the goal of Experiment 2 was to determine whether it is possible to train infants to perceive the pitch of IRN stimuli.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Sixteen healthy 8-month-old infants (13 males; mean age = 258 +/- 2.65 days) participated. All infants were born within 2 weeks of full term, were healthy at the time of testing, and no parent reported a history of chronic ear infection or hearing impairment. An additional two infants did not pass the training phase of the experiment and one infant failed to complete the experiment due to fussiness. All research protocols were approved by the McMaster Research Ethics Board.

3.1.2. Stimuli

The stimuli in the experimental phase were identical to those of Experiment 1. In the training phase, the IRN stimuli were also taken from Experiment 1, but on the target (change) IRN stimuli, which had a perceived pitch of 200 Hz, a 200 Hz sine tone was added. Both the IRN and sine tone components had amplitudes of 61 dB (A) such that the overall presentation level of the target stimulus remained at 64 dB (A).

3.1.3. Procedure

The experimental phase was identical to that of Experiment 1. The training phase was also identical to that of experiment 1, with the following exception: During the training phase, the stimulus on change trials was the 200 Hz IRN/pure tone stimulus rather than the 200 Hz IRN stimulus. This deviant stimulus was presented at the same amplitude as the background stimulus (58 dB[A]) in an effort to ensure that infants

learned the contingency between visual reinforcement and the 200 Hz pitch sensation rather than between reward and a change in stimulus amplitude.

3.2. Results

Figure 1 shows a box plot of the d' values obtained in the current study. The median individual d' across infants was 0.69, and all but three infants had d' values greater than zero, suggesting that the majority of 8-month-olds were able to detect a change from 167 Hz to 200 Hz in the perceived pitch of IRN stimuli. A one-sample Wilcoxon Signed Rank Test revealed that this median value was significantly above chance levels ($p=0.01$). Combining the data across listeners yielded a group sensitivity of $d' = 0.58$.

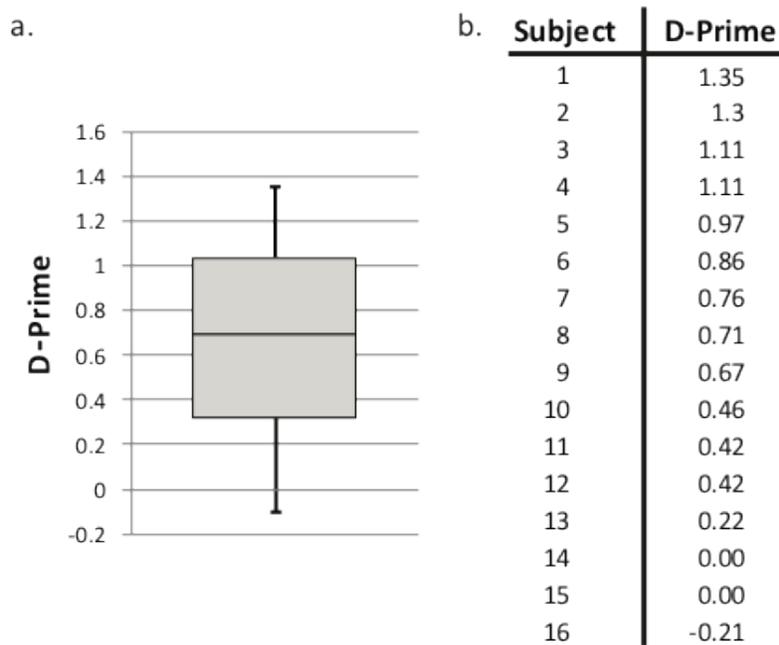


Figure 1. Panel a) shows a box plot of the d' sensitivity scores observed for the discrimination of a change in the perceived pitch of an IRN stimulus from 167 Hz to 200 Hz. Panel b) shows the individual d' sensitivity scores for each infant in Experiment 2.

Figure 2 shows the hit and false alarm rates for each infant. It is important to note that, while the majority of infants had hit rates that exceeded their false alarm rates, the performance of the 8-month-olds tested here was far worse than that of the pilot adults described in Experiment 1, who were at ceiling levels.

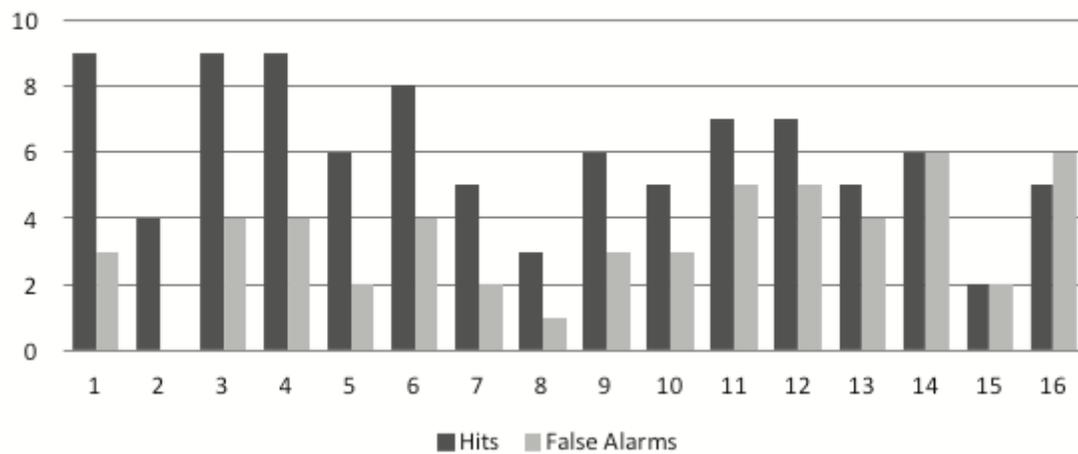


Figure 2. Hit rates (dark bars) and false alarm rates (light bars) for each infant in the current study, arranged in order of decreasing sensitivity. Thirteen of sixteen infants showed hit rates that exceeded their rate of false-alarms.

4. Discussion

In the current experiments we used a conditioned head-turn procedure to measure the sensitivity of 8-month-old listeners to a pitch change in IRN stimuli containing information only outside the region of spectral resolution. These IRN stimuli typically elicit a weak pitch sensation in adult listeners, but one that can be readily detected. Four of the ten infants tested in experiment 1 failed to complete the training phase of the experiment, suggesting they were unable to learn the contingency between the change in

the pitch of IRN stimuli and the visual reinforcement provided, even in the presence of a loudness cue. For those infants who completed the experimental phase, the low group average sensitivity ($d' = 0.24$) suggests that the IRN stimuli were not being successfully discriminated.

IRN stimuli have a salient noisy timbral quality. In the training phase of experiment 2, the target stimulus was designed to draw attention to pitch in an effort to prime the infants to attend to pitch differences. The group average sensitivity ($d' = 0.58$) was statistically above chance levels, and exceeded the sensitivity previously reported in complex tones containing only unresolved harmonics ($d' = 0.44$; Clarkson & Rogers, 1995). However, this sensitivity is well below that reported for infants for similar pitch changes in harmonic stimuli with spectrally resolved components, even when the fundamental frequency is missing ($d' = 1.25$, Clarkson & Rogers, 1995; $d' = 1.80$, Clarkson & Clifton, 1995). Collectively, the results reported in experiment 1 and 2 suggest that 8-month-old infants can perceive the pitch of IRN stimuli with no cues in the region of resolvable harmonics, when first primed to attend to stimulus pitch. However, the perceived pitch sensation for such stimuli is weak compared to complex tones with resolvable harmonics.

It is possible that the spectral ripples present in IRN stimuli might create distortion products at or near the fundamental frequency of the sound as a result of cochlear nonlinearities. However, for harmonic tones, these distortions depend on simple phase relationships between components (Pressnitzer & Patterson, 2001) that are weak or absent in IRN stimuli (Sayles & Winter, 2008). Moreover, evidence suggests that if IRN

produces audible distortion products, they are at such a low level as to be essentially negligible (Winter, Wiegrebe, and Patterson, 2001). This, combined with the low overall presentation level and high cutoff value for stimulus filtering used in the current experiment (high-passed at the 12th harmonic of the 200 Hz stimulus), suggests that the discriminations observed in Experiment 2 were based on temporal pitch cues in the stimulus, rather than on cochlear distortions in the region of resolved spectral information.

The perception of pitch in inharmonic tones and IRN stimuli relies to a large extent on the temporal mechanism for pitch extraction. While pitch discrimination performance degrades for both infants and adults as inharmonicity increases, infant performance appears to drop off much more rapidly than does adult performance (Clarkson & Clifton, 1995). Similarly, the results reported here with the IRN stimuli suggest very low performance in infants compared to adults (e.g., our adult pilot subjects showed perfect performance on the stimuli of the present paper; Barker et al., 2011, showed 70.7% accuracy when discriminating between 100 Hz and 160 Hz IRN stimuli in a 3AFC paradigm [see their Figure 3]). These observations are consistent with Clarkson and Clifton's (1995) suggestion that the mechanism responsible for extracting and/or interpreting temporal pitch cues may take longer to develop compared to the mechanism acting on resolved, spectral cues.

Examining individual differences, it appears that the sample of infants tested in experiment 2 may consist of two subgroups of infants; one that can perceive the pitch of high-pass filtered IRN stimuli and one that cannot (see Figure 2). If it is indeed the case

that pitch perception given unresolved spectral content is limited by the development of the mechanism responsible for extracting and/or interpreting temporal pitch cues, this split may reflect differential development of this mechanism across infants. Further suggestion that the temporal mechanism might be improving during this period of development is that the infants in the present study (mean= 258 days) were somewhat older than those tested using harmonic complex stimuli by Clarkson and Rogers (1995; mean=220.71 days). Thus, the somewhat better performance of infants in the present study compared to the infants in the Clarkson and Rogers study might reflect an increased sensitivity of the temporal mechanism with increased age.

Perhaps one of the most interesting findings of the present study was that a very short period of training that highlighted the pitch of the IRN stimuli facilitated infants' ability to discriminate pitch. This finding raises the possibility that infant performance on a number of tasks, including discriminating the pitch of harmonic complexes without resolvable harmonic, might similarly be improved through focused training.

In summary, the present study provides behavioral evidence that following a training period designed to emphasize stimulus pitch, 8-month-old infants can discriminate the pitches of IRN stimuli whose spectral content is limited to the region beyond which individual harmonics can be resolved by the cochlea. As in adults, pitch perception in infants is degraded when resolvable spectral content is absent, but this degradation appears to be more marked in infants than adults. Taken in conjunction with the results of previous studies, it appears that the extraction of pitch without information

in the resolvable region, and the temporal mechanism for pitch extraction, are not yet mature in 8-month-old infants.

