

Research Article

Infant Pitch and Timbre Discrimination in the Presence of Variation in the Other Dimension

BONNIE K. LAU^{1,2} , ANDREW J. OXENHAM³ , AND LYNNE A. WERNER⁴

¹ *Institute for Language and Brain Sciences, University of Washington, 1715 NE Columbia Rd, Box 357988, Seattle, WA 98195, USA*

² *Department of Otolaryngology – Head and Neck Surgery, University of Washington, 1701 NE Columbia Rd, Box 357923, Seattle, WA 98195, USA*

³ *Department of Psychology, University of Minnesota, 75 East River Parkway, Minneapolis, MN 55455, USA*

⁴ *Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Box 354875, Seattle, WA 98105, USA*

Received: 22 February 2021; accepted: 2 July 2021; Online publication: 14 September 2021

ABSTRACT

Adult listeners perceive pitch with fine precision, with many adults capable of discriminating less than a 1 % change in fundamental frequency (F0). Although there is variability across individuals, this precise pitch perception is an ability ascribed to cortical functions that are also important for speech and music perception. Infants display neural immaturity in the auditory cortex, suggesting that pitch discrimination may improve throughout infancy. In two experiments, we tested the limits of F0 (pitch) and spectral centroid (timbre) perception in 66 infants and 31 adults. Contrary to expectations, we found that infants at both 3 and 7 months were able to reliably detect small changes in F0 in the presence of random variations in spectral content, and vice versa, to the extent that their performance matched that of adults with musical training and exceeded that of adults without musical training. The results indicate high fidelity of F0 and spectral-envelope coding in infants, implying that fully mature cortical processing is not necessary for accurate discrimination of these features. The surprising difference in performance between infants and musically untrained adults may reflect a developmental trajectory for learning natural statistical covariations between pitch and timbre that improves coding efficiency but results in

degraded performance in adults without musical training when expectations for such covariations are violated.

Keywords: cortical maturation, auditory development, pitch perception, timbre perception

INTRODUCTION

Pitch is a fundamental perceptual attribute of sound that can be ordered on a scale from high to low, and is most closely related to the fundamental frequency (F0) or repetition rate of a stimulus (ANSI 2013). Some adult listeners perceive pitch with extremely fine precision, with the ability to discriminate less than a 1 % change in F0, particularly after training (e.g., Lau et al. 2017b; Micheyl et al. 2006). Studies investigating the neural correlates of pitch in humans and other primates have indicated the involvement of the auditory cortex in the processing of pitch-evoking sounds (Zatorre et al. 1992; Penagos et al. 2004; Bendor and Wang 2005; Norman-Haignere et al. 2013).

Infant pitch perception is interesting in this regard because of the protracted development of the central auditory system, including the auditory cortex (Moore and Guan 2001). Although infants begin responding to sound in the third trimester of gestation (Birnholtz and Benacerraf 1983), it is hypothesized that brainstem processing supports early responses to sound because

Correspondence to: Bonnie K. Lau · Institute for Language and Brain Sciences · University of Washington · 1715 NE Columbia Rd, Box 357988, Seattle, WA, 98195, USA. email: blau@uw.edu

of cortical immaturity (Eggermont and Moore 2012; Lau and Werner 2012). Given the hypothesized cortical involvement in pitch, and its slow developmental trajectory, it may be that infant pitch perception is similarly slow to develop.

Many studies have shown that infants are sensitive to large frequency or F0 changes in both speech and music during the first few months of life. Infants can discriminate the frequency of pure tones (Olsho 1984), the F0 of complex tones (Montgomery and Clarkson 1997; He and Trainor 2009; Lau and Werner 2012, 2014), and musical melodies (Trehub et al. 1985; Plantinga and Trainor 2009; Lau et al. 2017a), as well as lexical tones (Mattock et al. 2008) and F0 contours of syllables and words (Karzon and Nicholas 1989). Although infants perceive pitch, it is unclear whether they are able to discriminate subtle changes with the acuity of adult listeners. For pure tones, infants discriminate changes in frequency as well as adults by 6 months at high frequencies (Olsho 1984) but at low frequencies, discrimination thresholds do not seem to reach adult-like levels until 13–14 years of age (Maxon and Hochberg 1982). Immature pure-tone frequency-discrimination thresholds have also been reported for school-aged children at both high and low frequencies (Buss et al. 2014). However, as Buss et al. point out, children's performance is likely influenced by non-sensory factors such as memory, sustained attention, and testing method (Buss et al. 2014). Furthermore, with pure tones, it is difficult to determine whether participants are discriminating on the basis of pitch or a change in timbre.

To investigate the influence of cortical maturation on pitch perception of complex tones, we compared F0 discrimination in 3-month-olds, 7-month-olds, and adults. To prevent participants from using the lowest spectral edge or the spectral centroid of the stimulus rather than the F0 (e.g., Houtsma and Smurzynski 1990), random variations in the spectral content were introduced, which resulted in changes in the sounds' perceived brightness, one dimension of timbre — the perceptual aspect of sound that allows us to distinguish between different instruments (e.g., a piano and a violin) or voices (e.g., male and female) when they produce the same pitch at the same loudness. Infants at 3 and 7 months were tested because the organization of the auditory pathways appears to be distinctly different at these two ages. At 3 months, the auditory cortex is markedly immature, with activation of only the most superficial layer of the cortex by the reticular activating system; by 7 months, thalamocortical connections show increasing axonal conduction velocity, although significant immaturities in the auditory cortex persist (Eggermont and Moore 2012). To the extent that the thalamocortical pathways are required for precise F0 perception, we predicted a trajectory of improving F0 discrimination abilities from 3-month-olds to 7-month-olds to adults.

EXPERIMENT 1: F0 DISCRIMINATION WITH SPECTRAL VARIATION

This experiment tested the limits of infants' and adults' ability to detect a change in the F0 (the physical correlate of pitch) within a sequence of complex tones, each containing a random selection of consecutive harmonics, leading to random changes in brightness from tone to tone. In some sequences, a change in the F0 was introduced ("change" trial; Fig. 1A); in others, the F0 remained constant throughout the sequence ("no-change" trial; Fig. 1B). The same observer-based psychophysical procedure (Werner 1995) was used for both infants and adults, so that their discrimination abilities could be compared directly.

