# Masked Speech Perception in Infants, Children and Adults

**Lori J. Leibold**[1], **Angela Yarnell Bonino**[2], and **Emily Buss**[2]

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

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

## Abstract

**Objective—**The primary goal of this study was to compare infants' susceptibility to making produced by a two-talker speech and a speech-shaped noise masker. It is well documented that school-age children experience more difficulty recognizing speech embedded in two-talker speech than spectrally matched noise, a result attributed to immaturity in the ability to segregate target from masker speech, and/or to selectively attend to the target while disregarding the perceptually similar speech masker. However, findings from infant psychophysical studies suggest that infants are susceptible to auditory masking even when target and competing sounds are acoustically distinct.

**Design—**Listeners were infants (8-10 mos), children (8-10 yrs) and adults (18-33 yrs). The task was an observer-based, single-interval disyllabic word detection, in the presence of either a speech-shaped noise or a two-talker masker. The masker played continuously at 55 dB SPL, and the target level was adapted to estimate threshold.

**Results—**As observed previously for closed-set consonant and word identification as well as open-set word and sentence recognition, school-age children experienced relatively more masking than adults in the two-talker than the speech-shaped noise masker. The novel result of this study was that infants' speech detection thresholds were about 24 dB higher than those of adults in *both* maskers. While response bias differed between listener groups, it did not differ reliably between maskers.

**Conclusions—**It is often assumed that speech perception in a speech masker places greater demands on a listener's ability to segregate and selectively attend to the target than a noise masker. This assumption is based on results showing larger child/adult differences for speech perception in a speech masker composed of a small number of talkers than in spectrally matched noise. The observation that infants experience equal masking for speech and noise maskers suggests that infants experience informational masking in both maskers and raises the possibility that the cues which make the steady noise a relatively ineffective masker for children are learned.

Address correspondence to Lori J. Leibold, Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE, USA. lori.leibold@boystown.org.

## Introduction

The primary objective of this study was to evaluate and compare infants' speech detection in a two-talker speech and a speech-shaped noise masker. Understanding how masked speech perception develops is important because infants must learn about speech and language in natural environments that often contain multiple sources of competing sounds. For instance, a number of studies have shown that infants tend to spend more time in environments containing competing sounds than they do in quiet (e.g., van de Weijer 1998; Barker and Newman 2004; Lapierre et al. 2012; Ambrose et al. 2014).

Despite the prevalence of competing sounds, we know little about the masked speech perception challenges faced by infants in their everyday lives. One reason for this gap in knowledge is that infants are exposed to different types of competing sounds. While some sounds are relatively steady-state, such as the noise produced by a ventilation system, others are complex and dynamic, such as speech. At least for adults, the masking produced by competing noise and speech appears to reflect different auditory processes. Relatively steady-state noise is thought to produce primarily *energetic* masking (Fletcher 1940), caused by overlapping patterns of excitation on the basilar membrane (see also Stone and Moore 2014). Thus, masking sounds that produce energetic masking limit the extent to which elements of the target speech are encoded by the peripheral auditory system. Competing speech composed of a small number of talkers is thought to produce both energetic and *informational* masking (e.g., Carhart et al. 1969; Brungart 2001; Freyman et al. 2004; Brungart et al. 2006). Informational masking is dependent upon higher-order perceptual processing, including the ability to segregate and/or selectively attend to target versus masker speech (e.g., Brungart 2001; Freyman et al. 2004). Increased susceptibility to either energetic or informational masking could result in infant-adult differences in masked speech perception.

Several studies have examined infants' masked speech perception abilities in the presence of Gaussian or speech-shaped noise (e.g., Trehub et al. 1981; Nozza et al. 1990; 1991; Newman 2009). The results of these studies provide evidence that infants require a more advantageous signal-to-noise ratio (SNR) than adults to detect (e.g., Trehub et al. 1981; Nozza et al. 1988) or discriminate (Nozza et al. 1990; 1991) masked speech. For example, Trehub et al. (1981) measured infants' detection thresholds in the free-field for speech embedded in Gaussian noise. Average thresholds for 6-, 12-, 18-, and 24-month-old infants were 9-16 dB higher than for adults. No significant differences in speech detection thresholds were observed between the four age groups of infants.

It is unlikely that infants' pronounced speech-in-noise difficulties relative to adults' are to the result of immature sensory encoding. Results from anatomical, physiological, morphological, and histological experiments provide converging evidence that the cochlea is developed and functionally mature by at least 3 months following term birth (e.g., Lavigne-Rebillard and Pujol 1987; Bargones and Burns 1988; Lasky et al. 1992; Kalluri and Abdala 2015). Furthermore, adult-like behavior has been observed by six months of age for indices of basic auditory processing such as frequency resolution (e.g., Olsho 1985; reviewed by Werner 2007a).

One explanation for infant/adult differences in performance on speech-in-noise tasks is that infants are susceptible to both energetic and informational masking under conditions in which adults are affected primarily by energetic masking, and that infants are susceptible to informational masking under conditions in which adults do not exhibit any masking. Specifically, it has been suggested that infants are more susceptible to informational masking than adults, reflecting immaturities of the higher-order processing required to separate or selective attend to sounds originating from different sources (e.g., Werner 2007b; Newman 2009). Results from infant psychophysical studies provide evidence in support of this hypothesis, showing elevated detection thresholds for a pure-tone signal in the presence of remote-frequency masking sounds, conditions for which robust peripheral encoding of the target speech is expected (e.g., Werner and Bargones 1991; Leibold and Werner 2006). For example, Werner and Bargones (1991) measured thresholds for a 1000-Hz tone in quiet and in the presence of high-frequency band of noise that was two octaves higher in frequency than the target tone (4000-10000 Hz). Listeners were 6-month-old infants and adults. For adults, thresholds for the 1000-Hz tone were similar in quiet and in the presence of the remote-frequency noise. In contrast, thresholds for infants were about 10 dB higher in the remote-frequency noise than in quiet. These findings suggest that adults listen in a frequency-selective manner during pure-tone detection, but that infants listen over a broad range of frequencies. As a consequence of this unselective listening strategy, infants are susceptible to informational masking by remote-frequency noise even though the peripheral auditory system provides sufficient resolution to detect the pure-tone signal.

