Published in final edited form as:

Ear Hear. 2018 ; 39(5): 935–945. doi:10.1097/AUD.0000000000000554.

Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch

Lori J. Leibold¹, Emily Buss², and Lauren Calandruccio³

¹Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA

²Department of Otolaryngology/Head and Neck Surgery, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA

Abstract

Objective—The purpose of this study was to evaluate the extent to which infants, school-age children, and adults benefit from a target/masker sex mismatch in the context of speech detection or recognition in a background of 2 competing talkers. It was hypothesized that the ability to benefit from a target/masker sex mismatch develops between infancy and the early school-age years, as children gain listening experience in multi-talker environments.

Design—Listeners were infants (7–13 months), children (5–10 years), and adults (18–33 years) with normal hearing. A series of 5 experiments compared speech detection or recognition in continuous 2-talker speech across target/masker conditions that were sex matched or sex mismatched. In Experiments 1 and 2, an observer-based, single-interval procedure was used to estimate speech detection thresholds for a spondaic word in a 2-talker speech masker. In Experiments 3, 4, and 5, speech recognition thresholds were estimated in continuous 2-talker speech using a 4-alternative, forced-choice procedure. In Experiment 5, speech reception thresholds were estimated for adults using the forced-choice recognition procedure after ideal time-frequency segregation processing was applied to the stimuli.

Results—Speech detection thresholds for adults tested in Experiments 1 and 2 were significantly higher when the target word and speech masker were matched in sex than when they were mismatched, but thresholds for infants were similar across sex-matched and sex-mismatched conditions. Results for Experiments 3 and 4 showed that school-age children and adults benefit from a target/masker sex mismatch for a forced-choice word recognition task. Children, however, obtained greater benefit than adults in 1 condition, perhaps due to greater susceptibility to masking overall. In Experiment 5, adults had substantial threshold reductions and more uniform performance across the 4 conditions evaluated in Experiments 3 and 4 following the application of ideal time-frequency segregation to the stimuli.

Conclusions—The pattern of results observed across experiments suggests that the ability to take advantage of differences in vocal characteristics typically found between speech produced by male and female talkers develops between infancy and the school-age years. Considerable child-

adult differences in susceptibility to speech-in-speech masking were observed for school-age children as old as 11 years of age in both sex-matched and sex-mismatched conditions.

INTRODUCTION

Infants and children spend much of their days in settings with multiple people talking at the same time, either in person (van de Weijer 1998; Barker and Newman 2004) or via television and other electronic media (Lapierre et al. 2012; Ambrose et al. 2014). Children must learn about speech and language in these multi-talker environments. The ability to disentangle speech produced by different talkers is therefore critical for the development of communication skills in the real world.

Adults take advantage of differences in vocal characteristics between talkers to segregate target from masker speech (e.g., Bregman 1990; Bronkhorst 2000; Brungart 2001; Darwin et al. 2003). The most well established example of this phenomenon is that speech-in-speech recognition is markedly better for adults when target and masker talkers differ in sex than when they are the same sex (e.g., Festen and Plomp 1990; Brungart 2001; Helfer and Freyman 2008). Male and female speech productions diverge across several acoustic features associated with the length of the vocal folds and the size and shape of the vocal tract, including fundamental frequency (F0) and formant frequencies (Fitch and Giedd 1999). These sex-related acoustic differences facilitate segregation of male from female speech (and vice versa), thus minimizing informational masking relative to when target and masker speech are produced by talkers of the same sex (e.g., Brungart 2001; Helfer and Freyman 2008). For example, Brungart (2001) assessed adults' recognition of a target phrase masked by a competing masker phrase. Across conditions, target and masker phrases were produced by the same talker, by different talkers of the same sex, or by talkers not matched in sex. Speech intelligibility was 15–20 percentage points better when target and masker phrases were spoken by different talkers of the same sex than when target and masker phrases were spoken by the same talker. Speech intelligibility increased by an additional 15–20 percentage points, however, when target and masker phrases were produced by talkers not matched in sex. These results indicate that, while adults benefit from acoustic differences in voice characteristics when talkers are sex matched, it is substantially easier to segregate target from masker speech when talkers are sex mismatched.

There is mounting evidence that the ability to segregate target from background speech matures substantially over the first decade and a half of life. Infants and children have considerable difficulty perceiving target consonants, words, or sentences in the presence of speech produced by a small number of talkers of the same sex (e.g., Hall et al. 2002; Wightman and Kistler 2005; Newman 2009; Leibold and Buss 2013; Leibold et al. 2016). For example, Leibold and Buss (2013) observed a 35-percentage-point disadvantage in consonant identification performance for school-age children relative to adults when target speech produced by a female was embedded in 2-female-talker speech. While these maturational effects are largest for infants and young school-age children, substantial child/adult differences in susceptibility to speech-in-speech masking persist into the teenage years (e.g., Wightman et al. 2010; Corbin et al. 2015).

It is well established that infants and children are more susceptible to speech-in-speech masking than adults for sex-matched conditions, but data on age effects in the ability to benefit from target/masker sex mismatches are less conclusive. On one hand, even young infants can discriminate between male and female voices in quiet (e.g., Jusczyk et al. 1992; Kuhl 1983; Masapollo et al. 2016). On the other hand, infants require a more favorable signal-to-noise ratio (SNR) than adults to detect target speech when it is embedded in competing speech or noise maskers (e.g., Newman 2009; Leibold et al. 2016). Leibold et al. (2016) reported that speech-in-speech detection thresholds were more than 20 dB higher for 8- to 10-month-old infants than for adults in either speech-shaped noise or sex-matched 2-talker speech. Given extensive evidence that the peripheral encoding of sound is mature by at least 6 months postnatal age (reviewed by Werner 2007), the observation that both noise and 2-talker speech produced equivalently high amounts of masking suggests that infants listen to sound mixtures in a fundamentally different way than older children and adults.

Recent findings reported by Newman and Morini (2017) provide insight into the question of when in development the ability to benefit from target/masker sex mismatches emerges. Using a preferential-looking paradigm, recognition of target speech produced by a female talker was assessed at a fixed SNR in a single-stream of masker speech produced by either a female or a male talker. Listeners were 30-month-olds (Experiment 1) and 16- to 17-month-olds (Experiment 2). Better performance was observed for the older age group (30-month-olds) when there was a target/masker sex mismatch than when both were produced by female talkers. In contrast, similar performance was observed for the younger listeners (16- to 17-month-olds) between sex-matched and sex-mismatched conditions. These findings suggest that the ability to benefit from differences in talker sex emerges during the toddler years, although it is not clear whether young children benefit to the same extent as adults.

There are limited data on the degree to which school-age children benefit from target/masker sex mismatches, and the results are somewhat mixed. Wightman and Kistler (2005) evaluated speech-in-speech recognition in 4- to 16-year-olds and adults using the Coordinate Response Measure test (CRM; Bolia et al. 2000). Consistent with previous findings (e.g., Brungart 2001), adults showed an average improvement in speech intelligibility of about 30 percentage points for sex-mismatched compared with sex-matched target/masker conditions when target and masker sentences were presented to the same ear. Although children performed more poorly than adults overall, they showed a sex-mismatch benefit similar to that observed for adults. Note, however, that age effects in error patterns were observed in the sex-mismatched condition. While errors made by adults and older children were unrelated to the masker sentences, most errors made by children younger than 6 years of age were intrusions from the sex-mismatched masker speech. This pattern of results suggests that young children have difficulty segregating and/or selectively attending to the target speech, even when the target and masker are mismatched in sex.