Method

Participants

Participants were 26 infants, including 13 3-month-olds (7 female and 6 male) and 13 7-month-olds (8 female and 5 male), and 21 adults, including 10 (9 female and 1 male) with little or no musical training and 11 (6 female and 5 male) with 2 or more years of formal musical training, including school band or choir. Adults with musical training can typically discriminate smaller changes in F0 than non-musicians prior to explicit training (Micheyl et al. 2006; Madsen et al. 2019) and were included in order to obtain a full range of adult F0 discrimination abilities. A sample size of 10 per group was determined a priori based on estimates of variability obtained from our previous studies of infant pitch perception (Lau and Werner 2012, 2014; Lau et al. 2017a) and was sufficient to detect a difference between groups when one group reliably reaches criterion performance on the task (80 % or more meet criterion) and another group does not (20 % or fewer meet criterion) with a power (1- β) of 0.8 and probability (α) of 0.05 (Rosner 1995, p. 384).

All infants were born full term, had no history of otitis media within 2 weeks of testing, had no history of health or developmental concerns, and had passed their newborn hearing screening. Infants in each group completed testing within 10 days of the specified age. Adult participants were between 18 and 30 years of age, reported normal hearing, and had no prior participation in psychoacoustic experiments. Adult participants were required to pass an audiometric screening at 20 dB hearing level at octave frequencies between 250 and 8000 Hz at the time of testing. Data were excluded from five additional infants (4 failed training due to fussiness or inattention, 1 failed tympanometric screen) and five additional adults (4 failed training, 1 failed to complete testing). Adult participants were tested over an average of 4 sessions

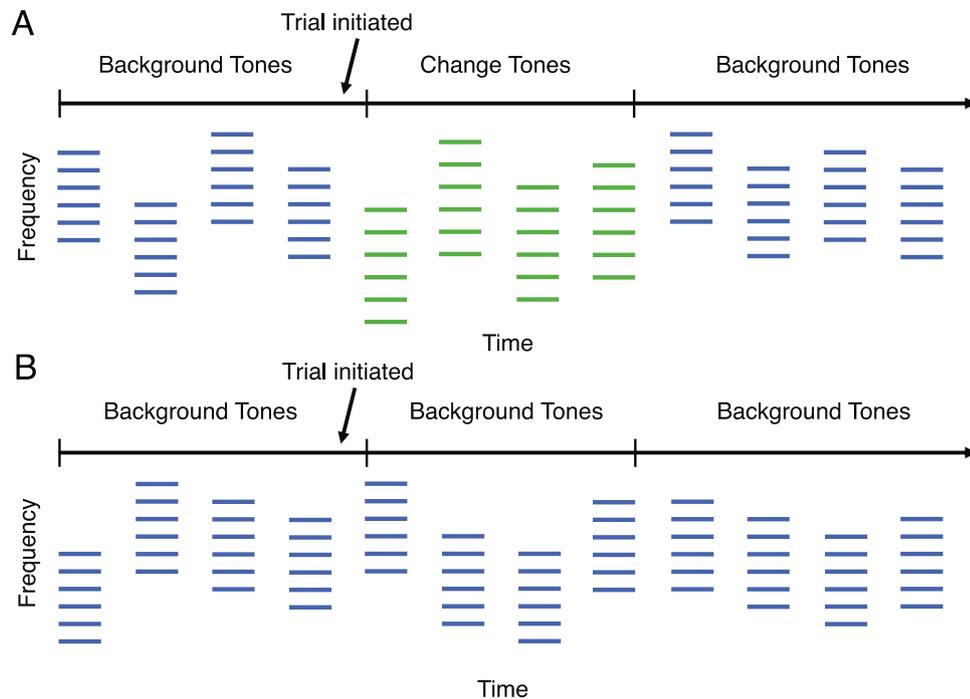


FIG. 1 Schematic diagram of a change (A) and a no-change trial (B) in experiment 1. **A** At the start of a session, background tones were played repeatedly. Random variation of which harmonics were included in the tone was used to produce changes in brightness but

not a change in pitch (blue tones). On change trials, a pitch change was introduced and the changed tones were played four times (green tones; $F_0 + \Delta F_0$). **B** On no-change trials, the background tones continued to play throughout the trial.

within a single 1 h visit. Infants were tested over an average of 7 sessions in up to three 1 h visits (mean = 2.3 visits) because they required more breaks. All procedures were conducted according to protocols approved by the Institutional Review Board at the University of Washington (protocol #: 29,813, approved May 17, 2017) and informed consent was obtained from all participants or their legal guardians prior to testing.

Stimuli

The stimuli consisted of sequences of 650 ms harmonic complex tones, including 50 ms linear rise/fall times, separated by 500 ms gaps. Six consecutive harmonics of each complex tone were generated and then individually scaled to produce slopes of 12 dB per octave, spectrally centered on the complex tone, with no flat bandpass region. The components were combined in random phase. The random variation of harmonic numbers on each presentation (with no energy at the F_0 itself) was incorporated to limit participants' ability to respond to spectral changes as opposed to F_0 (e.g., Moore and Glasberg 1990; Moore and Moore 2003; Micheyl and Oxenham 2004; Micheyl et al. 2012); the 12 dB per octave slopes were incorporated in order to reduce any salient pitch cues related to the spectral edges of the stimulus (e.g., Kohlrausch & Houtsma, 1992).

The baseline F_0 was around 200 Hz, similar to that of an average female voice. Seven differences in F_0 (ΔF_0) were used, ranging from 0 to 5 %. The tones at F_0 had lowest harmonic numbers ranging from 3 to 7, and the tones at $F_0 \pm \Delta F_0$ had lowest harmonic numbers ranging from 2 to 6, thereby introducing random variations in brightness. The standard F_0 was varied in the range from 195 to 200 Hz between blocks with different ΔF_0 values to prevent repeated exposure to the same background tones.

The complex tones were presented at an overall level of 70 dB SPL and were embedded in a 65 dB SPL pink noise (1 to 12,000 Hz) to reduce the audibility of any distortion products. The stimuli were then presented via an Etymotic ER-2 insert earphone to the right ear. Sound pressure levels were calibrated in a Zwislocki coupler and checked in the subject's ear canal with an Etymotic ER-7 probe microphone system at the start of testing. A 1 kHz pure tone was presented and if the sound pressure level was not measured to be within 2 dB of the intended sound level, the ER-2 insert earphone was removed and replaced in the ear canal. Foam ear tips were shaved to fit infant ear canals, as needed. Sound pressure levels were subsequently rechecked before the start of testing. Testing was conducted in a sound-attenuating booth. Half the participants heard F_0 increases ($+ \Delta F_0$) while the other half heard F_0 decreases ($- \Delta F_0$), randomly determined.

