Published in final edited form as:
Ear Hear. 2019 ; 40(4): 927–937. doi:10.1097/AUD.0000000000000673.

Developmental effects in children's ability to benefit from F0 differences between target and masker speech

Mary M. Flaherty[1], **Emily Buss**[2], and **Lori J. Leibold**[1]

[1]Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska, USA.

[2]Department of Otolaryngology/Head and Neck Surgery, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA.

Abstract

Objectives: The objectives of this study were to (1) evaluate the extent to which school-age children benefit from fundamental frequency (F0) differences between target words and competing two-talker speech, and (2) assess whether this benefit changes with age. It was predicted that while children would be more susceptible to speech-in-speech masking compared to adults, they would benefit from differences in F0 between target and masker speech. A second experiment was conducted to evaluate the relationship between frequency discrimination thresholds and the ability to benefit from target/masker differences in F0.

Design: Listeners were children (5–15 years) and adults (20–36 years) with normal hearing. In the first experiment, speech recognition thresholds (SRTs) for disyllabic words were measured in a continuous, 60-dB-SPL two-talker speech masker. The same male talker produced both the target and masker speech (average F0 = 120 Hz). The level of the target words was adaptively varied to estimate the level associated with 71% correct identification. The procedure was a four-alternative forced-choice with a picture-pointing response. Target words either had the same mean F0 as the masker or it was shifted up by 3, 6 or 9 semitones. To determine the benefit of target/masker F0 separation on word recognition, masking release was computed by subtracting thresholds in each shifted-F0 condition from the threshold in the unshifted-F0 condition. In the second experiment, frequency discrimination thresholds were collected for a subset of listeners to determine whether sensitivity to F0 differences would be predictive of SRTs. The standard was the syllable /ba/ with an F0 of 250 Hz; the target stimuli had a higher F0. Discrimination thresholds were measured using a three-alternative, three-interval forced choice procedure.

Results: Younger children (5–12 years) had significantly poorer SRTs than older children (13–15 years) and adults in the unshifted-F0 condition. The benefit of F0 separations generally increased with increasing child age and magnitude of target/masker F0 separation. For 5- to 7-year-olds, there was a small benefit of F0 separation in the 9-semitone condition only. For 8- to 12-year-olds, there was a benefit from both 6- and 9-semitone separations, but to a lesser degree than what was observed for older children (13–15 years) and adults, who showed a substantial benefit in the 6-

Address correspondence to Mary M. Flaherty, Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, 901 South Sixth Street, Champaign, IL 61820, USA. maryflah@illinois.edu.

CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

and 9-semitone conditions. Examination of individual data found that children younger than 7 years of age did not benefit from any of the F0 separations tested. Results for the frequency discrimination task indicated that, while there was a trend for improved thresholds with increasing age, these thresholds were not predictive of the ability to use F0 differences in the speech-in-speech recognition task after controlling for age.

Conclusions: The overall pattern of results suggests that children's ability to benefit from F0 differences in speech-in-speech recognition follows a prolonged developmental trajectory. Younger children are less able to capitalize on differences in F0 between target and masker speech. The extent to which individual children benefitted from target/masker F0 differences was not associated with their frequency discrimination thresholds.

INTRODUCTION

Children struggle more than adults when listening to speech in the context of other talkers (e.g., Hall et al. 2002; Corbin et al. 2016). An immature ability to segregate speech streams and selectively attend to the target speech is thought to limit speech-in-speech recognition in young children (e.g., Werner 2012). In adults, these processes appear to be aided by utilizing perceptually salient acoustic differences between target and masker speech. Differences that adults appear to rely on to separate multiple talkers include cues related to spatial location, sentence onset asynchrony, and voice characteristics associated with talker sex (e.g., Brungart 2001; Freyman et al. 2001; Lee & Humes 2012). One voice characteristic that is particularly beneficial for speech segregation in adults is a difference in voice pitch, determined by the fundamental frequency (F0) of the talker's vocal fold vibrations (Darwin et al. 2003; Mackersie et al. 2011). Adults are able to take advantage of F0 differences to separate targets and maskers of the same sex as well as targets that differ in sex. To more closely examine the development of children's ability to benefit from talker voice differences in the context of speech-in-speech recognition, the current study investigated the extent to which children across a wide range of ages benefit from target/masker differences in F0.

Despite having a functionally mature peripheral auditory system by 6 months of age (e.g., Werner 2012), children are more susceptible than adults to interference from competing sounds. Compared to adults, children require a more advantageous signal-to-noise ratio (SNR) to achieve similar levels of performance on speech detection or recognition in the presence of competing noise or speech maskers (e.g., Nittrouer & Boothroyd 1990; Hall et al. 2002; Litovsky 2005; Wightman & Kistler 2005). Poorer speech understanding in a speech masker is thought to reflect children's immature sound segregation abilities as well as a reduced ability to understand speech based on sparse glimpses relative to adults (Buss et al. 2017b). Speech-in-speech recognition performance does not appear to reach maturity until the teenage years (Wightman et al. 2003; Wightman & Kistler, 2005; Corbin et al. 2016).

Given the pronounced developmental effects observed for speech-in-speech recognition, there is considerable interest in better understanding how children differ from adults in their ability to use acoustic differences between talkers to segregate and selectively attend to a target speech stream. Children benefit from some, but not all, of the acoustic cues shown to improve adults' performance in competing speech (Litovsky 2005; Yuen & Yuan 2014; Calandruccio et al. 2016; Corbin et al. 2017). For adults, voice differences between talkers of the same sex can improve intelligibility of the target speech by more than 20 percentage points; voice differences between talkers of different sexes can increase intelligibility by 50 percentage points (Festen & Plomp 1990; Brungart 2001). Whether children are able to take advantage of a sex mismatch between target and masker speech appears to depend on several factors, including listener age. Children 16 months of age and younger do not show a benefit of a target/masker sex mismatch (Leibold et al. 2013; Newman & Morini 2017), but recent evidence suggests that 30-months-olds do show improved word recognition when the target voice is female and the masker voice is male compared to when the masker is female (Newman & Morini 2017). Using a Coordinate Response Measure (CRM; Bolia et al. 2000) task, Wightman and Kistler (2005) investigated 4- to 16-year old's ability to use a sex mismatch. The CRM uses sentences of the form: "Ready <call sign> go to <color> <number> now". Listeners are asked to report the color and number named by the talker who used a specified call sign. In Wightman and Kistler (2005), the target phrase was always spoken by a male talker, and distracter phrases were spoken by a different talker that was either male or female. Children as young as 4 years of age benefitted from a sex mismatch between target and masker speech, although performance did not reach adult-like levels, even in the oldest children (aged 16 years). Using a four-alternative, forced-choice (4AFC) picture-pointing word recognition procedure, Leibold et al. (2018) examined the effects of a sex mismatch using a two-talker masker and found a similar benefit for 5- to 10- year old children and adults.

There are a variety of differences in the acoustic signal that can arise due to differences between male and female talkers' anatomy and physiology. The relative spacing of the formants in vowel production, for example, is determined by the overall length of the vocal tract (VTL) and the size of the mouth cavity relative to the pharyngeal cavity (Chiba & Kajiyama 1941; Fant 1960). In addition, there are physical differences in the structure of the larynx that can contribute to differences in source characteristics, such as F0, spectral tilt, and breathiness (Klatt & Klatt 1990). Two primary voice differences between male and female talkers are F0 and formant frequency dispersion. Males have a somewhat longer vocal tract and tend to have a lower F0 and lower formant frequencies compared to females. Male voices are typically just under one octave (12 semitones) lower in F0 than female voices (Peterson & Barney 1952). Differences in F0 or VTL between talkers of the same sex can improve adults' sentence recognition performance, but F0 differences tend to improve performance to a greater degree than differences in VTL (Darwin et al. 2003; Mackersie et al. 2011). The present study focused just on F0 and its role in understanding target speech in the presence of competing masker speech. Differences in F0 across talkers can substantially improve intelligibility for both double-vowel stimuli and concurrent sentences for adults (Brokx & Nooteboom 1982; Assmann & Summerfield 1987; Bird & Darwin 1998; Assmann 1999; Baskent & Gaudrain 2016). Sentence recognition performance improves with F0 differences up to 10 to 12 semitones (Bird & Darwin 1998). The benefit of F0 differences for sentence recognition is thought to reflect grouping by harmonic relations (Bregman 1990). That is, harmonics across different frequency regions for a single talker are grouped together based on common F0, which in turn assists in segregating the target from the competing sentences.

Children's ability to segregate competing speech streams based solely on F0 differences between talkers has not been directly studied in the context of speech-in-speech recognition, but there are limited data showing that the ability to utilize frequency differences to segregate tones remains immature into the school-age years. Sussman et al. (2007) tested the ability of 5- to 11-year olds to segregate two streams composed of brief (20 ms) tones differing in frequency. Children were capable of segregating the streams based on frequency differences, but they needed a larger frequency separation than the adults before reporting that they heard two streams. Children ages 9- to 11-years required a 9-semitone separation, while children ages 5- to 8-years required an 11-semitone separation. Adults, by comparison, only required a separation of 7 semitones to achieve the same level of performance. Although frequency differences used to segregate non-speech stimuli presented in quiet is not directly comparable to F0 differences in speech-in-speech recognition, these results suggest that pitch-based segregation is still developing in school-age children.