Additional data pertinent to the question of whether school-age children take advantage of target/masker sex mismatches come from several experiments investigating age effects in spatial release from masking (e.g., Litovsky 2005; Johnstone and Litovsky 2006). While these studies did not include sex-matched conditions, masked speech recognition was assessed using a closed-set word recognition task in which target words produced by a male

talker were presented in a background of sentences produced by 1 or 2 female talkers. In some conditions, target and masker speech sources were co-located (i.e., presented via the same loudspeaker). Children consistently required a more advantageous SNR than adults to achieve the same criterion level of performance. One explanation offered by Johnstone and Litovsky (2006) for children's increased susceptibility to speech-in-speech masking relative to adults is that they might be immature in their ability to use F0 differences between male and female talkers to segregate target from masker speech.

The objective of the present study was to evaluate the extent to which infants, school-age children, and adults benefit from a target/masker sex mismatch when detecting or recognizing speech in a background of other talkers. Specifically, a series of experiments compared speech detection (Experiments 1 and 2) and recognition (Experiments 3, 4, and 5) in a 2-talker masker across sex-matched and sex-mismatched target/masker conditions. The over-arching hypothesis for this study was that the ability to benefit from target/masker sex mismatches develops between infancy and the early school-age years, as children gain experience listening in multi-talker environments.

EXPERIMENT 1: EFFECT OF A TARGET/MASKER SEX MISMATCH ON INFANTS' AND ADULTS' SPEECH DETECTION IN A 2-MALE-TALKER MASKER

This experiment evaluated whether infants and adults benefit from a target/masker sex mismatch in the context of speech-in-speech detection. An observer-based psychoacoustic procedure (e.g., Olsho et al. 1997) was used to estimate infants' and adults' detection threshold for a spondee word produced by a male or female talker in the presence of 2-male-talker speech. The hypothesis was that the ability to segregate speech produced by different talkers is immature during infancy, even when the target and masker speech are mismatched in sex. The prediction was that no difference in infants' detection performance would be observed between sex-matched and sex-mismatched conditions.

Method

Listeners—Listeners were 18 infants (7–13 months) and 18 adults (18–26 years). The average age at the initial testing session was 10.4 months ($SD = 1.6$ months) for infants and 22.7 years ($SD = 3.8$ years) for adults. Data from 9 additional infants were excluded from analysis: 1 did not reach the training criterion; 2 did not provide sufficient test data; and 6 completed testing but were excluded because of a high response rate on no-signal trials ($>40\%$). Data from 1 adult were excluded due to experimenter error.

Selection criteria for infants were: (1) no risk factors for hearing loss as assessed by parental report; (2) English-speaking home environment; (3) no more than 2 episodes of otitis media; (4) not under treatment for otitis media within the prior week; and (5) no signs of a cold or other illness on the test date. In addition, screening tympanometry using a 226 Hz probe tone was performed on every infant at the end of each session. Peak admittance of at least 0.2 mmhos at a pressure between -200 and 50 daPa was required to pass the screening. Selection criteria for adults were: (1) no risk factors for hearing loss as assessed by self-

report; (2) no more than 2 years of formal musical training; (3) no previous participation in psychoacoustic studies; and (4) English as a first language. In addition, adults were required to pass a hearing screening prior to testing, with thresholds less than or equal to 20 dB HL bilaterally for octave frequencies between 250 and 8000 Hz (ANSI 2010).

Stimuli and conditions—Target stimuli were recordings of the spondee word *playground*. The target word was recorded in isolation from 1 adult female and 1 adult male using a condenser microphone (AKG 212C1000S) mounted approximately 6 inches from the talker's mouth. Both talkers were native speakers of American English. A single recording of the target word was selected for each talker. The mean F0 for the target word produced by the female talker was 182 Hz, compared with 120 Hz for target word produced by the male talker. Productions were amplified (Tucker Davis Technologies MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (CardDeluxe). Prior to the experiment, the 2 target words were equated for root-mean-square (rms) level and then resampled at a rate of 24.414 kHz using MATLAB.

Following Bonino et al. (2013), the masker was composed of recordings of 2 males that were obtained while they read aloud different passages from a popular children's novel. Both talkers were native speakers of American English, and were different from the male talker who produced the target words. The same methods described above for recording target stimuli were used to obtain the masker speech recordings. The mean F0 was 102 Hz for the first male talker and 109 Hz for the second male talker, based on their sustained vowel productions. Thus, the mean F0 for the male target word was 1.7 and 2.8 semitones above the masker talkers' mean F0s, and the mean F0 for the female target word was 8.9 and 10.0 semitones above the masker talkers' mean F0s. The individual speech streams were manually edited to reduce silent pauses longer than 300 ms, resulting in samples that were both greater than 3 minutes in length. Each sample was repeated without discontinuity for 60 minutes. The 2 speech streams were balanced for overall rms level and then mixed to create the 2-male-talker masker.

A custom MATLAB script controlled the selection and presentation of the stimuli. The target and masker stimuli were mixed (Tucker Davis Technologies SM3), amplified (Techtron 5507) and presented through a loudspeaker (Monitor Audio, Monitor 4). During testing, the listener was positioned approximately 1 m from the loudspeaker in the sound field of a 7 × 7 ft, double-walled sound-treated booth. The height and position of the listener's chair was adjusted so that stimuli were presented at approximately 0° azimuth and 0° elevation.

There were 2 conditions: (1) target/masker sex matched and (2) target/masker sex mismatched. In the matched condition, the target and masker speech were produced by males. In the mismatched condition, the target spondee was produced by a female, and the masker speech was produced by males. Each listener completed testing in either the matched or mismatched condition, with an equal number of listeners tested in each condition. Adults were tested in a single visit to the laboratory. Infants were tested in 2 visits occurring within a 2-wk period. For both age groups, each visit was approximately 45 min in duration.

Procedure—A single-interval, observer-based psychophysical procedure was used to test infants (Olsho et al. 1987). Each infant was tested while sitting on a parent's lap. An assistant sat inside the booth with the parent and infant, manipulating toys in order to keep the infant facing toward the midline. To prevent the assistant and the parent from hearing the stimuli and influencing the infant's response, the adults wore noise-isolating earphones (Etymotic MC5) that delivered masking sounds as well as noise-reduction earmuffs (Bilsom Thunder T3). To the listener's right were 2 mechanical toys with lights in a box made of dark acrylic glass. An observer sat outside of the booth and initiated trials when the listener was quiet and facing midline. The procedure for testing adults was the same as that for testing infants, but adult listeners were alone in the booth during testing. Adults were instructed to raise their hand when they heard the "sound that makes the toy come on."

The masker was presented continuously throughout testing at a fixed overall level of 50 dB SPL. Trials were either *signals*, in which the target word was presented once, or *catch trials*, in which no target word was presented. The observer was seated outside the booth and initiated trials, but did not know which type of trial occurred. For signal trials, the target word was presented at the start of the trial, immediately after the observer initiated the trial. The observer was required to decide the trial type based on the listener's behavior within 4 seconds of trial onset. The listener was provided with reinforcement if the observer correctly identified a signal trial. Reinforcement was the activation and illumination of a mechanical toy. The observer was provided with feedback after every trial.