Procedure

The task was to listen to the sequence of complex tones and respond when there was an F0 change. Only the participant heard the stimuli, and the experimenter (who was blind to the trial type) had to judge whether a change trial was presented, based solely on the observed behavior of the participant. This observer-based procedure has a long history in both visual (Teller 1979) and auditory (Werner 1995) psychophysics and has been widely used in developmental studies (Lau and Werner 2012; Benasich et al. 2014; Horn et al. 2017).

During testing, infants sat on a caregiver's lap inside a sound-attenuated booth. An assistant in the booth manipulated toys to keep the infants facing midline. There were two mechanical toys in a dark Plexiglas box and a monitor to the participant's right. Either the toys were activated or a video was presented to reinforce the infant's response to a change in F0. The experimenter sat outside the sound booth and observed through a window. The adults in the booth were unable to hear the sounds because the stimuli were presented to the infant via an insert earphone. As an extra precaution, the caregiver and assistant both wore circumaural headphones, with the caregiver listening to music and the assistant listening to the experimenter's instructions. The experimenter outside the sound booth, who also could not hear the experimental stimuli, wore headphones and a microphone to communicate with the assistant inside the booth.

At the start of testing, background tones with the same F0 were played repeatedly. The experimenter initiated a trial when the participant was quiet and facing midline. To control for response bias, both change trials and no-change trials were presented to calculate hit and false-alarm rates, respectively (Green and Swets 1966). On change trials, four consecutive tones at $F0 \pm \Delta F0$ were played before the tones reverted to F0 (Fig. 1A); on no-change trials, F0 tones continued to repeat (Fig. 1B). The experimenter had 4 s from trial onset to decide which trial type had occurred. Typical behaviors used to make judgments included infants' eye-darts and head-turns towards reinforcers (mechanical toys or video), as well as facial expressions, like eye-widening. Computer feedback was provided to the experimenter after each trial. Adult participants sat alone in the booth and were instructed simply to raise their hand when they heard the change in sound that activated the toys. In all other respects, the stimuli and procedure were the same for the adult and infant participants.

Conditioning and Training

Participants were trained to respond during a conditioning phase with the pairing of a large F0 change (5 % $\Delta F0$) and the activation of mechanical animals or a video. The probability of a change trial was 0.8 and the reinforcer was activated after every change trial regardless

of the experimenter's response. The experimenter had to respond correctly on 4 of 5 consecutive change trials and 1 no-change trial within a maximum of 15 trials to progress to the training phase.

In the training phase, the task and stimuli were the same, but the probability of a change trial was 0.5 and the reinforcer was activated only when the experimenter correctly identified a change trial. Participants were required to achieve a hit rate of at least 80 % and a false-alarm rate of at most 20 % over the last 5 change and no-change trials within 40 trials in order to pass the training phase. If the criterion was not met, the session was discontinued, and participants were given a break. Participants had up to four attempts to reach criterion to progress to the test phases.

After passing the training phase, participants were tested on up to 6 additional phases that presented progressively smaller F0 changes, with the probability of a change trial remaining at 0.5. The values of $\Delta F0$ were 2.5, 1.5, 1, 0.5, 0.25, and 0 % of the F0. In order to progress to the next $\Delta F0$, participants were required to reach the pass criterion of at least an 80 % hit rate with a false-alarm rate of at most 20 % on the last 5 change trials and the last 5 no-change trials, corresponding to a sensitivity (d') of about 1.68 or better. This equates to correct responses on at least 4 of the last 5 change trials and at least 4 of last 5 no-change trials before a step down was taken. The pass criterion was chosen to depend on both hit rate and false-alarm rate over multiple trials to reduce the effects of any potential bias differences between infants and adults on the outcome of the procedure.

To account for potential inattention, the $\Delta F0$ reverted to the previous (larger) $\Delta F0$ value following four consecutive incorrect responses. If participants responded correctly on 5 of 6 consecutive trials in a maximum of 15 trials at the previous $\Delta F0$, demonstrating that they were again attending to the task, testing resumed at the smaller $\Delta F0$ value. Otherwise, the session was discontinued. Participants had a maximum of four attempts at each $\Delta F0$, at which point testing was discontinued. Statistical analyses of data collected were conducted using SPSS Version 19 and R 4.0.3 with the mosaic package.

Results

To address our initial question of whether 7-month-old infants outperform 3-month-old infants in F0 acuity, we compared the proportion of each group that successfully performed the task at the smallest non-zero F0 difference (0.25 %). As shown in Fig. 2A, there were no clear differences between the two infant groups, with 11 of 13 of both the 3-month-olds and 7-month-olds (~85 % of each group; 95 % binomial exact confidence interval, CI [55 %, 98 %]) achieving criterion performance. In contrast, none of the 10 adult participants without musical training

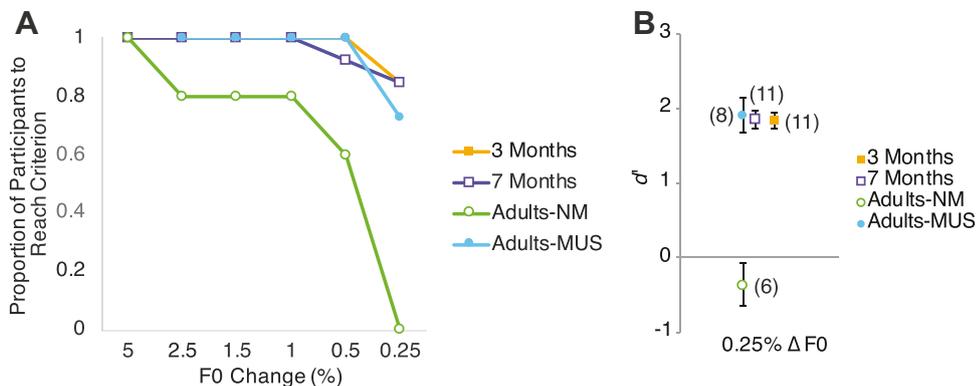


FIG. 2 **A** Proportion of participants that reached the pass criterion as a function of F0 change in experiment 1. Different symbols and colors represent the different groups, as shown in the legend. **B** Mean d' values (± 1 SEM) at the 0.25 % $\Delta F0$ for the different groups, as indicated by the symbols and colors. The number of participants in each group is shown in parentheses by each symbol.