Acknowledgements

This research was supported by grants to LJT from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) and an NSERC graduate scholarship to BEB. The authors wish to thank Andrea Unrau for assisting with data collection.

References

- Barker, D., Plack, C. J., & Hall D. A. (2011). Reexamining the evidence for a pitch-sensitive region: A human fMRI study using iterated rippled noise. *Cereb. Cortex*. doi:10.1093/cercor/bhr065
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436, 1161-1165.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (pp. 1-790). Cambridge, MA: MIT Press.
- Clarkson, M. G., & Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *J. Acoust. Soc. Am.*, 77, 1521-1528.
- Clarkson, M. G., & Clifton, R. K. (1995). Infants' pitch perception: Inharmonic tonal complexes. *J. Acoust. Soc. Am.*, 98, 1372-1379.

- Clarkson, M. G., & Rogers, C. (1995). Infants require low-frequency energy to hear the pitch of the missing fundamental. *J. Acoust. Soc. Am.*, *98*, 148-154.
- Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psych. Bull.*, *97*, 412-429.
- He, C., & Trainor, L. J. (2009). Finding the pitch of the missing fundamental in infants. *J. Neurosci.*, *29*, 7718-7722.
- Houtsma, A. J. M., & Goldstein, J. L. (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *J. Acoust. Soc. Am.*, *54*, 520-529.
- Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.*, *87*, 304-310.
- Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, *50*, 1202-1211.
- Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception, *Trends Cog. Sci.*, *9*, 578-584.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psych. Bull.*, *98*, 185-199.
- McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.*, *18*, 452-463.
- Montgomery, C. R., & Clarkson, M. G. (1997). Infants' pitch perception: Masking by low- and high-frequency noises. *J. Acoust. Soc. Am.*, *102*, 3665-3672.

Moore, B. C. J. (2008). *An Introduction to the Psychology of Hearing* (5th ed.) (pp. 1-413). London, England: Emerald Group Publishing.

Olsho, L. W., Schoon, C., Sakai, R., Terpin, R., and Sperduto, V. (1982). Auditory frequency discrimination in infancy, *Dev. Psych.*, *18*, 721-726.

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.*, *100*, 3286-3294.

Plomp, R. (1967). Pitch of complex tones. *J. Acoust. Soc. Am.*, *41*, 1526-1533.

Pressnitzer, D., & Patterson, R. D. (2001). Distortion products and the perceived pitch of complex tones. In D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 97-104). Maastricht, The Netherlands: Shaker.

Ritsma, R. J., (1962). Existence region of the tonal residue. I. *J. Acoust. Soc. Am.*, *34*, 1224-1229.

Ritsma, R. J., (1967). Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Am.*, *42*, 191-198.

Sayles M., & Winter, I. M. (2008). Ambiguous pitch and the temporal representation of inharmonic iterated rippled noise in the ventral cochlear nucleus. *J. Neurosci.*, *28*, 11925-11938.

Trainor, L. J., & Corrigan, K. A. (2010). Music acquisition and effects of musical experience. In M. Riess-Jones & R. R. Fay (Eds.), *Springer Handbook of Auditory Research: Music Perception* (pp. 89-128). Heidelberg: Springer.

Winter, I. M., Wiegrebe, L., & Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *J. Physiol.*, *537*, 553-566.

Yost, W. A. (1996). Pitch of iterated rippled noise. *J. Acoust. Soc. Am.*, *100*, 511-518.

CHAPTER 4: PREFACE

Butler, B. E., and Trainor, L. J. (submitted). Brief pitch-priming facilitates infants' discrimination of pitch-evoking noise: Evidence from event-related potentials. *Brain and Cognition*, Manuscript ID BRCG-12-271

Chapter 3 provided behavioural evidence that 8-month-old infants can discriminate between IRN stimuli of different pitches, despite the absence of resolved spectral pitch cues. However, the sensitivity values reported were quite low compared to those reported in the literature for pitch-evoking stimuli with robust spectral cues.

This chapter uses event-related potentials to examine a cortical representation of pitch discrimination for IRN stimuli. Experiment 1 uses a traditional oddball paradigm to determine whether an infrequent change in the pitch of IRN stimuli from 167 Hz to 200 Hz can elicit a mismatch negativity (MMN) response. In Experiment 2, a brief training period in which pure tone stimuli were added to the IRN sounds to emphasize stimulus pitch preceded the testing phase.

While IRN stimuli have been used across a number of studies in adult listeners, this represents the first cortical evidence of infant pitch discriminations for stimuli without resolvable spectral cues. Moreover, this chapter provides evidence that the behavioural data presented in Chapter 3 represented true discriminations between stimuli of different pitches, despite the weak sensitivity values observed therein. Finally, the

results of this chapter reinforce that a period of pitch-priming is necessary for infants to accurately discriminate between IRN stimuli.

**Brief pitch-priming facilitates infants' discrimination of pitch-evoking noise:
Evidence from event-related potentials**

Blake E. Butler¹ and Laurel J. Trainor^{1,2}

¹Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario L8S 4K1, Canada, and ²Rotman Research Institute, Baycrest Centre, University of Toronto, Toronto, Ontario M6A 3E1, Canada

Running Title: Pitch perception and priming in infants

Correspondence:

Dr. Laurel J. Trainor
McMaster University
Department of Psychology, Neuroscience and Behaviour
Auditory Development Lab
1280 Main Street West
Hamilton, Ontario, Canada, L8S 4L8
ljt@mcmaster.ca
Phone: (905) 525-9140 ext. 23007
Fax: (905) 529-6225

ABSTRACT

Pitch is derived by the auditory system through complex spectrotemporal processing. Pitch extraction is thought to depend on both spectral cues arising from lower harmonics that are resolved by cochlear filters in the inner ear, and on temporal cues arising from the pattern of action potentials contained in the cochlear output. Adults are capable of extracting pitch in the absence of robust spectral cues, taking advantage of the temporal cues that remain. However, recent behavioral evidence suggests that infants have difficulty discriminating between stimuli with different pitches when resolvable spectral cues are absent. In Experiment 1, we used the mismatch negativity (MMN) component of the event related potential derived from electroencephalographic (EEG) recordings to examine a cortical representation of pitch discrimination for iterated rippled noise (IRN) stimuli in 4- and 8-month-old infants. IRN stimuli are pitch-evoking sounds generated from repeatedly adding a segment of white noise to itself at a constant delay. We created IRN stimuli (delays of 5 and 6 msec creating pitch percepts of 200 and 167 Hz) and high-pass filtered them to remove all resolvable spectral pitch cues. We did not find EEG evidence that infants could detect the change in the pitch of these IRN stimuli. However, in Experiment 2, after a brief period of pitch-priming during which we added a sine wave component to the IRN stimulus at its perceived pitch, infants did show significant MMN in response to pitch changes in the IRN stimuli with sine waves removed. This suggests that (1) infants can use temporal cues to process pitch, although such processing is not mature and (2) that a short amount of pitch priming experience can alter pitch representations in auditory cortex during infancy.

Keywords: Pitch, Development, Event-related potentials, Mismatch negativity, Iterated rippled noise (IRN)

1. INTRODUCTION

Pitch perception is central to musical development (see Trainor and Corrigan, 2010 for review), and conveys a wealth of semantic and prosodic information that is crucial for language acquisition (see Moore, 2008 for review). Additionally, pitch provides information vital to the identification of environmental sounds, and aids in the perceptual separation of co-occurring sounds (Bregman, 1990). Stimuli with pitch are typically complex tones with energy at a fundamental frequency, and at integer multiples of that frequency, known as harmonics, and it is this harmonic structure that provides cues to pitch. The basilar membrane of the cochlea is responsible for translating the mechanical energy of sound waves into a pattern of action potentials. Physical characteristics of this membrane, such as width and rigidity, differ along its length, such that high frequencies cause maximal displacement at the basal end, while lower frequencies cause maximal displacement at the apex of the membrane (Von Bekesy, 1960). This is referred to as tonotopic organization, and it is this arrangement that gives rise to spectral, or place cues to pitch. A complex tone with many harmonics is heard as a single entity, and its pitch is derived from the relations between the harmonics present and not just from the frequency of the fundamental. This is evidenced by the fact that removing the fundamental from a complex tone does not alter its pitch, a phenomenon known as hearing the pitch of the missing fundamental.

The basilar membrane can be thought of as a series of bandpass filters. In contrast to harmonics, which are linearly spaced, basilar membrane filters are logarithmically spaced such that at low frequencies, individual harmonics fall into separate filters and are thus individually resolved. However, at higher harmonics, the bandwidth of cochlear filters exceeds harmonic spacing and multiple harmonics fall into the same filterband, causing interfering patterns on the basilar member. These harmonics are thus unresolved, and place cues cannot provide an accurate pitch estimate. A second, temporal mechanism is thought to compensate by taking advantage of temporal regularity in the pattern of action potentials in the cochlear output. Because action potentials are generated at the point of maximal displacement of the basilar membrane, the pattern of potentials across a population of auditory nerve fibers approximates the frequency of an auditory stimulus (e.g., Delgutte and Cariani, 1992; Cariani and Delgutte, 1996; Meddis and O'Mard, 1997). Current models of pitch perception typically include contributions from both spectral and temporal cues (e.g. Cedolin and Delgutte, 2007; Larsen et al., 2008).

A number of studies have demonstrated that infants perceive pitch-evoking stimuli with resolved spectral content in much the same way as adults. While the limits of complex tone discrimination have not been reported, 8-month-old infants have been shown to discriminate behaviorally between complex stimuli, with and without energy at the fundamental frequency, that differ in pitch by 20% (e.g. 160 and 200 Hz; Clarkson and Clifton, 1985). Montgomery and Clarkson (1997) verified that 8-month-old infants' discrimination of missing-fundamental stimuli is not impaired by the addition of a low-frequency noise masker covering the region of the fundamental; this suggests that, like

adults, infants discriminate these stimuli based on integration of harmonic content into a pitch percept, rather than based on low-frequency combination tones resulting from nonlinearities of the inner ear. Furthermore, He et al. (2009) used event-related potentials (ERPs) to demonstrate that a cortical representation of the pitch of the missing fundamental emerges between 2 and 4 months of age, suggesting that by this age infants are able to integrate harmonics into a single percept with pitch. Finally, Clarkson and Clifton (1995) have demonstrated that 7-month-old infants can discriminate pitch changes in inharmonic complexes (where spectral content is resolved, but the harmonics do not fit exactly to the expected template of a complex tone), and that their performance is related to the degree of inharmonicity in a manner qualitatively similar to adult performance.