Remote-frequency sounds also appear to interfere with infants' speech perception (Polka et al. 2008; Newman et al. 2013). Based on the psychophysical study by Werner and Bargones (1991), Polka et al. (2008) used a habituation paradigm to test 6- to 8-month-olds' discrimination of /bu/ and /gu/ in the presence of competing sounds that were not expected to overlap in frequency with the target phonemes. While the target syllables were low-pass filtered to remove energy above 4000 Hz, the remote-frequency masker was comprised of bird and cricket sounds that were high-pass filtered to remove energy below 5000 Hz. Three groups of children completed the two-phase protocol, comprising habituation and subsequent testing. Group 1 completed both phases in quiet, Group 2 completed both phases in the presence of the masker, and Group 3 habituated with the masker, but was tested in quiet. All but one infant in the group that completed the habituation phase in quiet (Group 1) was able to reliably discriminate between the two phonemes in the subsequent testing phase. In contrast, only about half of the infants who completed the habituation phase in the presence of the masker (Groups 2 and 3) showed evidence of phoneme discrimination in the testing phase, regardless of whether testing was completed in quiet or in noise. These findings suggest that infants were unable to discriminate between the phonemes in the presence of remote-frequency noise. Polka et al. (2008) interpreted their findings as support for the hypothesis that listening to speech in the presence of competing signals is more cognitively demanding for infants than it is for adults, perhaps due to immature selective attention.

The results described in the previous paragraph suggest that infants are susceptible to informational masking even when the target and masker are not similar (i.e., target speech presented in steady noise) and excite different neural populations along the basilar

membrane (i.e., target speech presented in remote-frequency noise). However, far less is currently known about infants' speech-in-speech perception, or how this ability develops between infancy and the school-age years. Given observations that infants are exposed to speech the majority of their waking hours (e.g., van de Weijer 1998; Barker and Newman 2004; Lapierre et al. 2012), this lack of knowledge restricts our understanding of infants' functional communication skills.

What is known is that recognizing speech embedded in a perceptually similar speech masker is often difficult even for most adults. Previous findings from multiple laboratories have consistently demonstrated that speech maskers tend to produce substantial informational masking (e.g., Carhart et al. 1969; Brungart 2001; Hall et al. 2002; Freyman et al. 2004). In the case of speech maskers composed of a small number of competing talkers, informational masking is often sufficient to prevent listeners from taking advantage of transient improvements in signal-to-noise ratio, which may occur at different time points across frequency. Consequently, performance is generally similar to or poorer than observed in a spectrally matched steady noise masker (e.g., Carhart et al. 1975; Freyman et al. 2004).

Speech-in-speech recognition has been evaluated in school-age children via consonant identification (Leibold and Buss 2013), forced-choice spondee identification (e.g., Hall et al. 2002), open-set monosyllabic word recognition (e.g., Corbin et al. 2015), and sentence recognition (e.g., Wightman and Kistler 2005). Consistent results have been observed across these studies, indicating a larger child/adult difference in speech maskers composed of a small number of competing talkers than in speech-shaped noise. For example, Hall et al. (2002) examined word recognition in the presence of a speech-shaped noise or a two-talker speech masker using an adaptive, closed-set paradigm. Listeners were 5- to 10-year-old children and adults. On average, children required a 3-dB higher signal than adults in the noise masker. In constrast, the average child/adult difference was 7 dB in the two-talker masker. Note also that, while adult-like performance is typically observed before 10 years of age for measures of masked speech perception in steady noise (e.g., Eisenberg et al. 2000; Nishi et al. 2010; but see McCreery and Stelmachowicz 2011), speech-in-speech recognition remains immature for some adolescence as old as 16 years of age (e.g., Wightman et al. 2003; Wightman and Kistler 2005). These data on school-age children support the idea that mastering the perceptual skills required to segregate and/or selectively attend to target from masker speech follows a more prolonged time course of development compared to when the background is relatively steady-state noise.

Although the data on speech-in-speech perception during infancy are limited, a series of studies conducted by Newman and her colleagues provide evidence that a single stream of competing speech has larger detrimental effects on speech perception during infancy compared to steady-state noise or multi-talker babble (e.g., Newman and Jusczyk 1996; Barker and Newman 2004; Newman 2005, 2009). For example, Newman (2009) found that two age groups of infants (5.5 and 8 months) preferred to listen to their own name presented at a +10 dB SNR in the presence of a multi-talker babble, but exhibited no preference at the same SNR in either single-talker speech or time-reversed single-talker speech. Newman (2009) posited that these findings may reflect infants' immature ability to segregate and/or selectively attend to the target speech in the presence of acoustically similar speech maskers.

One caveat is that a single stream of competing speech fluctuates in both amplitude and frequency over time. It is well established that speech perception in adults with normal hearing benefits from the introduction of masker level fluctuation, presumably due to the epochs of improved SNR associated with modulation minima (e.g., Dirks and Bower 1970). This benefit of modulation can be observed even when modulation minima occur at different times at low and high frequencies (e.g., Howard-Jones and Rosen 1993). Thus, an alternative explanation for infants' poorer performance in single-talker or time-reversed signal-talker speech compared to multi-talker babble discussed by Newman (2009) is that infants are poorer at listening in the minima of a fluctuating masker as compared to adults (also see Werner 2013).