(0 %; 95 % binomial exact CI [0 %, 31 %]) achieved criterion performance at this F0 difference. Thus, the proportion of successful infants at the smallest F0 change was significantly greater than the proportion of adults without musical training, $\chi^2_{(1, N=23)} = 16.22, P < 0.0001, \Phi_c = 0.84$ for each infant group. Of the adult group with musical training, 8 of 11 (73 %, 95 % binomial exact CI [39 %, 94 %]) reached criterion. This proportion was significantly higher than that of the musically untrained adults, $\chi^2_{(1, N=21)} = 11.75, P = 0.001, \Phi_c = 0.75$, but was not significantly different from the proportion for either the 3- or 7-month-old infants, $\chi^2_{(1, N=24)} = 0.51, P = 0.48, \Phi_c = 0.15$ in both cases. As expected, no participants achieved criterion performance at the 0 % F0 difference.

In a secondary analysis to determine whether there were systematic differences in sensitivity between the groups for those achieving criterion performance at a given F0 difference, we calculated mean d' values at

each value of $\Delta F0$. Of the participants that reached the 0.25 % $\Delta F0$ value, average d' values for the 3-month-old, 7-month-old, and musician adult groups on the last five change and no-change trials were greater than 1 (3 months: $N = 11$, mean $d' = 1.84$, SEM = 0.11; 7 months: $N = 11$, mean $d' = 1.85$, SEM = 0.12; musicians: $N = 8$, mean $d' = 1.91$, SEM = 0.23), whereas the average d' value for the musically untrained participants who reached the smallest $\Delta F0$ value was -0.37 ($N = 6$, SEM = 0.29), which was not significantly different from zero (one-sample t -test, $t_{(5)} = -1.30, P = 0.25, 95\% \text{ CI} [-1.11, 0.36]$, Cohen's $d = -0.53$), consistent with chance-level performance (Fig. 2B). The four musically untrained adults who did not reach the smallest $\Delta F0$ value failed either at $\Delta F0 = 2.5\%$ ($N = 2$) or at $\Delta F0 = 0.5\%$ ($N = 2$). Similar values of d' between groups were also found at all other values of $\Delta F0$ (Fig. 3). The fact that sensitivity was similar between groups for participants achieving the criterion level of performance suggests that any differences between the groups, in terms of the number of participants actually achieving the criterion level of performance, reflect true differences in sensitivity rather than simply between-group differences in response bias.

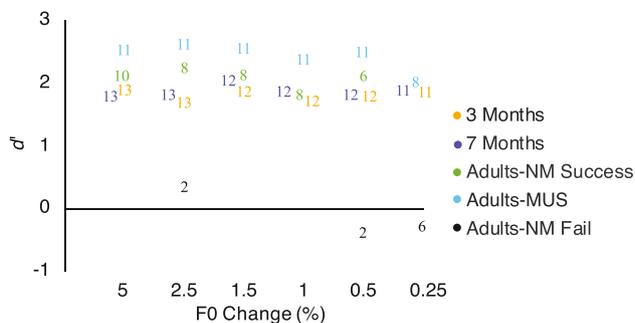


FIG. 3 Mean d' as a function of F0 change for participants that reached the pass criterion for each $\Delta F0$. The number of participants that contributed to each average is shown by the color-coded numbers. Non-musician adults who failed to reach criterion are shown in black.

To assess the stability of the results, the proportions of infants and adults reaching criterion at the 0.25 % $\Delta F0$ value were resampled using bootstrapping with replacement to generate 1000 replications. The 95 % CIs using the percentile method were then computed based on the bootstrap sampling distribution. The bootstrapped CI for adults without musical training (95 % percentile bootstrap CI [0 %, 0 %]) did not overlap with the 3-month-olds' (95 % percentile bootstrap CI [62 %, 100 %]) or the 7-month-olds' (95 % percentile bootstrap CI [62 %, 100 %]), confirming the significant difference in performance between the adults without musical training and both groups of infants.

EXPERIMENT 2: SPECTRAL CENTROID DISCRIMINATION WITH F0 VARIATION

The results from experiment 1 did not support our initial prediction that 7-month-olds would outperform 3-month-olds in an F0-discrimination task. In fact, the two groups of infants performed similarly and a higher proportion of infants in each group reached the pass criterion at the smallest value of $\Delta F0$ than was found for the musically untrained adults. One possible explanation for this surprising outcome is that the variations in spectrum (due to the randomly selected harmonics in each trial) interfered with adults', but not infants', ability to discriminate F0. That might occur because the infants perceived the spectral variations, but the variations did not interfere with their F0 discrimination, or because the infants simply did not perceive the spectral variations. Experiment 2 was designed to distinguish between these two possibilities by testing the limits of infants' and adults' ability to perceive changes in the spectral envelope of harmonic complex tones. The task was the same as in experiment 1 except that the parameters were reversed: Participants were required to detect a change in spectral envelope (induced by varying the spectral centroid of the stimulus) while ignoring random and uninformative variations in F0.

Method

Participants

The new participants were 20 3-month-olds (12 female, 8 male), 20 7-month-olds (7 female, 13 male), and 10 musically untrained adults (5 female, 5 male). As in experiment 1, a minimum sample size of 10 per group was determined, based on an ability to detect a difference between two groups in which proportions of 0.8 and 0.2 achieve the performance criterion. All recruitment and participant inclusion criteria were the same as for experiment 1.

Stimuli

All tones were 500 ms harmonic complexes, including 20 ms raised-cosine rise/fall ramps, separated by silent gaps of 500 ms. All harmonics up to 10 kHz were generated and individually scaled to produce slopes of 24 dB per octave around the appropriate center frequency (CF), with no flat bandpass region. The components were combined in sine phase. Although the harmonics in experiment 1 were combined in random phase, we prioritized maintaining stimulus parameters consistent with Allen and Oxenham (2014) to allow for comparison of adult thresholds. This phase difference across experiments is not expected to have any effects on performance because phase effects are known to affect F0 discrimination and

the timbre of unresolved harmonics but not the mostly spectrally resolved harmonics used here (Plomp and Steeneken 1969; Houtsma and Smurzynski 1990).