While infant pitch discrimination may be qualitatively adult-like for stimuli containing robust spectral pitch cues, infants appear to be significantly impaired relative to adults when discriminating stimuli that do not contain such cues to pitch, consistent with Werner's (1992) suggestion that spectral mechanisms mature earlier in development than temporal mechanisms. Understanding how infants perform in the absence of resolved spectral information is important to the study of auditory development in general and language acquisition in particular because in everyday noisy environments, low frequency components are often masked by background noise, such that only higher frequency, unresolved components remain as cues to vocal pitch. When presented with complex stimuli containing only higher, unresolved harmonics, infants do not appear to be able to successfully categorize them according to pitch in a behavioral conditioned head turn paradigm in which infants are rewarded for turning their head in response to a

change in pitch (Clarkson and Rogers, 1995). Consistent with this, we (Butler et al., 2012; under review) found that without any pitch priming (i.e., training on how to perceive the pitch of such stimuli), 8-month-old infants did not show behavioral (conditioned head turn) discrimination of changes in the pitch of iterated rippled noise (IRN) stimuli.

IRN stimuli are created by repeatedly adding a sample of frozen white noise to itself following a short, fixed delay. In this way, temporal regularity is introduced, and a (weak) pitch percept is created that is equal to the inverse of the delay (e.g., a delay of 5 msec produces a perceived pitch of 200 Hz; Yost, 1996). The stimuli used in Butler et al. (2012; under review) were high pass filtered so as to contain no resolved spectral cues, so the sensation of pitch was predominantly dependent on the temporal mechanism.

Interestingly, Butler et al. (2012; under review) found that if infants were given a period of training in which a sine tone was added to the IRN stimulus at the frequency of its perceived pitch, infants were able to behaviorally discriminate a pitch change from 167 Hz to 200 Hz above change levels. However, performance was still quite poor ($d' = .69$) under conditions where adults were 100% correct. It is therefore important to gather converging evidence that infants are able to process the pitch of IRN stimuli and to investigate how pitch priming affects representations for pitch in the infant nervous system. In the present paper, we look for evidence using event-related potentials (ERPs) that (1) infants 4 months and 8 months of age can use temporal cues to process IRN stimuli, and (2) that pitch representations in auditory cortex are enhanced after pitch priming experience.

Exactly how and where pitch percepts are formed in auditory cortex is not entirely clear. However, functional imaging studies in adults suggest that a common pitch-processing centre is located beyond primary auditory cortex, along the lateral aspect of Heschl's gyrus (Griffiths et al., 1998; Patterson et al., 2002; Penagos et al., 2004; Hall et al., 2005; Puschmann et al., 2010). Moreover, an event-related potential study in adults has shown that pitch changes in IRN stimuli similar to those used in the current study elicit a mismatch negativity (MMN) component, the source of which is consistent with these imaging studies (Butler and Trainor, 2012). The MMN reflects automatic detection of an infrequent deviant stimulus, and can be recorded from both adults and infants. For example, a mismatch negativity (MMN)-like deflection is elicited by changes in the frequency of synthesized piano tones in 2-month-old infants, and that deflection increases in amplitude and decreases latency in the months that follow (He et al., 2007; 2009). A number of other studies have also successfully used the MMN to study deviation in pure tone frequency (e.g., Alho et al., 1990; Leppänen et al., 1999; 2004; Cheour et al., 1999; Cēponienė et al., 2000, Hirasawa et al., 2003) and the pitch of harmonic tones (e.g., Cēponienė et al., 2002; Kushnerenko et al., 2002; Fellman et al., 2004).

No study to date has demonstrated evidence of a cortical basis for pitch discrimination in infants when the pitch-evoking stimuli contain no resolved spectral cues. Experiment 1 examines whether pitch changes in high-pass filtered IRN stimuli elicit a mismatch response in 4- and 8-month-old infant listeners. Experiment 2 asks whether brief priming of the pitch of IRN stimuli can enhance the representation of pitch in auditory cortex.

2. EXPERIMENT 1

2.1. MATERIALS AND METHODS

2.1.1. Participants

Fifteen 4-month old infants (8 males; mean age = 19.7 +/- .17 weeks) and fifteen 8-month old infants (5 males; mean age = 36.0 +/- .36 weeks) participated. All infants were born within 2 weeks of full term, were healthy at the time of testing, and no parent reported a history of chronic ear infection or hearing impairment. An additional 13 infants were unable to complete the minimum number of trials due to fussiness, while 4 infants were excluded from data analysis because excessive movement during testing left too few artifact-free trials.

2.1.2. Stimuli

The IRN stimuli used in this experiment were created by generating a sample of frozen white noise and adding it to itself following a delay equal to the inverse of the desired pitch. The standard stimulus had a perceived pitch of 167 Hz (corresponding to a 6 msec delay) and was presented on 85% of trials. The deviant stimulus was presented on the remaining 15% of trials and had a perceived pitch of 200Hz (corresponding to a 5 msec delay). The delay-and add process was performed a total of 16 times, as the strength of the pitch percept has been shown to plateau following 16 iterations (Patterson et al., 1996). Following this many iterations, some peaks appear in the power spectra of IRN stimuli. In order to eliminate the contribution of resolvable spectral pitch cues, IRN stimuli were high-pass filtered at 2600 Hz (high-order Butterworth filter), representing

the 13th harmonic of the 200 Hz stimulus. Stimuli were 450 msec in length, had 10 msec onset and offset ramps, and were presented with a stimulus onset asynchrony (SOA) of 800 msec at 70 dBA. Sometimes high-pass filtered stimuli are presented in white noise in order to mask any potential low frequency distortion products arising from nonlinearities in the inner ear. However, these distortions depend on simplistic phase relationships between components (Pressnitzer & Patterson, 2001) that are weak, or not present in the IRN stimuli used in the present experiment (Sayles & Winter, 2008). Moreover, Winter et al. (2001) have suggested that if IRN produces audible distortion products, they are at such a low level as to be essentially negligible. Thus, a white noise masker was not used in the present experiment to avoid creating any further ambiguity in an already weak pitch percept.

2.1.3. *Procedure*

Experimental procedures were explained to parents, who gave informed consent to have their child participate. Each infant was seated on his or her parent's lap in a sound-attenuated room, facing a loudspeaker and a computer monitor. In order to keep the infant still and awake during the experiment, he or she watched a silent movie and a puppet show provided by an experimenter seated in the room. Sounds were presented using E-Prime software through a loudspeaker located 1 m in front of the infant. The experiment consisted of 1600 trials and lasted approximately 21 minutes. Stimuli were presented in a quasi-random fashion, such that at least two standard stimuli were presented between deviant stimuli.

EEG data were recorded at a sampling rate of 1000 Hz from 124-channel HydroCel GSN nets (Electrical Geodesics, Eugene, OR) referenced to Cz. Impedance at each electrode was maintained below 50 k Ω during recording. EEG data were band-pass filtered offline between .5 and 20 Hz using EEProbe software to remove slow wave activity. Data were then resampled at 200 Hz and artifacts resulting from muscle activity such as eye blinking and head movements were removed using an Artifact Blocking paradigm in MATLAB (Mourad et al., 2007). Finally, the data were re-referenced offline to an average reference, and then segmented into 500 msec epochs that included a 100 msec baseline.

2.1.4. *Analysis*

Responses to standard and deviant stimuli were averaged, and difference waveforms were computed for each participant by subtracting their response to the standard stimulus from their response to the deviant stimulus. Grand average standard and deviant waveforms and difference waveforms (deviant – standard) were then computed for each age group. Subsequently, for statistical analysis, 90 electrodes were selected and divided into 5 groups for each hemisphere representing frontal, central, parietal, occipital, and temporal scalp regions (See Figure 1). Electrodes near the face and periphery of the net were excluded in order to further reduce the impact of muscle artifacts from the eyes, face, and neck. Midline electrodes were excluded to allow for comparisons between hemispheres.

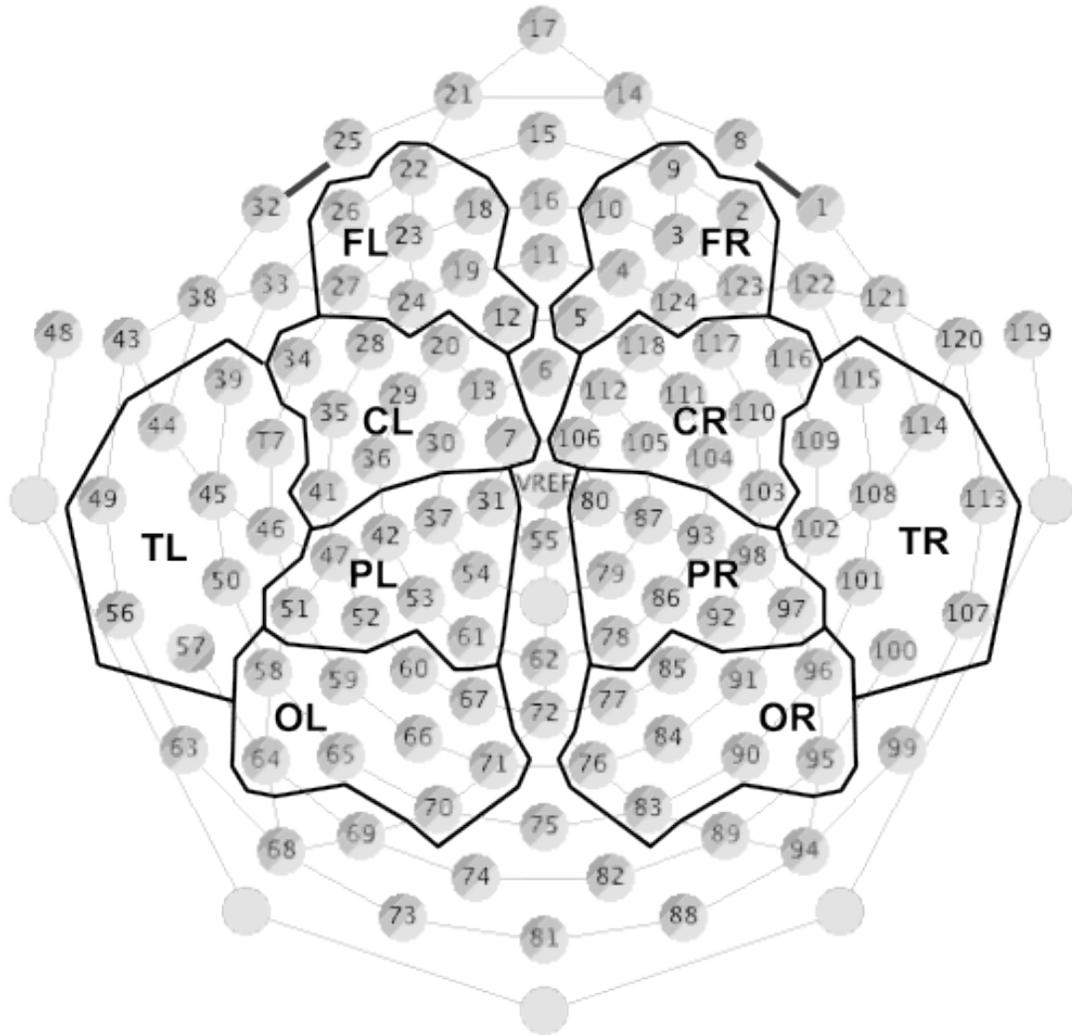


Figure 1. Electrode groupings (see Materials and Methods section for details). Ninety of 124 electrodes were divided into 5 groups (frontal, central, parietal, occipital and temporal) for each hemisphere. Each group contained between 16 and 20 electrodes that were averaged together to represent EEG responses 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.