The ability to discriminate pitch does not necessarily predict the ability to utilize it in the context of competing speech. Infants as young as 3 months can discriminate complex tones on the basis of their pitch (Lau & Werner 2012, 2014; Lau et al. 2017), and 4- to 6-month-olds can discriminate voices of the same or different sex (Masapollo et al. 2016). However, infants as old as 16 months of age do not appear to be able to use this information to detect speech in the presence of competing sounds (Leibold et al. 2013; Newman & Morini 2017). Results from infant studies provide information for listeners who are at one end of the developmental continuum, while adults are at the other end. The present study evaluates performance in school-age children, to better understand the entire developmental progression of these abilities.

The purpose of the current study was to determine whether children between 5 and 15 years of age benefit from differences in ð between target words and competing two-talker speech for speech-in-speech recognition, and to evaluate whether this benefit changes with age. Experiment 1 tested whether F0 separations of 3, 6, and 9 semitones between target and masker speech improved word recognition. The F0 separations used were chosen to cover a range of F0 differences that may occur between talkers, from small (3 semitones) to large (9 semitones, which corresponds to the average male/female F0 difference). Adults were also tested to provide an estimate of mature performance. It was hypothesized that, while children would be more susceptible to speech-in-speech masking than adults, they would benefit from differences in F0 between target and masker speech. This hypothesis is based on results demonstrating children's ability to benefit from a sex mismatch between talkers (Wightman & Kistler 2005; Newman & Morini 2017; Leibold et al. 2018). If children do not benefit from F0 differences, that suggests that children may rely on redundant acoustic information to benefit from talker sex mismatches (combined effects of VTL and F0 differences). Experiment 2 was conducted to evaluate the relationship between frequency discrimination thresholds and the ability to benefit from target/masker differences in F0. It was hypothesized that individual differences in children's F0 discrimination thresholds would be predictive of their ability to benefit from F0 differences for speech-in-speech recognition, such that children with poorer discrimination would be less likely to benefit from differences in F0, and children with better discrimination would be more likely to benefit.

Experiment 1 PARTICIPANTS AND METHODS

Listeners

Listeners were 51 children (5.0–15.9 years; 27 females and 24 males) and 18 adults (18.8–36.3 years; 13 females and 5 males). Three age groups of children were recruited: (1) 17 children ages 5 to 7 years, including 7 females and 10 males, (2) 19 children ages 8 to 12 years, including 13 females and 6 males, and (3) 15 children ages 13 to 15 years, including 7 females and 8 males. These age groups were chosen based on studies showing development of speech-in-speech recognition within and across these age ranges (Leibold & Buss 2013; Corbin et al. 2016; Leibold et al. 2016). The youngest age group (5- to 7-year-olds) was chosen based on data showing greater speech-on-speech masking for children younger than 7 years compared to older children (e.g., Leibold & Buss 2013). Inclusion of 8- to 12-year-olds was based on observations of immature behavior persisting into adolescence for masked speech perception (e.g., Wightman & Kistler 2005). Inclusion of 13- to 15-year-olds was based on previous data showing rapid maturation in speech-on-speech recognition around 13 years of age (Corbin et al. 2016). Adults were tested using the same stimuli and procedures to provide an estimate of mature performance. All listeners were native speakers of American English and had normal hearing, with thresholds ≤ 20 dB HL for octave frequencies between 250 and 8000 Hz (ANSI 2010). Listeners provided written informed consent and were paid for their participation.

Stimuli and conditions

Target words and masker speech were both recorded from the same male talker. He was a native English speaker, aged 44 years, with a standard American-English dialect. The speech was recorded in a sound-treated booth using a condenser microphone (Shure-KSM42) mounted approximately 6 inches from the talker’s mouth. The target tokens were recordings of 30 disyllabic words that were unambiguously represented by pictures (Calandruccio et al. 2014; see Figure 1 for the list of words and sample pictures). The words were all meaningful nouns that were in the vocabulary of a typical 5-year-old. Stimulus details regarding word frequency, number of phonological neighbors, phonetic content, and word familiarity can be found in Calandruccio et al. (2014). Token durations ranged from 323 to 597 ms (mean = 439 ms), and the mean F0 was 120 Hz (SD = 8 Hz). Prior to the experiment, the 30 disyllabic tokens were scaled to have equal total rms levels and resampled at a rate of 24.414 kHz using MATLAB.

The masker was two-talker speech composed of sentences from the Revised Bamford-Kowal-Bench (BKB) Standard Sentence Test (Bench et al. 1979). The BKB corpus includes 21 lists of 16 sentences each. Lists 1–6 were used for one masker stream and lists 7–12 were used for the other. The individual masker streams were manually edited to reduce silent pauses longer than 200 ms, resulting in samples that were 2.67 and 2.71 minutes in duration.

The two masker speech streams were balanced for overall rms level and then mixed to create the two-male-talker masker, which was played continuously during testing. The mean masker F0 was 124 Hz (SD = 28 Hz). The level of the two-talker speech masker was fixed at 60 dB SPL.

There were four target conditions. The mean F0 of the target was either matched to the masker's mean F0 (i.e., unshifted), or shifted up by 3, 6, or 9 semitones. The F0-shifted words were resynthesized from the original productions using the pitch-synchronous overlap-add (PSOLA) algorithm (Moulines & Charpentier 1990) implemented in Praat (Boersma & Weenink 2016). The target F0 was incremented relative to the masker based on pilot data and published results which tend to show a larger effect for increments in F0 than decrements in F0 (e.g., Assmann 1999; Mackersie et al. 2011). The magnitudes of the F0 shifts were chosen based on previous data showing that adults can benefit from semitone differences in this range (Summers & Leek 1998; Mackersie et al. 2011), and to evaluate age effects in the degree of target/masker F0 separation required to improved speech-in-speech recognition performance. A 9-semitone separation also corresponds to the average male/ female F0 difference (Peterson & Barney, 1952). All items were verified for clarity before being used in the experiment.

A custom MATLAB script controlled the selection and presentation of the stimuli. The target and masker stimuli were mixed (Avid Technology Inc., Pro Tools Fast Track Solo, Burlington, MA), amplified (Applied Research and Technology, SLA4, Niagara Falls, NY), and presented through a loudspeaker (JBL-1, Northridge, CA). During testing the listener was positioned approximately 1 m in front of and facing the loudspeaker in the sound field of a 7 X 8 ft, sound-treated booth.

Procedure

Each listener completed testing in all four conditions. An adaptive 4AFC disyllabic word identification task was used. Listeners held a touchscreen (iPad Mini 2, Apple, Cupertino, California) during testing. On each trial, one of the 30 disyllabic words was randomly selected as the target. Four different pictures were displayed on the touchscreen in black and white approximately simultaneously with acoustic presentation of the selected target word. Each picture was randomly assigned to appear in one quadrant of the touchscreen. One of the four pictures corresponded to the target word. The other three pictures were drawn without replacement from the 29 remaining possibilities, with the caveat that pictures were not repeated in sequential trials. After the target word was presented, the pictures turned from black and white to color (see Figure 1, panel B). Listeners were instructed to touch the picture that corresponded to the word they heard. After each response, correct-answer feedback was provided on the touchscreen by displaying the picture corresponding to the target word in isolation.

Each listener completed a familiarization phase in quiet prior to testing, in which they listened to and identified each of the disyllabic words by pointing to the associated picture in a laminated book. Following this familiarization phase, each listener completed a practice phase to ensure task comprehension, verify familiarity with the response illustrations, and provide practice inputting responses on the touchscreen. The practice phase used the unshifted target words presented in the two-talker speech masker, with the target presented at 10 dB SNR (70 dB SPL). Listeners were required to correctly identify 8 of the last 10 target words before they could go on to testing. All participants successfully completed the practice phase within 8–10 trials.

Speech reception thresholds (SRTs) were measured by adjusting the level of the target using a two-down, one-up tracking procedure (Levitt 1971), which estimates the level associated with 71% correct identification. The starting level for the target word was approximately 10 dB above the expected SRT for each condition, adjusted for individual listeners. The initial step size for the adaptive track was 4 dB, reduced to 2 dB after the second reversal. Runs were stopped after eight reversals, and the average target level at the final six reversals was used to estimate the SRT for each condition. At least two runs were completed for each target condition. A third run was obtained if the first two estimates differed by 6 dB or more. A third estimate was obtained for four children in the unshifted condition, eight children in the 3-semitone condition, nine children in the 6-semitone condition, and ten children in the 9-semitone condition. A third estimate was obtained for three adults in the unshifted condition, two adults in the 3-semitone condition, six adults in the 6-semitone condition, and four adults in the 9-semitone condition. For listeners requiring an additional run, the final SRTs reported below correspond to the average of the two most similar SRTs.

The order of the four conditions was counterbalanced across listeners; listeners completed all runs in a condition before moving on to the next one. Informed consent, hearing screening, and testing were completed in a single 1-hr visit to the laboratory.

RESULTS

Age effects in the unshifted-F0 condition

The data were first analyzed by comparing performance in the unshifted-F0 condition across the four age groups of listeners: 5 to 7 years (n = 17), 8 to 12 years (n = 19), 13 to 15 years (n = 15), and adults (n = 18). Estimates of the masked SRT for each listener in the unshifted-F0 condition are shown in Figure 2. Higher SRTs indicate poorer performance than lower SRTs.