A complete session included 2 training phases and 1 testing phase. Target words were presented at a supra-threshold level in both training phases, depending on age group. The goal of the first training phase was to establish the relationship between the presentation of the target word and the mechanical toy reinforcement. The probability of a signal trial was 0.80, and the probability of a catch trial was 0.20. Listeners were reinforced after each signal trial, regardless of the observer's response. The first training phase was completed when the observer correctly responded on 4 of 5 consecutive trials, including at least 1 catch trial. The goal of the second training phase was to demonstrate to the listener that he/she was required to respond to signal trials in order to turn on the mechanical toy reinforcer. The probability of both signal and catch trials in this phase was 0.50. Reinforcement was only provided to the listener if the observer correctly identified a signal trial. The second training phase was completed when the observer/listener team maintained a hit rate of 0.80 or higher, and a catch trial rate of 0.20 or lower, on the last 10 sequential trials.

During the testing phase, the level of the target spondee was adaptively varied using a 2-down, 1-up procedure that estimated the SNR corresponding to 70.7% correct detection performance (Levitt 1971). The probability of a signal trial was 0.75, and the probability of a catch trial was 0.125. In addition, probe trials were presented with a probability of 0.125. Probe trials were presentations of the target word at the supra-threshold level used during training. Infants required an average of 48 trials to complete the testing phase (range = 26 – 61), corresponding to approximately 36 signal trials, 6 catch trials, and 6 probe trials per adaptive track.

Only signal trials affected the adaptive track. Based on pilot data, the starting level for the target word was about 10 dB higher than the expected threshold value. The initial step size was 4 dB. After the first two reversals, the step size reduced to 2 dB. Eight reversals were obtained, and threshold was based on the level at the last 6 reversals. Thresholds were only accepted if the response rate to probe trials was 0.60 or higher, and the catch trial rate was 0.40 or lower.

Results and Discussion

Figure 1 presents thresholds in dB SNR corresponding to 70.7% detection performance for infants (left) and adults (right). Unfilled boxes show the range of performance spanning the 25th to the 75th percentile for listeners tested in the matched condition (male target/2-male-talker masker). Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines. Lower thresholds indicate greater sensitivity. Shaded boxes show the same range of performance for listeners tested in the mismatched condition (female target/2-male-talker masker).

Results for adults are consistent with data reported in previous studies that have compared adults' speech-in-speech recognition between sex-matched and sex-mismatched conditions (e.g., Festen and Plomp 1990; Brungart 2001). For example, Festen and Plomp (1990) measured adults' masked sentence recognition thresholds using an adaptive procedure. Target sentences produced by a male or a female talker were presented in continuous speech produced by either one female or one male. The average sex-mismatch benefit ranged from 6 to 10 dB across conditions. The present detection data are in agreement with their pattern of results; adults' average speech detection threshold was 6.4 dB better in the mismatched than the matched condition.

In contrast to the adult data, there was no evidence that thresholds for infants tested in the mismatched condition were lower than for infants tested in the matched condition. Thresholds ranged from -2.3 to 10.3 dB SNR (mean = 3.5) and from -4.7 to 10.3 dB SNR (mean = 7.5) for infants tested in the matched and mismatched conditions, respectively. Thus, the average infant threshold was 4.0 dB *higher* when there was a target/masker sex mismatch than when both target and masker speech were produced by males.

An independent-samples *t-test* was performed on the data for each age group to evaluate the trends observed in Figure 1. The rationale for performing a separate analysis on the data for each age group was to guard against the possibility of infant/adult differences in response bias influencing the results (e.g., Leibold and Werner 2006; Leibold et al. 2016), and because the goal of this experiment was to determine whether listeners within each age group benefit from a target/masker sex mismatch.¹ No significant difference in threshold between infants tested in the matched condition and infants tested in the mismatched condition was observed [$t_{16} = -1.44$, $p = 0.17$]. There was a significant difference in threshold between adults tested in the matched condition and adults tested in the mismatched

¹On this task, a d' of around 1 is consistent with 70% correct responses if listeners were unbiased. While we cannot rule out the possibility that response bias differed across condition within age group, an examination of false alarm rate indicated no significant difference in false alarm rate between matched- and mismatched-sex conditions for infants [$F(1,16) = 1.54$; $p = 0.23$; $\eta^2_p = 0.09$] or adults (all adults had a false alarm rate of 0) who participated in Experiment 1.

condition [$t_{16} = 3.02$, $p < 0.01$]. This significant group effect reflects a difference in average SNR of 6.4 dB between the adults tested in the matched condition (mean = -9.5 dB SPL) and adults tested in the mismatched condition (mean = -15.9 dB SPL).

It is evident from Figure 1 that between-subjects variability was substantial for both age groups. Thus, one important question is whether the effects observed in group comparisons would extend to individual listeners. To address this question, supplemental data were collected on 5 of the 18 infants and 9 of the 18 adults who completed testing in their assigned condition (matched or mismatched) quickly enough that data could be collected in the other condition. The group of infants included 3 listeners who completed testing in the matched condition first and 2 listeners who completed testing in the mismatched condition first. The group of adults included 5 listeners who completed testing in the matched condition first and 4 listeners who completed testing in the mismatched condition first. The small number of infants who completed testing in both conditions is insufficiently powered to make inferences about the larger population. Nonetheless, it is interesting to note that the individual data were in general agreement with the group data. While the average adult threshold was 5.1 dB lower in the mismatched compared with the matched condition, the average infant threshold difference was 1.9 dB higher in the mismatched compared with the matched condition.

EXPERIMENT 2: EFFECT OF TARGET/MASKER SEX MISMATCH ON INFANTS' AND ADULTS' SPEECH DETECTION IN A 2-FEMALE-TALKER MASKER

Infants tested in Experiment 1 did not benefit from a target/masker sex mismatch in the context of word detection in a 2-male-talker masker. Previous investigations have demonstrated that adults take advantage of differences between males and females in the acoustic voice characteristics of F0 and vocal tract length to segregate target from sex-mismatched masker speech (e.g., Brungart 2001; Darwin et al. 2003; Helfer and Freyman 2008). Thus, the lack of a sex-mismatch benefit for infants is consistent with the idea that infants are immature in their ability to use sex-based differences in acoustic voice characteristics to segregate auditory streams. Another possible interpretation, however, is that the perceptual salience of the two-male-talker masker used in Experiment 1 may not be the same for infants and adults. It is well documented that infants spend more time listening to female than male voices (reviewed by Soderstrom 2007), showing a clear preference for female versus male voices in quiet (e.g., Werker and McLeod 1989; Pegg et al. 1992). The novelty of using a masker composed of 2 streams of speech produced by unfamiliar male talkers may have distracted attention away from the target talker. This possibility was evaluated in Experiment 2 by testing new infant and adult listeners in the presence of masker speech composed of 2 female talkers. The procedures, hypotheses and predictions were the same as in Experiment 1.

