Development of Simultaneous Pitch Encoding: Infants Show a High Voice Superiority Effect

Céline Marie[1,2] and Laurel J. Trainor[1,2]

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario L8S 4K1, Canada and [2]McMaster Institute for Music and the Mind, Department of Psychology, Neuroscience & Behaviour, Hamilton, Ontario L8S 4K1, Canada

Address correspondence to Laurel J. Trainor, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada. Email: ljt@mcmaster.ca.

Infants must learn to make sense of real-world auditory environments containing simultaneous and overlapping sounds. In adults, event-related potential studies have demonstrated the existence of separate preattentive memory traces for concurrent note sequences and revealed perceptual dominance for encoding of the voice with higher fundamental frequency of 2 simultaneous tones or melodies. Here, we presented 2 simultaneous streams of notes (15 semitones apart) to 7-month-old infants. On 50% of trials, either the higher or the lower note was modified by one semitone, up or down, leaving 50% standard trials. Infants showed mismatch negativity (MMN) to changes in both voices, indicating separate memory traces for each voice. Furthermore, MMN was earlier and larger for the higher voice as in adults. When in the context of a second voice, representation of the lower voice was decreased and that of the higher voice increased compared with when each voice was presented alone. Additionally, correlations between MMN amplitude and amount of weekly music listening suggest that experience affects the development of auditory memory. In sum, the ability to process simultaneous pitches and the dominance of the highest voice emerge early during infancy and are likely important for the perceptual organization of sound in realistic environments.

Keywords: auditory development, auditory scene analysis (ASA), infant, mismatch negativity (MMN), music, polyphonic sound

Introduction

Natural auditory environments contain multiple overlapping sounds, such as those made by human voices, musical instruments, animals, wind, water, appliances, cars, doors slamming, and so on. The complex sound wave that reaches the ear is made up of a mixture of these sounds. Typically, each sound source contains many frequency components that vary over time, and the compositions of different sounds overlap. In order to determine what auditory objects are present, the auditory system must perform a spectrotemporal analysis of the incoming sound wave in order to determine which components belong together (e.g., the harmonics of a single sound source, such as a musical instrument or a voice, or the successive sounds of an instrument or voice) and which groups of components belong to separate objects (e.g., 2 different instruments or 2 different talkers). These processes are known as auditory stream integration and auditory stream segregation, respectively, and together, they constitute auditory scene analysis (Bregman 1990). In polyphonic music, in which more than one stream (streams are referred to as "voices") is present at the same time, both integration and segregation work together. For instance, in Western tonal music, different

streams or voices typically fit together harmonically, and at the same time, they can also be perceived as separate voices (Huron 2001).

Behavioral studies indicate that a number of sound features are involved in auditory scene analysis. For example, sequential sounds tend to be perceptually grouped together if they are similar in pitch and timbre (e.g., Singh 1987; Bregman et al. 1990; Iverson 1995; Cusack and Roberts 2000). Pitch and timbre interact with temporal factors such that the faster the tempo, the more likely successive nonidentical sounds are to be perceived as belonging to separate streams or voices (for reviews, see Bregman 1990; Darwin and Carlyon 1995). Presumably, this phenomenon occurs because in the real world, most sounding objects do not change dramatically in pitch or timbre over a short time interval. These sound features are also involved in the perception of simultaneous sounds. For example, the auditory system tends to group harmonics together that are at integer multiples of a fundamental such that a single complex sound is perceived, whereas components without such frequency relations are more likely to be segregated into different streams or voices. When one harmonic in a complex sound is mistuned (i.e., not at an integer multiple of the fundamental) or its onset is shifted in time relative to other harmonics, this harmonic tends to be heard as a separate auditory object, while the rest of the harmonics integrate into a second auditory object (Hartmann et al. 1990; Alain et al. 2001, 2003; Alain and McDonald 2007; Folland et al. 2012).

Bregman (1990) proposed that much of auditory scene analysis occurs automatically and preattentively, and this is corroborated by recent brainstem and several event-related potential (ERP) studies (Winkler et al. 1992; Ritter et al. 1995; Sussman et al. 1999; Shinozaki et al. 2000; Yabe et al. 2001; Brattico et al. 2002; Nager et al. 2003; Lee et al. 2009). Fujioka et al. (2005) examined preattentive processing of simultaneous musical melodies by examining the mismatch negativity (MMN) component of the ERP recorded with magnetoencephalography (MEG). In particular, they asked whether preattentive evidence of 2 different memory traces in auditory cortex could be found for 2 simultaneously presented melodies. MMN occurs in response to occasional unexpected changes (deviants) in an ongoing stream of sounds (for reviews, see Picton et al. 2000; Näätänen et al. 2007). MMN is generated in auditory cortex and manifests at the surface of the head as a frontal negativity peaking between 120 and 250 ms after onset of a deviant, accompanied by a polarity reversal at the back of the head. Although MMN can be modified by attention, in adults, it is present whether or not subjects are attending to the stimulus (e.g., Näätänen and Michie 1979; Alho et al. 1989). MMN reflects an auditory cortex response to any violation of regularity

in the auditory scene based on a memory representation of the auditory input from the past few seconds (e.g., Picton et al. 2000; Ulanovsky et al. 2004; Näätänen et al. 2007; Winkler et al. 2009). MMN appears to reflect the updating of a memory trace as it increases in amplitude the lower the probability of a deviant. Another relevant property is that when standard and deviant sounds are each presented 50% of the time, no MMN is seen. Previous studies have shown that when several deviants are presented that affect different sound features such as frequency, duration, and timbre, MMN can still be elicited with a 50% overall deviance rate (with adults: Näätänen et al. 2004; Pakarinen et al. 2007; and with newborns: Sambeth et al. 2009). Fujioka et al. (2005) tested whether separate memory traces are formed for simultaneous melodies by embedding different deviants to the "same" feature, frequency, in either the higher or the lower of 2 simultaneous melodies. On each trial, they presented 2 simultaneous 5-note melodies that fit together harmonically, composed from the first 5 notes of the Western major scale (C-D-F-E-G and G-F-D-C-E in the key of C major). Which melody was in the high voice (C5–G5) and which was in the low voice (C4–G4) varied across conditions. A final deviant note was presented on 50% of trials. Although the overall deviance rate was 50%, 25% of trials had a pitch deviant final note in the upper melody voice and 25% in the lower melody voice. The logic was that if each melody had a separate memory representation, MMN would be expected, as the deviance rate would be 25% for each voice. However, if there is only a unified representation, a small or no MMN would be expected, as the pitch deviance rate would be 50%. As they found significant MMN with an overall deviance rate of 50%, they reasoned that separate memory traces must have been formed for the upper and lower voices. This result was replicated with more simple stimuli in Fujioka et al. (2008). In this study, each voice consisted of a repeating single note (high and low voices were separated by the interval of a minor 10th, or 15 semitones, frequency ratio of 2.3) rather than a melody. Otherwise, the study was similar, with the high and low tones presented simultaneously and with 25% of trials containing a pitch deviant note in the upper voice and 25% containing a pitch deviant in the lower voice. Again, they found significant MMN elicited by deviants in each voice, providing further evidence for 2 separate memory traces in auditory cortex.