On average, the two younger age groups (5- to 7-year-olds and 8- to 12-year-olds) required a more advantageous SNR than older children (13- to 15-year-olds) and adults to achieve 71% correct recognition of target words in the unshifted-F0 condition. SRTs were similar for adults and 13- to 15-year-olds. Mean SRTs were −3.8 dB SNR (SD = 2.2) for the adults and −1.9 dB SNR (SD = 2.0) for the 13- to 15-year-olds. In contrast, mean SRTs were 1.0 dB SNR (SD = 1.7) for 8- to 12-year-olds, and 3.5 dB SNR (SD = 2.3) for 5- to 7-year-olds, resulting in child-adult differences of 4.8 and 7.3 dB, respectively. A one-way analysis-of-variance (ANOVA) comparing masked SRTs in the unshifted-F0 condition found a significant difference in SRTs between the four age groups $[F(3, 65) = 42.69, p < 0.001, \eta_p^2 = 0.66]$. *Post hoc* comparisons (Bonferroni adjusted) confirmed that 5- to 7-year-olds had a higher average baseline SRT in the unshifted-F0 condition than all other age groups, and that SRTs for 8- to 12-year-olds were higher than those for 13- to

15-year-olds and adults $(p < 0.01)$. Masked SRTs did not differ significantly between the 13- to 15-year-olds and the adults (p = 0.057). A linear regression model, using SRT as the dependent variable, demonstrates a strong negative association between listener age and SRT in the unshifted condition (r^2= 0.52, p < 0.001). Including listener sex in the model indicates no significant effect (p = 0.385). Individual differences in SRTs were consistent with what has previously been reported in studies investigating children's speech-in-speech recognition abilities (Bonino et al. 2013; Wightman et al. 2003; Wightman & Kistler 2005; Wightman et al. 2006) [1]

Benefit of F0 separation

To determine the benefit of target/masker F0 separation on word recognition, release from masking was computed by subtracting SRTs in each of the shifted-F0 conditions from the SRT in the unshifted-F0 condition. Estimates of the group average release from masking for target/masker separations of 3, 6, and 9 semitones are shown in Figure 3. Estimates of average release from masking are provided for 5- to 7-year-old children (squares), 8- to 12-year-old children (triangles), 13- to 15-year-old children (diamonds), and adults (circles). Error bars indicate ±1 standard error of the mean.

The average masking release tended to increase with listener age and F0 separation. Adults and 13- to 15-year-olds achieved similar benefits from F0 separation, with a maximum benefit in the 9-semitone separation of 10.8 dB (SD = 4.8) and 8.7 dB (SD = 3.8), respectively. The 8- to 12-year-olds obtained less benefit overall, with only 5.5 dB (SD = 4.6) of benefit for the 9-semitone separation. The 5- to 7-year-olds obtained little or no benefit in any condition, with a mean benefit of only 2.1 dB (SD = 2.2) for the 9-semitone condition.

A repeated-measures ANOVA was conducted to evaluate the statistical significance of the trends observed in Figure 3. This analysis evaluated release from masking using the within-subjects factor of F0-separation condition (3 semitones, 6 semitones, and 9 semitones) and the between-subjects factor of Age Group (5- to 7-year-olds, 8- to 12-year-olds, 13- to 15-year-olds, and adults). There was a significant main effect of F0-separation condition on masking release, $[F(2, 130) = 85.92, p < 0.001, \eta_p^2 = 0.57]$, and a significant main effect of Age Group, $[F(3, 65) = 15.58, p < 0.001, \eta_p^2 = 0.42]$. The interaction between condition and Age Group was also significant, $[F(6, 130) = 8.38, p < 0.001, \eta_p^2 = 0.28]$.Separate tests of the simple main effect of F0-separation condition within each age group were conducted, using one-

[1] One potential question that arose during the review process was whether there were any perceptual differences introduced during the processing and resynthesis of the shifted stimuli that were not present in the unshifted stimuli. Unlike the target stimuli in the 3-, 6- , and 9-semitone conditions, the targets in the unshifted-F0 condition were not processed using the PSOLA algorithm. To probe whether signal processing could explain differences in SRTs observed between the unshifted stimuli and the shifted stimuli, two adults (ages 28.0 and 30.3 years) and two teenagers (ages 13.3 and 15.8 years) were tested on the unprocessed, unshifted target stimuli (used in Experiment 1) and on a processed version of the unshifted stimuli. The processed stimuli maintained the F0 of the original stimulus but processed/resynthesized the stimuli through the PSOLA algorithm in Praat. Procedures were the identical to those described for Experiment 1, except that only two conditions (unprocessed/unshifted and processed/unshifted stimuli) were tested, with two runs per condition. Order of conditions was counterbalanced across listeners. There were no differences in SRTs observed between the processed and unprocessed stimuli. The average SRT was -0.7 dB SNR (SD = 0.9) for the unprocessed/unshifted stimuli and -0.6 dB SNR (SD = 1.5) for the processed/unshifted stimuli. This would suggest that signal processing was not responsible for the observed effects of F0-separation in Experiment 1.

way ANOVAs, to further probe this interaction. The results of this follow-up analysis revealed that, for 5- to 7-year-olds, there was an effect of F0-separation condition $[F(2, 32) = 6.36, p < 0.05, \eta_p^2 = 0.28]$. Follow-up pairwise comparisons found that the masking release in the 9-semitone condition was significantly different from masking release in the 3-semitone condition ($p < 0.05$), but there was not a significant difference between the 3- and 6-semitone conditions ($p = 1.0$) or between the 6- and 9-semitone conditions ($p = 0.91$). For 8- to 12-year-olds, there was an effect of F0 separation $[F(2, 36) = 18.99, p < 0.001, \eta_p^2 = 0.51)$. Pairwise comparisons revealed that the amount of masking release in the 9-semitone separation condition was significantly different from both the 3-semitone and 6-semitone conditions ($p < 0.001$). There was no significant difference in masking release between the 3- and 6-semitone conditions ($p = 0.104$). Both the 13- to 15-year-old age group $[F(2, 28) = 25.08, p < 0.001, \eta_p^2 = 0.64]$, and the adults $[F(2, 34) = 40.63, p < 0.001, \eta_p^2 = 0.71]$ showed an effect of F0-separation. In contrast to the younger age groups, the pairwise comparisons revealed that masking release was not significantly different for the 6-semitone and 9-semitone conditions for either 13- to 15-year-olds or adults $(p \geq 0.100)$. Masking release in both the 6- and 9-semitone conditions was significantly different from the 3-semitone condition for the 13- to 15-year-olds ($p < 0.05$) and the adults ($p < 0.001$). Including listener sex in the model resulted in no main effect of sex or interaction with sex ($p \geq 0.214$).

Individual Data

Figure 4 shows amount of masking release for all individuals in the 3-semitone condition (top panel), the 6-semitone condition (middle panel), and 9-semitone condition (bottom panel), plotted as a function of listener age. Individual data are in general agreement with mean results. In the 9-semitone condition, all of the adults showed a benefit of 5 dB or greater. In contrast, no listener younger than 7 years of age showed an improvement > 5 dB for any target/masker F0 separation. A 3-semitone separation had little effect on SRTs for children of any age, while introducing a 6- or 9-semitone difference improved SRTs, often dramatically for older children and adults. The majority of listeners in the 13- to 15-year old age group demonstrated a benefit greater than 5 dB in either the 6- or 9-semitone condition.

DISCUSSION

The purpose of Experiment 1 was to determine the extent to which children benefit from F0 differences between target and masker speech in the context of speech-in-speech recognition, and whether this ability develops during childhood. The pattern of results observed across listener age for the unshifted-F0 condition is consistent with previous findings that, despite substantial individual differences within and between age groups, young children are generally more susceptible to speech-in-speech masking than older children and adults, even for relatively simple closed-set recognition tasks (Hall et al 2002; Leibold et al. 2013; Buss et al. 2016). Older children (ages 13- to 15-years) and adults had similar thresholds, consistent with prior research demonstrating that susceptibility to masking becomes adult-like in adolescence (Corbin et al. 2015; Leibold & Buss 2013).

When considering masking release in the shifted-F0 conditions, the results indicate that the ability to benefit from target/masker F0 differences was dependent upon listener age. There was a substantial (8.7 dB) improvement in mean SRTs for 13- to 15-year-olds when a 9-semitone difference in F0 was introduced between the target and masker, but the benefit was smaller (5.5 dB) for the 8- to 12-year-old children, and for the 5- to 7-year-olds (2.1 dB). This age effect was unexpected given that previous studies have consistently demonstrated that most children over the age of 4 years are able to benefit to a similar degree as adults from a sex mismatch between target and masker speech (Wightman & Kistler 2005; Leibold et al. 2018).

One potential explanation for young children's inability to benefit from F0 differences in the present study could be related to their sensitivity to pitch differences associated with voice F0. In the absence of other acoustic differences between the target and masker speech, young children could be at a disadvantage due to poorer frequency discrimination abilities relative to older children and adults (Maxon & Hogberg 1982; Jensen & Neff 1993). As children mature, improvements in frequency discrimination could lead to an increased ability to take advantage of F0 differences between target and masker speech.

Results describing children's frequency discrimination abilities are mixed, but most indicate immature frequency discrimination thresholds until about 8 to 12 years of age (Jensen & Neff 1993; Halliday et al. 2008; Buss et al. 2014). For example, pure-tone frequency discrimination thresholds as high as 10% of the standard frequency have been reported for children under 8 years of age, compared to 0.5–1% in adults (Halliday et al. 2008). While some data indicate more adult-like performance in young children when discrimination is measured with harmonic complexes (Deroche et al. 2012, 2014), other data indicate similar frequency discrimination for pure-tone and voice stimuli (Buss et al. 2017a). Although young children's frequency discrimination thresholds tend to be poor relative to adults', sensitivity to a 10% change in frequency would support discrimination of the F0 shifts used in the present experiment (e.g., 3 semitones = 18.8%). However, it is unclear whether discriminability is sufficient to support segregation, in that more supra-threshold pitch differences could provide stronger segregation cues. If that is the case, then poor frequency discrimination for speech in young children could limit their ability to benefit from differences in F0 between target and masker speech.