In both age groups, when the waveforms were averaged across all infants (grand average waveforms) a small negative-going component resembling the mismatch negativity response (MMN) was observed at the frontal and central electrode sites (with reversing polarity at occipital and temporal electrodes). This component was not present at parietal sites, consistent with the inversion of morphology typically observed between frontal/central regions and occipital regions for responses generated in auditory cortex. Thus, parietal responses were eliminated from further analysis. Initially, a series of one-tailed t-tests were performed at each time point within this analysis window to determine whether the mean amplitude of this component differed significantly from zero. To analyze the amplitude of the negative-going deflection, the component peak was determined from the grand average difference waveforms for each age group, and a 50 msec time window was constructed centered at this latency. The mean amplitude across this 50 msec window was computed for each subject and each region, and was taken as the measure of amplitude. An analysis of variance (ANOVA) was conducted to examine effects of age, scalp region and hemisphere, and Greenhouse-Geisser corrections were applied where necessary. Finally, paired samples t-tests were conducted to examine significant main effects and interactions, with p-values adjusted for multiple comparisons.

2.2. RESULTS

Figure 2 shows the grand average difference waves recorded from 4-month-old (solid lines) and 8-month-old listeners (dashed lines) in response to a change in the pitch of IRN stimuli. One-tailed t-tests at each time point within the analysis window revealed

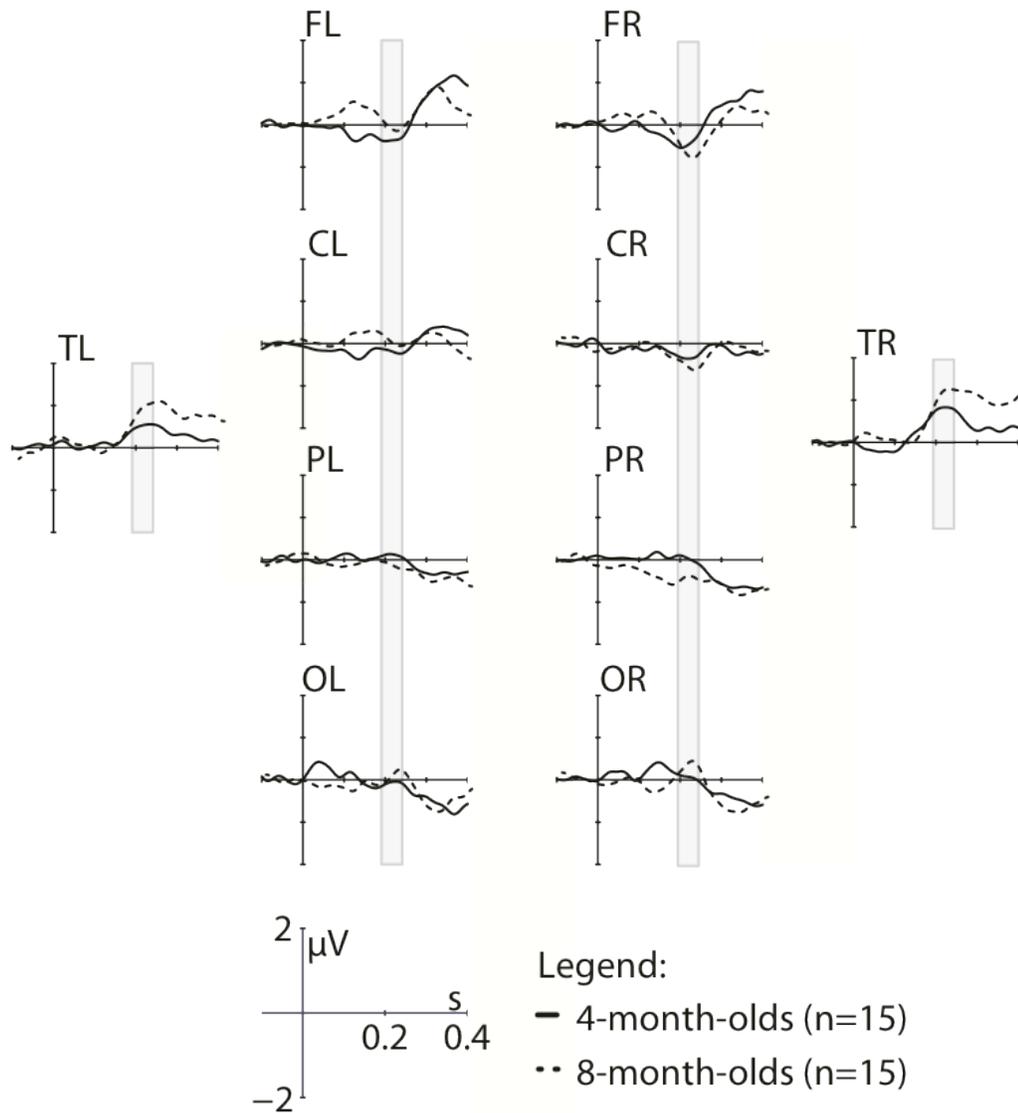


Figure 2. Grand average difference waveforms (deviant-standard) for pitch changes recorded in Experiment 1. Responses from each of the 10 electrode regions are presented for both 4-month-old (solid lines) and 8-month-olds (dashed lines). Vertical bars illustrate the analysis window over which peak amplitude was measured.

that the MMN failed to reach significance in any electrode region for either the 4-month-olds or 8-month-olds (all $ps > .05$). The test using the average amplitude for each subject in the 50-msec windows around the grand average peak also revealed no significant MMN response at any electrode in either age (all $ps > .05$).

In Experiment 2 we investigated whether infants' representations of the pitch of IRN stimuli could be changed and their performance improved with a brief priming of the pitches they were to listen for.

3. Experiment 2

3.1. MATERIALS AND METHODS

3.1.1. Participants

Fifteen, 4-month-old infants (5 males; mean age = 20.0 +/- .17 weeks) and fifteen, 8-month old infants (12 males; mean age = 36.5 +/- .27 weeks) participated. None of the infants had participated in Experiment 1. All infants were born within 2 weeks of full term, were healthy at the time of testing, and no parent reported a history of chronic ear infection or hearing impairment. An additional 8 infants were unable to complete the minimum number of trials due to fussiness, while 5 infants were excluded from data analysis because excessive movement during testing left too few artifact-free trials.

3.1.2. Stimuli

Two types of stimuli were used in this experiment. The priming phase consisted of IRN stimuli with a pure tone of equivalent pitch added. The experimental phase

consisted of the IRN stimuli alone and was identical to Experiment 1. As in Experiment 1, in both the priming and experimental phases, the standard stimulus had a perceived pitch of 167 Hz and was presented on 85% of trials. The deviant stimulus was presented on the remaining 15% of trials and had a perceived pitch of 200Hz. All stimuli were 450 msec in length, had 10 msec onset and offset ramps, and were presented with a stimulus onset asynchrony (SOA) of 800 msec at 70 dBA.

3.1.3. *Procedure*

The experimental procedure was identical to that of Experiment 1 with the following exceptions. The experiment included a priming phase consisting of 600 trials (lasting approximately 8 minutes), followed by the experimental phase consisting of 1600 trials (lasting approximately 21 minutes) that was identical to that of Experiment 1.

EEG data collected during the experimental phase were recorded, filtered, resampled, artifact blocked and re-referenced as in Experiment 1. Data collected during the priming phase were not analyzed as there were too few trials.

3.1.4. *Analysis*

Data were analyzed as in Experiment 1.

3.2. RESULTS

Figure 3 shows the grand average difference waves recorded from 4-month-old (solid lines) and 8-month-old listeners (dashed lines) in response to a change in the pitch

of IRN stimuli. The results of one-tailed t-tests at each time point within the analysis window are found below each grand average waveform (shaded areas within solid and dashed boxes representing significance in 4- and 8-month-olds, respectively). As seen in Figure 3, the mean amplitude of the MMN was significant in the right frontal and right temporal regions in 4-month-olds. This was confirmed in a second analysis whereby one-tailed t-tests showed that the average amplitude measured across the 50 msec analysis window was significant in these same electrode regions (frontal: $t[14]=3.6$, $p=.01$; temporal: $t[14]=5.1$, $p=.001$), as well as in the left temporal region ($t[14]=3.4$, $p=.002$).

In the 8-month-olds, the one-tailed t-tests at each time point reached significance in the right frontal and right central regions, and in the temporal regions bilaterally. One-tailed t-tests confirmed that the amplitude measured across the 50 msec analysis window was significant in the right frontal region ($t[14]=3.5$, $p=.02$) as well as the left ($t[14]=3.6$, $p=.01$) and right temporal regions ($t[14]=3.4$, $p=.02$).

To further analyze absolute MMN amplitude in both age groups, a mixed-model ANOVA was conducted with region (frontal, central, occipital, temporal), and hemisphere (left, right) as within-subject factors, and age as a between-subject factor. There was no significant effect of age ($p=.66$), and the difference wave amplitudes were very similar across listening conditions for both age groups. Overall, MMN amplitude was greater in the right hemisphere than in the left ($F(1,28)=16.0$, $p<.001$).