One interesting finding in Fujioka et al. (2005) was that musicians showed much larger MMN responses to deviants in simultaneous melodies compared with nonmusicians. This finding corroborates a number of other studies indicating superior auditory processing in musicians for single tones (e.g., Koelsch et al. 1999; Tervaniemi et al. 2001; Marie et al. 2012) and single melodies (Fujioka et al. 2006), suggesting that specific experience and training may improve auditory scene analysis as well as general musical processing. A second interesting finding was that consistently in musicians and nonmusicians, larger MMN was found for deviants in the higher compared with in the lower voice, whether the voices were melodies (Fujioka et al. 2005) or single tones (Fujioka et al. 2008). This finding is consistent with behavioral studies showing superior processing of the highest of several voices even in school-aged children (Zenatti 1969). As well, a recent study of brainstem responses indicates that musicians show stronger encoding of the harmonics of the upper than the lower tone of a musical interval (Lee et al. 2009). Note that this effect is consistent with the widespread compositional practice of putting the melody into the highest voice in Western music.

However, in tasks involving focused attention, experienced listeners are better able to discriminate deviants in the lower voices in polyphonic contexts compared with less experienced listeners, which suggests that these processes are somewhat open to learning (Palmer and Holleran 1994; Crawley et al. 2002).

In the present study, we address 2 main questions. First, if auditory scene analysis is automatic, preattentive, and emerges early in development, MMN showing separate memory traces for 2 simultaneous tones should be seen in infancy. Second, if there is an innate tendency for superior processing of the higher of 2 simultaneous note sequences, this effect should also be seen in infancy. Its absence in infancy would suggest that it is learned.

Frequency resolution in the cochlea is relatively mature at birth (Teas et al. 1982; Abdala and Chatterjee 2003), and at 6 months of age, high-frequency discrimination (above 4000 Hz) appears to be mature. Using the head turn procedure, researchers have shown that frequency discrimination improves greatly between 5 and 9 months of age for frequencies below 2000 Hz. At 6 months of age, infants respond to 2% changes in 500 Hz sine tones under conditions where adults respond to changes of 1% (Olsho et al. 1982; Sinnott and Aslin 1985; Werner 2002). Frequency discrimination thresholds continue to mature, especially for low frequencies (under 1000 Hz) reliant on temporal processing, until 10 or 11 years of age (Maxon and Hochberg 1982; Werner 2007). Nonetheless, pitch discrimination in infancy is good enough for musical perception and auditory scene analysis (for reviews, see Trainor and Corrigall 2010; Trainor and Unrau 2012). Several studies indicate that principles of auditory scene analysis are operative in young infants for sequential sounds (Demany 1982; Fassbender 1993; McAdams and Bertoncini 1997; Winkler et al. 2003; Smith and Trainor 2011). For instance, Fassbender (1993) found that infants could not discriminate a sequence of rising tones from its retrograde inversion (forming a falling sequence) if random tones in the same pitch range were interleaved between the tones of the sequence. However, the infants were able to do this task when the random interleaved tones were in a different frequency range and therefore segregated perceptually into a different auditory stream. As far as the development of auditory scene analysis for simultaneous tones, the integration of harmonics into a single pitch percept may not have a cortical representation until after 3 months of age as evidenced by ERP studies on the pitch of the missing fundamental (He and Trainor 2009). Moreover, one study suggests that at 6 months, infants can use harmonic structure to segregate simultaneous auditory objects, noticing when one harmonic of a complex tone is mistuned (Folland et al. 2012). Under these circumstances, adults perceive 2 auditory objects, the mistuned harmonic and a complex tone composed of the remaining in-tune harmonics. Despite young infants' abilities with respect to auditory scene analysis, it should be noted that they are not yet enculturated to the pitch structure of the music in their environment (Trainor and Trehub 1992; Trainor 2005; Hannon and Trainor 2007; Trainor and Corrigall 2010; Trainor and Unrau 2012). Specifically, Western infants do not yet expect musical melodies to contain only the notes of a Western scale, and they perform equally well at detecting changes that remain within or go outside the key of a melody, whereas adults and 5 years old are better able to detect out-of-key changes (Trainor and Trehub 1994; Corrigall and Trainor 2009).

Here, we use MMN to test directly whether 7-month-old infants have separate memory traces for 2 simultaneous tones and whether the memory trace is more robust for the higher than for the lower tone. This is an appropriate measure with infants as a number of studies indicate that MMN is robust to pitch changes at this age (e.g., Kushnerenko et al. 2002; He et al. 2007, 2009a, 2009b; Tew et al. 2009). In the present study, we repeatedly present the 2 simultaneous tones of Fujioka et al. (2008) such that on 25% of trials, there is either an upward or a downward change of one semitone (1/12 octave) to the higher tone, and on 25% of trials, there is a similar change to the lower tone. If infants can form 2 memory traces for 2 simultaneous tones, we expect to find MMN in response to both changes. If encoding of the higher tone is more robust, we expect larger MMN for changes to the higher than to the lower tone. In addition, we compare performance for each tone in isolation with performance for that tone in the context of the other tone.

Materials and Methods

Participants
Twenty 7-month-old infants were tested. Three were excluded due to excessive movement or fussiness during the recording and one for having artifacts due to pacifier use, leaving 16 infants in the final sample (8 males; mean age = 234.8 days, range = 219–244 days). 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 = 9 h/week, range = 2–20 h/week). Parents of 6 infants had played an instrument before having children but they reported having stopped playing by the time of testing. Finally, 6 families were bilingual (English and French or Italian), and the other 10 families spoke only English.

Stimuli
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 = 600 ms) at approximately 60 dB(A) measured at the location of the infant's head.

Each condition was 11 min long, containing 1088 individual notes presented in pseudorandom order, with the constraint that a deviant could not be followed immediately by an identical deviant. Figure 1 shows the 3 conditions of the experiment: Two-Voice (2V), High-Voice-alone (HV-alone), and Low-Voice-alone (LV-alone). Following Fujioka et al. (2008), in the 2V condition, the standard tones had fundamental frequencies of 466.2 Hz (B-flat$_4$, international standard notation) and 196.0 Hz (G$_3$), which are 15 semitones apart (frequency ratio = 2.3) and form a minor 10th interval (octave displaced minor third). Deviants were created by a one-semitone (1/12 octave) pitch deviation, going up or down from each tone of the dyad (i.e., B$_4$ and A$_4$ for the High voice deviants, G#$_3$ and F#$_3$ for the Low voice deviants). The HV-alone condition was identical to the 2V condition except that the lower tones were omitted. Similarly, the LV-alone condition was identical to the 2V condition except that the higher tones were omitted.

