Early development of polyphonic sound encoding and the high voice superiority effect

C’line Marie*a,b, Laurel J. Trainora,b,c,n

a Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
b McMaster Institute for Music and the Mind, Hamilton, Ontario, Canada
c Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada

Article history:
Received 11 September 2013
Received in revised form 25 February 2014
Accepted 26 February 2014
Available online 13 March 2014

Keywords:
Auditory development
Auditory scene analysis (ASA)
Infant
Mismatch negativity (MMN)
Polyphonic sound
Music

Abstract

Previous research suggests that when two streams of pitched tones are presented simultaneously, adults process each stream in a separate memory trace, as re ected by mismatch negativity (MMN), a component of the event-related potential (ERP). Furthermore, a superior encoding of the higher tone or voice in polyphonic sounds has been found for 7-month-old infants and both musician and non-musician adults in terms of a larger amplitude MMN in response to pitch deviant stimuli in the higher than the lower voice. These results, in conjunction with modeling work, suggest that the high voice superiority effect might originate in characteristics of the peripheral auditory system. If this is the case, the high voice superiority effect should be present in infants younger than 7 months. In the present study we tested 3-month-old infants as there is no evidence at this age of perceptual narrowing or specialization of musical processing according to the pitch or rhythmic structure of music experienced in the infant’s environment. We presented two simultaneous streams of tones (high and low) with 50% of trials modi ed by 1 semitone (up or down), either on the higher or the lower tone, leaving 50% standard trials. Results indicate that like the 7-month-olds, 3-month-old infants process each tone in a separate memory trace and show greater saliency for the higher tone. Although MMN was smaller and later in both voices for the group of sixteen 3-month-olds compared to the group of sixteen 7-month-olds, the size of the difference in MMN for the high compared to low voice was similar across ages. These results support the hypothesis of an innate peripheral origin of the high voice superiority effect.

Copyright © 2014 Elsevier Ltd. All rights reserved.

1. Introduction

From birth, infants are immersed in a noisy world containing simultaneous and overlapping sounds (e.g., multiple talkers, environmental sounds, music). The capacity to process the complex sound wave reaching the ear and determine what auditory objects (sources emitting streams of sounds) are present, and where they are located in the environment, is crucial for following voices and learning linguistic and musical structure. Although several studies have demonstrated infants ability to perceptually organize sequential, non-overlapping sounds (e.g., Demany, 1982; McAdams & Bertocini, 1997; Smith & Trainor, 2011; Trainor & Adams, 2000; Winkler, H den, Ladinig, Sziller, & Honing, 2009), only a few experiments have examined the perception of simultaneous sounds in infancy (Folland, Butler, Smith, & Trainor, 2012; Marie & Trainor, 2013). In the present study we examine whether 3-month-old infants can encode two streams of tones with simultaneous note onsets in auditory cortex and, if so, whether they are like older infants and adults in showing a more robust encoding of the stream with higher pitch.

Auditory scene analysis (Bregman, 1990) requires spectrotemporal analysis of the incoming sound wave in order to integrate components that belong to single objects and separate components that belong to different objects. These processes of integration and separation apply to both simultaneous and sequential aspects of the sound input. Bregman (1990) hypothesized that although learning can affect some aspects of auditory scene analysis, these processes largely occur automatically and without conscious awareness. Given the importance of relatively low level, bottom up processes in auditory scene analysis, it is plausible to predict that the ability to segregate simultaneous sounds should be present early in development.