The primary goal of the present study was to compare infants' susceptibility to making produced by speech with that produced by spectrally matched noise. In order to accomplish this goal, masked detection thresholds were estimated adaptively for disyllabic words presented in a continuous background of two-talker speech or speech-shaped noise. Infants (8-10 months), school-age children (8-10 years), and adults (18-26 years) were tested using the same observer-based psychoacoustic procedure, with the primary goal of comparing infants' performance across the two masker conditions. Adults were tested to obtain an estimate of mature performance. School-age children were tested to evaluate whether masker effects for speech detection are similar to previous findings for open- and closed-set speech recognition (e.g., Hall et al. 2002). Based on these previous results, we predicted that speech detection thresholds would be more similar for adults and school-age children in the speech-shaped noise masker than in the two-talker speech masker. Infants were expected to perform more poorly than school-age children and adults in both masker conditions. Two competing predictions were evaluated with respect to the pattern of infants' masked speech detection performance between the two masker conditions. The first prediction was that infant/adult differences in masked speech detection would be greater in the two-talker speech than the speech-shaped noise masker. The rationale for this prediction is that school-age children show greater immaturity in two-talker speech than in speech-shape noise (e.g., Hall et al. 2002), so infants' immature auditory processing could result in a greater effect of masker type. Alternatively, infants appear to have difficulty segregating and/or selectively attending to target sounds in the presence of *any* type of competing background sound (e.g., Werner and Bargones 1991; Polka et al. 2008). Thus, the alternative prediction was that thresholds would be more similar across the two-talker speech and speech-shaped noise masker for infants than for school-age children.

## Materials and Methods

### Listeners

Data were collected from seven infants (8.3 to 10.1 months), ten school-age children (8.2 to 10.9 years) and eight adults (18.8 to 33.1 years). Recruitment for each age group was based on a power analysis with a desired power of 80% and alpha level of 0.05. The average age at the initial testing session was 9.1 months ($SD = 0.8$ months) for infants, 9.3 years ($SD = 0.9$ years) for children, and 23.3 years ($SD = 4.8$ years) for adults. Data from four additional infants were excluded from analysis: one infant did not reach the training criterion; two

infants did not provide sufficient test data; and one infant completed testing but was excluded because of a high response rate on no-signal trials (>40%). Data from one adult were excluded due to experimenter error. An additional six infants completed testing in only one condition after three test visits; their average age was 9.3 months ($SD = 0.9$ their average age was 9.3 months (SD = 0.9 months). Three infants in this group completed testing in the speech-shaped noise masker, and the other three infants completed testing in the two-talker speech masker.

Participant selection criteria were: (1) no risk factors for hearing loss as assessed by parental report or, in the case of adults, self-report; (2) no more than two reported episodes of otitis media; (3) not under treatment for otitis media within the prior week; (4) healthy on the test day; and (5) no more than two years of musical training. In addition, screening tympanometry was performed on every listener using a 226 Hz probe tone. Peak admittance of at least 0.2 mmhos at a pressure between -200 and 50 daPa was required to pass the screening.

### Stimuli

Target stimuli were the disyllabic words "baby", "tiger", and "ice-cream". The words were recorded in isolation from an adult female speaker using a condenser microphone (AKG Acoustics) mounted approximately three inches from the speaker's mouth. Productions were amplified (TDT MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (Digital Audio Labs, CardDeluxe). Overall word durations were 946 ms (ice-cream), 979 ms (baby), and 1237 ms (tiger). The three target words were scaled to have equal total root-mean-square (rms) levels and resampled at a rate of 24.4 kHz using MATLAB. Pilot data collected from adults indicated equivalent speech detection thresholds across the three target tokens within individual listeners for each of the masker conditions described below.

Target words were presented in a continuous background of two-talker speech or speech-shaped noise. The two-talker masker consisted of two streams of meaningful speech. Each stream of speech was recorded from a different female talker reading aloud from popular children's books. The two individual speech streams were manually edited to ensure silent pauses did not exceed 300 ms, resulting in samples that were 3.5 and 3.1 minutes in duration. Each sample was repeated without discontinuity for 60 minutes. The two individual streams were balanced for overall rms level, mixed, and then down-sampled from 44.1 to 24.4 kHz with a resolution of 32 bits. The spectral envelope of this two-talker masker was used to create the speech-shaped noise masker. A 95.1-sec Gaussian noise was transformed into the frequency domain and multiplied by the magnitude spectrum of an equal-duration sample of the two-talker masker. The result was transformed back into the time domain, generating a 95.1-sec sample of noise that could be repeated without discontinuities at the beginning and end of the array.

Custom software (MATLAB) was used to control the selection and presentation of stimuli. The assigned target token and masker were mixed (TDT SM3), sent to a headphone buffer (TDT HB6), and presented to the listener's left ear via an insert earphone (Etymotic ER-1). Listeners were tested inside a double-walled, sound-treated room.

## Procedure

Listeners were randomly assigned to be tested with one of the three target words. Masked speech detection thresholds for the selected target word were estimated in each of the two masker conditions (speech-shaped noise and two-talker speech), with testing order counterbalanced across listeners. The same target word was used to estimate thresholds in both masker conditions for a given listener. Adults and school-aged children were tested in a single visit to the laboratory. Infants were tested in three separate visits occurring within a 2-week period. Visits for all three age groups were approximately 45 minutes in length.

A single-interval, observer-based psychophysical procedure was used to assess performance for all listeners (Olsho et al. 1987). Infants were tested while sitting on their parents' laps. 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 target words and inadvertently influencing the infant's response, the assistant and parent wore circumaural headphones that delivered speech-shaped noise that masked the stimuli presented to the infant. To the infants' left were two mechanical toys, each housed in a dark Plexiglas box with a computer-controlled light. An observer sat outside of the booth and initiated trials when the infant was quiet and facing midline.

The masker was presented continuously throughout testing at an overall level of 55 dB SPL. Trials were either *signals*, in which the target word was presented, or *catch trials*, in which no target word was presented. The observer sitting outside the booth did not know which type of trial occurred and was required to decide the trial type based on the infants' behavior within 4 sec of trial onset. The most common infant behaviors that observers used to decide whether or not a signal occurred were head turns, eye movements, and changes in general motor activity. Listener reinforcement was the activation and illumination of a mechanical toy after presentation of the signal. The observer was provided with feedback after every trial.

The procedure used to test school-age children and adults was the same as that used for infants. However, children and adults were alone in the booth during testing and were asked to raise their hand when they heard the target word. Consistent with the infant testing procedure, an observer seated outside the booth indicated when a response was observed to end the trial and provide the toy reinforcement.