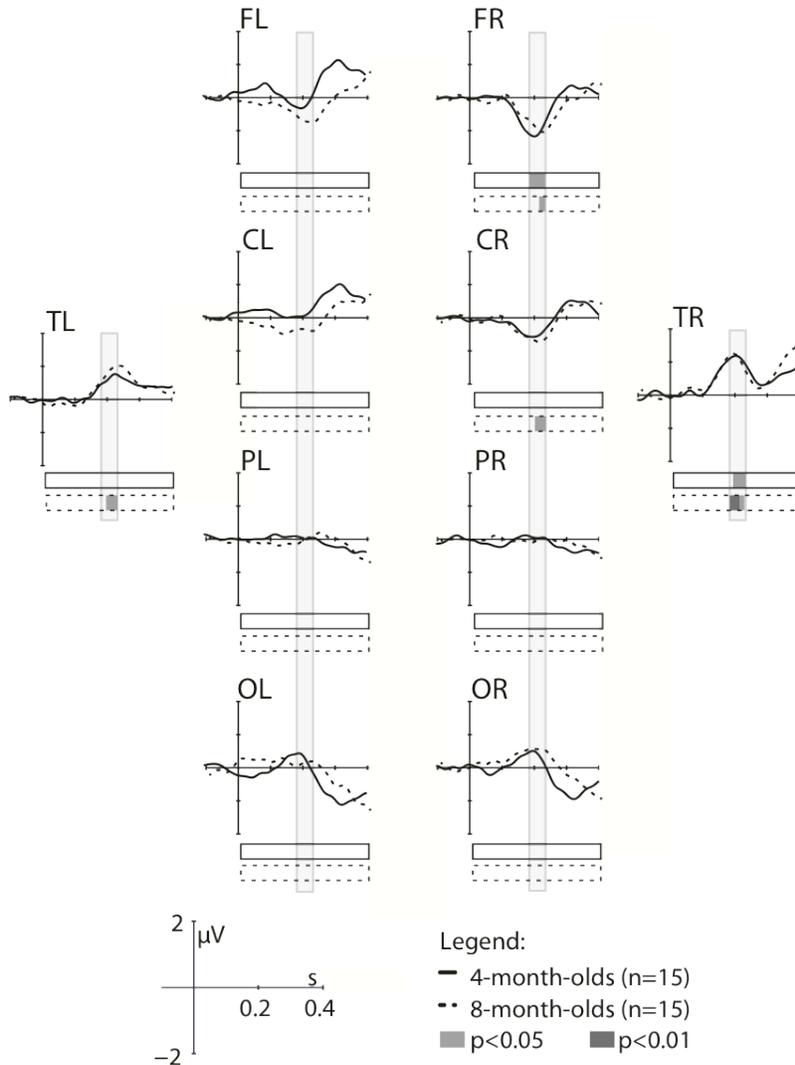


Figure 3. Grand average difference waveforms (deviant-standard) for pitch changes recorded in Experiment 2. Responses from each of the 10 electrode regions are presented for both 4-month-old (solid lines) and 8-month-olds (dashed lines). The bars under the waveforms show latencies at which the amplitude of the difference wave differed significantly from zero for each age group. Vertical bars illustrate the analysis window over which peak amplitude was measured.

Additionally, there was a significant effect of electrode region ($F(3,84)=6.4, p=.006$), with the absolute magnitude of the response in the frontal regions exceeding responses from the central ($p=.006$) and occipital regions ($p<.001$), and responses from temporal regions exceeding those recorded in central regions ($p=.018$).

4. DISCUSSION

A previous behavioral study found that 8-month-old infants showed no evidence of the ability to discriminate the pitch of IRN stimuli that contain no resolvable harmonics, unless previously primed with stimuli in which a sine tone was added to the IRN stimuli at the frequency of the pitch percept (Butler et al., 2012; under review). The electrophysiological evidence presented here corroborates this behavioral evidence. In Experiment 1, infants did not show a significant MMN response to a pitch change with IRN stimuli, but in Experiment 2, after pitch priming, infants did show a significant MMN to pitch changes using the same stimuli as in Experiment 1. Furthermore, the present study indicates that the failure of infants in Butler et al. (2012; under review) to show pitch discrimination prior to priming does not reflect the inability of those infants to respond behaviorally to a pitch change that they were, in fact, perceiving. Rather, the electrophysiological responses analyzed in the present study indicate no evidence of cortical encoding of pitch differences before priming, but significant evidence of cortical pitch coding after priming. Finally, to our knowledge, the current study is the first to demonstrate that infants as young as 4 months of age can discriminate pitch-evoking auditory stimuli that contain no resolvable spectral cues.

Infants have been shown to process pitch-evoking stimuli in an adult-like manner in the presence of robust spectral cues (Cėponienė et al., 2002; Kushnerenko et al., 2002; Fellman et al., 2004; He et al., 2007; 2009). For deviations in the pitch of the missing fundamental, an MMN-like negativity emerges by 4 months of age (He and Trainor, 2009). This suggests that 4-month-old infants are capable of integrating the harmonic structure of pitch-evoking stimuli into a single pitch percept. Moreover, it suggests that the cortical generators of the MMN are in place by 4 months of age for stimuli that contain resolved spectral content.

While resolved spectral information may provide the most salient cues to pitch, temporal cues also make a functionally important contribution to pitch perception. For example, the ability to extract pitch in the absence of resolved spectral cues allows one to communicate over the telephone, where band-pass filters often remove resolvable harmonics, or to understand speech against a noisy background that often masks those same harmonics. Thus, in situations where spectral cues to pitch are unavailable, both infants and adults, rely more heavily on temporal information. The fact that behavioral performance is low for discriminating pitch-evoking stimuli without resolved spectral cues (Butler et al., 2012, under review; Clarkson and Rogers, 1995), and that electrophysiological responses are only significant after priming, suggests that the temporal mechanism for pitch is slower to mature than the spectral mechanism. This is consistent with the literature on infant pure tone frequency discrimination, which suggests a similar pattern of development (see Werner, 1992 for review). However, the presence of a significant MMN response in the difference wave of the infants in the present study

provides evidence that infants can extract pitch from stimuli where spectral cues are limited to the unresolvable region.

The pitch-priming phase of experiment 2 provided resolved spectral pitch cues (a pure tone at the fundamental frequency of the IRN stimulus) in an effort to make clear to infants the basis on which the different stimuli could be most easily discriminated. Following this priming period, the mismatch response recorded during the test period (where pure tones were removed) reached significance in both 4- and 8-month-old listeners. The MMN is thought to reflect the passive updating of auditory memory traces. Thus, it appears that the experience acquired during the pitch-priming phase impacts the formation of these auditory traces in the subsequent testing phase. Similar effects of learning on evoked components have been observed previously in adults. MMN responses to unattended deviants have been shown to emerge slowly across blocks in a single experimental session (Näätänen et al., 1993). However, in this case, passive listening blocks were interspaced with blocks requiring active discrimination; repeated, unattended exposure to the stimuli alone was insufficient to affect passive discrimination. This suggests that, in adults, the process of attending to, and/or actively discriminating stimuli sharpens encoding in auditory memory such that the representation of the standard stimulus is precise enough to allow for the passive discrimination of a deviant. In a related study, Schulte et al. (2001) demonstrated that adults learn to use complex pitch-cues following repeated exposure to a melody. Learning was indexed by an increase in evoked gamma band responses, suggesting increased neural synchrony and/or an enlargement of the cortical network generating the response. Both of these studies

demonstrate an effect of exposure on cortically-evoked responses. However, each depends to some extent on attention to the stimuli being discriminated. Attention is not easily manipulated in infant listeners, as it can be difficult to capture and maintain. Moreover the maturational states of attentional networks and working memory during early infancy are poorly understood. However, the present study demonstrates that a short amount of passive exposure may be sufficient to modulate cortically-evoked, processing-related ERP components in infants aged 4 to 8 months of age.

Across ages in the current study, mismatch responses to a change in IRN pitch were much larger in amplitude in right hemisphere electrodes than in left hemisphere electrodes. This is in agreement with recent imaging studies in adults suggesting that the right hemisphere is selectively activated in tasks related to pitch perception (Hyde et al., 2008), and production (Perry et al., 1999). Moreover, lesion studies suggest that the right hemisphere dominates perception of the pitch of the missing fundamental (Zatorre, 1988), processing of complex spectral structures (Sidtis and Volpe, 1988), and discriminating melodic pitch patterns (Zatorre, 1985; 1988). Functional lateralizations have been previously demonstrated in infants in response to a variety of acoustic properties including: speech signals (e.g. Dehaene-Lambertz et al., 2010), pitch accents (Sato et al., 2009), and temporal structures (Telkemeyer et al., 2009). In many cases, these asymmetries resemble those recorded from adult listeners. Thus, the hemisphere effect observed in the present study may represent an early right-hemisphere specialization for pitch processing.

Behavioral evidence suggests that spectral cues dominate pitch perception in infancy (Clarkson and Rogers, 1995; Butler et al., 2012, under review); discrimination is drastically impaired in the absence of resolved spectral pitch cues. However, the present study demonstrates that, following a period of pitch priming, infants as young as 4-months-old produce a significant mismatch component in response to a pitch change in IRN stimuli that lack resolvable spectral information. This represents the first evidence for a cortical representation of pitch discrimination by infant listeners in the absence of spectral cues. In future studies, it would be of interest to use ERPs to examine infant pitch discrimination for other types of auditory stimuli that evoke a pitch sensation in the absence of resolved spectral cues (e.g. high-pass filtered complex tones, Huggins pitch, click trains etc.) and whether priming can similarly lead to better pitch processing for such stimuli in infancy.

ACKNOWLEDGEMENTS:

This research was supported by grants to LJT from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) and an NSERC graduate scholarship to BEB. The authors wish to thank Elaine Whiskin and Cathy Chen for assisting with data collection.

REFERENCES

- Alho K, Sainio K, Sajaniemi N, Reinkainen K and Näätänen R. Event-related brain potentials of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*. 77(2):151-155, 1990.
- Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: MIT Press, 1990.
- Butler BE, Folland NA and Trainor LJ. Development of pitch processing: Infants' discrimination of iterated rippled noise stimuli with unresolved spectral content. *Journal of the Acoustical Society of America*. Under Review.
- Butler BE, Folland NA and Trainor LJ. Pitch discrimination in infancy: Examining the development Conference on Infant Studies. 2012. In Proceedings of the 18th International Conference on Infant Studies.
- Butler BE and Trainor LJ. Sequencing the cortical processing of pitch-evoking stimuli using EEG and source estimation. *Frontiers in Psychology*. 3(180): 1-13, 2012.
- Cariani PA and Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*. 76(3):1698-1716, 1996.
- Cheour M, Cěponienė R, Hukki J, Haapanen ML, Näätänen R and Alho K. Brain dysfunction in neonates with cleft palate revealed by the mismatch negativity. *Electroencephalography and Clinical Neurophysiology*. 110(2):324-328, 1999.
- Cedolin L and Delgutte B. Spatio-temporal representation of the pitch of complex tones in the auditory nerve. In Kollmeier B, Hohmann V, Langemann U, Mauermann

M, Uppenkamp S, and Verhey J (Eds.), *Hearing – From Sensory Processing to Perception*. Berlin: Springer-Verlag, Berlin, 2007: 61-70.