Method

Listeners—Listeners were 16 infants (7–13 months) and 16 adults (18–33 years). None of the listeners participated in Experiment 1. The average age at the initial testing session was

10.1 months ($SD = 1.2$ months) for infants and 22.8 years ($SD = 3.2$ years) for adults. Data from 6 additional infants were excluded from analysis: 4 did not provide sufficient test data and 2 completed testing but were excluded because of a high response rate on no-signal trials ($>40\%$). The selection criteria for both age groups were the same as in Experiment 1.

Stimuli, conditions and procedure—The target stimuli were recordings of the spondee word *hotdog* produced by the same male and female talkers who produced the target word for Experiment 1. Recall that the mean F0s of the target word were 182 Hz and 120 Hz for the female and male target talker, respectively. The masker was continuous 2-female-talker speech. Recordings of 2 female speakers of English were obtained while each was reading aloud different passages from a popular infant and toddler books. The mean F0 for the first female talker was 227 Hz and the mean F0 for the second female talker was 243 Hz, based on their sustained vowel productions. The talkers who produced the masker speech were not given explicit instructions regarding speaking style. However, both spontaneously produced child-directed speech, characterized by exaggerated prosodic contour and slow speaking rate. The mean F0 of the target word produced by the female talker was 3.8 and 5.0 semitones below the masker talkers' mean F0s. The mean F0 of the target word produced by the male target talker was 11.0 and 12.2 semitones below the masker talkers' mean F0s. Stimuli were recorded and digitally edited using the same procedures described for Experiment 1.

Each listener completed testing in a single condition. In the sex-matched condition, the target and masker speech were produced by females. In the sex-mismatched condition, the target speech was produced by a male, and the masker speech was produced by females. Infants were tested in two separate visits occurring within a 2-wk period. Each visit lasted about 45 minutes. Adults were tested in a 45-minute visit to the laboratory. The observer-based testing procedure was the same as in Experiment 1. Infants required an average of 43 total trials to complete the testing phase for Experiment 2 (range = 22 – 51).

Results and Discussion

Figure 2 shows masked speech detection thresholds in dB SNR for infants and adults, following the format used for Figure 1. The pattern of results using the 2-female-talker was the same as that observed for Experiment 1 using the 2-male-talker masker. While thresholds were similar for infants tested in the matched condition (range = 0.0 to 23.5 dB; mean = 8.4 dB) and infants tested in the mismatched condition (range = -4.8 to 25.3 dB; mean = 6.2 dB), adults tested in the mismatched condition (range = -23.2 to -17.0; mean = -21.3) were more sensitive to the target word than adults tested in the matched condition (range = -21.0 to -4.1 dB; mean = -14.2).

An independent-samples *t*-test was performed on the data for each age group to evaluate the trends observed in Figure 2.² No significant difference in threshold between infants tested in the matched condition and infants tested in the mismatched condition was observed [$t_{14} = 0.48$, $p = 0.64$]. There was, however, a significant difference in threshold between adults

²No significant difference in false alarm rate between matched- and mismatched-sex conditions was observed for infants [$F(1,14) = 0.57$; $p = 0.46$; $\eta^2_p = 0.04$] or adults [$F(1,14) = 1.00$; $p = 0.33$; $\eta^2_p = 0.07$] who participated in Experiment 2.

tested in the matched condition and adults tested in the mismatched condition [$t_{14} = 3.25$, $p < 0.01$]. This significant effect reflects a difference in average SNR at threshold of 7.0 dB between the 2 groups of adults.

A major question raised by the results of the first two experiments is why don't infants benefit from a target/masker sex mismatch in the context of speech-in-speech detection? Adults showed a robust sex-mismatch benefit in both experiments, but thresholds for infants were not significantly different across matched and mismatched conditions in either 2-male-talker or 2-female-talker speech. One possible explanation for this pattern of results is that infants have not yet mastered the ability to segregate streams of speech based on acoustic differences between male and female talkers. Darwin et al. (2003) demonstrated that the primary acoustic differences responsible for adults' improved speech-in-speech recognition in sex-mismatched versus sex-matched conditions are voice pitch (i.e., F0) and perceived vocal tract length (i.e., formant frequencies). Infants may not yet have learned the full range of acoustic cues that differentiate male and female voices due to their limited exposure to different talkers and acoustic environments. Several lines of evidence are consistent with this idea. First, early listening experience appears to be crucial for identifying talkers in quiet (e.g., Johnson et al. 2011). Second, familiarity with the target talker's voice improves speech perception in noise for adults (e.g., Nygaard et al. 1994) and young children (White and Aslin 2011). Finally, Newman and Morini (2017) demonstrated that 30-month-olds, but not 16- to 17-month-olds, show better speech-in-speech recognition performance on a preferential-looking when there is a target/masker sex mismatch than when both target and masker are produced by female talkers.

The lack of a sex-mismatch benefit for infants also raises the question, when in development does the ability to benefit from a target/masker sex mismatch mature? While recent data reported by Newman and Morini (2017) suggest this ability emerges between 17 and 30 months of age, it is not clear whether children benefit to the same degree as adults. The following 2 experiments were designed to address this question by evaluating speech-in-speech recognition in school-age children for sex-matched and sex-mismatched conditions using the same 2-talker maskers used to test infants and adults in the first 2 experiments.

EXPERIMENT 3: EFFECT OF TARGET/MASKER SEX MISMATCH ON CHILDREN'S AND ADULTS' SPEECH RECOGNITION IN A 2-MALE-TALKER MASKER

The purpose of this experiment was to determine the extent to which 5- to 10-year-old children and adults benefit from a target/sex mismatch in the context of a speech-in-speech recognition task. It was hypothesized that the ability to segregate speech produced by different talkers using male/female differences in voice characteristics emerges between infancy and the school-age years. This hypothesis was based on data reported by Wightman and Kistler (2005), who observed a similar sex-mismatch benefit in the presence of a single stream of competing speech for 4- to 16-year-old children and adults in the context of the CRM test. School-age children were expected to be more susceptible to speech-in-speech

masking than adults in the present experiment, but the magnitude of the sex-mismatch benefit was predicted to be similar for children and adults.

Method

Listeners—Listeners were 10 school-age children (5–10 years) and 8 adults (19–28 years). The average age at the initial testing session was 7.3 years ($SD = 1.6$ years) for children and 22.5 years ($SD = 3.2$ years) for adults. All listeners were native speakers of American English, with a self or parental report of normal hearing, speech and language. Listeners passed a hearing screening on the day of testing, with thresholds less than or equal to 20 dB HL bilaterally at octave frequencies between 250 and 8000 Hz (ANSI 2010).

Stimuli and conditions—Target stimuli were 25 spondee words, recorded by one male and one female talker. Both talkers were native speakers of American English. The following words, originally chosen for their visual unambiguity (Hall et al., 2002), were used: airplane, armchair, baseball, bathtub, birthday, bluebird, cowboy, cupcake, doormat, flashlight, football, hotdog, ice-cream, mailman, mousetrap, mushroom, playground, popcorn, sailboat, seesaw, shoelace, sidewalk, snowman, toothbrush, and toothpaste. Recordings were created in a sound-isolated room. Individual recordings were obtained while the talker was seated in a comfortable chair so that his/her mouth was positioned 6 inches from a condenser microphone (AKG-C1000S). Talkers produced each spondee using a carrier phrase (“Say the word X again”, where X was 1 of the 25 spondees). The mean F0s across all target words produced by the male and female talkers who recorded the target words were 144 Hz and 231 Hz, respectively. Recordings were amplified (TDT MA3) and digitized (CardDeluxe) using a 44.1 kHz sampling rate (32 bits). Target spondees were digitally excised from the carrier phrase, scaled to normalize the RMS level across tokens, and down sampled to a rate of 24.414 kHz.