A complete session included a conditioning phase, a training phase, and a testing phase. The target word was presented at a level that was expected to be clearly audible for both the conditioning and training phases, depending on the masker type and age group of the listener. The goal of the conditioning phase was to familiarize the listener with the relationship between the presentation of the target word and the mechanical toy reinforcement. In this phase 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 conditioning phase was completed when the observer correctly responded to four of five consecutive trials, including at least one catch trial. The goal of the 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 the training phase was 0.50. Reinforcement was only provided to the listener if the observer correctly identified a signal trial. The training phase was completed after a run of 10 sequential trials associated with a hit rate of 0.80 or higher, and a false alarm rate of 0.20 or lower. The average number of trials required to complete the training phase for the noise masker condition was 20.0 for infants ($SD$ = 11.4), 11.0 for children ($SD$ = 1.6), and 11.9 for adults ($SD$ = 2.2). The average number of trials required to complete the training phase for the two-talker masker condition was 29.2 for infants ($SD$ = 22.7), 10.2 for children ($SD$ = 1.5), and 10.0 for adults ($SD$ = 1.2).

During the testing phase, thresholds for the target word were measured adaptively using a 2-down, 1-up procedure (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 training level. Only signal trials were included in the adaptive track. The starting level for the target word was about 10 dB higher than the expected threshold value, based on pilot data. The initial step size was 4 dB. The step size was reduced to 2 dB after the second track reversal. Eight reversals were obtained, and threshold was computed as the mean signal level at the last six reversals. Thresholds were only accepted if the proportion of responses to probe trials was 0.60 or higher, and the proportion of responses to catch trials was 0.40 or lower. A single threshold estimate was obtained in each masker condition.

## Results

### Within-subjects data

Figure 1 shows average thresholds for listeners completing testing in both masker conditions for each age group. The filled circles and open squares show estimates in the presence of the speech-shaped noise and two-talker speech maskers, respectively. The average infant threshold was similar in speech-shaped noise (68.7 dB SPL) and in two-talker speech (69.4 dB SPL). The average adult threshold was about 24 dB lower than the average infant threshold in both maskers. Similar to the infant data, the average adult threshold was about the same in the presence of speech-shaped noise (44.9 dB SPL) and two-talker speech (45.2 dB SPL). A discrepancy in threshold estimates across the two masker conditions was observed in the children's data. The average child threshold was similar to the average adult threshold in the speech-shaped noise masker (42.9 dB SPL), but was 6.5 dB higher than the adult threshold in the two-talker masker (51.7 dB SPL).

A repeated-measures analysis of variance (ANOVA) confirmed the trends observed in Figure 1. All of the effects in the Masker Type X Age Group analysis were significant: Masker Type [$F(1,22)$ = 7.67; $p$ = 0.01; $\eta_p^2$ = 0.26], Age Group [$F(2,22)$ = 164.81; $p$ <0.001; $\eta_p^2$ = 0.94], and Masker Type X Age Group [$F(2,22)$ = 6.08; $p$ < 0.01; $\eta_p^2$ = 0.36]. The significant Masker Type X Age Group interaction confirms that the difference between masker conditions is not the same for the three age groups. To explore the nature of the interaction, a paired-samples $t$-test was performed within each age group. Thresholds for school-age children were higher in the two-talker than in the noise masker [$t(9)$ = -4.75; $p$ = 0.001]. No significant difference in thresholds was observed for infants [$t(6)$ = -0.47; $p$ = 0.66] or for

adults [$t$(7) = -0.12; $p$ = 0.91]. Overall, these results support the conclusion that masker type affects the performance of school-age children, but not that of infants or adults.

Figure 2 shows threshold estimates for individual listeners tested in both masker conditions, with infant data presented in the left panel, child data presented in the middle panel, and adult data presented in the right panel. The filled circles and open squares show thresholds for the speech-shaped noise and two-talker masker conditions, respectively. The listeners are ordered by age from youngest to oldest. The effects observed in the individual data are generally consistent with those in the group data. For the two-talker masker condition, thresholds for all seven infants were higher than thresholds for school-aged children and adults. Thresholds for children were generally higher than those for adults in the two-talker masker, although one child (C4) performed comparable to adults in this condition, and one adult (A6) had a threshold higher than the mean threshold for children. For the speech-shaped noise condition, thresholds for all infants were higher than those for children and adults. No systematic differences between children and adults were observed in the individual data for the speech-shaped noise condition.

The difference in threshold between the two masker conditions was 5 dB or less for all but one infant (I2), who performed 6.6-dB worse in the speech-shaped noise than the two-talker speech masker. Similar to infants, only two adults showed a threshold difference between the two masker conditions larger than 3 dB: one adult performed 6-dB worse in the speech-shaped noise than the two-talker condition (A5), and one adult performed 17-dB worse in the two-talker than the speech-shaped noise condition (A6). A different pattern of individual results was observed for children than for infants and adults. Thresholds for all ten children were higher in the two-talker than the speech-shaped noise masker, with differences ranging from 1.4 to 18.6 dB.

### Data for infants who completed testing in only one condition

A subset of six infants completed tested in only a single masker condition. Because of the large individual differences in masking often observed for infants in masking experiments, it is important to evaluate whether performance for these infants is consistent with the data from the seven infants who provided data points in both masker conditions. Thus, a between-subjects analysis of threshold was performed on the data of 13 infants. This analysis included data from the six infants who completed a single masker condition (n=3 in each masker), as well as threshold estimates in the first masker condition completed by the seven infants who provided data points in both masker conditions (n=4 in speech-shaped noise, n=3 in two-talker speech). Figure 3 shows the average infant thresholds in dB SPL, with each listener contributing data in only one of the two masker conditions. Data for listeners completing both masker conditions, shown in Figures 1 and 2, are re-plotted for comparison. Consistent with the previous analysis, the average threshold for infants tested in the speech-shaped noise masker in their first or only condition (67.9 dB SPL) was similar to the average threshold for infants tested in the two-talker speech masker in their first or only condition (68.6 dB SPL). Moreover, thresholds differed by less than 1 dB for the listeners providing data in one vs. two masker conditions. Recall that the average threshold for infants tested in both masker conditions was 68.7 dB SPL in speech-shaped noise and 69.4 dB SPL in two-

talker speech. A one-way ANOVA of threshold confirmed that that the main effect of masker condition was not significant [$F(1, 11) = 0.56$; $p = 0.47$].