We investigated whether two simultaneous tones elicit two memory traces in auditory cortex in adults using the mismatch negativity (MMN) component of event-related potential (ERP) electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005;
In general, MMN is elicited when there is an occasional change or deviant stimulus (deviants can re ect a change in pitch, intensity, timbre, or pattern, etc.) in an ongoing sequence of standard stimuli (see N - t n, Pauvliainen, Rinne, & Alho, 2007; Picton et al., 2000 for reviews). MMN is generated in secondary auditory cortex. It appears at the scalp as a frontal negativity, reverses in polarity at posterior sites when using an average reference, and peaks between 120 ms and 250 ms after stimulus onset, depending on stimulus complexity. MMN is thought to re ect the formation of memory traces in auditory cortex. Importantly, it occurs in response to unpredicted sounds, is only present when deviant sounds in an ongoing stream are relatively rare, and increases in amplitude with decreasing probability of a deviant sound. Fujioka et al. (2005) presented binaurally two simultaneous melodies (called voices or streams) on each trial. The two melodies t together harmonically, and were each composed from the rst 5 notes of the Western major scale. Which melody was in the high voice and which was in the low voice varied across conditions of the experiment. On 25% of trials, one note (deviant stimulus) was changed in the higher-pitched melody and on another 25% of trials, one note (deviant stimulus) was changed in the lower-pitched melody, leaving 50% standard trials. Signi cant MMN was found for both changes, suggesting that separate memory traces are formed for each melody in auditory cortex. This result was replicated with simpler stimuli in which a single high tone (462.6 Hz, B- at, international standard nota- tion) repeated in one voice and a single low tone (196.0 Hz, G3) in another, with 23% of repetitions containing a pitch change (two semitones) in the higher tone and 25% in the lower tone (Fujioka et al., 2008). The stimuli were presented binaurally and, again, MMN was found for both deviant stimuli. Furthermore, we tested 7-month-old infants on the same simultaneous tone stimuli (Marie & Trainor, 2013) and found that they also showed MMN responses to pitch changes in both the high and low tones, suggesting that they also form two memory traces for simultaneous tones in auditory cortex.

Interestingly, the MMN amplitude in both Fujioka et al. (2005) and Fujioka et al. (2008) was consistently larger for deviant stimuli in the higher-pitched compared to lower-pitched voice. This high voice superiority effect is consistent with the musical practice of most often placing the most important melody line in the highest-pitched voice. It is also consistent with behavioral studies indicating that listeners nd it easiest to detect pitch changes in the highest-pitched of several simultaneous streams (Crawley, Acher-Mills, Pastore, & Weil, 2002; Palmer & Holleran, 1994; Zenatti, 1969). In order to test whether the high voice superiority effect results from experience with Western music, we tested adult musicians who played either a soprano range or a bass range instrument (Marie, Fujioka, Herrington, & Trainor, 2012). The logic was that if this effect reects long-term exposure to Western musical compositional practice, musicians who have played an instrument in the bass register for many years should show a low voice superiority effect. We found that musicians who played a soprano range instrument showed a high voice superiority effect as expected; however, musicians who played a bass range instrument did not show a low voice superiority effect, but rather a nonsigni cant trend for a high voice superiority effect. This suggests that although the high voice superiority effect might be somewhat modi able by many years of experience, it is dif cult, if not impossible, to reverse the high voice superiority effect. Furthermore, the 7-month-old infants in Marie and Trainor (2013) also showed a high voice superiority effect.

We investigated a possible peripheral innate origin to the high voice superiority effect using a computational model of the ear (Ibrahim & Bruce, 2010; Zilany, Bruce, Nelson, & Carney, 2009) whose input is a sound wave and whose output is a representation of neural ring patterns in the auditory nerve (Trainor, Marie, Bruce, & Bidelman, 2014). Inputting the stimuli of Fujioka et al. (2008) and Marie and Trainor (2013), we found that the pitch salience (Bidelman & Heinz, 2011) of the higher tone was greater than that of the lower tone, particularly in mid and lower pitch ranges. These results suggest that the high voice superiority effect might have an innate origin in physiological properties of the cochlea.