Procedure
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 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. Each of the 3 experimental conditions consisted of 1088 trials and lasted 11 min. In the 2V condition, 50% (or 544) of trials were standards and 50% (or 544) were deviants, 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 deviants, with 12.5% (or 136) of each deviant type (up and down). All infants were run on the 2V condition first. If the infant completed the 2V condition and was not fussy, they began either the HV-alone or the LV-alone condition (counterbalanced across infants). If they completed the second condition and were not fussy, they were then run in the remaining condition. All 16 infants included in the analyses completed 2 conditions, but only 4 completed all 3 conditions. Some analyses were conducted with all 16 infants. For other analyses, 2 subgroups were formed. Infants in Group 1 had completed the 2V and HV-alone conditions (n = 10). Infants in Group 2 had completed the 2V and LV-alone conditions (n = 10).

EEG Recording and Processing
Electroencephalography (EEG) data were recorded at a sampling rate of 1000 Hz from 124-channel HydroCel GSN nets (Electrical Geodesics, Eugene, OR) referenced to Cz. The impedances of all electrodes were below 50 KΩ during the recording in accordance with Electrical
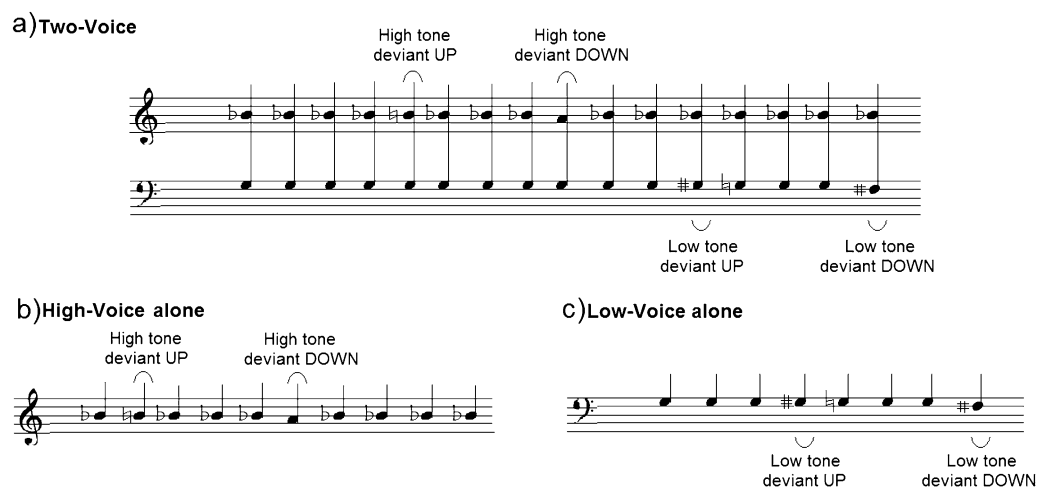
Figure 1. Description of the stimulus sequences illustrated in musical notation. (a) Two-Voice condition, (b) High-Voice-alone condition, and (c) Low-Voice-alone condition.

Geodesics' guidelines (note that the amplifiers have an input impedance of about 200 MΩ). After recording, EEG data were band-pass filtered between 1.6 and 20 Hz (roll-off = 12 dB/oct) using EEprobe software in order to remove slow wave activity. The sampling rate was modified to 200 Hz in order to run the Artifact Blocking program with Matlab so as to remove artifacts from muscle activity, such as eye blinks and movements (AB; artifact removal technique, Mourad et al. 2007; Fujioka et al. 2011). Recordings were re-referenced off-line using an average reference, including all electrodes, and then segmented into 600 ms epochs (−100 to 500 ms relative to note onset).

ERP Data Analysis

Standards and deviants were averaged and difference waveforms were computed for each condition and participant by subtracting ERPs elicited by the standards from those elicited by each deviant. In order to quantify MMN amplitude, the grand average difference waveform was computed for each electrode for each deviant type (2V-High-tone up, 2V-High-tone down, 2V-Low-tone up, 2V-Low-tone down, HV-alone up, HV-alone down, LV-alone up, LV-alone down). Subsequently, for statistical analysis, 88 electrodes were selected and divided into 5 groups for each hemisphere (Left and Right) representing frontal, central, parietal, occipital, and temporal regions (FL, FR, CL, CR, PL, PR, OR, OL, TL, and TR; see Fig. 2). Thirty-six electrodes were excluded from the groupings due to the following considerations: electrodes on the forehead near the eyes in order to further reduce the contamination of eye movement artifacts; electrodes at the edge of the Geodesic net to reduce contamination of face and neck muscle movement; and electrodes in the midline to enable comparison of the EEG response across hemispheres.

Difference waves (deviant–standard) were computed for each deviant type for each condition. Initially, the presence of MMN was tested with *t*-tests to determine where the difference waves were significantly different from zero. As expected, there were no significant effects at parietal sites (PL, PR) so these regions were eliminated from further analysis. To analyze MMN amplitude, first, the most negative peak in the right frontal region (FR) between 150 and 250 ms poststimulus onset was determined from the grand average difference

waves for each condition, and a 50 ms time window was constructed centered at this latency. For each subject and each region, the average amplitude in this 50-ms time window for each condition was used as the measure of MMN amplitude. Finally, for each subject, for each condition, the latency of the MMN was measured as the time of the most negative peak between 150 and 250 ms at the FR region (see Table 1) since visual inspection showed the largest MMN amplitude at this region. Analyses of variance (ANOVAs) were conducted on amplitude and latency data. Greenhouse–Geisser corrections were applied where appropriate and Tukey post hoc tests were conducted to determine the source of significant interactions. Finally, Pearson correlations were used to explore the relation between amount of music listening reported (number of hours per week) and the amplitude of the MMN component in the Two-Voice condition.

Results

Two-Voice (2V) Condition

Amplitude

A four-way ANOVA was conducted with Voice (high, low), Deviance Direction (up, down), Hemisphere (Left, Right), and Region (frontal: FL and FR; central: CL and CR; temporal: TL and TR; and occipital: OL and OR) as within-subject factors and MMN amplitude as the dependent measure. There was an interaction between Voice and Region, $F_{3,45} = 16.31$, $P < 0.001$. Tukey post hoc tests revealed larger responses for the high than for the low voice which were significant at frontal (−1.40 and −0.76 μV, respectively; post hoc $P = 0.02$), central (−1.31 and −0.62 μV, respectively; post hoc $P = 0.007$), and occipital (1.03 and 0.42 μV, respectively; post hoc $P = 0.03$) regions and marginally significant at temporal (1.12 and 0.57 μV, respectively; post hoc $P = 0.059$) regions (see Fig. 3). No other main effects or interactions were significant.