### Response bias

Adults and school-age children typically exhibit a conservative response bias during single-interval procedures in which the temporal interval is undefined (e.g., Marshall and Jesteadt 1986; Bonino and Leibold 2008); this is true whether the listener responds directly (e.g., by pressing a button) or an observer-based method is used to identify signal-present responses (e.g., Werner and Marean 1991; Leibold and Werner 2006). In contrast, data from infants tested in the observer-based method tend to be unbiased (e.g., Werner and Marean 1991; Leibold and Werner 2006). These age effects in response bias can result in differences in d′ at threshold between the infant-observer team and the older age groups (e.g., Leibold and Werner 2006). Thus, an additional analysis was performed to evaluate whether the age effects reported here were the result of differences in response bias across age groups and/or across masker conditions within each age group. Estimates of d′ near threshold were compared across the three age groups and two masker conditions. Estimates of d′ were calculated for each listener in each masker condition based only on trials near threshold (within ± 1 SD of threshold for each listener). For cases in which the false alarm rate was 0, a value of 0.5 was added to all cells (Snodgrass and Corwin 1988).

Table 1 shows the group average estimates of d′ near threshold for both masker conditions. As expected based on previous studies (e.g., Leibold and Werner 2006), the average d′ estimates for infants were close to 1.0, consistent with 71%-correct performance for an unbiased observer. The average d′ estimates for school-age children and adults were higher than those for infants, with average values ranging from 1.5 to 1.8. Based on this analysis, signals presented at each listener's 71%-correct threshold would have been identified 80% of the time if child and adult listeners had been unbiased. No evidence of systematic differences in estimates of d′ was observed between the two masker conditions for the group of school-age children. Note, however, the average estimate of d′ in speech-shaped noise was slightly higher than in two-talker speech for infants, whereas the opposite trend was observed for adults. The results of a repeated-measures ANOVA performed on individual listener's estimates of d′ were generally consistent with the trends shown in Table 1. The main effect of Age was significant [$F(2, 22) = 30.48$; $p < 0.001$; $\eta_p^2 = 0.74$], indicating lower d′ estimates for infants compared to the two older age groups. The main effect of Masker Type [$F(1,22) = 0.33$; $p = 0.57$] was not significant. The Masker Type X Age interaction [$F(2, 22) = 3.25$; $p = 0.06$] also failed to reach statistical significance. These results are inconsistent with the possibility that differences in response bias are responsible for the pattern of threshold differences between masker conditions within each age group.

## Discussion

The results of this study indicate that 8- to 10-month-old infants have substantially more difficulty detecting disyllabic words in the presence of either speech-shaped noise or two-talker speech than 8- to 10-year-old children or adults. For example, average masked detection thresholds were 24-dB higher for infants than adults for both masker conditions

using the same psychophysical testing procedure. These age effects are consistent with previous studies of infants' tone (e.g., Bargones et al. 1995; Leibold and Werner 2006) and speech (e.g., Trehub et al. 1981; Nozza et al. 1988) detection in noise, providing converging evidence that infants are more susceptible to auditory masking than older listeners.

Results for 8- to 10-year-old children are in line with data reported in earlier studies of masked speech recognition involving school-age children (e.g., Hall et al. 2002; Wightman and Kistler 2005; Bonino et al. 2013). These data suggest different developmental trajectories for speech perception in competing noise versus competing speech. The average detection threshold for children was the same as that observed for adults in the speech-shaped noise masker, but was 6.5-dB higher in the two-talker speech masker. Hall et al. (2002) observed a comparable child-adult difference in spondee identification thresholds in a continuous two-talker masker using a forced-choice recognition task. Consistent findings across studies involving school-age children tested on different speech perception measures provide evidence that mastering the perceptual skills required to segregate and/or selectively attend to target speech in a speech masker requires more extensive listening experience and/or neural maturation than listening in relatively steady-state noise.

Infants performed more poorly than school-age children and adults, but infants' speech detection thresholds were similar for the two masker conditions. Comparable thresholds across the two maskers were obtained in the data on infants who completed both conditions, as well as data in the first (or only) condition completed by each infant. These results appear to be in conflict with data reported by Newman (2009), who found that that 5.5- and 8.5-month-old infants failed to show a preference to listen to their own name at a +10 dB SNR in a single stream of competing speech (original or time reversed), but did show a preference to listen to their own name in a nine-talker babble. While the babble was composed of individual streams of speech, several investigators have shown a sharp reduction in informational masking as the number of talkers added to the masker stream exceeds about three or four (e.g., Freyman et al. 2004). Presumably, the multi-talker masker begins to approximate steady-state noise as additional talkers are added, decreasing target/masker similarity. The most obvious difference between the present study and Newman (2009) relates to the one- and two-talker masker stimuli. The silent gaps present in a two-talker masker tend to be briefer than those in a one-talker masker (e.g., Freyman et al. 2004; Rosen et al. 2013), reducing opportunities for listeners to take advantage of epochs with a favorable SNR. As discussed in the introduction to this paper, infants appear immature in their ability to listen in the minima of a modulated masker (e.g., Werner 2013). We did not include a single-talker masker condition in the present study; thus, we cannot rule out the possibility that infants' speech detection thresholds would be higher in a single-talker compared to a two-talker masker.