Although the data from 7-month-old infants is consistent with an innate origin for the high voice superiority effect, there is evidence of enculturation or perceptual narrowing by this age, so it is possible that by 7 months of age, infants have already learned the Western cultural norm of placing the most important melody in the highest voice. Perceptual narrowing refers to the phenomenon whereby very young infants process perceptual features that are relevant as well as those that are not relevant in their environment, whereas older infants become specialized for features that are relevant in their environment (e.g., Lewkowicz & Ghazanfar, 2009; Scott, Pascalis & Nelson, 2007). Specifically, older infants, like adults in their culture, become worse at processing perceptual features that only matter in foreign contexts. For example, at 6 months of age, infants are equally good at processing linguistic consonant contrasts from their native and foreign languages, but by 10 months of age they, like adults, become worse at perceiving consonant contrasts that are not used in their native language (e.g., Japanese adults and older infants have dif cult distinguishing /l/ and /r/ whereas younger Japanese infants do not; Werker, 1989). Similarly, 4-month-olds are able to discriminate monkey faces as well as human faces, and foreign-race faces as well as own-race faces, but between 6 and 8 months infants are becoming specialized for discriminating faces from their own species and own race (e.g., Slater et al., 2010). As well, the ability to discriminate voices from a foreign species (rhesus monkey) decreases between 6 and 12 months of age (Friendly, Rendall, & Trainor, 2013). In the musical domain, more studies are needed, but the evidence suggests that infants become specialized for processing the tonal and rhythmical structures in the music of their culture after 6 months of age (Gerry et al., 2012; Hannon & Trehub, 2005; Hannon & Trainor, 2007; Lynch, Eilers, Oller, & Urbano, 1990; Lynch & Eilers, 1992; Trainor & Trehub, 1992; Trainor, 2005; Trainor & Corrigall, 2010; Trainor & He, 2013; Trainor, Marie, Gerry, Whiskin, & Unrau, 2012; Trainor & Unrau, 2012; Trainor & Hannon, 2013). Therefore, in the present study, we chose to test whether infants younger than 6 months show a high voice superiority effect, as there is little evidence of perceptual narrowing in any domain before this age.

By 3 months of age, adult-like MMN responses to pitch changes can be measured. The ERP responses at 2 months of age to occasional changes (deviant stimuli) in pitch in a repeating sequence of sounds are dominated by an increase in the amplitude of a frontally positive slow wave (e.g., Friederici, Friedrich, & Weber 2002; He, Hotson, & Trainor, 2007, 2009a,b; Lepp nen, Guttorm, Ptkko, Takkinen, & Lytinen, 2004; Morr, Shafer, Kreuzer, & Kurtzberg, 2002), a response that decreases with increasing age. However, by 3 months of age an MMN with adult-like morphology is clearly evident, which decreases in latency and increases in amplitude with increasing age (e.g., Choudhury & Benasich, 2011; He et al., 2007; Kushnerenko et al., 2002; Trainor, 2012).

Related to the development of auditory scene analysis for simultaneous tones, the integration of harmonics into a single pitch percept appears not to have a cortical representation until after 3 months of age as evidenced by ERP studies on the pitch of the missing fundamental (He & Trainor, 2009), and the ability to use harmony as a cue for auditory object separation also appears to emerge around this age as evidenced by ERP studies of object-related ERP responses to mistuned harmonics.
(Folland et al., 2012). Based on this literature, we hypothesized that the ability to process two simultaneous tones as separate auditory objects should be present around 3 months of age. We therefore chose to study 3-month-olds infants. With respect to the high voice superiority effect, we hypothesized that, despite a lack of specialization for Western musical structure at this age, 3-month-old infants would nonetheless show a high voice superiority effect, consistent with the idea that it arises at the level of the cochlea (Trainor, Marie, Bruce, & Bidelman, 2014).

2. Materials and methods

2.1. Participants

Twenty-ve 3-month-old infants were tested. All were born between 38 and 42 weeks gestation and were healthy at birth. Six were excluded due to excessive movement or fussiness during the recording and 3 for having artifacts due to paci er use, leaving 16 infants in the nal sample (8 males; mean age 15.24 weeks; S.D. 0.83 weeks, see Table 1). All infants showed normal hearing by the Ontario Newborn Screening protocol (Hype, 2005). In Ontario, normal hearing screening is universal. All infants are given an automated otoacoustic emissions (AOE) test and a follow up auditory brainstem response (ABR) test if necessary. In our experiment we included only infants who passed this screening. After providing informed consent to participate, parents completed a brief questionnaire for auditory screening purposes and to assess musical background. According to the questionnaire, no infants had a history of frequent ear infections or a history of hearing impairment in the family, and all infants were healthy at the time of testing. All parents reported that infants listened to music every week (mean 12 h/week, range 3 28 h/week). Parents of 10 infants had played an instrument before having children but they reported having stopped playing by the time of testing. Finally, 3 families were bilingual (English with French or Croatian or Korean) and the other 13 families spoke only English. The data from the sixteen 3-month-old infants were compared in certain analyses to those from the 7-month-old infants tested in Marie and Trainor (2013). The group of 7-month-olds was composed of sixteen infants including 8 males (mean age 234.8 days, range 219 244 days).