As no effect of Deviance Direction (up or down) was found in the ANOVAs, we collapsed across this variable and tested correlations between the number of hours per week of music listening and the amplitude of the MMN. Significant correlations were found at the left Central region for the High voice ($r = -0.57$; $P = 0.02$) and at the left FR for the Low voice ($r = -0.54$; $P = 0.03$) revealing that the more infants listened to music at home, the larger the MMN in both voices over left frontocentral regions. Note that the correlation between amount of music listened to and the difference between MMN amplitude in response to changes in the higher versus lower voice was not significant, suggesting that listening experience did not affect the superiority of the representation of the higher voice.
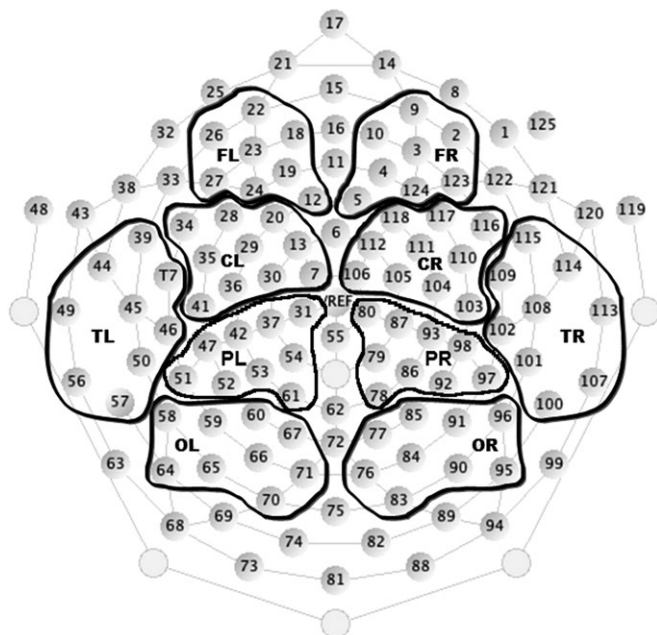

Figure 2. The grouping of electrodes in the Geodesic net (for details, see Materials and Methods). Eighty-four of 124 electrodes were selected to be divided into 5 groups (frontal, central, parietal, occipital, and temporal) for each hemisphere. The waveforms for all channels in each region were averaged together to represent EEG responses from that scalp region. The other 36 channels were excluded from further analysis to avoid artifacts and enable comparisons across hemispheres.

Table 1

Latency (ms) of the peak MMN amplitude in the grand average waveforms at FR

Conditions	Peak MMN latency (ms)	Time window	Peak MMN amplitude (μV)	Number of subjects
2V-High down	210	185–235	−2.13	16
2V-Low down	215	190–240	−1.39	16
2V-High up	210	185–235	−1.46	16
2V-Low up	215	190–240	−0.96	16
HV-alone down	180	155–205	−1.00	10
HV-alone up	185	160–210	−0.40	10
LV-alone down	200	175–225	−2.51	10
LV-alone up	195	170–220	−2.33	10

Note: A plus and minus 25 ms window was defined around each latency peak of the grand average to obtain amplitude values for the MMN in each condition for each subject.
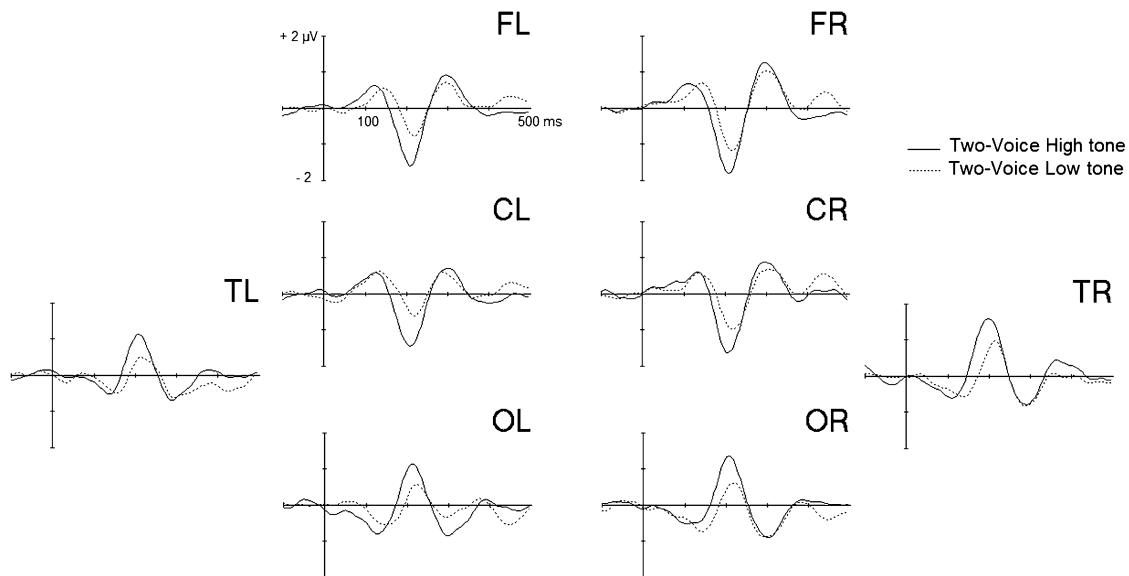
Figure 3. Difference waveforms (deviant–standard) in the Two-Voice condition with all infants ($n = 16$) for the High voice (black line) and the Low voice (dashed line). Deviants up and down were combined for each voice.
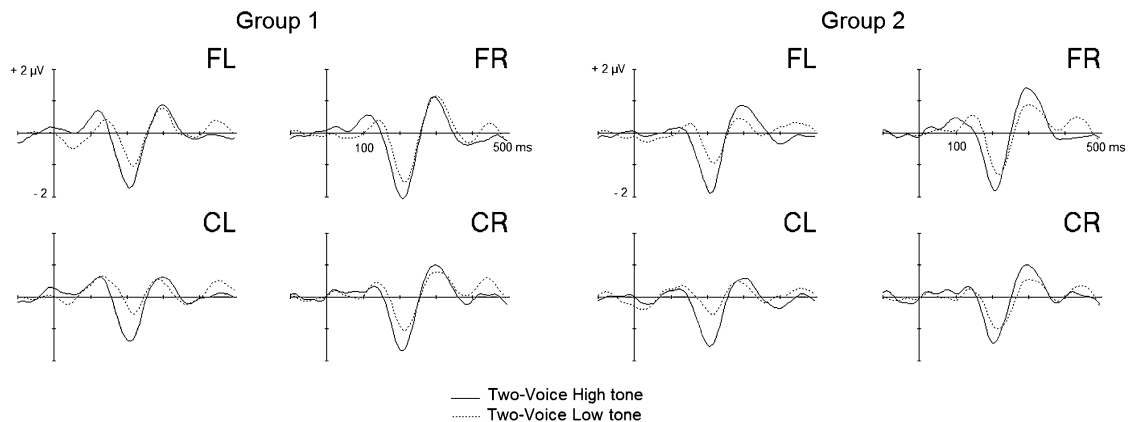
Figure 4. Difference waveforms in the Two-Voice condition for each group: Group 1 (Left side, see text) and Group 2 (Right side, see text) for the High voice (black line) and the Low voice (dashed line). Deviants up and down were combined for each voice.

In order to examine the reliability of the results, the same analysis was run separately on the infants who completed the 2V and HV-alone conditions, Group 1 ($n = 10$), and on the infants who completed the 2V and LV-alone conditions, Group 2 ($n = 10$, see Fig. 4). There was an interaction between Voice and Region for each group (Group 1: $F_{3,27} = 9.24$, $P = 0.002$ and Group 2: $F_{3,27} = 7.78$, $P = 0.009$). When the analysis included only the frontal and the central regions, a main effect of Voice was found in each group indicating larger MMN amplitude to deviants in the High voice than to deviants in the Low voice (Group 1: $F_{1,9} = 9.47$, $P = 0.01$ and Group 2: $F_{1,9} = 7.89$, $P = 0.02$). No other interactions reached significance.