The masker was the same 2-male-talker recording used with infants and adults in Experiment 1. Recall that the mean F0 was 102 Hz for the first male talker and 109 Hz for the second male talker. Thus, the mean F0 of the male target talker was 4.8 and 6.0 semitones above the masker talkers’ mean F0s, and the mean F0 of the female target talker was 13.0 and 14.2 semitones above the masker talkers’ mean F0s.

Listeners were tested in each of 2 conditions: (1) target/masker sex matched using target spondees produced by the male target talker and (2) target/masker sex mismatched using target spondees produced by a female target talker. Testing order was counterbalanced across listeners, but testing was completed for the first condition before commencing testing for the second condition.

Procedure—Testing was completed in a single-walled sound-isolated room (IAC). Listeners sat in a comfortable chair in front of a desk with a computer monitor and a mouse. Testing was controlled using custom software (MATLAB). All listeners completed a familiarization task in quiet prior to experimental testing; listeners were asked to point to the appropriate illustration of each of the 25 target spondee word shown in sets of 4 using a laminated picture book. This task was performed without errors by all listeners.

Listeners performed a 4-alternative, forced-choice (4AFC) task. One of the 25 spondee words was randomly presented on each trial. Three other spondee illustrations, selected from the remaining 24 words, were randomly selected without replacement to serve as foils. The 4 pictures were randomly positioned; with 1 picture appearing in each of the quadrants of the computer monitor about 20 ms prior to the audio target presentation. After the target spondee was presented, listeners were prompted to select the picture that represented the word they heard. Visual feedback was provided after the listener provided a response, indicating the target word by flashing the associated illustration.

The overall rms level of the target spondees was fixed at 50 dB SPL throughout testing. The level of the continuous masker speech was adapted using a 2-up, 1-down rule (Levitt 1971) to estimate the SNR associated with 70.7% correct spondee identification. While experiments 1 and 2 fixed the masker and adjusted the target, this procedural difference was not expected to affect the results³. The starting level for the first threshold estimation run of each condition was +10 dB SNR for children and 0 dB SNR for adults. For subsequent runs, the starting level was 10 dB above the SNR associated with threshold for the first run. A step size of 4 dB was used for the first 2 reversals, reduced to 2 dB for the remaining 6 reversals of the run. Threshold was computed by averaging the SNR at the final 6 reversals. Two runs were completed in succession for each condition. A third run was completed if the 2 runs differed by more than 4 dB; this occurred for 3 children and 6 adults. The final threshold for each condition was obtained by averaging the 2 thresholds that were within 4 dB of each other, represented in dB SNR. Testing was completed in a visit lasting 1.5 hours for adults and children older than 8 years of age, and in 2 visits lasting 1 hour each for children younger than 8 years of age.

Results and Discussion

Individual speech reception thresholds (SRTs) in dB SNR are shown in Figure 3, plotted as a function of listener age. Circles show SRTs in the matched condition and Xs show SRTs in the mismatched condition. These data add to the growing number of studies reporting considerable child/adult differences in susceptibility to speech-in-speech masking (e.g., Hall et al. 2002; Wightman and Kistler 2005; Bonino et al. 2013). Group SRTs are shown on the left side of each panel in Figure 5. In agreement with data collected on school-age children by Hall et al. (2002), who used the same 4AFC procedure to evaluate speech recognition for male spondees in a 2-male-talker masker, 5- to 10-year-old children required a more advantageous SNR than adults to achieve criterion performance in the sex-matched condition (mean difference = 11.5 dB). The average SRT for children in the matched condition was -4.8 dB SNR ($SD = 3.1$), compared with -16.3 dB SNR ($SD = 2.3$) for adults.

³There is at least one example in the literature where results differ depending on whether the masker level is fixed and the target level is varied, or vice versa. Hall and Grose (1991) measured detection thresholds for a tone in notched noise to estimate frequency selectivity in school-age children and adults. Results were consistent with poorer frequency selectivity in 4-year-olds than adults when the masker level was fixed, but not when the target level was fixed. This outcome was attributed to reduced efficiency in children and differences in growth of loudness in the notch and no-notch condition when the masker level was fixed. While reduced efficiency in children is likely to impact speech-in-speech recognition, growth of loudness is not expected to differ for sex-matched and sex-mismatched talkers, apart from effects related to differences in perceptual similarity.

The goal of this experiment was to determine whether school-age children benefit from a target/masker sex mismatch and, if so, to compare the magnitude of this benefit between children and adults. Lower SRTs were observed for all listeners in the mismatched compared to the matched condition. The vertical lines in Figure 3 indicate the target/masker sex-mismatch benefit, defined as the difference in between SRTs in the matched and mismatched conditions. This benefit ranged from 3.8 to 11.2 dB (mean = 7.4) for children and from 2.8 to 10.3 dB (mean = 7.5) for adults.

A two-way repeated-measures analysis-of-variance (ANOVA) on SRT confirmed the trends observed in Figure 3. The analysis included the within-subjects factor of Condition (matched, mismatched) and the between-subjects factor of Age Group (children, adults). The main effect of Condition was significant [$F(1,16) = 108.96$; $p < 0.001$; $\eta^2_p = 0.87$], indicating listeners benefitted from the sex-mismatch. The main effect of Age Group was also significant [$F(1,16) = 62.1$; $p < 0.001$; $\eta^2_p = 0.80$], indicating better performance overall for adults than for children. The Condition $\times$ Age Group interaction was not significant [$F(1,16) = 0.01$; $p = 0.99$; $\eta^2_p = 0.00$], indicating children and adults benefitted to a similar degree from the target/masker sex mismatch.

The data indicate that school-age children can take advantage of acoustic differences between male and female talkers to facilitate segregation of target words from a 2-male-talker masker. Moreover, the magnitude of this benefit is adult-like for school-age children. These findings are in agreement with previous data reported by Wightman and Kistler (2005). In that study, 4- to 16-year-olds and adults showed a similar improvement in performance when a single target phrase and a single masker phrase were mismatched relative to matched in sex.

EXPERIMENT 4: EFFECT OF TARGET/MASKER SEX MISMATCH ON CHILDREN'S AND ADULTS' SPEECH RECOGNITION IN A 2-FEMALE-TALKER MASKER

While differences in the salience of male versus female masker speech and/or speaking style did not appear to impact speech detection thresholds in Experiments 1 and 2, it is possible these factors influence performance on the more challenging 4AFC recognition task. Thus, this experiment was a replication of Experiment 3, but new listeners were tested in the presence of 2-female-talker speech. The target stimuli, procedures, hypotheses, and predictions were the same as in the previous experiment.

Method

Listeners—Fifteen school-age children (5–10 years) and 15 adults (18–26 years) participated in Experiment 4. None of the listeners participated in Experiment 3. The mean age of the children and adults was 7.8 ($SD = 1.6$) and 21.3 ($SD = 2.3$) years old, respectively. The inclusion criteria were identical to those described for children and adults tested in Experiment 3.