A second major difference between the present study and Newman (2009) is that we used an adaptive, observer-based psychophysical procedure to estimate detection thresholds (e.g., Olsho et al. 1987). In contrast, Newman (2009) examined infants' masked speech perception at a fixed SNR using a preferential looking paradigm; infants in that study may have found the single-talker maskers more interesting than the nine-talker babble. That is, the looking time associated with the target speech tokens may have been influenced by the perceptual

salience of the masker stimuli. Interestingly, Barker and Newman (2004) demonstrated in a later study that infants prefer to listen to their mother's voice compared to the voice of an unfamiliar woman in the presence of background talkers. Future investigations using psychophysical procedures are needed to determine the role of preference in the preferential looking versus the observer-based method, as well as the role of preference on infants' overall speech-in-speech perception abilities.

The major question raised by the results of the present study is why infants' speech detection thresholds were the same in both masker conditions, whereas children's thresholds were different. One possible explanation is that infants, at least those younger than 10 months of age, do not have enough listening experience to be proficient at segregating sounds, selectively attending to relevant sounds, and/or knowing what competing background sounds they should ignore in a particular listening environment. The limited experience of school-age children may be sufficient for performing the detection task in speech-shaped noise, but not in the more perceptually similar two-talker masker. Reduced listening experience may present a more severe limitation for infants than for school-age children, making it difficult for infants to selectively attend, and thus detect, target speech both masker conditions. Consistent with this idea, Werner (2007b) suggested that infants' failure to listen selectively in the frequency dimension reflects an innate strategy that supports rapid learning about the important features of speech across different languages, speakers, and environments. While an unselective or "broadband" listening strategy may facilitate rapid speech and language acquisition, it could come at the cost of increased susceptibility to masking.

This study was focused on maturational effects in auditory masking, and we hypothesize that infants' pronounced susceptibility to both noise and speech maskers reflects immaturity in the higher-order processes that underlie auditory scene analysis (Bregman 1990). Another way of viewing these data, however, is in the context of how children develop the perceptual strategies needed to recognize the acoustic components of speech that are inherent to their native language. Nittrouer and colleagues have suggested that young school-age children are more "obliged" than adults to integrate across the spectrum of speech, at least for sounds that are likely to have been produced by a human vocal tract (e.g., Nittrouer and Crowther 2001; Nittrouer and Tarr 2011). The present results are consistent with this more restrictive argument, in that masked detection thresholds for school-age children were substantially elevated relative to adults in the speech, but not the noise masker. On the other hand, infants showed similar performance for the two masker conditions. It is possible that infants tend to integrate all sounds in their environment, including sounds that are unlikely to have been generated by a human vocal tract. It may not be until children gain experience with speech and language that more sophisticated, speech-specific perceptual strategies emerge.

One limitation of this study is the age difference in response bias between infants and both age groups of older listeners. Estimates of $d'$ near threshold suggest that school-age children and adults were conservative in their response bias, whereas infant-plus-observer teams were unbiased. Thus, it is likely that the infant/child and infant/adult differences in threshold measured using the adaptive, single-interval procedure underestimate differences in sensitivity between the age groups. Note, however, that response bias was similar across the

two masker conditions within each age group, so bias does not affect interpretation of within-listener effects of masker type.

In summary, the results of this study are in agreement with earlier work showing that infants have considerable difficulty perceiving speech in the presence of competing background sounds. School-age children, as predicted, showed elevated speech detection thresholds relative to adults in two-talker speech, but not in speech-shaped noise. Surprisingly, similar thresholds were observed for infants for the two masker conditions. One practical implication of these results is that a relatively simple listening situation for an adult may be a difficult listening situation for an infant, perhaps due to a lack of experience with sound that limits reconstruction of the auditory scene.

## Acknowledgments

## References

Ambrose SE, VanDam M, Moeller MP. Linguistic Input, Electronic Media, and Communication Outcomes of Toddlers with Hearing Loss. Ear Hear. 2014; 35:139–147. [PubMed: 24441740]

Bargones JY, Burns EM. Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults. J Acoust Soc Am. 1988; 83:1809–1816. [PubMed: 3403795]

Bargones JY, Werner LA, Marean GC. Infant psychometric functions for detection: Mechanisms of immature sensitivity. J Acoust Soc Am. 1995; 98:99–111. [PubMed: 7608411]

Barker BA, Newman RS. Listen to your mother! The role of talker familiarity in infant streaming. Cognition. 2004; 94:B45–B53. [PubMed: 15582622]

Bonino AY, Leibold LJ. The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex. J Acoust Soc Am. 2008; 124:321.

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. Cambridge, MA: MIT Press; 1990.

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, Simpson BD, Ericson MA, et al. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am. 2001; 110:2527–2538. [PubMed: 11757942]

Carhart R, Tillman T, Greetis R. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am. 1969; 45:694–703. [PubMed: 5776931]

Carhart R, Johnson C, Goodman J. Perceptual masking of spondees by combinations of talkers. J Acoust Soc Am. 1975; 58:S35.

Corbin NE, Bonino AY, et al. Development of Open-Set Word Recognition in Children: Speech-Shaped Noise and Two-Talker Speech Maskers. Ear Hear. 2015 published-ahead-of-print.

Dirks DD, Bower D. Effect of forward and backward masking on speech intelligibility. J Acoust Soc Am. 1970; 47:1003–1008. [PubMed: 5443146]

Eisenberg LS, Shannon RV, et al. Speech recognition with reduced spectral cues as a function of age. J Acoust Soc Am. 2000; 107:2704–2710. [PubMed: 10830392]

Fletcher H. Auditory patterns. Rev Mod Phys. 1940; 12:47.

Freyman RL, Balakrishnan U, Helfer KS. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J Acoust Soc Am. 2004; 115:2246–2256. [PubMed: 15139635]

Hall JW III, 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]

Howard-Jones PA, Rosen S. Uncomodulated glimpsing in "checkerboard" noise. J Acoust Soc Am. 1993; 93:2915–2922. [PubMed: 8315155]

Kalluri R, Abdala C. Stimulus-frequency otoacoustic emissions in human newborns. J Acoust Soc Am. 2015; 137:EL78–EL84. [PubMed: 25618103]

Lapierre MA, Piotrowski JT, et al. Background television in the homes of US children. Pediatrics. 2012; 130:839–846. [PubMed: 23027166]

Lasky R, Perlman J, Hecox K. Distortion-product otoacoustic emissions in human newborns and adults. Ear Hear. 1992; 13:430–441. [PubMed: 1487105]

Lavigne-Rebillard M, Pujol R. Surface aspects of the developing human organ of Corti. Acta Oto-laryngol. 1987; 104:43–50.