Latency

A two-way ANOVA with Voice and Deviance Direction as within-subject factors and latency at region FR as the dependent variable revealed a main effect of Voice, $F_{1,15} = 10.29$, $P < 0.006$, with significantly shorter MMN latency to deviants in the High voice (204 ms) than to deviants in the Low voice (214 ms, see Fig. 3). No other main effects or interactions were significant. Separate analyses for the 2 groups defined above

revealed a similar latency difference of about 10 ms between MMN to High voice compared with Low voice deviants (see Fig. 4). In Group 1, MMN latency to deviants in the High voice (201 ms) was shorter than in the Low voice (211 ms, main effect of Voice: $F_{1,9} = 4.7$, $P = 0.05$). In Group 2, MMN latency to deviants in the High voice (206 ms) was also shorter than MMN to deviants in the Low voice (216 ms, main effect of Voice: $F_{1,9} = 9.04$, $P = 0.01$). Note that the correlation between amount of music listened to and MMN latency was not significant, suggesting that listening experience did not affect the speed of processing of pitch deviants in polyphonic contexts.

Comparison of Two-Voice (2V) and One-Voice (HV-Alone, LV-Alone) Conditions

Amplitude

To compare MMN in the high voice when it was presented alone and when it was presented in the context of a lower voice, MMN amplitude for Group 1 infants was compared for high voice deviants in the 2V and HV-alone conditions (see Fig. 5). A four-way ANOVA was conducted with Number of Voices
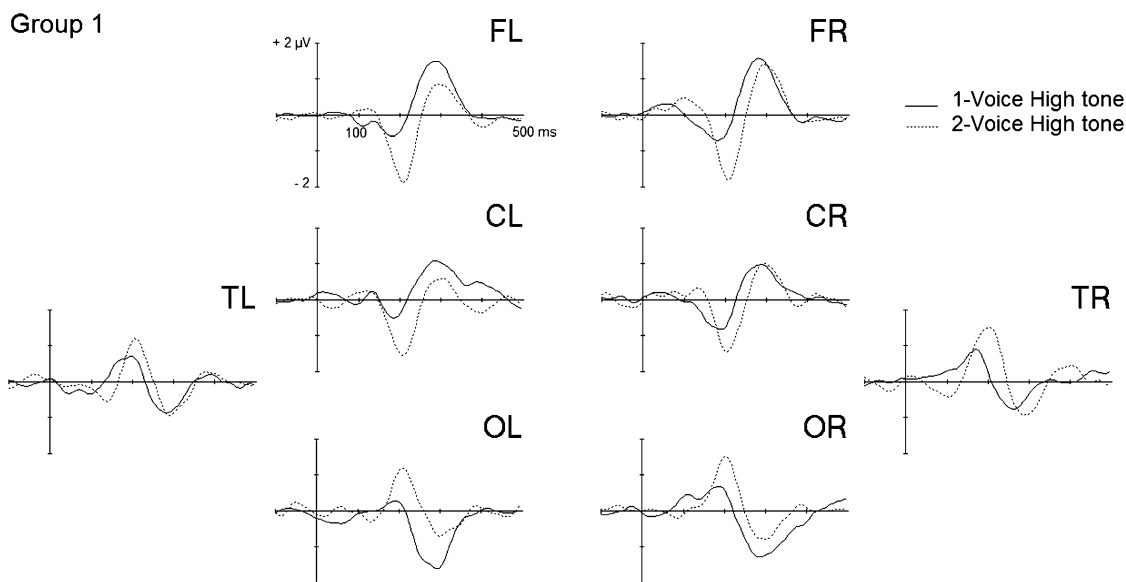
Group 1

Figure 5. Difference waveforms for Group 1 ($n = 10$) in the High-Voice-alone (black line) and Two-Voice High tone conditions (dashed line).
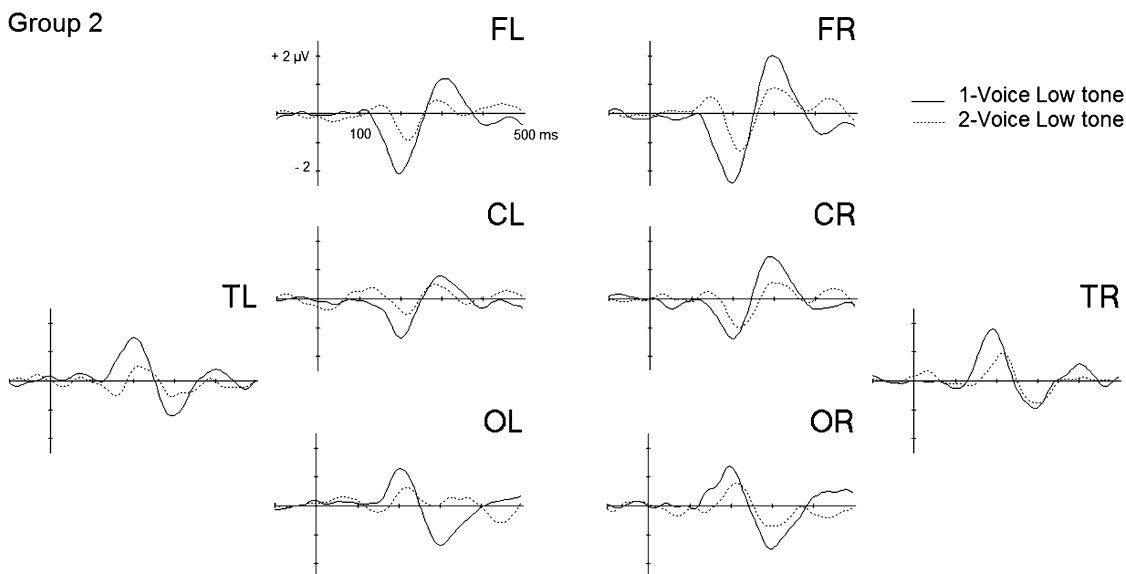
Group 2

Figure 6. Difference waveforms for Group 2 ($n = 10$) in the Low-Voice-alone (black line) and Two-Voice Low tone conditions (dashed line).

(one, two), Deviance Direction (up, down), Hemisphere (Left, Right), and Region (frontal: FL and FR; central: CL and CR; temporal: TL and TR; occipital: OL and OR) as within-subject factors and MMN amplitude as the dependent measure. There was a main effect of Number of Voices, $F_{1,9} = 4.91$; $P = 0.05$, with MMN amplitude larger when the high voice was presented in a two-voice context (-0.12 μV) than in a one-voice context (-0.02 μV). There was also an interaction between Number of Voices and Region, $F_{3,27} = 5.42$; $P = 0.02$, reflecting that this effect was larger at the FRs. No other effects were significant.