Stimuli, conditions and procedure—The stimuli, conditions and procedure followed those used in Experiment 3, except the masker was the same 2-female-talker speech recording used to test infants and adults in Experiment 2. The mean F0 of the female target talker was 0.3 semitones above the mean F0 of the first masker talker and 0.9 semitones below the mean F0 of the other masker talker. The mean F0 of the male target talker was 7.9 and 9.1 semitones below the mean F0s of the masker talkers.

Results and Discussion

SRTs in dB SNR for individual listeners are shown in Figure 4, following the format used for Figure 3. Group SRTs are shown on the right of each panel in Figure 5. The average child SRT was 7.6 dB higher in the matched (mean = −6.9 dB SNR) compared with the mismatched condition (mean = −14.5 dB SNR). The average adult SRT was also higher in the matched (mean = −21.0 dB SNR) compared with the mismatched condition (mean = −24.8 dB), but the average difference between the 2 conditions (3.8 dB) was smaller than observed for children. Individual differences were extensive; this was true in both conditions, in the magnitude of the sex-mismatch benefit, and for both age groups. Nonetheless, SRTs for 14/15 children and 13/15 adults were lower in the mismatched compared with the matched condition.

A 2-way repeated-measures ANOVA on SRT was performed to evaluate the trends observed in Figure 4. This analysis included the within-subjects factor of Condition (matched, mismatched) and the between-subjects factor of Age Group (children, adults). All of the effects in the analysis were significant: Condition [$F(1,28) = 53.29$; $p < 0.001$; $\eta^2_p = 0.66$], Age Group [$F(1,28) = 57.47$; $p < 0.001$; $\eta^2_p = 0.67$], and the Condition $\times$ Age Group interaction [$F(1,28) = 5.74$; $p < 0.05$; $\eta^2_p = 0.17$]. The significant interaction indicates the magnitude of the sex-mismatch benefit is not the same for children and adults. This interaction was further examined by performing a paired-samples *t*-test on the sex-mismatch benefit (SRT in the matched condition – SRT in the mismatched condition) with Age Group (children, adults) as an independent variable. The results of this analysis indicated a larger sex-mismatch benefit for children than for adults [$t(14) = 2.59$; $p < 0.05$].

Figure 5 presents boxplots summarizing SRTs for listeners tested in Experiment 3 (2-male-talker masker) and for listeners tested in Experiment 4 (2-female-talker masker). In agreement with the pattern of results observed for school-age children in the 2-male-talker masker (Experiment 3), 5- to 10-year-olds tested in the present experiment were able to take advantage of a target/masker sex mismatch in the 2-female-talker masker. The magnitude of this benefit appears to be similar for children tested in the 2-male-talker masker (average benefit = 7.4 dB) and children tested in the 2-female-talker masker (average benefit = 7.6 dB). In contrast to the previous experiment, however, adults showed a significantly smaller sex-mismatch benefit than children in the present experiment. While children's average SRT in the 2-female-talker masker threshold was 7.6 dB lower in the sex-mismatched than sex-matched condition, this difference was only 3.8 dB for adults. This masker-related discrepancy for adults was not observed in the context of speech-in-speech detection. The average sex-mismatch benefit for adults tested in Experiment 1 (2-male-talker masker) was 6.4 dB, compared with 7.1 dB for adults tested in Experiment 2 (2-female-talker masker).

One potential explanation for the apparent discrepancy in the magnitude of the sex-mismatch benefit for adults across Experiments 3 and 4 is that the 2-female-talker masker may produce less informational masking for adults than the 2-male-talker masker in the corresponding sex-matched conditions. As shown in Figure 5, the average SRT for children differed by only 2 dB across experiments for the sex-matched conditions, but the average SRT for adults was almost 5 dB lower in the female target/female masker condition (Experiment 4) than in the male target/male masker condition (Experiment 3). The reduced sex-mismatch benefit observed for adults in the 2-female-talker masker in the present study may be a byproduct of reduced informational masking in the sex-matched condition, thus reducing the potential mismatch benefit. The idea that different 2-talker masker samples are not equivalent with respect to the amount of informational masking they tend to produce is supported by previous research involving adults (e.g., Freyman et al. 2007; Calandruccio et al. 2010). For example, Freyman et al. (2007) evaluated differences in masking effectiveness for adults by assessing the recognition of nonsense sentences produced by a female in the presence of 5 different 2-female-talker maskers. At a -4 dB SNR, performance ranged from 20–70% correct across the masker samples. One possible source of the apparent difference in masking effectiveness observed between the 2-male-talker and 2-female-talker maskers in the present study relates to speaking style. In particular, the 2-female-talker masker used in Experiment 4 was composed of 2 streams of speech that were obtained from female talkers as they read aloud from familiar books designed for toddlers and preschoolers. While these talkers were not instructed to use a particular speaking style, the resulting productions shared several acoustic features with child-directed speech (e.g., exaggerated prosody).

To evaluate the possibility that differences in informational masking in the sex-matched conditions were responsible for the smaller target/masker sex-mismatch benefit observed in 2-female-talker compared with 2-male-talker speech, a fifth experiment was carried out on adults in which ideal time-frequency segregation (ITFS; Brungart et al. 2006) was applied to the stimuli used in Experiments 3 and 4. The rationale for evaluating performance with ITFS-processed speech stimuli was to estimate the contributions of energetic and informational masking to performance in both sex-matched and sex-mismatched conditions for each 2-talker masker.

EXPERIMENT 5: EVALUATING ENERGETIC AND INFORMATIONAL MASKING CONTRIBUTIONS TO PERFORMANCE FOR ADULTS USING IDEAL TIME-FREQUENCY SEGREGATION

The purpose of this experiment was to assess the relative contribution of informational and energetic masking on performance for the target/masker conditions evaluated in Experiments 3 and 4 using ideal time-frequency segregation (ITFS) processing. Previous studies have utilized ITFS to estimate the amount of energetic masking expected for various combinations of target and masker speech (e.g., Brungart et al. 2006), including sex-matched stimuli (Brungart et al. 2009). ITFS is based on the concept of the ideal binary mask, a computational technique in which an auditory mixture containing both target and masker speech is divided into time-frequency units based on estimates of peripheral time and frequency resolution. Units that exceed a pre-selected SNR are retained, but all other units

are eliminated. Applying a binary mask improves target speech intelligibility for adults, often dramatically (e.g., Brungart et al. 2006; Wang et al. 2009). Speech recognition using stimuli that have undergone ITFS is thought to reflect the limit of energetic masking, as omitting time-frequency units dominated by the masker effectively eliminates informational masking. It was hypothesized that differences in informational masking, rather than energetic masking, account for the variability in performance observed across the 4 target/masker combinations evaluated in Experiments 3 and 4. Thus, uniform speech recognition thresholds were expected across conditions once ITFS processing was applied.

Method

Listeners—Ten adults (18–29 years) participated in Experiment 5. None of the listeners participated in any of the previous experiments. The mean listener age was 23.1 years ($SD = 4.2$). All listeners met the same inclusion criteria used for adults in Experiments 3 and 4.