The CF of each complex determined its brightness. Six values of ΔCF , from 0 to 15 % of the CF, were used to test spectral-envelope discrimination. For each CF and $CF \pm \Delta CF$ value, tones were generated with F0s of 170, 180, 190, 200, 210, and 220 Hz. The standard CF was varied between 1000 and 1100 Hz between blocks of trials to prevent repeated exposure to background tones with the same CF. Half the participants heard CF increases while the other half heard CF decreases, randomly determined.

Procedure

Stimulus presentation, calibration, and procedures were the same as for experiment 1. Only three ΔCF s were tested for each infant, to reduce the testing time per infant. All adults and all but one infant completed testing in a single visit of approximately 1 h. Both infants and adults were conditioned and trained to categorize two sets of complexes that differed by 15 % in their CF. Infants were then randomly assigned to one of two groups, with group 1 tested on 10 %, 5 %, and 0 % ΔCF s and group 2 tested on 2 %, 0.5 %, and 0 % ΔCF s. Adults were tested with all ΔCF values.

Results

Infants at both ages were able to discriminate smaller changes in spectrum than the musically untrained adults (Fig. 4A). At the ΔCF of 0.5 %, only 1 of 10 adult listeners (10 %, binomial exact 95 % CI [3 %, 45 %]) reached the pass criteria, compared with 9 of 10 3-month-olds (90 %, 95 % CI [56 %, 100 %]) and all 10 7-month-olds (100 %, 95 % CI [69 %, 100 %]) in group 2. The proportion of successful adults was significantly different from the proportion of successful 3-month-olds, $\chi^2_{(1, N=20)} = 12.80$, $P < 0.0001$, $\Phi_c = 0.80$, and successful 7-month-olds, $\chi^2_{(1, N=20)} = 16.36$, $P < 0.0001$, $\Phi_c = 0.91$. Resampling the infants and adults passing and failing the criterion at the 0.5 % ΔCF using bootstrapping with 1000 replications again showed that the CI for adults (95 % percentile bootstrap CI [0 %, 33 %]) did not overlap with that for the 3-month-olds (95 % percentile bootstrap CI [67 %, 100 %]) or the 7-month-olds (95 % percentile bootstrap CI [100 %, 100 %]). As expected, none of the participants achieved criterion when the ΔCF was 0.

In a secondary analysis to check for any between-group differences in sensitivity, we assessed mean group d' values at each ΔCF change. Of the participants who progressed to the 0.5 % ΔCF , average 3-month-old and 7-month-old d' values for the last 10 trials (5 change and no-change trials) were greater than 1 (3 months: $N=9$, mean $d' = 1.77$, SEM = 0.09; 7 months: $N=10$,

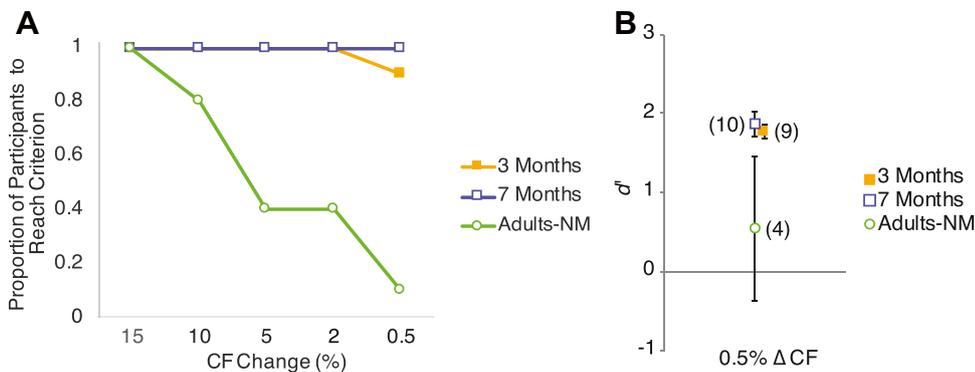


FIG. 4 **A** Proportion of participants that reached the pass criterion as a function of CF change in experiment 2. Different colors and symbols represent the three different groups, as indicated in the legend. **B** Mean d' (± 1 SEM) values at the 0.50 % Δ CF. Number of participants in each group is shown in parentheses by each symbol.

mean $d' = 1.86$, SEM = 0.16). However, the average adult d' value was less than 1 ($N = 4$, mean $d' = 0.55$, SEM = 0.91) and was not significantly greater than zero (one-sample t -test: $t_{(3)} = 0.60$, $P = 0.59$, 95 % CI [-2.36, 3.46], Cohen's $d = 0.50$), consistent with chance-level performance. Note that the data from the one adult participant who reached the pass criteria at the 0.5 % Δ CF with a d' of 2.49 were included in this average. The average d' values for this lowest Δ CF are shown in Fig. 4B; average d' values for all Δ CFs are shown in Fig. 5. As for experiment 1, no systematic differences in sensitivity were observed between groups for those participants achieving the criterion level of performance.

These results rule out the possibility that infants' superior F0 discrimination found in experiment 1 was due to an inability to perceive spectral changes. Indeed, as with F0, infants' ability to discriminate small changes in spectral envelope exceeded that of musically untrained adults.

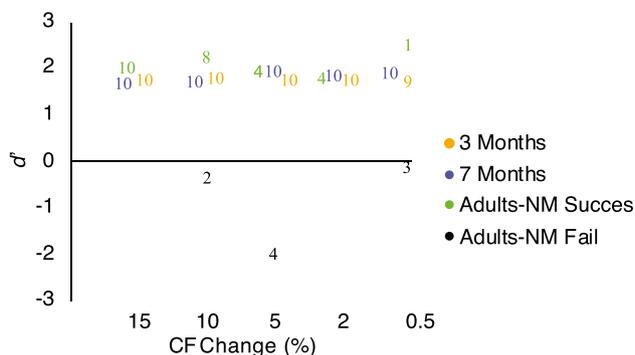


FIG. 5 Mean d' as a function of CF change for participants that reached the pass criterion on each Δ CF. The number of participants that contributed to each average is shown by the color-coded numbers. Adult participants who failed to reach criterion are shown in black.