Experiment 2

Experiment 2 examined whether the pronounced age effects observed in Experiment 1 were the result of development in the ability to discriminate voice pitch. Specifically, this experiment was designed to evaluate whether sensitivity to changes in F0 could account for age effects in children's ability to benefit from F0 differences in the context of speech-in-speech recognition. Stimuli and procedures were closely based on those used by Buss et al. (2017a). It was hypothesized children's F0 discrimination thresholds would be predictive of their ability to benefit from F0 differences for speech-in-speech recognition, controlling for age.

PARTICIPANTS AND METHODS

Listeners

Following collection of the data obtained in Experiment 1, 43 of the 51 child listeners (5.0–15.9 years) participated in a F0-discrimination task: 14 were 5 to 7 years old, 14 were 8 to 12 years old, and 15 were 13 to 15 years old. The remaining listeners did not complete the discrimination task due to scheduling constraints unrelated to individual performance. These measurements were completed during the same 1-hr session as Experiment 1. Testing lasted approximately 10–15 minutes. Due to their uniform ability to benefit from F0-separation in Experiment 1, adults were not included in Experiment 2.

Stimuli and Procedure

Stimuli were based on a recording from a female American English speaker producing the syllable /ba/, with a mean F0 of 243 Hz. The F0 of this sample was modified using the Praat implementation of the PSOLA algorithm so that the standard /ba/ had a monotonized F0 of 250 Hz, and the targets were modified to have a higher F0 than the standard. There were 40 target stimuli generated prior to the experiment, with frequencies ranging from 250.5 to 357.6 Hz, with equal spacing on a log scale. All stimuli were presented at 65 dB SPL.

Discrimination thresholds were collected using a custom Matlab (Mathworks) script. The stimuli were presented through a loudspeaker (JBL-1). During testing the listener was positioned approximately 1 m in front of and facing the loudspeaker in the sound field of a 7 × 8 ft sound treated booth. The experimenter sat in the booth, positioned behind the child. The task was a three-alternative, forced-choice (3AFC). Each trial contained three 500-ms intervals, separated by 500 ms. Two intervals contained the standard stimulus, and one randomly chosen interval contained the target. Intervals were indicated visually on a computer monitor showing three animated frogs. The monitor was positioned directly below the speaker so as not to interfere with the sound field. Each frog opened and closed its mouth, synchronous with the onset and offset of the associated stimulus. Listeners were instructed to listen for the sound that was different from the other two. This task was easily understood by even the youngest listeners. The listener used a handheld laser pointer to select the frog associated with the target and the experimenter entered the response using a computer mouse. After each response, an animation of the target frog catching a fly played, providing immediate feedback.

Thresholds were estimated using a 2-down, 1-up stepping rule, which converges on the frequency difference(Δf) associated with 71% correct performance. Each track began with a (Δf) of 10%, such that the target and standard were 275 and 250 Hz, respectively. At the outset of the track the (Δf) was adjusted by a factor of $2^{0.5}$. This factor was reduced $2^{0.25}$ after the second track reversal. Listeners completed at least two threshold estimation tracks. An additional threshold was collected if the first two spanned a range greater than 5%. The final threshold estimates are reported as the geometric mean of all thresholds obtained from each listener. Analyses that included child age as a continuous variable used log age.

RESULTS

Figure 5 shows discrimination thresholds for individual child listeners, plotted as a function of age. The geometric mean of the F0 discrimination threshold was 1.2 semitones (6.9% of the standard) for 5- to 7-year-olds, 0.4 semitones (2.6% of the standard) for 8- to 12-year olds, and 0.2 semitones (1.2% of the standard) for 13- to 15-year-olds. All but three listeners (ages 5.5, 6.8, and 6.9 years) had discrimination thresholds that corresponded to an F0 sensitivity of less than 3 semitones (18.9%). Thresholds for listeners generally improved as a function of child age. For the following analyses, both thresholds and listener age were transformed to log units, following the approach taken by Buss et al. (2014). Log transformed thresholds were used because log units are thoughts to reflect perceptually equivalent changes in frequency and equalize the variance in pitch discrimination data (Micheyl et al. 2006). The use of log age was based on findings that maturation progresses the most rapidly for younger children compared to older children (Mayer & Dobson 1982; Moller & Robbin 2002). All analyses used a one-tailed significance criterion, unless otherwise noted. When both frequency discrimination threshold and age are represented in log units, the correlation is $r = -0.54$ ($p < 0.001$). A linear regression model evaluating the effects of listener age and sex on frequency discrimination thresholds indicated a significant negative effect of age ($p < 0.001$) but no effect of listener sex ($p = 0.972$).

While there was a trend for greater benefit of F0 separation for masked speech recognition with increasing child age (Experiment 1) and a trend for improved frequency discrimination with increasing child age (Experiment 2), there was not compelling evidence of a strong relationship between frequency discrimination and the ability to benefit from F0 shifts. To assess the relationship between frequency discrimination thresholds and masking release, a linear mixed effects model was conducted. The model included the fixed effects of discrimination thresholds, listener age, and F0-separation condition (3, 6, or 9 semitones), as well as 2- and 3-way interactions between these variables. The model contained a random intercept for each participant, to capture differences in thresholds across participants. Table 1 includes the results from this analysis (Model A, left panel). The main effect of F0-separation condition was significant ($p < 0.001$), as was the interaction between age and condition ($p < 0.001$). Of interest to the present analysis, the main effect of frequency discrimination was not significant ($p = 0.383$), but the interaction between F0-separation condition and discrimination thresholds was significant ($p = 0.042$). No other significant interactions were observed. It is important to note that the significant interaction between F0-separation and frequency discrimination is driven by data of a single listener with particularly poor performance in both tasks (age 6.9 years; masking release in 9-semitone condition = −4.5 dB; discrimination threshold = 26.4%). When data from this listener were excluded, the interaction between frequency discrimination and F0-separation condition was no longer significant, ($p = 0.110$). Statistical results omitting this listener are in Table 1 (Model B, right panel). Overall, this pattern of results suggests that performance on both of the tasks improves with child age, but that frequency discrimination ability does not play a pivotal role in the ability to benefit from supra-threshold shifts in F0 for the masked speech recognition task.

The pattern of individual differences across the whole age range lends further support for the conclusion that F0 mismatch benefit is not limited by frequency discrimination. For example, one of the youngest listeners (5.1 years) had a relatively low frequency discrimination threshold of 0.5 semitones (2.7%). Despite having relatively good frequency discrimination, this listener derived a modest release from masking of 3.5 dB for a 9-semitone F0 separation. In contrast, two older children (11.3 and 12.8 years) with relatively high discrimination thresholds of 11.6% and 11.9% derived benefits of 5.5 and 5.8 dB in the 9-semitone separation condition.

DISCUSSION

The results from Experiment 2 are in general agreement with prior investigations of children's pure-tone frequency discrimination. Young school-age children tended to have higher discrimination thresholds compared to older children and adults, and mean performance tended to improve with increasing age. Adult-like frequency discrimination thresholds are often not seen until about 8–12 years of age (Jensen & Neff 1993; Thompson et al. 1999), and some studies have reported that performance continues to improve past 12 years of age (Maxon & Hochberg 1982). In 5- to 6-year olds, thresholds of 10% are commonly reported for a 1000-Hz standard frequency, compared to thresholds of less than 1% in adults (Halliday et al. 2008). Note, however, that not all studies have observed this age effect. Deroche et al. (2012, 2014) found adult-like performance in children as young as 6 years of age when testing F0 discrimination with 100-Hz and 200-Hz broadband harmonic complexes. The precise age at which children achieve mature frequency discrimination abilities appears to vary depending on the particular task and psychophysical paradigm being used. Results of the present study are consistent with those reported by Buss et al. (2017a), where frequency discrimination thresholds for the /ba/ stimulus improved from 15% at 5 years of age to 2% at 10 years of age. In that study, discrimination thresholds did not significantly differ between the 250-Hz /ba/ stimulus and the 250-Hz pure tone, indicating no advantage for children when the standard stimulus was speech instead of a pure tone. The mean thresholds in both the present study and in Buss et al. (2017) were similar to thresholds that have been reported for a 1000 Hz tone and were slightly higher than thresholds reported for 500 and 5000 Hz tones (Buss et al. 2014).

Regardless of when adult-like frequency discrimination performance is observed, it is clear that even young children can discriminate the relatively large F0 differences used in the present experiment. For all but three listeners (ages 5.5, 6.8, and 6.9 years), discrimination thresholds were well below 3 semitones, with thresholds averaging below 1–2 semitones. This indicates that the ability to discriminate the F0 differences used in Experiment 1 was likely not a limiting factor in children's performance. In addition, if children's ability to use F0 to segregate voices was dependent on their overall sensitivity to F0 differences, then we would expect that children with poorer F0 discrimination thresholds would also have poorer SRTs. Neither statistical analysis nor examination of individual data supported this prediction, as discrimination thresholds in Experiment 2 did not appear to be related to the ability to benefit from target/masker F0 differences in Experiment 1. Despite performance improving with age on both tasks, once age was controlled for, the analysis demonstrated only a weak association between discrimination thresholds and SRTs in any of the F0-

separation conditions, driven by a single listener. Therefore, the age effects observed in Experiment 1 are likely not related to the development of children’s ability to discriminate voice pitch.