To compare MMN in the lower voice when it was presented alone and when it was presented in the context of a higher voice, a similar ANOVA was conducted on the data from infants in Group 2 (see Fig. 6). As with the high voice, there was a main effect of Number of Voices, $F_{1,9} = 8.30$; $P = 0.02$, but in contrast

to the High voice condition, the MMN amplitude was smaller when the low tone was presented in a two-voice context (-0.08 μV) than in a one-voice context (-0.2 μV). There was also an interaction between Number of Voices and Region, $F_{3,27} = 5.56$; $P = 0.02$, reflecting that this effect was larger at the FRs. No other effects were significant.

Latency

MMN latency was compared for each voice (separate ANOVAs for the High voice and Low voice) alone and in the context of the other voice for the FR region using two-way ANOVAs with Number of Voices (one, two) and Deviance Direction (up, down) as within-subjects factors. In both cases, MMN latency was significantly longer when in the context of the other voice than when alone. Specifically, for the high voice (Group 1, see Fig. 5), MMN was earlier when the high tone was presented

alone (HV-alone, 184 ms) than when presented 1in the two-voice (2V) context (201 ms; main effect of Number of Voices, $F_{1,9} = 9.99$, $P = 0.01$). For the low voice (Group 2, see Fig. 6), MMN was also earlier when the low voice was presented alone (LV-alone, 193 ms) than in the two-voice (2V) context (216 ms; main effect of Number of Voices, $F_{1,9} = 14.80$, $P = 0.004$). No other main effects or interactions were significant.

Discussion

Infants must learn to make sense of auditory environments that contain multiple sound sources that overlap in time. In music, forming separate representations for simultaneous notes may be particularly difficult as 2 notes may have the same onset and offset timing, eliminating temporal cues, and they may also be harmonically related and thus share common harmonics or subharmonics. In the present paper, we demonstrated that infants can hold separate traces for 2 simultaneously presented complex tones in auditory working memory at the same time. Specifically, when 2 simultaneous streams of notes were presented, MMN was elicited by separate deviants in both the high and the low streams, even though the overall deviance rate was 50%. As no MMN response is expected when the deviance rate is 50%, we interpreted the emergence of MMN for frequency deviants in both the high and the low voices as indicating that separate memory traces were formed for each tone of the dyad. The MMN was morphologically similar to that found with simultaneous tones and melodies in adults using MEG (Fujioka et al. 2005, 2008), indicating that the finding is robust across age and measuring technique. This result adds to the literature indicating that MMN is readily elicited by small frequency changes during infancy (e.g., Kushnerenko et al. 2002; Carral et al. 2005; He et al. 2007, 2009a, 2009b; Trainor 2008; Tew et al. 2009; Trainor et al. 2011) and extends it to polyphonic contexts. Importantly, these results add to the small literature demonstrating auditory scene analysis in infants, a literature that before the present study has focused almost exclusively on sequential rather than simultaneous sounds (e.g., Demany 1982; Fassbender 1993; McAdams and Bertoncini 1997; Winkler et al. 2003; Smith and Trainor 2011). This result is consistent with one behavioral study indicating that 6-month-old infants can perceive a mistuned harmonic in a complex tone that, when not mistuned, integrates with the other harmonics into the percept of a single sound (Folland et al. 2012). Finally, it adds to a previous study showing that by 4 months, infants can integrate harmonics into a single sound percept and can perceive the pitch of the missing fundamental (He and Trainor 2009). The present study extends these findings by showing that when presented with 2 simultaneous complex tones, infants can segregate the harmonics belonging to one tone into one percept and those belonging to the other tone into a second percept and hold and process both percepts in auditory working memory at the same time with some degree of independence.

The second major novel finding of the present paper is that at 7 months, infants are already like adults (Fujioka et al. 2005, 2008) in having a more robust memory trace for the higher of 2 simultaneously presented tones. Specifically, MMN was larger and earlier to deviants in the higher compared with deviants in the lower voice when the voices were presented simultaneously. Furthermore, when MMN in response to deviants in each voice was compared between conditions where both voices were presented simultaneously or each voice was

presented alone, the presence of the second voice reduced the MMN amplitude of the lower voice but increased the MMN amplitude of the higher voice. With adults, Fujioka et al. (2008) also found decreased MMN amplitude for the lower voice when in the context of the higher voice than when alone, although they did not find any differences for the higher voice. In any case, these findings suggest that although 2 simultaneous memory traces are formed, they are not entirely independent. This interaction between voices is also reflected in longer MMN latencies in the case of 2 voices compared with the case of 1 voice.

The interval we used between the voices of 15 semitones is larger than what is typical in musical contexts, at least in the pitch range we used. It would presumably be more difficult to form separate memory traces for tones separated by smaller intervals. In addition, the interval of 15 semitones was chosen as it is neither highly consonant nor highly dissonant. The effects of consonance on the ability to form separate memory traces is also not known, but presumably, highly consonant intervals would more easily fuse into a single percept as the tones comprising them contain overlapping harmonics and/or subharmonics. Therefore, future studies should address how polyphonic representations are affected by interval size and consonance relations during development. It should also be noted that the ability to encode simultaneous sounds probably does not extend indefinitely. Behavioral studies of polyphony perception indicate that once there are more than 3 voices, even highly experienced adult listeners tend to underestimate their number, suggesting limitations on how many auditory objects can be represented at once, particularly if they interact as in musical harmony (Huron 1989). Thus, it would be interesting for future research to test for the nature and limit of parallel memory traces in infants and in adults by manipulating parametrically the number of simultaneous sounds presented and their harmonic relationships.

A major question concerns whether the more robust encoding of the higher voice is innate or whether it is the result of experience. Certainly, in Western music composition, it is most common to put the melody line in the highest voice. Further, the bias for better encoding of the higher voice has been shown previously in adults using behavioral indices, brainstem EEG, and cortical MEG recording (e.g., Zenatti 1969; Fujioka et al. 2005; 2008; Lee et al. 2009). However, to our knowledge, this is the first study to show this effect in infancy, raising the possibility that it is relatively immune to experience. Although studies with infants indicate that pitch discrimination thresholds continue to mature particularly for lower frequencies reliant on temporal processing until 10 or 11 years of age (Maxon and Hochberg 1982; Werner 2007), such differential maturation for high and low frequencies is unlikely to explain our results because the fundamental frequencies of the 2 voices both fall within the range reliant on temporal processing (Moore et al. 2008) and the harmonics of the 2 complex tones overlap in frequency range. If only pure tones are considered, one might actually predict a low voice superiority effect due to the asymmetric shape of tuning curves in the auditory nerve and the consequent upward spread of masking (Egan and Hake 1950), as pointed out by Fujioka et al. (2008). However, for complex tones, the harmonics of each tone likely play an important role in explaining the perceptual prominence of the higher notes. For example, the lower and the higher piano tones used as standards in our study had fundamental frequencies of 196 and 466 Hz, respectively.