DISCUSSION

A total of 66 infants were tested on F0 and spectral centroid discrimination in the presence of random variation in the other dimension. In experiment 1, 85 % of the infants tested and 73 % of adults with musical training were able to reach the performance criterion at the smallest F0 difference (0.25 %) in the test phase, whereas none of the adult participants without musical training reached this criterion. In experiment 2, 95 % of the infants were able to reach the performance criterion at the smallest CF difference (0.5 %), compared to only 10 % of the adult participants without musical training. These differences were highly significant in both experiments, with $p < 0.0001$ and an effect size, $\Phi_c > 0.8$ for each of the comparisons.

In terms of our primary research question, to investigate the early development of complex pitch perception, the results showed that infants at both 3 and 7 months performed as well as adults with musical training. This finding is in line with previous studies showing that infants can discriminate F0 and spectral differences in the first few months of life (Clarkson et al. 1988; Clarkson 1996; Montgomery and Clarkson 1997; Plantinga and Trainor 2009; He and Trainor 2009; Lau and Werner 2012). The surprising aspect of this finding is that no age-related improvements in discrimination were observed, as might be expected given the long developmental trajectory of auditory cortex. The fact that 3-month-olds performed as well as 7-month-olds and musically trained adults suggests that F0 and spectral-envelope discrimination is not dependent on a mature thalamocortical pathway or cortex.

The most unexpected finding — one that motivated experiment 2 — is that more infants reached criterion at the 0.25 % Δ F0 than adults without musical training. Adults often have difficulty distinguishing between F0 and the aspect of timbre manipulated in this experiment

known as brightness, describing increases in both as being “higher.” Indeed melodic contours can be recognized when they are conveyed by changes in spectral-envelope peak or CF, as well as by changes in F0 (McDermott et al. 2008). In studies of discrimination, interference between the two dimensions is common (Moore and Glasberg 1990; Borchert et al. 2011; Allen and Oxenham 2014; Lau and Werner 2014), and in functional imaging studies, substantial overlap between representations of pitch and brightness has been reported (Allen et al. 2017, 2019). Musically trained listeners typically demonstrate better F0 discrimination than non-musicians, at least prior to explicit training in the task (Micheyl et al. 2006; Madsen et al. 2019), but even musicians exhibit confusion when discriminating between F0 and spectral peak (Allen and Oxenham 2014). This perceptual interference may be an efficient coding strategy, as the two dimensions often covary in natural sounds (Whalen and Levitt 1995; Kitahara et al. 2005). Indeed recent studies have demonstrated that relatively brief exposure to covariations in sound features can lead to rapid perceptual learning in adults (Stilp et al. 2010; Stilp and Kluender 2012). Such experiments have shown that adult listeners readily exploit statistical regularity in stimulus attributes to improve task performance. One interpretation of our unexpected results is therefore that infants may not yet be able to exploit the expected statistical covariations between stimulus attributes, and that this inability actually enhances their perceptual performance, relative to musically untrained adults, in situations where expectations of covariation are violated.

Another related possible explanation is that adults are more susceptible to a confusion effect induced from the interfering irrelevant dimension. Allen and Oxenham (2014) showed that adult discrimination thresholds were better when the irrelevant variations in F0 or spectral peak were congruent with the change in the target dimension as opposed to when they were incongruent. In line with this observation, we would predict that if infants do not take into account the natural statistics of covariation, then they should not show the benefits of congruent variations in F0 and spectrum found for adults and should also not be adversely affected by incongruent variations in the two dimensions. This prediction remains to be tested.

It may also be that infants’ and adults’ processing strategies differ, even when they are faced with the same task. Specifically with F0, it has been shown that different species seem to rely on different acoustic cues to extract F0, with non-primates (e.g., ferrets and chinchillas) relying primarily on temporal-envelope cues provided by spectrally unresolved harmonics (Shofner and Chaney 2013; Walker et al. 2019), whereas humans and possibly other primates seem to rely more on the cues provided by spectrally resolved harmonics (Oxenham et al. 2011; Song et al. 2016). It is conceivable that different weighting of

such cues could lead to different patterns of interference, although our current experimental design cannot shed light on this question.

An important point to note is that while all aspects of the stimuli, task, and protocol remained the same between infants and adults, adults were explicitly instructed to raise their hand when they heard the sound change while infants were implicitly conditioned to respond. It is possible that this difference between implicit versus explicit task performance is in part responsible for the difference in behavioral performance between the infants and the musically untrained adults. Nevertheless, musically trained adults (who were instructed in the same way as the musically untrained adults) did perform at levels similar to the infants.

There are several limitations due to the design of the study that limit our comparisons of F0 and CF discrimination of infants and adults beyond group level probabilities. As we designed the study to detect whether participants in a given age group can or cannot discriminate a given change in F0 or spectral centroid, it is possible that there are more subtle differences in how well 3- and 7-month-olds can discriminate pitch. However, our study was not sufficiently powered to detect such subtle differences. The use of a pass criterion of at least an 80 % hit rate with a false-alarm rate of at most 20 % on the last 5 change trials and the last 5 no-change trials to obtain a broad measure of whether or not a listener can perceive a given change also allows for some difference in d' to exist between listeners who reach criterion (i.e., there are participants who reach criterion with a d' of greater than 1.68). Moreover, listeners did not hear the same fixed number of trials, with poorer performers given more opportunities (i.e., more trials) to meet the criterion than good performers, who met the criterion after the minimum number of trials.

In terms of stimuli, the filter slope was 12 dB per octave in experiment 1 versus 24 dB per octave in experiment 2; the effect of this difference on infant F0 and CF discrimination is unknown. In experiment 2, we ruled out the possibility that infants are unable to perceive changes in CF; however, it is also possible that infants perceive changes in spectral centroid but perceptually weight CF changes less than F0 changes (e.g., prioritize pitch over timbre) in comparison to adult listeners without musical training. Finally, in experiment 1, we recruited two groups of adults — those with musical training and those without. It is possible that innate differences in auditory or cognitive difference exist between these groups that go beyond the presence or absence of musical training (Corrigall et al. 2013; McKay 2021).