A potential limitation that may make it difficult to make comparisons across the two experiments is that the stimuli in Experiment 2 were from a female talker with an F0 of 250 Hz, while the stimuli in Experiment 1 were from a male talker with an F0 of 120 Hz. The stimuli for Experiment 2 was chosen primarily because these stimuli have been used successfully in previous research investigating the development of F0 discrimination abilities (Buss et al. 2017). Despite there being different-sex talkers for the two experiments in the present study, the standard F0 of the female talker in Experiment 1 (250 Hz) was similar to the shifted F0 in the 9-semitone condition of Experiment 1 (210 Hz).

GENERAL DISCUSSION

The results of these two experiments suggest a prolonged time course of development in the ability to benefit from F0 differences between target words and continuous masker speech, at least under conditions in which target/masker F0 differences are provided in isolation. This ability appears to be unrelated to children’s sensitivity to F0 differences, as measured in Experiment 2. Not only are younger children more susceptible than adults to masking associated with competing speech, they also differ in their ability to benefit from F0 differences between the target and masker. This was unexpected given that previous studies have consistently demonstrated that most children over the age of 4 years are able to benefit from a sex mismatch between target and masker speech (Wightman & Kistler 2005; Leibold et al. 2018). It is difficult to directly compare the present results to those of Wightman and Kistler (2005) because that study used a CRM task with only one competing talker. However, we can directly compare the present results with those of Leibold et al. (2018). That study examined the effects of a sex mismatch on word recognition using a picture-pointing task in a two-talker masker, similar to the present study. Results indicated that 5- to 10-year-olds benefitted to a similar degree as adults from a sex mismatch between the target and masker. In contrast, the current results suggest that 5- to 7-year-olds do not have the ability to benefit from F0 separations as large as 9 semitones. This suggests that the sex-mismatch benefit observed for children is not due to differences in F0-separations alone.

One consideration when evaluating the results of the present study is the fact that the cues available to the listener were tightly controlled. Using the same talker for both target and masker speech ensured that stimuli were consistent along dimensions other than mean F0. This does, however, create a contrived situation that unlike what listeners would experience in a natural acoustic environment. Stimuli produced by different talkers could support better performance in younger children.

Another consideration is that the target stimuli were disyllabic words, which lack the contextual cues that would normally be present in running speech. Compared to isolated words, natural sentences provide both acoustic (prosodic) and semantic information that binds auditory cues together (Brokx & Nooteboom 1982; Darwin 1975). There is also evidence that auditory streams build up over time (e.g., Sussman-Fort & Sussman 2014). It is possible that children might derive more benefit from F0 differences for longer duration targets, such as sentences, due to the greater opportunity for the buildup of segregation. However, a sex-mismatch benefit for children has been demonstrated for words rather than sentences (Leibold et al. 2018).

There are two possible explanations for why school-age children benefit to a similar degree as adults from talker sex mismatches but not from manipulation of F0 alone in a word recognition task. One is that children rely on acoustic characteristics that differentiate talker sex other than F0, such as those resulting from differences in VTL. Another is that children rely on a combination of acoustic characteristics, such as those provided by F0 *and* VTL, to capitalize on talker sex mismatches. Male and female voices differ along a number of dimensions, including F0, spectral features associated with VTL, breathiness, and speaking style (Lass et al. 1976; Klatt & Klatt 1990; Mullennix et al. 1995; Smith & Patterson 2005). Manipulating just F0 does not necessarily perception of talker sex. In adults manipulating just the F0 results in less release from masking than manipulating talker sex or manipulating F0 and VTL together (Darwin et al. 2003; Clarke et al. 2014). This suggests that more acoustic detail, such as differences in *both* F0 and vocal tract length, may improve the segregation of competing talkers. Numerous studies have found that children require more acoustic information than adults to achieve similar performance on a variety of speech perception tasks. For example, in speech recognition tasks, children have been shown to require greater temporal and spectral information than adults to recognize speech (Buss et al. 2017b; Elliot et al. 1987; Hazan & Barrett 2000; Mlot et al. 2010). These findings suggest a possible explanation for the large child/adult difference in the present experiment. Young children may struggle to use the F0 cue in isolation, whereas natural talker sex mismatches prove less challenging due to the presence of multiple segregation cues (provided by F0, VTL, and other voice qualities). As children's cognitive processing matures and they gain more experience with speech, they may be less reliant on the provision of multiple cues, which could explain why some 7-year-olds are able to take advantage of relatively larger target/masker F0 differences, but 5- to 6-year-olds are not.

Studies of coherence masking protection (CMP) in children and adults have investigated whether children are obliged to process speech signals as broad spectral patterns (Gordon 1997; Nittrouer & Tarr 2011; Tarr & Nittrouer 2013). These studies involved labeling a phonetically-important low frequency formant (F1) presented in noise either alone or in combination with a higher frequency cosignal (a second and third formant, F2 and F3) which does not provide any additional vowel information and is outside the critical band of the target. When the cosignal and target were both present, labeling accuracy improved for children and adults. However, adults were more likely to benefit from the presence of the cosignal if it was harmonically related to the target. Children, on the other hand, showed large CMP effects regardless of whether the target and cosignal were harmonically related or not. This could suggest that harmonicity is a not salient for children when integrating spectral components of speech. This finding has been interpreted as showing that children pay attention to broad spectral structure instead of the fine spectral detail present in F0 (Nittrouer & Miller, 1997; Nittrouer & Tarr, 2011; Tarr & Nittrouer, 2013).

Work is currently underway to differentiate between these two explanations for children's relative inability to benefit from F0 mismatches in a speech-in-speech task (reliance on converging cues vs. reduced attention to find spectral detail of F0). In either case, the present data suggest that F0 differences between streams of speech from a single talker may provide a weaker segregation cue for children than adults, irrespective of children's sensitivity to F0 differences. One implication of these findings is that the child/adult difference in speech recognition appears to depend critically on the speech materials used to assess performance. Materials that provide robust F0 cues would be expected to maximize the child/adult difference compared to materials with less prominent F0 differences. It is unclear what role children's failure to capitalize on F0 differences plays in the large child/adult difference observed in the literature on speech recognition in a two-talker masker.

Future work using sentence materials will evaluate the role of target duration in development of speech-in-speech recognition. Regardless of potential limitations, the present study demonstrates that unlike adults, young children are unable to take full advantage of F0 differences between target words and competing speech. The current study provides evidence of a prolonged time course of development for the ability to utilize F0 differences for speech-in-speech understanding in the absence of other acoustic differences.

Acknowledgments

FINANCIAL DISCLOSURE:

This research was supported by grants R01 DC011038 and T32 DC 000012 from the National Institute of Deafness and Other Communications Disorders. Subject recruitment efforts were supported by the Center for Perception and Communication in Children (NIGMS P20 DM109023).

REFERENCES

ANSI (2010). ANSI S3.6–2010, American National Standard Specification for Audiometers (American National Standards Institute, New York).

Assmann PF, Summerfield Q (1987). Perceptual segregation of concurrent vowels. J Acoust Soc Am, 82(S1), S120.

Assmann PF (1999). Fundamental frequency and the intelligibility of competing voices. Proceedings of the 14th International Congress of Phonetic Sciences, (i), 179–182.

Başkent D, & Gaudrain E (2016). Musician advantage for speech-on-speech perception. J Acoust Soc Am, 139(3), EL51–EL56. [PubMed: 27036287]

Bench J, Kowal A, Bamford J (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Brit J Audiol, 13(3), 108–112. [PubMed: 486816]

Bird J, Darwin CJ (1998). Effects of a difference in fundamental frequency in separating two sentences In: Palmer AR, Rees A, Summerfield AQ, Meddis R (Eds.), Psychophys Physiol Adv Hear. Whurr, London, pp. 263–269.

Boersma, P., Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.16, retrieved 6 April 2016 from http://www.praat.org/.

Bolia RS, Nelson WT, Ericson MA, et al. (2000). A speech corpus for multitalker communications research. J Acoust Soc Am, 107(2), 1065–1066. [PubMed: 10687719]

Bonino AY, Leibold LJ, Buss E (2013). Release from perceptual masking for children and adults: Benefit of a carrier phrase, Ear Hear 72(2), 181–204.

Bregman AS (1990). Auditory Scene Analysis. Cambridge, MA: MIT Press.

Brokx J, Nooteboom S (1982). Intonation and the perceptual separation of simultaneous voices. J Phonetics, 10, 23–36.

Brungart D (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am, 109(3), 1101–1109. [PubMed: 11303924]

Buss E, Taylor CN, Leibold LJ (2014). Factors affecting sensitivity to frequency change in school-age children and adults. J Speech Lang Hear, 57(5), 1972–1982.

Buss E, Leibold LJ, Hall III JW (2016). Effect of response context and masker type on word recognition in school-age children and adults. J Acoust Soc Am, 140, 968–977. [PubMed: 27586729]

Buss E, Flaherty MM, Leibold LJ (2017a). Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. J Acoust Soc Am, EL150–EL154. [PubMed: 28764444]

Buss E, Leibold LJ, Porter H, et al. (2017b). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of segregation and glimpsing. J Acoust Soc Am, 141(4), 2650–2660. [PubMed: 28464682]

Calandruccio L, Gomez B, Buss E, et al. (2014). Development and preliminary evaluation of a pediatric Spanish-English speech perception task. Am J Audiol, 23(2), 158–72. [PubMed: 24686915]

Calandruccio L, Leibold LJ, Buss E (2016). Linguistic masking release in school-age children and adults. Am J Audiol, 25(1), 34–40. [PubMed: 26974870]

Chiba T, & Kajiyama M (1941). The vowel: Its nature and structure. Tokyo: Kaiseikan.