Considering the lower note, its fundamental frequency should be well encoded in the peripheral auditory system as no other components are close to it in frequency. However, its second harmonic (392 Hz) is close to the fundamental frequency of the higher note. If these 2 components were of equal intensity, the lower component (second harmonic of the lower note) would be expected to suppress the higher component (fundamental of the higher note). However, because intensity falls off with increasing harmonic number in piano tones, and because the component with the higher intensity dominates when 2 components are close in frequency, the fundamental frequency of the higher note would be expected to suppress the second harmonic of the lower note. Following this reasoning, whenever harmonics from the 2 notes are close in frequency, the harmonic from the higher note will be more intense than the harmonic from the lower note (because it would be of a lower harmonic number) and therefore would be expected to dominate. In sum, because the harmonics contribute substantially to the pitch percept, this pattern of suppression would lead to a better percept for higher than lower notes when presented simultaneously. We are currently testing this idea with a model of the auditory periphery. If it is correct, it would suggest that the high voice superiority effect is innate in that it has a peripheral origin. Consistent with an innate origin was our finding that there was no correlation between amount of music listening and MMN latency or the degree of high voice superiority (i.e., the difference between MMN amplitude for the high and low voices).

Although we found no evidence that experience affected the high voice superiority effect, we did find evidence consistent with a role for experience-driven neural plasticity for pitch encoding in general. Specifically, we found a correlation between the amplitude of the MMN in each voice and the average amount of music listening per week. Previous work from our group has shown that at the behavioral level, Western infants are not yet sensitive to Western scale structure at the attentive level of processing (e.g., Trainor and Trehub 1992, 1994; Hannon and Trainor 2007; Trainor and Corrigall 2010; Trainor and Unrau 2012), although some culture-specific musical pitch processing is evident by 4 years of age in the absence of formal musical instruction (Corrigall and Trainor 2010). The correlations in the present study indicate that musical experience affects the ability to encode musical pitch before children become enculturated listeners. This effect of experience is also consistent with a study in which infants' experience with novel timbres was controlled (Trainor et al. 2011). This study reported greater ERP responses to pitch changes for tones in experienced compared with nonexperienced timbres. Finally, a very recent ERP study revealed larger and/or earlier brain responses to musical tones after 6 months of active compared with passive participatory music classes, beginning at 6 months of age during infancy (Trainor et al. forthcoming). A full answer to the question of the relative roles of experience and innate factors in the superior encoding of the higher of 2 voices will likely need to involve a consideration of cross-cultural differences in terms of music compositional practice.

Conclusion

At the level of preconscious memory, 2 concurrent pitches are encoded in separate memory traces in auditory cortex at 7 months of age. The adult bias for better encoding of the higher over the lower of 2 simultaneous voices is also evident during infancy. Moreover, significant correlations between amount of music listening and general memory trace strength suggest a role of experience-driven plasticity in the processing of polyphonic music and an important role for music in the development of auditory cortex. These findings support the view that by strengthening neural activation in response to sound differences in young infants, music-based active training might be useful for children with poor auditory and language skills (see also Trainor et al. forthcoming).

References

Abdala C, Chatterjee M. 2003. Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans. J Acoust Soc Am. 114:932–943.

Alain C, McDonald KL. 2007. Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. J Neurosci. 27:1308–1314.

Alain C, McDonald KL, Ostroff JM, Schneider B. 2001. Age-related changes in detecting a mistuned harmonic. J Acoust Soc Am. 109:2211–2216.

Alain C, Theunissen EL, Chevalier H, Batty M, Taylor MJ. 2003. Developmental changes in distinguishing concurrent auditory objects. Brain Res Cogn Brain Res. 16:210–218.

Alho K, Sams M, Paavilainen P, Reinikainen K, Naatanen R. 1989. Event-related brain potentials reflecting processing of relevant and irrelevant stimuli during selective listening. Psychophysiology. 26:514–528.

Brattico E, Winkler I, Näätänen R, Paavilainen P, Tervaniemi M. 2002. Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. Neuroreport. 13:1747–1751.

Bregman AS. 1990. Auditory scene analysis; The perceptual organization of sound. Cambridge (MA): MIT Press.

Bregman AS, Liao C, Levitan R. 1990. Auditory grouping based on fundamental frequency and formant peak frequency. Can J Psychol. 44:400–413.

Carral V, Huotilainen M, Ruusuvirta T, Fellman V, Näätänen R, Escera C. 2005. A kind of auditory "primitive intelligence" already present at birth. Eur J Neurosci. 21:3201–3204.

Corrigall KA, Trainor LJ. 2009. Effects of musical training on key and harmony perception. Ann N Y Acad Sci. 1169:164–168.

Corrigall KA, Trainor LJ. 2010. Musical enculturation in preschool children: acquisition of key and harmonic knowledge. Music Percept. 28:195–200.

Crawley EJ, Acker-Mills BE, Pastore RE, Weil S. 2002. Change detection in multi-voice music: the role of musical structure, musical training, and task demands. J Exp Psychol Hum Percept Perform. 28:367–378.

Cusack R, Roberts B. 2000. Effects of differences in timbre on sequential grouping. Percept Psychophys. 62:1112–1120.

Darwin CJ, Carlyon RP. 1995. Auditory grouping. In: Moore BC, editor. Hearing. San Diego (CA): Academic Press. p. 387–424.

Demany L. 1982. Auditory stream segregation in infancy. Infant Behav Dev. 5:261-276.

Egan JP, Hake HW. 1950. On the masking pattern of a simple auditory stimulus. J Acoust Soc Am. 22:622-630.

Fassbender C. 1993. Auditory grouping and segregation processes in infancy. Norderstedt (Germany): Kaste Verlag.

Folland NA, Butler BE, Smith NA, Trainor LJ. 2012. Processing simultaneous auditory objects: Infants' ability to detect mistunings in harmonic complexes. J Acoust Soc Am. 131:993-997.

Fujioka T, Mourad N, Trainor LJ. 2011. Development of auditory-specific brain rhythms in infants. Eur J Neurosci. 33:521-529.

Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ. 2006. One year of musical training affects development of auditory cortical-evoked fields in young children. Brain. 129:2593-2608.

Fujioka T, Trainor LJ, Ross B. 2008. Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. Neuroreport. 19:361-366.

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. 2005. Automatic encoding of polyphonic melodies in musicians and nonmusicians. J Cogn Neurosci. 17:1578-1592.

Hannon EE, Trainor LJ. 2007. Music acquisition: effects of enculturation and formal training on development. Trends Cogn Sci. 11:466-472.

Hartmann WM, McAdams S, Smith BK. 1990. Hearing a mistuned harmonic in an otherwise periodic complex tone. J Acoust Soc Am. 88:1712-1724.

He C, Hotson L, Trainor LJ. 2007. Mismatch responses to pitch changes in early infancy. J Cogn Neurosci. 19:878-892.

He C, Hotson L, Trainor LJ. 2009a. Development of infant mismatch responses to auditory pattern changes between 2 and 4 months of age. Eur J Neurosci. 29:861-867.

He C, Hotson L, Trainor LJ. 2009b. Maturation of cortical mismatch responses to occasional pitch change in infancy: effects of presentation rate and magnitude of change. Neuropsychologia. 47:218-229.

He C, Trainor LJ. 2009. Finding the pitch of the missing fundamental in infants. J Neurosci. 29:7718-8822.

Huron D. 1989. Voice denumerability in polyphonic music of homogeneous timbres. Music Percept. 6:361-382.

Huron D. 2001. Tone and voice: a derivation of the rules of voice-leading from perceptual principles. Music Percept. 19:1-64.