In studies of speech perception, differential sensitivity to native and non-native phonetic contrasts has been shown for adults but not young infants (Werker and Tees 1984). Although there are non-native phonemes that infants can discriminate but adults cannot, adults

generally outperform infants on native language speech discrimination (Kuhl et al. 2006). In other music-related studies that tested both infants and adults on comparable tasks such as musical structure and rhythm perception, absolute performance was also generally poorer for infants (Trainor and Trehub 1992; Hannon and Trehub 2005). Remarkably, our results suggest the reverse developmental trajectory — that F0 discrimination in the presence of spectral variations, and vice versa, becomes more difficult for adult listeners without formal musical training. This conclusion does not necessarily imply a continuous degradation in F0 discrimination from infancy. Instead, it may be that statistical covariations between F0 and spectrum are learned by 12 months (as are speech and musical regularities, see Werker and Tees 1984; Hannon and Trehub 2005). Thus, perceptual interference between F0 and spectrum may emerge at a relatively early age, degrading performance in older infants (12+ months) on tasks, such as ours, that violate these learned expectations. This prediction remains to be tested. Nevertheless, it is clear from our study that accurate F0 and spectral-peak discrimination can be achieved by 3- and 7-month infants, implying that accurate discrimination of these important auditory dimensions is not dependent on complete auditory cortical maturation, and thus suggesting a possible subcortical substrate for both F0 and spectral coding or that an immature auditory cortex is sufficient for F0 and spectral-peak discrimination.

Funding This work was supported by NIH grants R01 DC00396 and P30 DC04661 to L.A.W.

Declarations

Conflict of Interest The authors declare no competing interests.

REFERENCES

- ALLEN EJ, BURTON PC, MESIK J ET AL (2019) Cortical correlates of attention to auditory features. *J Neurosci* 39:3292–3300. <https://doi.org/10.1523/JNEUROSCI.0588-18.2019>
- ALLEN EJ, BURTON PC, OLMAN CA, OXENHAM AJ (2017) Representations of pitch and timbre variation in human auditory cortex. *J Neurosci* 37:1284–1293. <https://doi.org/10.1523/JNEUROSCI.2336-16.2016>
- ALLEN EJ, OXENHAM AJ (2014) Symmetric interactions and interference between pitch and timbre. *J Acoust Soc Am* 135:1371–1379. <https://doi.org/10.1121/1.4863269>
- ANSI (2013) American national standard acoustical terminology. American National Standards Institute, New York
- BENASICH AA, CHOUDHURY NA, REALPE-BONILLA T, ROESLER CP (2014) Plasticity in developing brain: active auditory exposure impacts prelinguistic acoustic mapping. *J Neurosci* 34:13349–13363. <https://doi.org/10.1523/JNEUROSCI.0972-14.2014>
- BENDOR D, WANG X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161–1165. <https://doi.org/10.1038/nature03867>
- BIRNHOLZ JC, BENACERRAF BR (1983) The development of human fetal hearing. *Science* 222:516–519
- BORCHERT EMO, MICHEYL C, OXENHAM AJ (2011) Perceptual grouping affects pitch judgments across time and frequency. *J Exp Psychol Hum Percept Perform* 37:257–269. <https://doi.org/10.1037/a0020670>
- BUSS E, TAYLOR CN, LEIBOLD LJ (2014) Factors affecting sensitivity to frequency change in school-age children and adults. *J Speech Lang Hear Res* 57:1972–1982. https://doi.org/10.1044/2014_JSLHR-H-13-0254
- CLARKSON MG (1996) Infants' intensity discrimination: spectral profiles. *Infant Behav Dev* 19:181–190. [https://doi.org/10.1016/S0163-6383\(96\)90017-X](https://doi.org/10.1016/S0163-6383(96)90017-X)
- CLARKSON MG, CLIFTON RK, PERRIS EE (1988) Infant timbre perception: discrimination of spectral envelopes. *Percept Psychophys* 43:15–20. <https://doi.org/10.3758/BF03208968>
- CORRIGALL KA, SCHELLENBERG EG, MISURA NM (2013) Music training, cognition, and personality. *Front Psychol* 4. <https://doi.org/10.3389/fpsyg.2013.00222>
- EGGERMONT JJ, MOORE JK (2012) Morphological and functional development of the auditory nervous system. In: Werner L, Fay RR, Popper AN (eds) *Human Auditory Development*. Springer, New York, pp 61–105
- GREEN DM, SWETS JA (1966) *Signal detection theory and psychophysics*. Wiley, New York
- HANNON EE, TREHUB SE (2005) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55. <https://doi.org/10.1111/j.0956-7976.2005.00779.x>
- HE C, TRAINOR LJ (2009) Finding the pitch of the missing fundamental in infants. *J Neurosci* 29:7718–8822. <https://doi.org/10.1523/JNEUROSCI.0157-09.2009>
- HORN DL, WON JH, RUBINSTEIN JT, WERNER LA (2017) Spectral ripple discrimination in normal hearing infants. *Ear Hear* 38:212–222. <https://doi.org/10.1097/AUD.0000000000000373>
- HOUTSMA AJM, SMURZYNSKI J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304–310. <https://doi.org/10.1121/1.399297>
- KARZON RG, NICHOLAS JG (1989) Syllabic pitch perception in 2- to 3-month-old infants. *Percept Psychophys* 45:10–14. <https://doi.org/10.3758/BF03208026>
- KITAHARA T, GOTO M, OKUNO HG (2005) Pitch-dependent identification of musical instrument sounds. *Appl Intell* 23:267–275. <https://doi.org/10.1007/s10489-005-4612-1>
- KOHLRAUSCH A, HOUTSMA AJ (1992) Pitch related to spectral edges of broadband signals. *Philos Trans R Soc Lond, B, Biol Sci* 336:375–381; discussion 381–382. <https://doi.org/10.1098/rstb.1992.0071>
- KUHL PK, STEVENS E, HAYASHI A ET AL (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci* 9:F13–F21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- LAU BK, LALONDE K, OSTER M-M, WERNER LA (2017) Infant pitch perception: missing fundamental melody discrimination. *J Acoust Soc Am* 141:65–72. <https://doi.org/10.1121/1.4973412>
- LAU BK, RUGGLES DR, KATYAL S ET AL (2017) Sustained cortical and subcortical measures of auditory and visual plasticity following short-term perceptual learning. *PLoS ONE* 12. <https://doi.org/10.1371/journal.pone.0168858>
- LAU BK, WERNER LA (2012) Perception of missing fundamental pitch by 3- and 4-month-old human infants. *J Acoust Soc Am* 132:3874–3882. <https://doi.org/10.1121/1.4763991>
- LAU BK, WERNER LA (2014) Perception of the pitch of unresolved harmonics by 3- and 7-month-old human infants. *J Acoust Soc Am* 136:760–767. <https://doi.org/10.1121/1.4887464>