Clarke J, Gaudrain E, Chatterjee M, & Başkent D (2014). T'ain't the way you say it, it's what you say - Perceptual continuity of voice and top-down restoration of speech. Hear Res, 315, 80–87. [PubMed: 25019356]

Corbin NE, Bonino AY, Buss E, et al. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. Ear Hear, 37(1), 55–63. [PubMed: 26226605]

Corbin NE, Buss E, Leibold LJ (2017). Spatial release from masking in children: Effects of simulated unilateral hearing loss. Ear Hear, 38(2), 223–235. [PubMed: 27787392]

Darwin CJ, Brungart DS, Simpson BD (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J Acoust Soc Am, 114(5), 2913–2922. [PubMed: 14650025]

Darwin CJ, (1975). On the dynamic use of prosody in speech perception In: Cohen A, Nooteboom SG (Eds.), Structure and Process in Speech Perception (pp. 178–194). Berlin: Springer-Verlag.

Deroche MLD, Zion DJ, Schurman JR, et al. (2012). Sensitivity of school-aged children to pitch-related cues. J Acoust Soc Am, 131(4), 2938–2947. [PubMed: 22501071]

Deroche MLD, Lu HP, Limb CJ, et al. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. Fron Neurosci, 8(282), 1–13.

Elliott LL, Hammer MA, Evan KE (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. Percept Psychophys, 42(2), 150–157. [PubMed: 3627935]

Fant G (1960). Acoustic theory of speech production. Hague, The Netherlands: Mouton, pp. 107–135.

Festen JM, Plomp R (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am, 88(4), 1725–1736. [PubMed: 2262629]

Freyman RL, Balakrishnan U, Helfer KS (2001). Spatial release from informational masking in speech recognition. J Acoust Soc Am, 109(5 Pt 1), 2112–2122. [PubMed: 11386563]

Gordon PC (1997). Coherence masking protection in speech sounds: The role of formant synchrony. Percept Psychophys, 59(2), 232–242. [PubMed: 9055618]

Hall JW, Grose JH, Buss E, et al. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear, 23(2), 159–165. [PubMed: 11951851]

Halliday LF, Taylor JL, Edmondson Jones, et al. (2008). Frequency discrimination learning in children. J Acoust Soc Am, 123(6), 4393–4402. [PubMed: 18537390]

Hazan V, Barrett S (2000). The development of phonemic categorization in children aged 6–12. J Phonetics, 28(4), 377–396.

Hillenbrand JM, Getty LA, Clark MJ, et al. (1995). Acoustic characteristics of American English vowels. J Acoust Soc Am, 97(5 Pt 1), 3099–3111. [PubMed: 7759650]

Jensen JK, Neff DL (1993). Development of basic auditory discrimination in preschool children. Psychol Sci, 4(2), 104–107.

Klatt DH, Klatt LC (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am, 87(2), 820–857. [PubMed: 2137837]

Lass NJ, Hughes KR, Bowyer MD, et al. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. J Acoust Soc Am, 59(3), 675–678. [PubMed: 1254794]

Lau BK, Werner LA (2012). Perception of missing fundamental pitch by 3- and 4-month-old human infants. J Acoust Soc Am, 132(6), 3874–82. [PubMed: 23231118]

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(2), 760–767. [PubMed: 25096110]

Lau BK, Lalonde K, Oster M, et al. (2017). Infant pitch perception: Missing fundamental melody discrimination. J Acoust Soc Am, 141(1), 65–72. [PubMed: 28147620]

Lee JH, Humes LE (2012). Effect fundamental-frequency and sentence-onset differences on speech identification performance of young and older adults in a competing-talker background. J Acoust Soc Am, 132(3), 1700–1717. [PubMed: 22978898]

Leibold LJ, Buss E (2013). Children's identification of consonants in a speech-shaped noise and two-talker masker. J Speech Lang Hear Res, 56(4), 1144–1155. [PubMed: 23785181]

Leibold LJ, Taylor CN, Hillock-Dunn A, et al. (2013). Effect of talker sex on infants' detection of spondee words in a two-talker or a speech-shaped noise masker. Proc Meet Acoust, 19, 060074.

Leibold LJ, Buss E, Calandruccio L (2018) Developmental effects in masking release for speech-on-speech perception due to a target/masker sex mismatch. Ear Hear. Advance online publication. DOI: 10.1097/AUD.0000000000000554

Leibold LJ, Bonino AY, Buss E (2016). Masked speech perception thresholds in infants, children, and adults. Ear Hear, 37(3), 345–353. [PubMed: 26783855]

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am, 49(2), 467–477.

Litovsky RY (2005). Speech intelligibility and spatial release from masking in young children. J Acoust Soc Am, 117(5), 3091–3099. [PubMed: 15957777]

Mackersie CL, Dewey J, Guthrie LA (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. J Acoust Soc Am, 130(2), 1006–1019. [PubMed: 21877813]

Masapollo M, Polka L, Ménard L (2016). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. Dev Sci, 19(2), 318–328. [PubMed: 25754812]

Maxon AB, Hochberg I (1982). Development of psychoacoustic behavior: sensitivity and discrimination. Ear Hear, 3(6), 301–308. [PubMed: 7152153]

Mayer DL, Dobson V (1982) Visual acuity development in infants and young children, as assessed by operant preferential looking. Vis Res, 22(9), 1141–1151. [PubMed: 7147725]

Micheyl C, Delhommeau K, Perrot X, Oxenham AJ (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hear Res, 219(1–2), 36–47 [PubMed: 16839723]

Mlot S, Buss E, Hall III JW (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. Ear Hear, 31(1), 56–62. [PubMed: 19816182]

Moller AR, Rollins PR (2002) The non-classical auditory pathways are involved in hearing in children but not in adults. Neurosci Lett, 319(1):41–44. [PubMed: 11814649]

Moulines E, Charpentier F (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun, 9(5–6), 453–467.

Mullennix JW, Johnson KA, Topcu-Durgun M, et al. (1995). The perceptual representation of voice gender. J Acoust Soc Am, 98(6), 3080–3095. [PubMed: 8550934]

Newman RS, Morini G (2017). Effect of the relationship between target and masker sex on infants' recognition of speech. J Acoust Soc Am, 141(2), EL164–EL169. [PubMed: 28253666]

Nittrouer S, Boothroyd A (1990). Context effects in phoneme and word recognition by young children and older adults. J Acoust Soc Am, 87(6), 2705–2715. [PubMed: 2373804]

Nittrouer S, Tarr E (2011). Coherence masking protection for speech in children and adults. Attent Percept Psychophys, 73(8), 2606–2623.

Peterson G, Barney H (1952). Control methods used in a study of the vowels. J Acoust Soc Am, 24(2), 175–184.

Smith DRR, Patterson RD (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. J Acoust Soc Am, 118(5), 3177–3186. [PubMed: 16334696]

Summers V, Leek M (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. J Speech Lang Hear, 41(6), 1294–1306.

Sussman E, Wong R, Horvath J, et al. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. Hear Res, 225(1–2), 117–127. [PubMed: 17300890]

Sussman-Fort J, & Sussman E (2014). The effect of stimulus context on the buildup to stream segregation. Front Neurosci, 8(93), 1–8. [PubMed: 24478622]

Tarr E, Nittrouer S (2013). Explaining coherence in coherence masking protection for adults and children. J Acoust Soc Am, 133(6), 4218–31. [PubMed: 23742373]

Thompson NC, Cranford JL, Hoyer E (1999). Brief-tone frequency discrimination by children. J Speech Lang Hear, 42(5), 1061–1068.

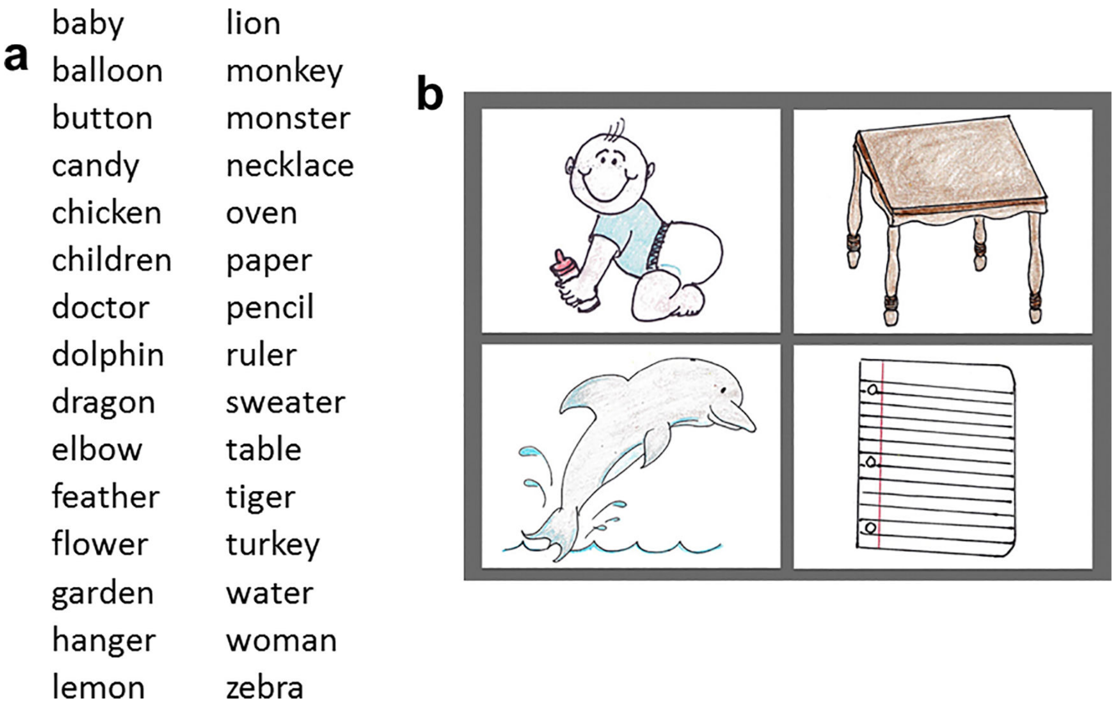
Werner LA, Fay RR, Popper AN (Eds.). (2012). Human Auditory Development (Vol. 42). New York, NY: Springer.