2.2. Stimuli

The stimuli from Marie and Trainor (2013) were used. Tones were 300 ms computer-synthesized piano tones (Creative Sound Blaster). The stimuli were equalized for loudness using the Equal-loudness function from Cool Edit Pro software (Group waveforms normalize). This normalization takes into account the sensitivity of the human auditory system across the frequency range. Notes were presented every 600 ms (stimulus onset asynchrony, SOA 600 ms) at approximately 60 dB(A) measured at the location of the infant's head. Each condition was 11 minutes long, containing 1088 individual notes presented in pseudorandom order, with the constraint that a deviant stimulus could not be followed immediately by an identical deviant. In the Two-Voice (2V) condition, the standard tones had fundamental frequencies of 466.2 Hz (B), 2 international standard notation) and 196.0 Hz (G), which are 15 semitones apart and form a minor tenth interval (octave displaced minor third). Deviant stimuli were created by a one-semitone (1/12 octave) pitch deviation, going up or down from each tone of the dyad (i.e., B to C# or D to E). For the high voice deviants, GF or 207.8 Hz and F# or 185 Hz for the low voice deviants). The High-Voice-alone (HV-alone) condition was identical to the 2V condition except that the lower tones were omitted. Similarly, the low-Voice-alone (LV-alone) condition was identical to the 2V condition except that the higher tones were omitted (See Fig. 1).

2.3. Procedure

The procedure was identical to that in Marie and Trainor (2013) so as to be able to directly compare the performance of 3- and 7-month-old infants. The procedure was explained to parents who gave consent for their infant to participate. The parents sat in the sound attenuated chamber (Industrial Acoustics Company) with their infant sitting on their lap, facing the loudspeaker and a screen. In order to keep them still, awake and happy, during the experiment the infant watched a silent movie and/or a puppet show provided by an experimenter who also sat in the room. Sounds were presented using Eprime software through a loudspeaker located 1 m in front of the infant's head. In the 2V condition, 50%(or 544) of trials were standards and 50%(or 544) were deviant stimuli, with 12.5%(or 136) of each deviant type (high-tone up, high-tone down, low-tone up and low-tone down). In the HV-alone and LV-alone conditions, 75% of trials were standards (or 816) and 25% were deviant stimuli, with 12.5%(or 136) of each deviant type (up, down). All infants were run on the 2V condition rst. If the infant completed the 2V condition and was not fussy, they began either the HV-alone or LV-alone condition (counter-balanced across infants). All 16 infants included in the analyses completed fully the 2V condition. However, not enough infants completed a full second block (HV-alone or LV-alone conditions) so we did not analyze the data in the HV-alone and LV-alone conditions.

2.4. EEG recording and processing

EEG data were recorded at a sampling rate of 1000 Hz from 124-channel HydroCel GSN nets (Electrical Geodesics, Eugene, OR) referenced to Cz during recording. The impedances of all electrodes were below 50 K during the recording in accordance with Electrical Geodesics guidelines. EEG data were bandpass (zero-phase) filtered off-line between 1.6 Hz and 20 Hz (roll-off 12 dB/oct) using Eprime software in order to remove slow wave activity. The sampling rate was reduced to 200 Hz in order to run the Artifact Blocking algorithm in Matlab (AB, artifact removal technique, Mourad, Reilly, De Bruin, Hasey, & MacCrimmon, 2007).

Table 1

<table>
<thead>
<tr>
<th>Infant</th>
<th>Age at testing (Weeks)</th>
<th>Music exposure (Hours/Week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>14.9</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>16.1</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>15.4</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>15.3</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>15.9</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>15.4</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>16.1</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>15.3</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>13.3</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>16.7</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>14.9</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>14.3</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>14.1</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>15.1</td>
</tr>
</tbody>
</table>