Iverson P. 1995. Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. J Exp Psychol Hum Percept Perform. 21:751-763.

Koelsch S, Schröger E, Tervaniemi M. 1999. Superior pre-attentive auditory processing in musicians. Neuroreport. 10:1309-1313.

Kushnerenko E, Ceponiene R, Balan P, Fellman V, Naatanen R. 2002. Maturation of the auditory change-detection response in infants: a longitudinal ERP study. Neuroreport. 13:1843-1848.

Lee KM, Skoe E, Kraus N, Ashley R. 2009. Selective subcortical enhancement of musical intervals in musicians. J Neurosci. 29:5832-5840.

Marie C, Kujala T, Besson M. 2012. Musical and linguistic expertise influence preattentive and attentive processing of non-speech sounds. Cortex. 48:447-457.

Maxon AB, Hochberg I. 1982. Development of psychoacoustic behavior: sensitivity and discrimination. Ear Hear. 3:301-308.

McAdams S, Bertoncini J. 1997. Organization and discrimination of repeating sound sequences by newborn infants. J Acoust Soc Am. 102:2945-2953.

Moore DR, Ferguson MA, Halliday LF, Riley A. 2008. Frequency discrimination in children: Perception, learning and attention. Hear Res. 238:147-154.

Mourad N, Reilly JP, De Bruin H, Hasey G, MacCrimmon D. 2007. A simple and fast algorithm for automatic suppression of high-amplitude artifacts in EEG data. In: ICASSP, IEEE international conference on acoustics, speech and signal processing—proceedings, Honolulu (HI). p. I393-I396. doi:10.1109/ICASSP.2007.366699.

Näätänen R, Michie PT. 1979. Early selective attention effects on the evoked potential. A critical review and reinterpretation. Biol Psychol. 8:81-136.

Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol. 118:2544-2590.

Näätänen R, Pakarinen S, Rinne T, Takegata R. 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin Neurophysiol. 115:140-144.

Nager W, Teder-Sälejärvi W, Kunze S, Münte TF. 2003. Preattentive evaluation of multiple perceptual streams in human audition. Neuroreport. 14:871-874.

Olsho LW, Schoon C, Sakai R, Turpin R, Sperduto V. 1982. Auditory frequency discrimination in infancy. Dev Psychol. 18:721-726.

Pakarinen S, Takegata R, Rinne T, Huotilainen M, Näätänen R. 2007. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). Clin Neurophysiol. 118:177-185.

Palmer C, Holleran S. 1994. Harmonic, melodic, and frequency height influences in the perception of multivoiced music. Percept Psychophys. 56:301-312.

Picton TW, Alain C, Otten L, Ritter W, Achim A. 2000. Mismatch negativity: different water in the same river. Audiol Neurootol. 5:111-139.

Ritter W, Deacon D, Gomes H, Javitt DC, Vaughan HG Jr. 1995. The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review. Ear Hear. 16:52-67.

Sambeth A, Pakarinen S, Ruohio K, Fellman V, van Zuijen TL, Huotilainen M. 2009. Change detection in newborns using a multiple deviant paradigm: a study using electroencephalography. Clin Neurophysiol. 120:530-538.

Shinozaki N, Yabe H, Sato Y, Sutoh T, Hiruma T, Nashida T, Kaneko S. 2000. Mismatch negativity (MMN) reveals sound grouping in the human brain. Neuroreport. 5:1597-1601.

Singh PG. 1987. Perceptual organization of complex-tone sequences: a trade-off between pitch and timbre. J Acoust Soc Am. 82:886-899.

Sinnott JM, Aslin RN. 1985. Frequency and intensity discrimination in human infants and adults. J Acoust Soc Am. 78:1986-1992.

Smith NA, Trainor LJ. 2011. Auditory stream segregation improves infants' selective attention to target tones amid distracters. Infancy. 16:655-668.

Sussman E, Winkler I, Ritter W, Alho K, Näätänen R. 1999. Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. Neurosci Lett. 264:161-164.

Teas DC, Klein AJ, Kramer SK. 1982. An analysis of auditory brainstem responses in infants. Hear Res. 7:19-54.

Tervaniemi M, Rytkönen M, Schröger E, Ilmoniemi RJ, Näätänen R. 2001. Superior formation of cortical memory traces for melodic patterns in musicians. Learn Mem. 8:295-300.

Tew S, Fujioka T, He C, Trainor L. 2009. Neural representation of transposed melody in infants at 6 months of age. J Acoust Soc Am. 1169:287-290.

Trainor LJ. 2005. Are there critical periods for music development? Dev Psychobiol. 46:262-278.

Trainor LJ. 2008. Event-related potential (ERP) measures in auditory developmental research. In: Schmidt LA, Segalowitz SJ, editors. Developmental psychophysiology: theory, systems and methods. New York: Cambridge University Press. p. 69-102.

Trainor LJ, Corrigall KA. 2010. Music acquisition and effects of musical experience. In: Riess-Jones M, Fay RR, editors. Springer handbook of auditory research: music perception. Heidelberg (Germany): Springer. p. 89-128.

Trainor LJ, Lee K, Bosnyak DJ. 2011. Cortical plasticity in 4-month-old infants: specific effects of experience with musical timbres. Brain Topogr. 24:192-203.

Trainor LJ, Marie C, Gerry D, Whiskin E, Unrau A. forthcoming. Becoming musically enculturated: effects of music classes for infants on brain and behaviour. Ann N Y Acad Sci.

Trainor LJ, Trehub SE. 1992. A comparison of infants and adults sensitivity to Western musical structure. J Exp Psychol Hum Percept Perform. 18:394-402.

Trainor LJ, Trehub SE. 1994. Key membership and implied harmony in Western tonal music: developmental perspectives. Percept Psychophys. 56:125-132.

Trainor LJ, Unrau AJ. 2012. Development of pitch and music perception. In: Werner LA, Fay RR, Popper AN, editors. Springer Handbook of Auditory Research: Human Auditory Development. New York: Springer. p. 223-254.

Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. J Neurosci. 24:10440-10453.

Werner LA. 2002. Infant auditory capabilities. Curr Opin Otolaryngol Head Neck Surg. 10:398-402.

Werner LA. 2007. Human auditory development. In: Hoy R, Dallos P, Oertel D, editors. The senses: a comprehensive reference, volume 3—Audition. New York: Academic Press. p. 871-894.

Winkler I, Denham SL, Nelken I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci. 13:532-540.

Winkler I, Kushnerenko E, Horváth J, Ceponiene R, Fellman V, Huotilainen M, ldots Sussman E. 2003. Newborn infants can organize the auditory world. Proc Natl Acad Sci U S A. 100: 11812-11815.

Winkler I, Paavilainen P, Näätänen R. 1992. Can echoic memory store two traces simultaneously? A study of event-related brain potentials. Psychophysiology. 29:337-349.

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Suto T, Hiruma T, Kaneko S. 2001. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res. 897:222-227.

Zenatti A. 1969. Le développement génétique de la perception musicale. Monographies Francaises de Psychologie, 17. Paris (France): CNRS.