Wightman FL, Callahan MR, Lutfi RA, et al. (2003). Children’s detection of pure-tone signals: Informational masking with contralateral maskers. J Acoust Soc Am, 113(6), 3297–3305. [PubMed: 12822802]

Wightman FL, Kistler DJ (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. J Acoust Soc Am, 118(5), 3164–3176. [PubMed: 16334898]

Wightman FL, Kistler DJ, Brungart D (2006). Informational masking of speech in children: auditory-visual integration. J Acoust Soc Am, 119(6), 3940–3949. [PubMed: 16838537]

Yuen KC, Yuan M (2014). Development of spatial release from masking in Mandarin-speaking children with normal hearing. J Speech Lang Hear, 57(5), 2005–2023.



baby
balloon
button
candy
chicken
children
doctor
dolphin
dragon
elbow
feather
flower
garden
hanger
lemon
lion
monkey
monster
necklace
oven
paper
pencil
ruler
sweater
table
tiger
turkey
water
woman
zebra

Figure 1.

Target word list (panel a) and one sample of the pictures that were used during picture-pointing task to represent the target words (panel b) from Experiment 1.

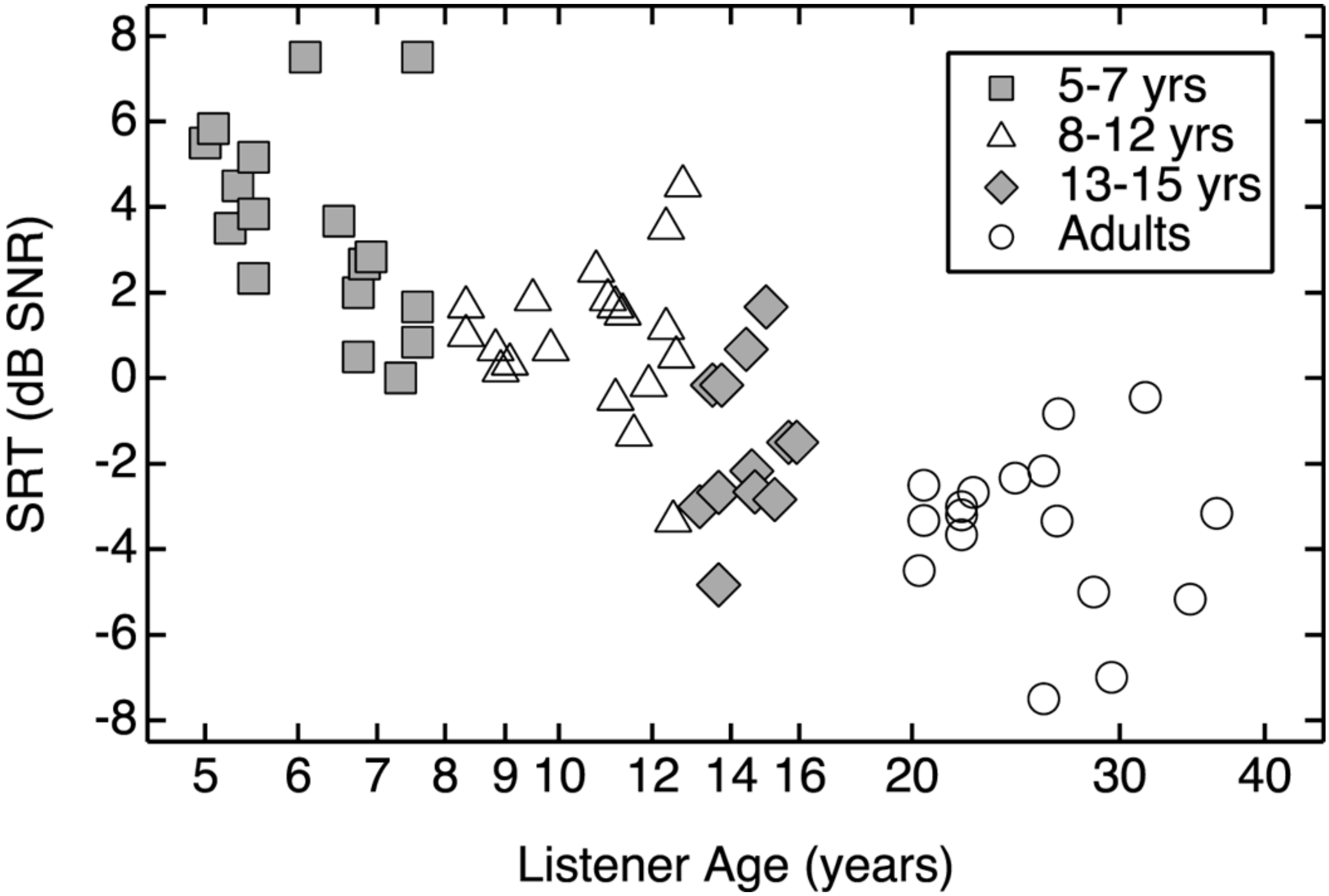


Figure 2.

Individual masked SRTs (dB SNR) in the unshifted-F0 condition for all listeners, plotted as a function of age on a log scale.

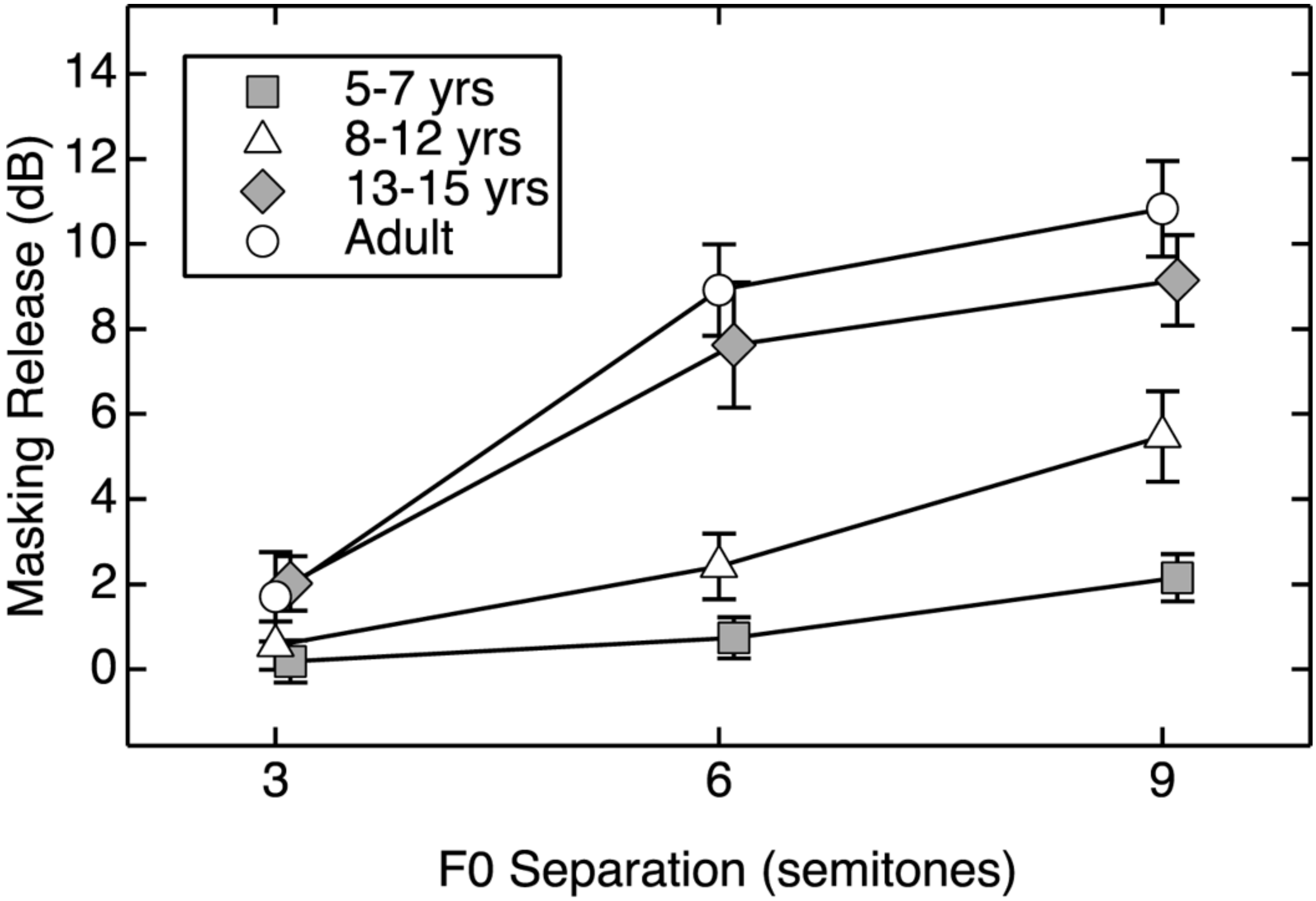


Figure 3.