- MADSEN SMK, MARSCHELL M, DAU T, OXENHAM AJ (2019) Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Sci Rep* 9:10404. <https://doi.org/10.1038/s41598-019-46728-1>
- MATTOCK K, MOLNAR M, POLKA L, BURNHAM D (2008) The developmental course of lexical tone perception in the first year of life. *Cognition* 106:1367–1381. <https://doi.org/10.1016/j.cognition.2007.07.002>
- MAXON AB, HOCHBERG I (1982) Development of psychoacoustic behavior: sensitivity and discrimination. *Ear Hear* 3:301
- McDERMOTT JH, LEHR AJ, OXENHAM AJ (2008) Is relative pitch specific to pitch? *Psychol Sci* 19:1263–1271. <https://doi.org/10.1111/j.1467-9280.2008.02235.x>
- McKAY CM (2021) No evidence that music training benefits speech perception in hearing-impaired listeners: a systematic review. *Trends in Hearing* 25:2331216520985678. <https://doi.org/10.1177/2331216520985678>
- MICHEYL C, DELHOMMEAU K, PERROT X, OXENHAM AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219:36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- MICHEYL C, OXENHAM AJ (2004) Sequential F0 comparisons between resolved and unresolved harmonics: no evidence for translation noise between two pitch mechanisms. *J Acoust Soc Am* 116:3038–3050. <https://doi.org/10.1121/1.1806825>
- MICHEYL C, RYAN CM, OXENHAM AJ (2012) Further evidence that fundamental-frequency difference limens measure pitch discrimination. *J Acoust Soc Am* 131:3989–4001. <https://doi.org/10.1121/1.3699253>
- MONTGOMERY CR, CLARKSON MG (1997) Infants' pitch perception: masking by low- and high-frequency noises. *J Acoust Soc Am* 102:3665–3672. <https://doi.org/10.1121/1.420153>
- MOORE BCJ, GLASBERG BR (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J Acoust Soc Am* 87:2163–2177. <https://doi.org/10.1121/1.399184>
- MOORE GA, MOORE BCJ (2003) Perception of the low pitch of frequency-shifted complexes. *J Acoust Soc Am* 113:977–985. <https://doi.org/10.1121/1.1536631>
- MOORE JK, GUAN YL (2001) Cytoarchitectural and axonal maturation in human auditory cortex. *J Assoc Res Otolaryngol* 2:297–311
- NORMAN-HAIGNERE S, KANWISHER N, McDERMOTT JH (2013) Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci* 33:19451–19469. <https://doi.org/10.1523/JNEUROSCI.2880-13.2013>
- OLSHO LW (1984) Infant frequency discrimination. *Infant Behav Dev* 7:27–35. [https://doi.org/10.1016/S0163-6383\(84\)80020-X](https://doi.org/10.1016/S0163-6383(84)80020-X)
- OXENHAM AJ, MICHEYL C, KEEBLER MV ET AL (2011) Pitch perception beyond the traditional existence region of pitch. *PNAS* 108:7629–7634. <https://doi.org/10.1073/pnas.1015291108>
- PENAGOS H, MELCHER JR, OXENHAM AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24:6810–6815. <https://doi.org/10.1523/JNEUROSCI.0383-04.2004>
- PLANTINGA J, TRAINOR IJ (2009) Melody recognition by two-month-old infants. *J Acoust Soc Am* 125:EL58–EL62. <https://doi.org/10.1121/1.3049583>
- PLOMP R, STEENEKEN HJM (1969) Effect of phase on the timbre of complex tones. *J Acoust Soc Am* 46:409–421. <https://doi.org/10.1121/1.1911705>
- ROSNER B (1995) *Fundamentals of biostatistics*, 4th edn. Duxbury Press, Belmont, CA
- SHOFNER WP, CHANEY M (2013) Processing pitch in a nonhuman mammal (*Chinchilla laniger*). *J Comp Psychol* 127:142–153. <https://doi.org/10.1037/a0029734>
- SONG X, OSMANSKI MS, GUO Y, WANG X (2016) Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc Natl Acad Sci* 113:781–786. <https://doi.org/10.1073/pnas.1516120113>
- STILP CE, KLUENDER KR (2012) Efficient coding and statistically optimal weighting of covariance among acoustic attributes in novel sounds. *PLoS ONE* 7:e30845. <https://doi.org/10.1371/journal.pone.0030845>
- STILP CE, ROGERS TT, KLUENDER KR (2010) Rapid efficient coding of correlated complex acoustic properties. *Proc Natl Acad Sci USA* 107:21914–21919. <https://doi.org/10.1073/pnas.1009020107>
- TELLER DY (1979) The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Infant Behav Dev* 2:135–153. [https://doi.org/10.1016/S0163-6383\(79\)80016-8](https://doi.org/10.1016/S0163-6383(79)80016-8)
- TRAINOR IJ, TREHUB SE (1992) A comparison of infants' and adults' sensitivity to Western musical structure. *J Exp Psychol Hum Percept Perform* 18:394. <https://doi.org/10.1037/0096-1523.18.2.394>
- TREHUB SE, THORPE LA, MORRONGIELLO BA (1985) Infants' perception of melodies: changes in a single tone. *Infant Behav Dev* 8:213–223. [https://doi.org/10.1016/S0163-6383\(85\)80007-2](https://doi.org/10.1016/S0163-6383(85)80007-2)
- WALKER KM, GONZALEZ R, KANG JZ, ET AL (2019) Across-species differences in pitch perception are consistent with differences in cochlear filtering. *Elife* 8. <https://doi.org/10.7554/eLife.41626>
- WERKER JF, TEES RC (1984) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7:49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- WERNER LA (1995) Observer-based approaches to human infant psychoacoustics. In: Klump GM, Dooling RJ, Fay RR, Stebbins WC (eds) *Methods in Comparative Psychoacoustics*. Birkhäuser Basel, pp 135–146.
- WHALEN DH, LEVITT AG (1995) The universality of intrinsic F0 of vowels. *J Phon* 23:349–366. [https://doi.org/10.1016/S0095-4470\(95\)80165-0](https://doi.org/10.1016/S0095-4470(95)80165-0)
- ZATORRE RJ, EVANS AC, MEYER E, GJEDDE A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849. <https://doi.org/10.1126/science.256.5038.846>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.