Stimuli, conditions and procedure—The stimuli used in Experiment 5 were created by applying an ideal binary mask to the target-plus-masker stimuli used in Experiments 3 and 4. The methods used to create the ideal binary mask were based on Brungart et al. (2006). A local criterion of -6 dB was used to specify which time-frequency epochs of the target-plus-masker combination to present to the listener. There were 2 differences between the present implementation of ITFS and that used by Brungart et al. (2006). First, Brungart et al. used 128 filters to implement the ideal-binary mask. In the present study, MATLAB was used to digitally implement the 4th order gammatone filter bank, including 35 filters with center frequencies from 50 to 11000 Hz, which overlapped at the 3-dB-down points. In contrast to the implementation of Brungart et al., reverse filtering was not performed.

As in the previous experiments, the SNR prior to ITFS processing was adjusted to estimate the SRT. However, in contrast to the previous experiments, the overall presentation level of the ITFS-processed stimulus was held constant at 70 dB SPL. The decision to fix overall level was made to account for the unfavorable SNRs expected for the adaptive track using ITFS-processed stimuli (e.g., Brungart et al. 2006).

Results and Discussion

SRTs in dB SNR prior to ITFS processing are presented in Figure 6 for each condition. Data for the 2-male-talker masker conditions are shown on the left, and data for the 2-female-talker conditions are shown on the right. Plotting conventions follow those used in Figures 1 and 2.

SRTs were more similar across the ITFS conditions of the present experiment than those observed in Experiments 3 and 4 using unprocessed recordings. In the 2-male-talker masker, the average SRT was -35.4 dB ($SD = 2.5$) in the sex-matched condition and -33.7 ($SD = 4.7$) in the sex-mismatched condition. In the 2-female-talker masker, the average SRT was -34.1 dB ($SD = 3.4$) in the sex-matched condition and -36.8 dB ($SD = 4.3$) in the sex-mismatched condition. A repeated-measures ANOVA on SRT was performed, including the within-subjects factors of Target Sex (male, female) and Masker Sex (male, female). The results of this analysis indicated a significant effect of Target Sex on threshold [$F(1,9) =$

5.61; $p = 0.04$; $\eta^2_p = 0.56$], indicating better performance with female than male target words. The main effect of Masker Sex [$F(1,9) = 0.85$; $p = 0.38$; $\eta^2_p = 0.13$] and the interaction between Target Sex and Masker Sex [$F(1,9) = 0.42$; $p = 0.53$; $\eta^2_p = 0.09$] were not significant.

Comparing results of the present experiment with adult data from Experiments 3 and 4, SRTs were substantially lower for all conditions when ITFS processing was applied compared with the corresponding conditions in Experiments 3 and 4 using unprocessed stimuli (>10 dB). In combination with the lack of a significant masker effect when ITFS processing was applied, the results of Experiment 5 are consistent with the a priori hypothesis that that better performance for adults tested in the unprocessed 2-female-talker masker (Experiment 4) relative to adults tested in the unprocessed 2-male-talker masker (Experiment 3) reflects differences in informational masking, consistent with previous data reported in the literature (e.g., Freyman et al. 2007; Calandruccio et al. 2010).

It is puzzling that school-age children did not likewise show better performance in the sex-matched 2-female-talker masker (Experiment 4) than in the sex-matched 2-male-talker masker (Experiment 3). Future studies are needed to determine the stimulus and listener factors responsible for school-age children's increased susceptibility to speech-in-speech masking relative to adults. One speculative explanation for the present pattern of results is that, while school-age children benefit from the acoustic differences between male and female talkers (e.g., relatively large differences in voice F0 and/or vocal tract length), they may have more difficulty taking advantage of more subtle acoustic differences in speech produced by different talkers of the same sex that are beneficial to experienced adult listeners. These more subtle differences may include variations in prosodic cues such as F0 contours (e.g., Binns and Culling 2007) or linguistic factors such as semantic context (e.g., Brouwer et al. 2012). It is also interesting to note that, while the average F0 of the target speech produced by the male talker (144 Hz) was higher than the average F0s of both male talkers who produced the masker speech (102 and 109 Hz) in Experiment 3, the average F0 of the target speech produced by the female talker (231 Hz) fell between the average F0s of the female talkers who produced the masker speech (227 and 243 Hz) in Experiment 4. Mackersie et al. (2011) showed asymmetric effects of F0 separation for adult listeners. When the target F0 was uncertain from one trial to the next, benefit was derived only when the target F0 was higher than the masker F0. In the absence of uncertainty, benefit of F0 separation was observed whether the target F0 was above or below the masker F0. This result suggests that the more natural listening strategy may be to attend to the voice with the higher F0. If children are less adept than adults at diverging from that natural strategy, then children may be more likely to attend to the voice with the highest F0 in the absence of large F0 differences. If that is the case, their performance should be particularly poor in the matched condition of Experiment 4, where the mean target F0 was intermediate between the two masker talker's F0s.

OVERALL SUMMARY AND CONCLUSIONS

1. Word detection (Experiments 1 and 2) and recognition (Experiments 3 and 4) thresholds for adults in a two-talker masker were substantially lower for

conditions in which the target and masker speech were mismatched in sex than when they were matched. These findings are consistent with previous data reported in the literature (e.g., Festen and Plomp 1990; Brungart 2001; Helfer and Freyman 2008).

2. Thresholds for infants were similar between sex-matched and sex-mismatched conditions. Consistent with recent data reported by Newman and Morini (2017), the lack of a sex-mismatch benefit for infants in the present study suggests that the ability to utilize target/masker sex mismatch is not established at birth.
3. School-age children showed a robust sex-mismatch benefit, indicating the ability to take advantage of differences between male and female talkers in the context of speech-in-speech recognition develops between infancy and the school-age years.
4. The procedures used to test infants (Experiments 1 and 2) differed from those used to test school-age children (Experiments 3 and 4). For infants, detection was measured using an observer-based paradigm and a fixed signal level. For school-age children, word recognition was measured in a forced choice with a fixed masker level. Future work is warranted to evaluate the possible effects of these differences on performance.
5. Although the magnitude of the sex-mismatch benefit was similar (Experiment 3) or larger (Experiment 4) for school-age children than adults, substantial child-adult differences in speech-in-speech thresholds were observed in both sex-matched and sex-mismatched conditions. In particular, children's thresholds were similar between the two-male-talker and two-female-talker sex-matched conditions, whereas adults' thresholds were substantially lower with the two-female-talker than the two-male-talker masker.

Acknowledgments

This work was supported from the National Institute of Deafness and Other Communication Disorders (R01 DC011038). Preliminary results for Experiments 1 and 2 were presented to the International Congress on Acoustics in Montreal, Canada in June 2013. Portions of the results for Experiments 3, 4, and 5 were presented to the American Auditory Society Annual Meeting in Scottsdale, AZ in March 2013. We are grateful to the members of the Human Auditory Development Laboratory, including Stephen Lockhart and Crystal Taylor.

This research was supported by the National Institutes of Health.