Average masking release (dB) across listeners (with SEs) for each of the four age groups are presented for the 3, 6, and 9 semitone separation conditions.

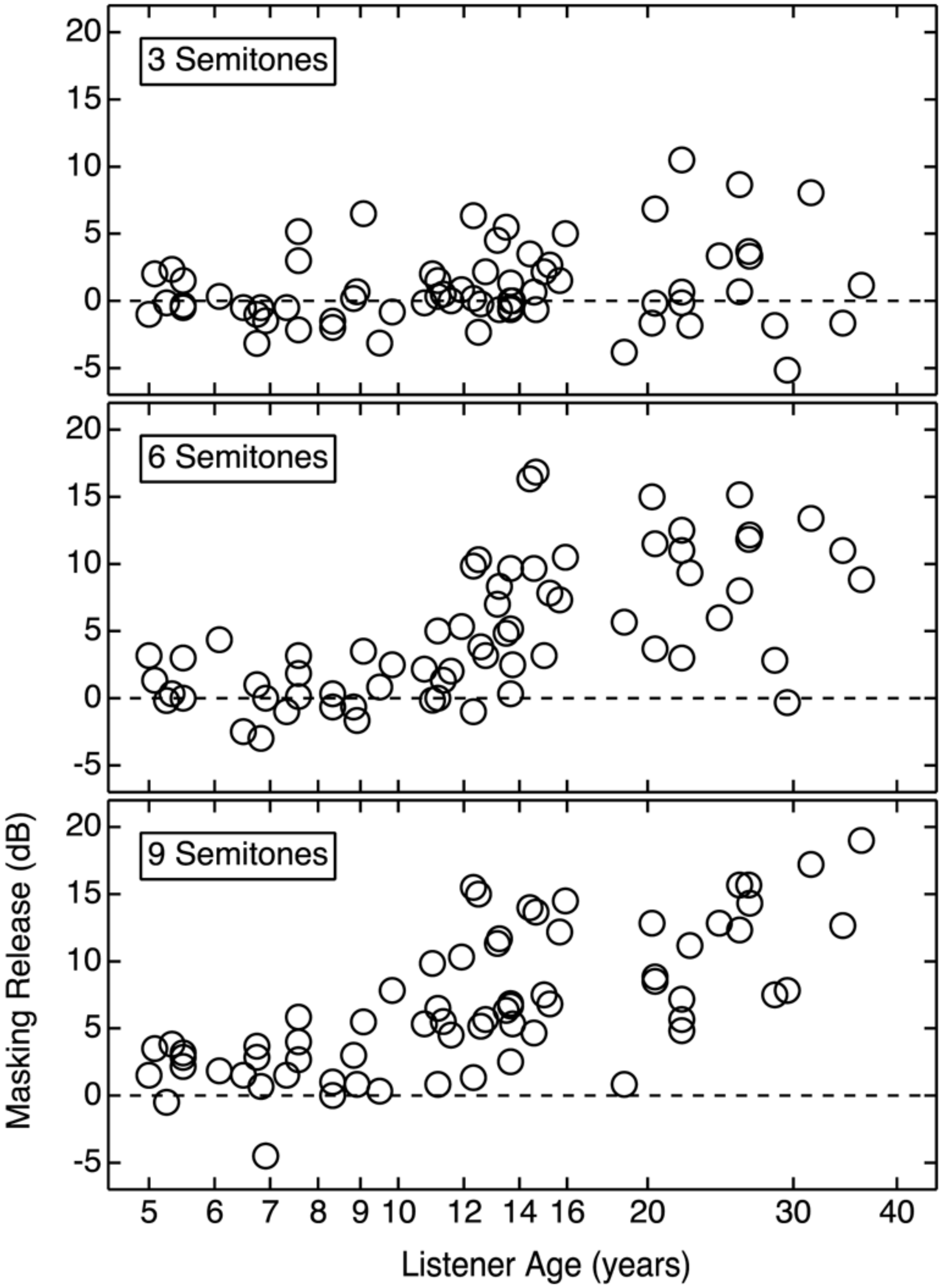


Figure 4.

Amount of masking release (dB) for individual listeners for each F0-shifted condition, plotted as a function of listener age on a log scale.

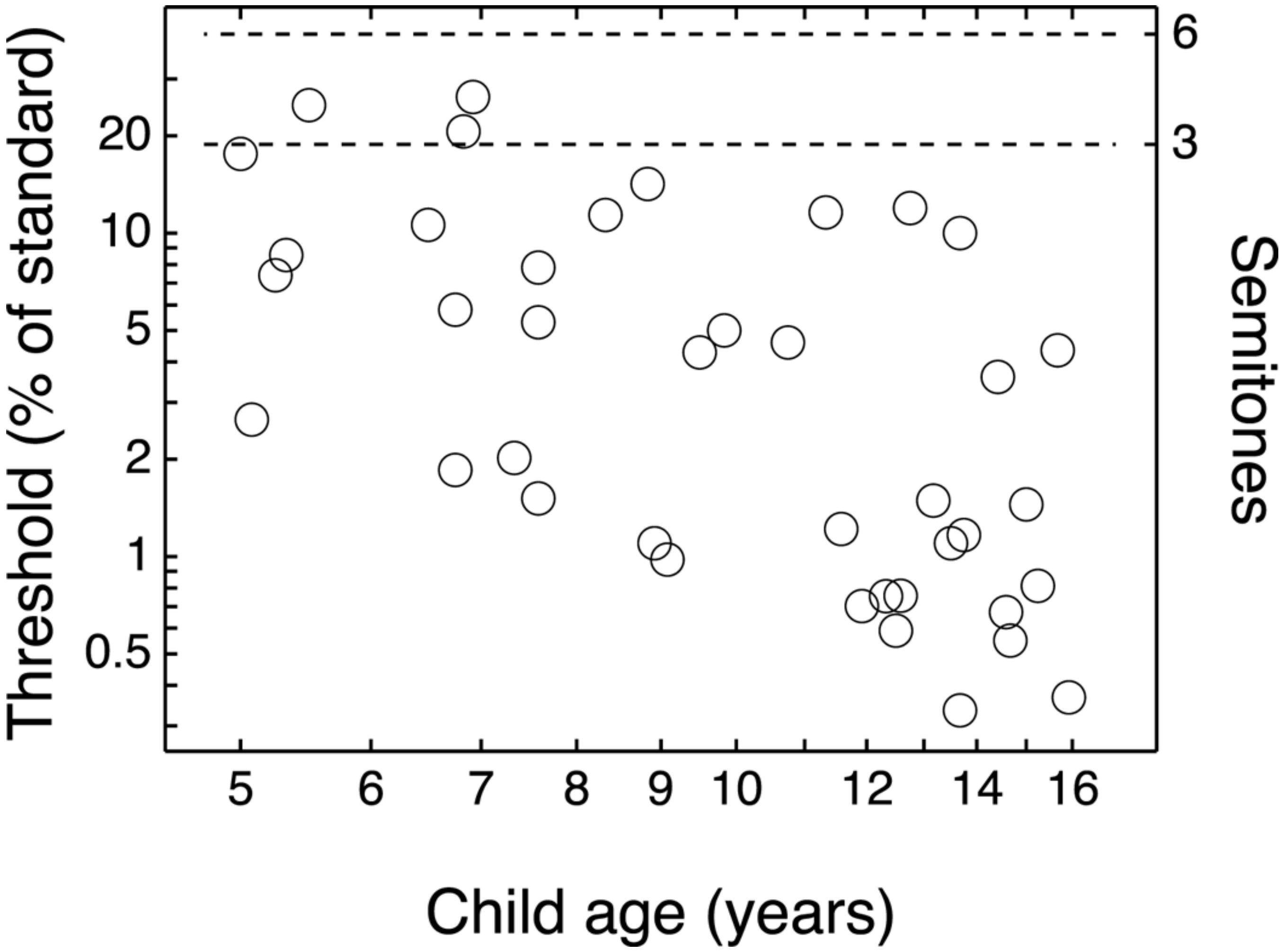


Figure 5.

Frequency discrimination thresholds for individual listeners, in percent of the standard (250 Hz) /ba/, plotted as a function of age. Both threshold and age are represented on a log scale. separations are shown on the right axis.

Table 1.

Parameter estimates for the mixed effects regression model analyzing data from all listeners (Model A) and from all listeners excluding one outlier participant (Model B).

	Model A					Model B				
	β	**SE**	**df**	***t***	***p***	**β**	**SE**	**df**	***t***	***p***
Intercept	−1.43	0.91	82	−1.56	0.607	−1.30	0.94	80	−1.38	0.085
F0 Condition	0.94	0.13	82	7.47	**0.000**	0.91	0.13	80	6.98	**0.000**
LogAge	0.08	2.62	39	0.03	0.489	−0.14	2.65	38	−0.05	0.479
Discrimination	−0.04	0.14	39	−0.30	0.383	−0.07	0.15	38	−0.47	0.322
F0 Condition x LogAge	1.17	0.36	82	3.24	**0.000**	1.23	0.37	80	3.38	**0.000**
F0 Condition x Discrim	−0.03	0.02	82	−1.74	**0.042**	−0.03	0.02	80	−1.23	0.110
LogAge x Discrim	−0.18	.32	39	−0.56	0.288	−0.17	0.33	38	−0.52	0.303
F0 Condition x LogAge x Discrim	−0.05	0.045	82	−1.03	0.154	−0.05	0.045	80	−1.11	0.135

β = coefficient estimate, **SE** = standard error, **df** = degrees of freedom, ***t*** = t-value, ***p*** = p-value.