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]

Levitt HCCH. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971; 49:467–477. [PubMed: 5541744]

Marshall L, Jesteadt W. Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures. J Speech Hear Res. 1986; 29:82–91. [PubMed: 3702383]

McCreery RW, Stelmachowicz PG. Audibility-based predictions of speech recognition for children and adults with normal hearing. J Acoust Soc Am. 2011; 130:4070–4081. [PubMed: 22225061]

Newman RS, Jusczyk PW. The cocktail party effect in infants. Percept Psychophys. 1996; 58:1145–1156. [PubMed: 8961826]

Newman RS. The cocktail party effect in infants revisited: listening to one's name in noise. Dev Psych. 2005; 41:352–362.

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, Chatterjee M. Infants' name recognition in on-and off-channel noise. J Acoust Soc Am. 2013; 133:EL377–EL383. [PubMed: 23656097]

Nishi K, Lewis DE, Hoover BM, et al. Children's recognition of American English consonants in noise. J Acoust Soc Am. 2010; 127:3177–3188. [PubMed: 21117766]

Nittrouer S, Crowther CS. Coherence in children's speech perception. J Acoust Soc Am. 110:2129–2140. [PubMed: 11681390]

Nittrouer S, Tarr E. Coherence masking protection for speech in children and adults. Att Percept Psychophys. 2011; 73:2606–2623.

Nozza RJ, Wagner EF, Crandell MA. Binaural release from masking for a speech sound in infants, preschoolers, and adults. J Speech Hear Res. 1988; 31:212–218. [PubMed: 3398495]

Nozza RJ, Rossman RN, Bond LC, et al. Infant speech-sound discrimination in noise. J Acoust Soc Am. 1990; 87:339–350. [PubMed: 2299044]

Nozza RJ, Miller SL, Rossman RN, et al. Reliability and validity of infant speech-sound discrimination-in-noise thresholds. J Acoust Soc Am. 1991; 34:643–650.

Olsho LW. Infant auditory perception: Tonal masking. Inf Beh Dev. 1985; 8:371–384.

Olsho LW, Koch EG, Haplin CF, et al. An observer-based psychoacoustic procedure for use with young infants. Dev Psychol. 1987; 23:627–640.

Polka L, Rvachew S, Molnar M. Speech perception by 6-to 8-month-olds in the presence of distracting sounds. Infancy. 2008; 13:421–439.

Rosen S, Souza P, Ekelund C, et al. Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. J Acoust Soc Am. 2013; 133:2431–2443. [PubMed: 23556608]

Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. J Exp Psychol Gen. 1988; 117:34–50. [PubMed: 2966230]

Trehub SE, Bull D, Schneider BA. Infants' detection of speech in noise. J Speech Hear Res. 1981; 24:202–206. [PubMed: 7265935]

Stone MA, Moore BCJ. On the near non-existence of "pure" energetic masking release for speech. J Acoust Soc Am. 2014; 135:1967–1977. [PubMed: 25234995]

van de Weijer, J. Unpublished PhD. University of Nijmegen; Max Planck Institute, The Netherlands: 1998. Language input for word discovery.

Werner LA, Bargones JY. Sources of auditory masking in infants: Distraction effects. Percept Psychophys. 1991; 50:405–412. [PubMed: 1788029]

Werner LA, Marean GC. Methods for estimating infant thresholds. J Acoust Soc Am. 1991; 90:1867–1875. [PubMed: 1960280]

Werner LA. Issues in human auditory development. J Comm Dis. 2007a; 40:275–283.

Werner LA. What do children hear? How auditory maturation affects speech perception. The ASHA Leader. 2007b Mar 27.

Werner LA. Infants' detection and discrimination of sounds in modulated maskers. J Acoust Soc Am. 2013; 133:4156–4167. [PubMed: 23742367]