References

- Ambrose SE, VanDam M, Moeller MP. Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear Hear.* 2014b; 35:139–147. [PubMed: 24441740]
- ANSI. ANSI S 3.6-2010, Specifications for Audiometers. American National Standards Institute; New York: 2010.
- Barker BA, Newman RS. Listen to your mother! The role of talker familiarity in infant streaming. *Cognition.* 2004; 94:B45–B53. [PubMed: 15582622]
- Binns C, Culling JF. The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am.* 2007; 122:1765–1776. [PubMed: 17927436]
- Bolia RS, Nelson WT, Ericson MA, et al. A speech corpus for multitalker communications research. *J Acoust Soc Am.* 2000; 107:1065–1066. [PubMed: 10687719]

- Bonino AY, Leibold LJ, Buss E. Release from perceptual masking for children and adults: benefit of a carrier phrase. *Ear Hear.* 2013; 34:3–14. [PubMed: 22836239]
- Bregman AS. *Auditory Scene Analysis: The perceptual organization of sound.* MIT Press; Cambridge: 1990.
- Bronkhorst AW. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust United Acust.* 2000; 86:117–128.
- Brouwer S, Van Engen KJ, Calandruccio L, et al. Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *J Acoust Soc Am.* 2012; 131:1449–1464. [PubMed: 22352516]
- Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am.* 2001; 109:1101–1109. [PubMed: 11303924]
- Brungart DS, Chang PS, Simpson BD, et al. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J Acoust Soc Am.* 2006; 120:4007–4018. [PubMed: 17225427]
- Brungart DS, Chang PS, Simpson BD, et al. Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers. *J Acoust Soc Am.* 2009; 125:4006–4022. [PubMed: 19507982]
- Calandruccio L, Dhar S, Bradlow AR. Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J Acoust Soc Am.* 2010; 128:860–869. [PubMed: 20707455]
- Corbin NE, Bonino AY, Buss E, et al. Development of Open-Set Word Recognition in Children: Speech-Shaped Noise and Two-Talker Speech Maskers. *Ear Hear.* 2016; 37:55–63. [PubMed: 26226605]
- Darwin CJ, Brungart DS, Simpson BD. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am.* 2003; 114:2913–2922. [PubMed: 14650025]
- Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am.* 1990; 88:1725–1736. [PubMed: 2262629]
- Fitch WT, Giedd J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J Acoust Soc Am.* 1999; 106:1511–1522. [PubMed: 10489707]
- Helfer KS, Freyman RL. Aging and speech-on-speech masking. *Ear Hear.* 2008; 29:87–98. [PubMed: 18091104]
- Freyman RL, Helfer KS, Balakrishnan U. Variability and uncertainty in masking by competing speech. *J Acoust Soc Am.* 2007; 121:1040–1046. [PubMed: 17348526]
- Hall JW, Grose JH, Buss E, et al. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear.* 2002; 23:159–165. [PubMed: 11951851]
- Johnson EK, Westrek E, Nazzi T, et al. Infant ability to tell voices apart rests on language experience. *Dev Sci.* 2011; 14:1002–1011. [PubMed: 21884316]
- Johnstone PM, Litovsky RY. Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *J Acoust Soc Am.* 2006; 120:2177–2189. [PubMed: 17069314]
- Juszyk PW, Pisoni DB, Mullennix J. Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition.* 1992; 43:253–291. [PubMed: 1643815]
- Kuhl PK. Perception of auditory equivalence classes for speech in early infancy. *Infant Behav Dev.* 1983; 6:263–285.
- Lapierre MA, Piotrowski JT, Linebarger DL. Background television in the homes of US Children. *Pediatrics.* 2012; 130:839–846. [PubMed: 23027166]
- Leibold LJ, Buss E. Children's identification of consonants in a speech-shaped noise or a two-talker masker. *J Speech Lang Hear Res.* 2013; 56:1144–1155. [PubMed: 23785181]
- Leibold LJ, Werner LA. Effect of masker-frequency variability on the detection performance of infants and adults. *J Acoust Soc Am.* 2006; 119:3960–3970. [PubMed: 16838539]
- Leibold LJ, Bonino AY, Buss E. Masked Speech Perception Thresholds in Infants, Children, and Adults. *Ear Hear.* 2016; 37(3):345–353. [PubMed: 26783855]

- Levitt HC. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971; 49:467–477.
- Litovsky RY. Speech intelligibility and spatial release from masking in young children. *J Acoust Soc Am*. 2005; 117:3091–3099. [PubMed: 15957777]
- Masapollo M, Polka L, Ménard L. When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Dev Sci*. 2016; 19:318–328. [PubMed: 25754812]
- Newman RS. Infants' listening in multitalker environments: Effect of the number of background talkers. *Atten Percept Psychophys*. 2009; 71:822–836. [PubMed: 19429961]
- Newman RS, Morini G. Effect of the relationship between target and masker sex on infants' recognition of speech. *J Acoust Soc Am*. 2017; 141:EL164–EL169. [PubMed: 28253666]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychol Sci*. 1994; 5:42–46. [PubMed: 21526138]
- Olsho LW, Koch EG, Halpin CF, et al. An observer-based psychoacoustic procedure for use with young infants. *Dev Psychol*. 1987; 23:627.
- van de Weijer J. Language Input for Word Discovery. Wageningen, the Netherlands: Ponsen & Looijen, BV; 1998.
- Wang D, Kjems U, Pedersen MS, et al. Speech intelligibility in background noise with ideal binary time-frequency masking. *J Acoust Soc Am*. 2009; 125:2336–2347. [PubMed: 19354408]
- Werner LA. Issues in human auditory development. *J Comm Dis*. 2007; 40:275–283.
- White KS, Aslin RN. Adaptation to novel accents by toddlers. *Dev Sci*. 2011; 14:372–384. [PubMed: 21479106]
- Wightman FL, Kistler DJ. Informational masking of speech in children: effects of ipsilateral and contralateral distracters. *J Acoust Soc Am*. 2005; 118:3164–3176. [PubMed: 16334898]
- Wightman FL, Kistler DJ, O'Bryan A. Individual differences and age effects in a dichotic informational masking paradigm. *J Acoust Soc Am*. 2010; 128:270–279. [PubMed: 20649222]

Figure 1.

Speech detection thresholds in the 2-male-talker masker are shown for infants (left) and adults. Unfilled and shaded boxes show the range of performance spanning the 25th to the 75th percentile for listeners tested in the matched (male target) and mismatched (female target) conditions, respectively. Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines.

Figure 2.

Speech detection thresholds in the 2-female-talker masker are shown, following the format used in Figure 1. Unfilled boxes show the range of performance in the matched condition (female target), and shaded boxes show the range of performance in the mismatched condition (male target).

Figure 3.

Individual speech recognition thresholds in the 2-male-talker masker are plotted in dB SNR as a function of listener age. Circles show thresholds in the matched condition (male target) and Xs show thresholds in the mismatched condition (female target).

Figure 4.

Individual speech recognition thresholds in the 2-female-talker masker are plotted in dB SNR as a function of listener age. Circles show thresholds in the matched condition (female target) and Xs show thresholds in the mismatched condition (male target).

Figure 5.

Speech reception thresholds are shown for children (left) and adults (right) tested in Experiments 3 (2-male-talker masker) and 4 (2-female-talker masker). Unfilled and shaded boxes show the range of performance spanning the 25th to the 75th percentile for listeners tested in the matched and mismatched conditions, respectively. Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines.

Figure 6.

Speech recognition thresholds after ITFS processing was applied are plotted for each of the 4 target/masker conditions, following the same format used in Figures 1 and 2.