Cěponienė R, Hukki J, Cheour M, Haapanen ML, Koskinen M, Alho K and Näätänen R. Dysfunction of the auditory cortex persists in infants with certain cleft types. *Developmental Medicine & Child Neurology*. 42(4): 258-265, 2000.

Cěponienė R, Kushnerenko E, Fellman V, Renlund M, Suominen K and Näätänen R. Event-related potential features indexing central auditory discrimination by newborns. *Brain Research*. 13(1):101-113, 2002.

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

Clarkson MG and Clifton RK. Infants' pitch perception: Inharmonic tonal complexes. *Journal of the Acoustical Society of America*. 98(3): 1372-1379, 1995.

Clarkson MG and Rogers C. Infants require low-frequency energy to hear the pitch of the missing fundamental. *Journal of the Acoustical Society of America*. 98(1): 148-154, 1995.

Dehaene-Lambertz G, Montavont A, Jobert A, Alliol L, Dubois J, Hertz-Pannier L and Dehaene S. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*. 114(2): 53-65, 2009.

- Delgutte B and Cariani P. Coding of the pitch of harmonic and inharmonic complex tones in the interspike intervals of auditory nerve fibers. In Schouten MEH (Ed.), *The Processing of Speech*. Berlin: Mouton-De Gruyter, 1992: 37–45.
- Fellman V, Kushnerenko E, Mikkola K, Čėponienė R, Leipala J and Näätänen R. Atypical auditory event-related potentials in preterm infants during the first year of life: A possible sign of cognitive dysfunction? *Pediatric Research*. 56(2): 291-197, 2004.
- Griffiths TD, Buchel C, Frackowski RSJ and Patterson RD. Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*. 1(5): 422-427, 1998.
- Hall DA, Barrett DJK, Akeroyd MA and Summerfield AQ. Cortical representations of temporal structure in sound. *Journal of Neurophysiology*. 94(5): 3181-3191, 2005.
- He C, Hotson L and Trainor LJ. Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*. 19(5): 878-892, 2007.
- He C, Hotson L and Trainor LJ. 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, 2009.
- He C and Trainor LJ. Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience*. 29(24): 7718-7722, 2009.
- Hirasawa K, Kurihara M and Konishi Y. The relationship between mismatch negativity and arousal level. Can mismatch negativity be an index for evaluating the arousal level in infants? *Sleep Medicine*. 3(S2):45-48, 2003.

- Hyde KL, Peretz I and Zatorre RJ. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*. 46(2): 632–638, 2008.
- Kushnerenko E, Cėponienė R, Balan P, Fellman V, Huotilainen M and Năätänen R. Maturation of the auditory change detection response in infants: A longitudinal ERP study. *Neuroreport*. 13(15): 1843-1848, 2002.
- Larsen E, Cedolin L and Delgutte B. Pitch representation in the auditory nerve: Two concurrent complex tones. *Journal of Neurophysiology*. 100(3):1301-1319, 2008.
- Leppänen PHT, Guttorm TK, Pihko E, Takkinen S and Lyytinen H. Maturation effects on newborn ERPs measured in the mismatch negativity paradigm. *Experimental Neurology*. 190(S1):91-101, 2004.
- Leppänen PHT, Pihko E, Eklund KM and Lyytinen H. Cortical responses of infants with and without a genetic risk for dyslexia: II. Group effects. *Neuroreport*. 10(5):969-973, 1999.
- Meddis R and O'Mard L. A unitary model of pitch perception. *Journal of the Acoustical Society of America*. 102(3):1811-1820, 1997.
- Montgomery CR and Clarkson MG. Infants' pitch perception: Masking by low- and high-frequency noises. *Journal of the Acoustical Society of America*. 102(6): 3665-3672, 1997.
- Moore BCJ. *An Introduction to the Psychology of Hearing 5th ed.* London: Emerald Group Publishing, 2008.
- Mourad N, Reilly JP, De Bruin H, Hasey G and MacCrimmon D. A simple and fast algorithm for automatic suppression of high amplitude artifacts in EEG data. In

IEEE international conference on acoustics, speech and signal processing – proceedings. Honolulu: ICASSP, 2007: 1393-1396.

Näätänen R, Schröger E, Karakas S, Tervaniemi M and Paavilainen P. Development of a memory trace for a complex sound in the human brain. *Neuroreport*. 4(5): 503-506, 1993.

Patterson RD, Handel S, Yost WA and Datta AJ. The relative strength of the tone and noise components in iterated rippled noise. *Journal of the Acoustical Society of America*. 100(5): 3286-3294, 1996.

Patterson RD, Uppenkamp S, Johnsrude IS and Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron*. 36(4): 767-776, 2002.

Penagos H, Melcher JR and Oxenham AJ. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*. 24(30): 6810-6815, 2004.

Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E and Evans AC. Localization of cerebral activity during simple singing. *Neuroreport*. 10(18): 3979–3984, 1999.

Pressnitzer D and Patterson RD. Distortion products and the perceived pitch of complex tones. In Breebart DJ, Houtsma AJM, Kohlrausch A, Prijs VF and Schoonhoven R (Eds.), *Physiological and psychophysical bases of auditory function*. Maastricht, The Netherlands: Shaker, 2001: 97-104.

Puschmann S, Uppenkamp S, Kollmeier B and Thiel CM. Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage*. 49(2): 1641-1649, 2010.

Sato Y, Sogabe Y and Mazuka R. Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*. 22(11): 2503-2513, 2008.

Sayles M and Winter IM. Ambiguous pitch and the temporal representation of inharmonic iterated rippled noise in the ventral cochlear nucleus. *Journal of Neuroscience*. 28(46): 11925-11938, 2008.

Schulte M, Knief A, Seither-Preisler A and Pantev C. Gestalt recognition in a virtual melody experiment. In Nenonen J, Ilmoniemi R and Katila T (Eds.), *Biomag2000, Proceedings of the 12th International Conference on Biomagnetism*. Espoo: Helsinki University of Technology, 2001: 107 -110.

Sidtis JJ and Volpe BT. Selective loss of complex-pitch of speech discrimination after unilateral lesion. *Brain and Language*. 34(2): 235–245, 1988.

Telkemeyer S, Rossi S, Koch SP, Nierhaus T, Steinbrink J, Poeppel D, Obrig H and Wartenburger I. Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience*. 29(47): 14726-14733, 2009.

Trainor LJ and Corrigan KA. Music acquisition and effects of musical experience. In Riess-Jones M and Fay RR (Eds.), *Springer Handbook of Auditory Research: Music Perception*. Heidelberg: Springer, 2010: 89-128.

Von Békésy G. *Experiments in Hearing*. New York: McGraw-Hill, 1960.

Werner LA. Interpreting developmental psychoacoustics. In Werner LA and Rubel EW (Eds.), *Developmental Psychoacoustics*. American Psychological Association: Washington, 1992:47-88.

Winter IM, Wiegrebe L, and Patterson RD. The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *Journal of Physiology*. 537(2): 553-566, 2001.

Yost WA. Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*. 100(5): 3329–3335, 1996.

Zatorre RJ. Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*. 23(1): 31–41, 1985.

Zatorre RJ. Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*.. 84(2): 566–572, 1988.

CHAPTER 5

Unique Contributions of this Thesis to the Literature

This thesis demonstrates for the first time that infants are sensitive to temporal pitch cues. This was demonstrated using iterated rippled noise (IRN) stimuli that are generated when an exact copy of a white noise stimulus is added to itself, following a short delay. When this delay-and-add process is repeated a number of times, the introduction of temporal regularity produces a stimulus that elicits a pitch percept equal to the inverse of the delay (e.g. a 5 ms delay results in a 200 Hz pitch sensation). While this process also introduces characteristic ripples in the frequency spectrogram of the stimulus, this information can be removed using a high-pass filter so that spectral pitch cues are confined to the region in which they cannot be resolved by cochlear filters. Thus, listeners must rely heavily on temporal cues to discriminate between IRN stimuli.

Infants showed sensitivity to the temporal cues contained in IRN stimuli in two chapters of this thesis. In Chapter 3, 8-month-old infants demonstrated behavioural sensitivity to a change in pitch from 167 Hz to 200 Hz in a visually-reinforced, conditioned head-turn procedure. Following a training period in which the pitch of the target IRN stimulus (200 Hz) was emphasized by the addition of a pure tone, infants were shown to discriminate that target from the 167 Hz background stimulus. Chapter 4 provides electrophysiological evidence for infant discrimination of IRN stimuli. Following a period of pitch-priming, mismatch negativity (MMN) components were elicited from both 4- and 8-month-old infants in response to infrequent changes in the pitch of IRN stimuli. This provides evidence of a cortical representation of infant pitch

discrimination based on temporal cues, and demonstrates a neural basis for the behavioural discriminations observed in Chapter 3.

This evidence extends previous studies showing that infants can use spectral information to extract pitch. For example, using behavioural techniques, 8-month-old infants have been shown to differentiate between complex harmonic sounds, with or without energy at the fundamental frequency (Clarkson and Clifton, 1985). Moreover, infant perception of missing fundamental stimuli is unaffected by a low-pass noise masker, suggesting that their percepts are based on the harmonic structure of the sound rather than on cochlear distortions. Using event-related potentials, He and Trainor (2009) extended these findings to demonstrate that infants perceive the pitch of the missing fundamental by 4 months of age. Thus, although simple frequency discrimination does not fully mature until 10 years of age (Jensen and Neff, 1993; Thompson, Cranford and Hoyer, 1999), it appears as though pitch cues can be consolidated into a singular pitch representation in early infancy.

Although the evidence presented in this thesis suggests that infants can extract pitch based on temporal cues, there are reasons to believe that this temporal percept is much weaker than one based on robust spectral cues. For example, the behavioural discriminations described in Chapter 3 were only observed following a training period in which the pitch of the target stimulus was emphasized by the addition of a pure tone that provided spectral information. Even then, performance was weak relative to sensitivity values described in the literature for pitch-evoking stimuli with robust spectral cues (e.g. Clarkson and Clifton, 1985). Similarly, the MMN elicited by IRN pitch changes reported

in Chapter 4 only reached significance when infants first heard a priming period in which IRN pitch was emphasized through the addition of pure tones. Thus, while infants as young as 4 months of age are sensitive to temporal pitch cues, their pitch percepts appear to be dominated by spectral pitch information. This is not surprising, as adult pitch percepts are stronger when spectral cues are present than when they are absent (Houtsma and Smurzynski, 1990), and are dominated by resolved spectral information provided by harmonics three through five (where available; Plomp 1967; Ritsma 1967).