Mean 15.24 12.75
SD 0.85 7.10

Fig. 1. The stimulus sequence illustrated in musical notation. Each of four deviant stimulus types (up or down shift of one semitone (1/12 octave)) in either the lower or higher of two simultaneous piano tones (fundamental frequencies of 196.0 and 466.2 Hz when not shifted) occurred on 12.5% of trials.
Table 2  
Latency (ms) of the peak MMN amplitude in the grand average waveforms at FR in both groups (3 months and 7 months). A plus and minus 25 ms window was deduced around each latency peak of the grand average to obtain amplitude values for the MMN in each condition for each subject.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Peak MMN latency (ms)</th>
<th>Time-window</th>
<th>Peak MMN amplitude (µV)</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2V-High Down_3m</td>
<td>230</td>
<td>205 255</td>
<td>0.68</td>
<td>16</td>
</tr>
<tr>
<td>2V-Low Down_3m</td>
<td>240</td>
<td>215 265</td>
<td>0.52</td>
<td>16</td>
</tr>
<tr>
<td>2V-High Up_3m</td>
<td>230</td>
<td>205 255</td>
<td>0.58</td>
<td>16</td>
</tr>
<tr>
<td>2V-Low Up_3m</td>
<td>240</td>
<td>230 280</td>
<td>0.39</td>
<td>16</td>
</tr>
<tr>
<td>2V-High Down_7m</td>
<td>210</td>
<td>185 235</td>
<td>2.13</td>
<td>16</td>
</tr>
<tr>
<td>2V-Low Down_7m</td>
<td>215</td>
<td>190 240</td>
<td>1.39</td>
<td>16</td>
</tr>
<tr>
<td>2V-High Up_7m</td>
<td>210</td>
<td>185 235</td>
<td>1.46</td>
<td>16</td>
</tr>
<tr>
<td>2V-Low Up_7m</td>
<td>215</td>
<td>190 240</td>
<td>0.96</td>
<td>16</td>
</tr>
</tbody>
</table>

3. Results
3.1. 3-Month-old infants
3.1.1. Amplitude

Waveforms on standard trials and on high voice and low voice deviant trials (collapsed across up/down deviants) are shown in Fig. 3. T-tests on the average amplitude in the 50 ms time window centered on the MMN peak in the grand average difference wave revealed that MMN amplitude was significantly different from zero for all conditions at frontal, central and occipital regions but not at parietal and temporal regions. Given the sign, we conducted a four-way ANOVA with the factor Voice (high, low), Deviance direction (up, down), Hemisphere (Left, Right), and Region (frontal: FL, FR, central: CL, CR, occipital: OL, OR) as within-subject factors. The ANOVA revealed a main effect of Voice [F(1,15) = 4.29, p < 0.05] indicating larger MMN in the High than in the Low voice. No other main effects or interactions reached sign cance. Fig. 4 shows the waveforms separately by high/low voice and up/down conditions and Fig. 5 compares high/low voice conditions collapsed across up/down.

3.1.2. Latency

MMN latency was tested at the FR region using a two-way ANOVA with Voice (High, Low) and Deviance direction (up, down) as within-subjects factors. Results revealed a signcant main effect of Voice [F(1,15) = 25.87, p < 0.001] showing shorter MMN latency for deviant stimuli in the High voice (231 ms, SE 4.9) than in the Low voice (247 ms, SE 5.9; see Fig. 5). The main effect of Deviance direction was signcant [F(1,15) = 6.31, p = 0.02] showing shorter MMN latency for deviants going Downward (234 ms, SE 6.1) than Upward (243 ms, SE 5.1). The interaction between Voice and Deviance direction was also signcant [F(1,15) = 19.02, p = 0.001] showing that MMN latency in response to downward deviant stimuli (236 ms, SE 4.0) was signcanly faster than in response to upward deviants in the low voice only (258 ms, SE 5.6, difference 22 ms, p = 0.001; Fig. 4B). For the high voice the effect was not signcant (difference 3 ms; p = 0.7; Fig. 4A).
3.1.3. Correlations

We tested for a correlation between the number of hours per week of music listening at 3 month of age and the amplitude of the MMN. In contrast with 7-month-olds (Marie & Trainor, 2013), this correlation was not significant. However, as for 7-month-old infants, the correlation between musical exposure and the high
3.2. Comparison between 3- and 7-month-old infants

3.2.1. Amplitude

In order to investigate the development of the MMN during infancy, we compared the data collected at 3 month of age with that collected at 7 month from Marie and Trainor (2013). A two-way ANOVA was conducted with Group (3-month-olds, 7-month-olds) as a between-subject factor and Voice (high, low), Deviance direction (up, down), Hemisphere (Left, Right), and Region (frontal: FL&FR; central: CL&CR; occipital: OL&OR) as within-subject factors.

The ANOVA revealed a significant main effect of Group [F(1,30) 4.68, p = 0.04] showing that MMN amplitude was overall larger in 7-month-olds (0.26 mV, SE 0.057; see Fig. 6). The main effect of Voice was significant [F(1,30) 18.37, p < 0.001] showing larger MMN in the high voice (0.44 mV, SE 0.14) than in the low voice (0.26 mV, SE 0.17). Moreover, the interaction between Voice and Region was significant [F(2,60) 53.8, p < 0.001]. Post-hoc Tukey tests of this interaction revealed that across the three brain regions examined, the high voice superiority effect (i.e., MMN high-voice minus MMN low-voice) was larger in both groups over Central (p = 0.004, high-voice low-voice = 0.49 V) than Frontal (p = 0.01, high-voice low-voice = 0.45 V) and Occipital regions (p = 0.03, high-voice low-voice = 0.41 V, see Fig. 6B). No other main effects or interactions were significant. Importantly, the interaction between Voice and Group was not significant (p = 0.12), indicating that the high voice superiority effect was of similar size at both ages (Fig. 7).

3.2.2. Latency

MMN latency was compared between groups at the FR region using three-way ANOVAs with Group (3-month-olds, 7-month-olds) as a between-subject factor and Voice (High, Low) and Deviance Direction (up, down) as within-subjects factors. Results revealed a significant main effect of Group [F(1,30) 39.03, p < 0.001] showing shorter MMN latency in 7-month-olds (205.5 ms, SE 3.3) than in 3-month-old infants (239 ms, SE 3.3; d = 2.95 ms; see Fig. 6). The main effect of Voice was also significant [F(1,30) 34.46, p = 0.001] revealing shorter MMN latency for deviant stimuli in the high-voice (218 ms, SE 3.2) than in the low-voice (231 ms, SE 4.1; d = 1.9 ms; see Fig. 6). Finally, there was a significant interaction involving Group by Voice by Direction [F(1,30) 9.68, p = 0.004]. A post-hoc Tukey test revealed no difference between upward and downward pitch deviant stimuli for both voices in 7-month-old infants (high-voice p = 1; low-voice p = 0.99); however, as indicated above, in 3-month-old infants the MMN latency in response to downward deviants was significantly shorter than in response to upward deviants in the low voice only (see Fig. 4A). No other main effects or interactions were significant. Importantly, there was no Group by Voice interaction (p = 0.18), consistent with a similar high voice superiority effect at both ages (see Fig. 7).

4. Discussion

We investigated the development of cortical representations for simultaneous sound streams and found that the high voice superiority effect is present in 3-month-old infants, before there is substantial in-utero experience with the native music system on perceptual representations. Specifically, at 3-months of age, the evidence indicates that perceptual narrowing for musical tonality and metrical structure has not yet begun (Hannon & Trainor, 2007; Trainor & Corrigall, 2010; Trainor & Hannon, 2013). Indeed our data are consistent with this evidence in that the size of MMN responses in general in the 7-month-olds reported in Marie and Trainor (2013), who were tested with the identical stimuli as in the present paper, was correlated with the amount of musical listening they experienced per week as reported by parents, whereas the MMN responses in the 3-month-olds reported here were not correlated with musical experience.