This thesis also provides new information about adults' processing of temporal versus spectral information. Chapter 2 demonstrates that harmonic complex stimuli and IRN stimuli, which differ in the relative strength of spectral and temporal pitch cues, elicit exogenous cortical components (P1, N1) that differ in latency and place of generation. Despite best efforts to equate these stimuli across a number of domains (e.g. pitch saliency, high-pass filter cutoff and slope), differences in stimulus features, including timbre, remained. Differences in the P1 and N1 components evoked by these stimuli reveal that featural differences are represented in cortically evoked ERP components that are related to feature extraction. Conversely, no differences were found in the MMN components elicited by the two stimulus types. The MMN is elicited by perceivable auditory objects (Näätänen and Winkler, 1999). Thus, the absence of latency and source differences in the MMN component suggests that pitch is processed in the same cortical area, whether pitch cues in the stimulus are predominantly spectral, temporal, or some combination of the two.

Source estimates in Chapter 2 demonstrated that the MMN elicited in response to a pitch change was common for IRN and complex harmonic stimuli. Moreover, this source was estimated to arise from an area of cortex located between the source estimates for P1 and N1. The finding of an MMN component that is common to both stimulus types suggests that the pitches of these different stimuli are being extracted in a common location. Moreover, the location of the source of the MMN component is in accordance with electrophysiological (Bendor and Wang, 2005) and functional imaging data (Griffiths et al., 1998; Patterson et al., 2002; Penagos et al., 2004; Hall et al., 2005; Puschmann et al., 2010) suggesting that a common pitch centre is located beyond primary auditory cortex, along the lateral aspect of Heschl's gyrus.

This evidence extends previous studies examining the cortical location of a common pitch-processing centre in adults. This body of research has overwhelmingly relied upon the impressive spatial resolution of techniques like functional magnetic resonance imaging (fMRI) to pinpoint an area of cortex that is sensitive to pitch. However, Chapter 2 represents the first study of adults' perception of pitch-evoking stimuli with differing contributions from spectral and temporal cues with sufficient temporal resolution to examine the timing of pitch percept formation, relative to other auditory processes (e.g. basic feature extraction). While Chapter 2 affirms that pitch-evoking stimuli are indeed processed in a common cortical area, regardless of the relative contributions of spectral and temporal cues, it also suggests that stimulus features elicit distinct cortical representations prior to object formation.

Together, the data presented in Chapters 2 through 4 make a number of significant contributions to the pitch perception literature. Current models of pitch perception acknowledge that temporal cues (derived from the pattern of action potentials contained in the cochlear output) make a meaningful contribution to the formation of pitch percepts. Adults are quite capable of extracting pitch from stimuli that rely on temporal cues. For example, occasional changes in the pitch of IRN stimuli elicit robust MMN responses from adults without pitch-priming (see Chapter 2), while they perform the behavioural discrimination described in Chapter 3 with 100% accuracy. The behavioural and electrophysiological evidence presented in Chapters 3 and 4, respectively, suggests that the mechanisms responsible for pitch extraction from temporal cues are present in listeners as young as 4 months of age. However, infants' weak performance on IRN pitch discrimination relative to their discrimination of stimuli with robust spectral pitch cues supports the ideas that temporal mechanisms are immature even at 8 months of age, and that infant pitch percepts are based primarily on spectral cues.

Limitations

Stimulus construction created limitations across each of the experiments in this thesis. IRN stimuli have been described as being comprised of two perceptual components: one that sounds similar to white noise, and one that has a perceivable pitch. It has been demonstrated that the pitch salience of the latter is related to the number of iterations of the delay-and-add process used to create IRN stimuli, reaching a plateau after sixteen repetitions (Yost, 1996). However, creating IRN stimuli with this number of

iterations introduces some spectral cues to the signal, in addition to temporal regularity. This spectral information is diffuse relative to spectral cues contained in a traditional harmonic complex (see Chapter 2, Figure 2) and can be confined, using high-pass filtering, to the region where spectral cues are unresolved. However, the presence of spectral information limits our ability to conclude that behavioural discriminations observed in Chapter 3 are based solely on temporal pitch cues; the same is true of the cortical representations of those discriminations observed in Chapter 4. The differences observed in Chapter 2 between the P1 and N1 components elicited by IRN stimuli and harmonic complex stimuli suggest that the pitch cues provided by these two stimulus types are indeed different, with perception of IRN stimuli likely relying to a greater extent on temporal pitch cues. While it is tempting to make conclusions regarding temporal pitch mechanisms based on the present results, one must consider that even the high-pass filtered IRN stimuli used herein do not contain *purely* temporal cues to pitch.

The fit of the source estimates for the MMN component presented in Chapter 2 were fairly accurate, accounting for a mean variance of 92%. However, without structural MRI scans of the individuals who participated in the current study, we cannot determine the exact location of this source in auditory cortex. Instead, we are limited to determining a location relative to the P1 and N1 components whose source locations have been well documented using intracerebral evoked potentials (Celesia, 1976; Richer, Alain, Achim, Bouvier, and Saint-Hilaire, 1989; Godey et al., 2001; Yvert et al., 2005). Thus, while the data are suggestive, the limited spatial resolution of the ERP technique

does not allow us to conclude that the MMN component is being generated along lateral Heschl's gyrus.

An additional limitation related to source modeling concerns the number of artifact-free trials required to compute a reasonable estimate. Two-thirds of the adults tested in Chapter 2 provided sufficient data to compute source estimates for each of the components of interest (P1, N1, and MMN). However, none of the 4- or 8-month-olds tested in Chapter 4 provided a sufficient number of artifact-free trials. This is a common limitation of infant ERP studies, and is related to the inherent constraints involved with electrophysiological testing in infancy: overall testing time is limited with infant listeners, and the number of trials containing movement-related artifacts is often quite large. Moreover, if reliable source estimates could be computed for infant listeners, the limitations related to specifying cortical location would be amplified relative to those describe above in adults. During infancy, large-scale changes in cortical folding and cortical thickness would make the accurate interpretation of source analysis a challenge, even if individual structural MRI scans were available. Thus, the lack of spatial resolution is a limitation to the interpretation of data collected using ERP in infants.

Finally, while the data presented in Chapters 3 and 4 suggest that infants are capable of discriminating between IRN stimuli, they do not show conclusively that these discriminations are based on pitch percepts. However, in both cases, performance increased following a period of pitch training/priming, suggesting they are indeed sensitive to temporal pitch cues.

Future Directions

Chapter 3 demonstrates that 8-month-old infants can discriminate between stimuli of different pitch with unresolved spectral content. This appears discordant with prior research suggesting that, given only the unresolved harmonics of a complex tone, infants are unable to classify stimuli according to pitch (Clarkson and Clifton, 1995). While I pointed out in Chapter 3 that the infants in the current study are somewhat older than those tested by Clarkson and Clifton (1995), it is possible that differences between stimuli also contribute to this differential performance. Future research should address the question of at what age infants can discriminate the stimuli used by Clarkson and Clifton (1995), and should extend this research to other pitch-evoking stimuli that rely primarily on temporal cues (e.g. Huggins pitch, binaural band pitch, etc.).

If the data presented in Chapter 3 are considered on an individual basis, it appears that there may be two separate subgroups: thirteen infants who were able to discriminate these stimuli, and an additional group of three infants who could not. It is possible that the ability to behaviourally discriminate IRN sounds has only recently come ‘online’ at 8 months of age, and that these two subgroups simply represent infants who have or have not passed this developmental milestone. Thus, it would be of interest to test infants of different age groups to track the developmental progression of behavioural discrimination. Alternatively, these subgroups may differ along a dimension that is not simply related to age. Performance was not related to any of the factors measured on the

survey administered in Chapter 3 (e.g. parental music experience, first/second language experience, etc.). However, it would be of interest to examine potential within-subject correlations between performance on IRN pitch discrimination and other challenging pitch tasks (e.g. discrimination of inharmonic tones), as well as non-pitch auditory tasks in the temporal domain (e.g. gap detection).

Each of the studies in the current thesis employed IRN tones created following 16 iterations of the delay-and-add process. In the future, it would be of benefit to extend this work using stimuli with fewer iterations. Firstly, this would introduce less spectral information into the stimuli, allowing for more conclusive statements regarding the processing of pitch-evoking stimuli in the absence of spectral cues. Additionally, varying the number of iterations used to create the IRN stimuli would provide a simple means of manipulating pitch saliency. With similar stimuli of different saliency, one could address whether cortical responses from the pitch centre in infants are related to pitch saliency, as they are in adults (Hall, Barrett, Akeroyd, and Summerfield, 2005; Hall and Plack, 2009; Puschmann et al., 2010; Barker et al., 2011a). Moreover, ERP analysis and source localization in adults may help to reveal whether cortical processes involved in pitch perception differ depending on the saliency of the pitch percept.

Final Conclusions

In this thesis, I examined how adult and infant listeners perceive pitch-evoking stimuli when spectral cues are limited to the region where they cannot be resolved by a place-based frequency mechanism. High-pass filtered IRN stimuli and more traditional

harmonic stimuli provide different pitch cues, as evidenced by differences in feature extraction-related evoked potentials. However, in adult listeners, it appears that the pitch percept resulting from these cues, as measured by MMN, is independent of the type of cue present. Moreover, source estimates for the MMN component elicited by pitch-changes were consistent with functional imaging studies suggesting a common pitch-processing centre along lateral Heschl's gyrus.

When robust spectral cues are present, 8-month-old infants perceive pitch-evoking stimuli in an adult-like manner. However, when these spectral cues could not be resolved, infants showed very poor discrimination between stimuli with different pitch. Indeed, I found that a period of pitch-priming, during which stimulus pitch was emphasized using spectral cues, was necessary prior to successful discrimination. However, ERP evidence suggested that following this priming period, infants as young as 4 months of age were capable of discriminating high-pass filtered IRN stimuli, although their performance was weak relative to discrimination for stimuli with resolvable spectral content.

The work in this thesis suggests that infants as young as 4 months of age are sensitive to temporal pitch cues. However, while the adult auditory system may perceive a single pitch percept regardless of the type of cues present, infant pitch percepts appear to be based predominantly on resolved spectral cues. This indicates that it is important to consider that infant pitch perception may be particularly impaired when listening to complex acoustic signals, such as speech, in difficult listening environments, such as in the presence of background noise.

References

- Alain, C., Woods, D.L., & Knight, R.T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research*, *812*, 23-37.
- Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potentials of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*, 151-155.
- Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology*, *31*, 469-479.
- Alho, K., Woods, D. L., Algazi, A., Knight, R. T., & Näätänen, R. (1994). Lesions of frontal cortex diminish auditory mismatch activity. *Electroencephalography and Clinical Neurophysiology*, *91*, 353-362.
- Arnott, S. R., & Alain, C. (2002). Stepping out of the spotlight: MMN attenuation as a function of distance from the attended location. *NeuroReport*, *13*(17), 2209-2212.
- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music specific disorder. *Brain*, *125*, 238-251.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*, 1161-1165.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.

- Brugge, J. F., Anderson, D. J., Hind, J. E., & Rose, J. E. (1969). Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *Journal of Neurophysiology*, *32*, 386-401.
- Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America*, *60*, 863-869.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 1698-1716.
- Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, *76*, 1717-1734.
- Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain: A Journal of Neurology*, *99*(3), 403-414.
- Cheour, M., Alho, K., Ceponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., et al. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, *29*, 217-226.
- Clarkson, M. G., & 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.
- Clarkson, M. G., & Clifton, R. K. (1995). Infants' pitch perception: Inharmonic tonal complexes. *Journal of the Acoustical Society of America*, *98*, 1372-1379.

- Clarkson, M. G., & Rogers, C. (1995). Infants require low-frequency energy to hear the pitch of the missing fundamental. *Journal of the Acoustical Society of America*, 98, 148-154.
- Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America* 30(5), 413-417.
- Dittmann-Balcar, A., Jüptner, M., Jentzen, W., & Schall, U. (2001). Dorsolateral prefrontal cortex activation during automatic auditory duration-mismatch processing in humans: A positron emission tomography study. *Neuroscience Letters*, 308, 119-122.
- Evans, E. F. (1977). Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus. In E. F. Evans & J. P. Wilson (Eds.), *Psychophysics and Physiology of Hearing* (pp. 185-192). Academic Press: London.
- Formisano, E., Kim, D., Di Salle, F., van de Moortele, P., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40, 859-869.
- Frick, R.W. (1985). Communicating emotion: The role of prosodic features. *Psychol. Bull.* 97, 412-429.
- Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, 13, 1251–1254.

Friedrich, M., Weber, C., & Friederici, A. D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment.

Psychophysiology, *41*, 772–782.

Giard, M. H., Lavikainen, J., Reinikainen, K., Bertrand, O., Pernier, J., & Näätänen, R.

(1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model study.

Journal of Cognitive Neuroscience. *7*, 133-143.

Gockel, H. E., Carlyon, R. P., Mehta, A., and Plack, C. J. (2011). The frequency

following response for dichotic pitch stimuli: No evidence for pitch encoding.

Journal of the Acoustical Society of America, *129*, 2592-2592.

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. *Clin. Neurophysiol.*

112, 1850-1859.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the

pitch of complex tones. *Journal of the Acoustical Society of America*, *54*, 1496-

1516.

Griffiths, T.D., Buchel, C., Frackowski, R.S.J., and Patterson, R.D. (1998). Analysis of

temporal structure in sound by the human brain. *Nature Neuroscience*, *1*, 422-427.

Hall, D. A., Barrett, D. J. K., Akeroyd, M. A., and Summerfield, A. Q. (2005). Cortical

representations of temporal structure in sound. *Journal of Neurophysiology*, *94*,

3181-3191.

Hall, D. A., and Plack, C. J. (2009). Pitch processing sites in the human auditory brain.

Cerebral Cortex, 19, 576-585.

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. (2009a). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *The European Journal of Neuroscience*, 29, 861-867.

He, C., Hotson, L., & Trainor, L. J. (2009b). 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.

Hepper, P. G., and Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood*, 71, F81-F87.

Houtsma, A. J. M., & Goldstein, J. L. (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *Journal of the Acoustical Society of America*, 51, 520-529.

Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, 87(1), 304-310.

Humphries, C., Liebenthal, E., and Binder, J.R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage* 50, 1202-1211.

- Javitt, D. C., Steinschneider, M., Schroeder, C. E., & Arezzo, J. C. (1996). Role of cortical *N*-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia. *Proceedings of the National Academy of Sciences of the USA*, *93*, 11962-11967.
- Jensen, J. K., and Neff, D. L. (1993). Development of basic auditory discrimination in preschool children, *Psychological Science*, *4*, 104-107.
- Kushnerenko, E., Ceponiené, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory change detection response in infants: A longitudinal ERP study. *NeuroReport*, *13*, 1843–1848.
- Lang, H. A., Nyrke, T., Ek, M., Aaltonen, O., Raimo, I., & Näätänen, R. (1990). Pitch discrimination performance and auditory event-related potentials. In C. H. M. Brunia, A. W. K. Gaillard, A. Kok, G. Mulder, & M. N. Verbaten (Eds.), *Psychophysiological brain research* (Vol. 1, pp. 294-298). Tilburg, The Netherlands: Tilburg University Press.
- Langers, D. R. M., & van Dijk, P. (2012). Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex*, *22*, 2024-2038.
- Møller, A. R. (1970). Two different types of frequency selective neurons in the cochlear nucleus of the rat. In R. Plomp, & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 168-174). Leiden, The Netherlands: A. W. Sijthoff.

- Morlet, T., Collet, L., Salle, B., & Morgon, A. (1993). Functional maturation of cochlear active mechanism and of the medial olivocochlear system in humans. *Acta Otolaryngologica*, *113*(3), 271-277.
- Müller, B. W., Jüptner, M., Jentzen, W., & Müller, P. (2002). Cortical activation to auditory mismatch elicited by frequent deviant and complex novel sounds: A PET study. *Neuroimage*, *17*, 231-239.
- Näätänen, R. (1979). Orienting and evoked potentials. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 61-75). Hillsdale, NJ: Erlbaum.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313-329.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1980). Brain potential correlates of voluntary and involuntary attention. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motor and sensory processes of the brain: Electrical potentials, behaviour and clinical use* (Vol. 54, pp. 343-348). Amsterdam: Elsevier.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544-2590.
- Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*, 436-450.

- Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375-425.
- Novitski N., Huotilainen M., Tervaniemi, M., Näätänen, R., & Fellman, V. (2007). Neonatal frequency discrimination in 250– 4000-Hz range: electrophysiological evidence. *Clinical Neurophysiology*, *118*, 412-419.
- Nozza, R. J. (1995). Estimating the contribution of non-sensory factors to infant-adult differences in behavioural thresholds. *Hearing Research*, *91*, 72-78.
- Olsho, L. W., Schoon, C., Sakai, R., Terpin, R., and Sperduto, V. (1982). Auditory frequency discrimination in infancy, *Developmental Psychology*, *18*(5), 721-726.
- Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination, *Journal of the Acoustical Society of America*, *82*(2), 454-464.
- Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right-hemisphere dominance of different mismatch negativities. *Electroencephalography & Clinical Neurophysiology*, *78*, 466-479.
- Pantev, C., Elbert, T., Ross, B., Eulitz, C., and Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in human auditory cortex. *Hearing Research*, *100*, 164-170.
- Patterson, R. D. (1973). The effect of relative phase and the number of components on residue pitch. *Journal of the Acoustical Society of America*, *53*, 1565-1572.

- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767-776.
- Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24*, 6810-6815.
- Pincze Z., Lakatos, P., Rajkai, C., Ulbert, I., & Karmos, G. (2001). Separation of mismatch negativity and the N1 wave in the auditory cortex of the cat: A topographic study. *Clinical Neurophysiology*, *112*, 778-784.
- Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, *41*, 1526-1533.
- Pujol, R., & Uziel, A. (1988). Auditory development: peripheral aspects. In E. Meisami, & P.S. Timiras (Eds.), *Handbook of human growth and developmental biology*, Vol. IB (pp. 109-130). Boca Raton, Florida: CRC Press, Inc.
- Puschmann, S., Uppenkamp, S., Kollmeier, B., and Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage*, *49*, 1641-1649.
- Richer, F., Alain, C., Achim, A., Bouvier, G., & Saint-Hilaire, J.-M. (1989). Intracerebral amplitude distributions of the auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, *74*(3), 202-208.
- Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America*, *42*, 1637-1644.

- Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, *11*(1), 61-63.
- Sams, M., Alho, K., & Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology*, *21*(4), 434-441.
- Schwarz, D.W., and Tomlinson, R.W. (1990). Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J. Neurophysiol.* *64*, 282-298.
- Schönweisner, M., von Cramen, D. Y., & Rübsem, R. (2002). Is it tonotopy after all? *Neuroimage*, *17*, 1144-1161.
- Schouten, J. F. (1940). The residue, a new component in subjective sound analysis. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, *43*, 356-365.
- Shahidullah, S., and Hepper, P. G. (1994). Frequency discrimination by the fetus, *Early Human Development*, *36*, 13-26.
- Song, J., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience*. *23*(9), 2268-2279.
- Spetner, N. B., & Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Development*, *61*, 632-652.
- Striem-Amit, E., Hertz, U., & Amedi, A. (2011). Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One*, *6*, e17832.

Sussman, E., Wikler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, *40*, 430-435.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, *1*, 155-182.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994).

Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*, 157-167.

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

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *370*, 90-92.

Trainor, L. J., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., et al. (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., Desjardins, R. N., & Sonnadara, R. R. (2001). Measuring temporal resolution in infants using mismatch negativity. *NeuroReport*, *12*, 2443–2448.

van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, *33*(3), 259-272.

Von Békésy, G. (1960). *Experiments in Hearing*. New York: McGraw-Hill.

- Walker, K. M. M., Bizley, J. K., King, A. J., & Schnupp, J. W. H. (2011). Cortical encoding of pitch: Recent results and open questions. *Hearing Research*, 271, 74-87.
- Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting*, 6, 171-178.
- Wightman, F. L. (1973). Pitch and stimulus fine structure. *Journal of the Acoustical Society of America*, 54, 397-406.
- Winkler I., Kushnerenko E., Horváth J., Ceponiene R., Fellman V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Science*, 100, 11812-11815.
- Winkler, I., Tervaniemi, M., & Näätänen, R. (1997). Two separate codes for missing-fundamental pitch in the human auditory cortices. *Journal of the Acoustical Society of America*, 102, 1072-1082.
- Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28, 30-42.
- Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, 8, 1971–1974.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35, 615– 619.

Yost, W.A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, *100*, 3329-3335.

Yost, W. A., Patterson, R. D., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, *99*, 1066-1078.

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.