With respect to cortical representations for simultaneous sounds, we found that an adult like MMN response was present at 3 months of age for deviant stimuli in each of the two simultaneous voices presented (see Figs. 4 and 5), indicating that infants can process polyphonic sounds and that, more specifically, they can store each voice in a separate memory trace. At the same time, our findings indicate that this ability improves between 3 and 7 months of age, as direct comparison to the data from the 7-month-old infants reported in Marie and Trainor (2013) indicated that the MMN increased in amplitude and decreased in latency over this time period (see Figs. 6 and 7). Decreases in latency likely reflect increased neural synchrony and increased amplitude both increased neural synchrony and increases in the
number of neurons responding. These findings are consistent with previous work on pitch perception in single streams indicating that an adult-like frontally negative MMN emerges between 2 and 4 months of age, and that it increases in amplitude and decreases in latency with increasing age (He et al., 2007; He, Hotson, & Trainor, 2009a,b). Our results extend these findings by showing that pitch change detection for polyphonic contexts appears to develop over the same time period as for single stream contexts.
One interesting difference between 3-month-olds and 7-month-olds was that the younger age group showed larger and earlier MMN for downward than for upward deviant stimuli in the lower voice, and no difference in the higher voice, whereas the older age group showed no effect of deviant direction in either voice (see Fig. 4). Previous studies have shown that newborns are able to discriminate downward from upward pitch contours (Carral et al., 2005) and that this ability improves over the 1st year after birth. Given that no difference between upward and downward deviant stimuli in either voice was found in older infants or adults (Marie & Trainor, 2013; Fujikawa et al., 2005, 2008) and knowing that the larger and the shorter the MMN, the easier the detection process (N. ten, Pakarinen, Rinne, & Takegata, 2004), this result suggests that not only does the lower voice generate a less robust representation, but also that the processing of upward deviant stimuli in the lower of two voices has a more protracted developmental trajectory in comparison to the processing of downward deviants.

With respect to the high voice superiority effect, we found that already at 3 months of age, MMN to deviant stimuli in the high voice was larger and earlier than MMN to deviants in the lower voice (see Fig. 5), similar to the case for 7-month-olds and adults (Fujikawa et al., 2008; Marie et al., 2012; Marie & Trainor, 2013). Indeed, analyses directly comparing the size of the high voice superiority effect revealed no signi cant difference between 3 and 7 months of age (see Fig. 7), despite increases in the size of the MMN itself across these ages. Furthermore, as with the 7-month-old data there was no signi cant correlation in the 3-month-old data between the size of the high voice superiority effect and musical experience as measured by the amount of musical exposure per week reported by parents, suggesting that music listening experience does not affect the strength of the encoding of polyphonic sounds. The lack of correlation between music experience and the high voice superiority effect is in contrast to the signi cant correlation between musical experience and the size of MMN responses in general in the 7-month-old data. Together, these data are consistent with the idea that music enculturation does not appear to affect the high voice superiority effect during infancy, and that the high voice superiority effect re eks innate properties of the auditory system.

The high voice superiority effect is important in that it affects how music is composed and how it is perceived, as well as how well speech will be understood in different noise environments. We are exploring the hypothesis that the high voice superiority effect has an innate origin using a model of the auditory periphery (Trainor, Marie, Bruce, & Bidelman, 2014). Speci cally, this model re eks middle-ear t ering and cochlear dynamics (Zilany et al., 2009; Ibrahim & Bruce, 2010). Simulations in which we presented stimuli similar to those of the present paper, but across a wide range of interval spacings and pitch registers, to the model indicate that the high voice superiority effect can be seen in the pitch salience of sound representations at the level of the auditory nerve. A peripheral origin for the high voice superiority effect would explain the data of the present experiment, speci cally, why the high voice superiority effect is present in young infants before enculturation to the musical system in their environment and why the high voice superiority effect does not change between 3 and 7 months of age despite large developmental changes in the MMN response itself during this time period.

5. Conclusion

The results from the present study provide evidence for two main conclusions. First, as early as 3-months of age, infants are able to process streams of simultaneous pitch-differentiated tones in separate memory traces. This result, taken in conjunction with previous research detailing the ability to perceive monophonic pitch variations of one semitone (He et al., 2007, 2009a,b) suggests that the ability to process polyphonic sounds develops in conjunction with the ability to process simple sounds. Second, by 3-month of age infants already show the high voice superiority effect previously observed in older infants and adults (Fujikawa et al., 2008; Marie & Trainor, 2013) and this effect is not correlated with musical experience in the infant groups. These results support our modeling work that indicates an innate peripheral origin to the high voice superiority effect (Trainor et al., 2014).

Funding

This research was supported by a grant to Laurel J. Trainor from the Canadian Institutes of Health Research. Cline Marie was supported by a postdoctoral fellowship from the National Science and Engineering Research Council of Canada Create Grant in Auditory Cognitive Neuroscience.

Acknowledgments

We thank Harrison Cruikshank and Elaine Whiskin for their help with infant testing.

References