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, Callahan M, Kistler DJ. A cocktail-party listening experiment with children. J Acoust Soc Am. 2003; 113:2208–2209.
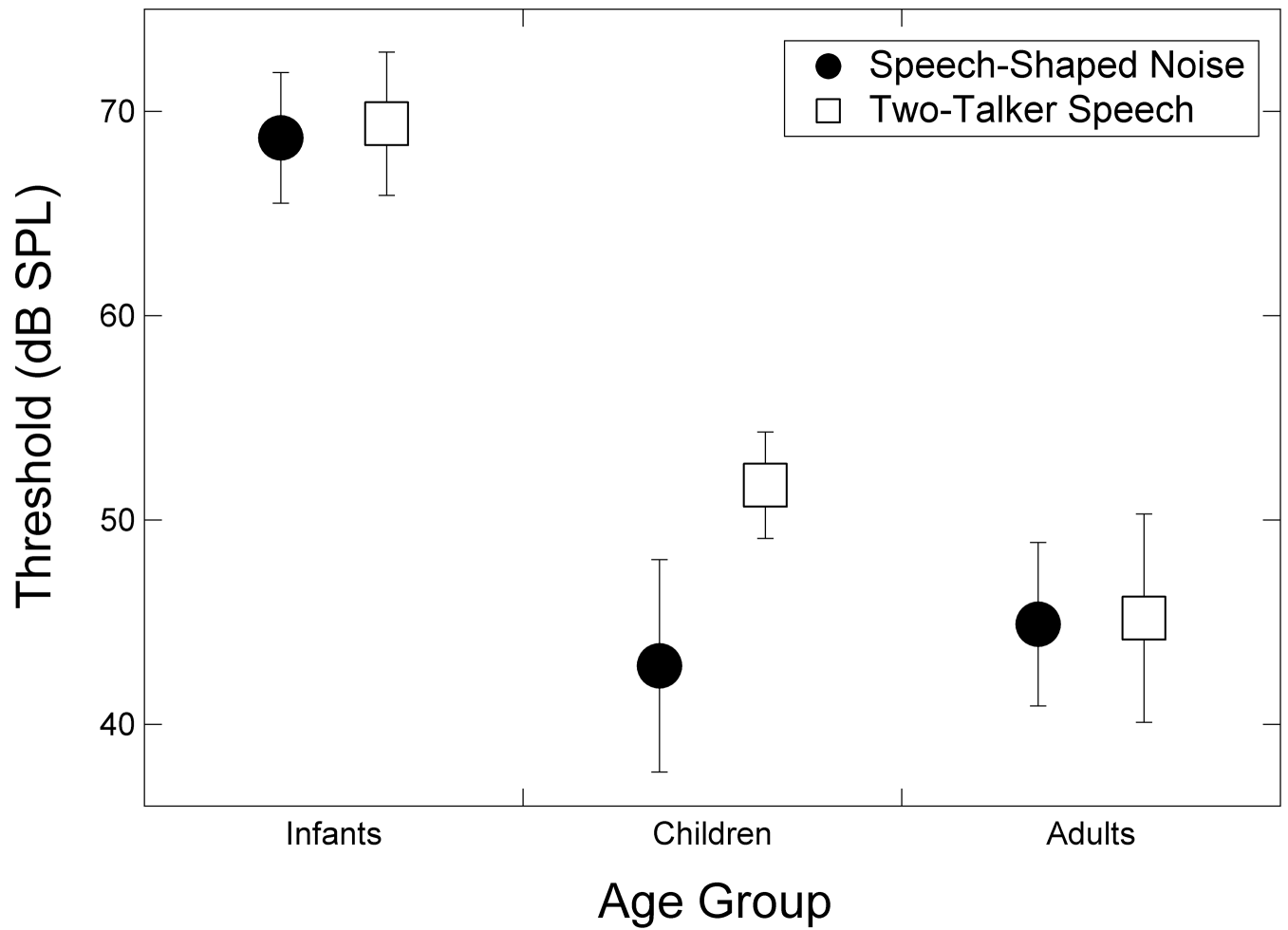
**Figure 1.**
Group average speech detection thresholds (in dB SPL) are shown for the infants (n=7), school-age children (n=10) and adults (n=8) who completed testing in both masker conditions. Thresholds in speech-shaped noise and two-talker speech are shown by the filled circles and open squares, respectively. Error bars represent plus or minus one standard error of the mean.
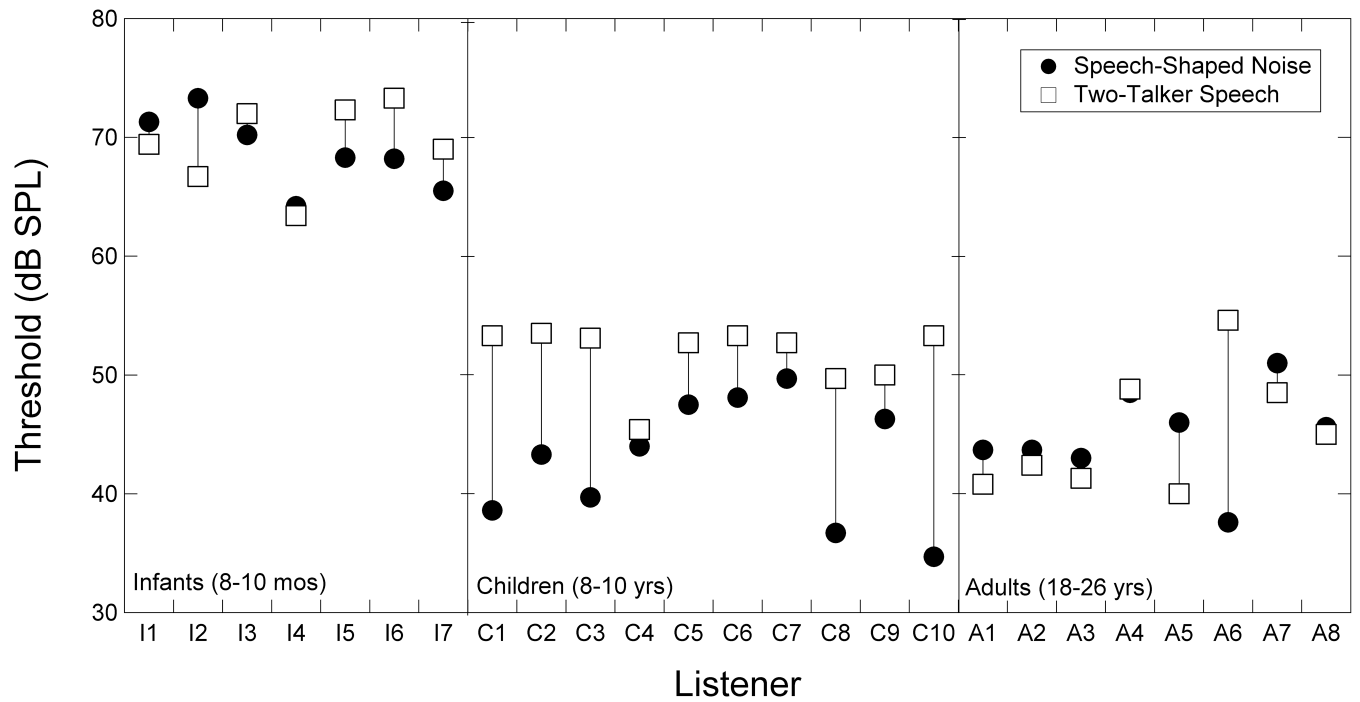
**Figure 2.**
Speech detection thresholds (in dB SPL) are shown for the individual infants (n=7; left panel), school-age children (n=10; middle panel) and adults (n=8; right panel) who completed testing in both masker conditions. Thresholds in speech-shaped noise and two-talker speech are shown by the filled circles and open squares, respectively.
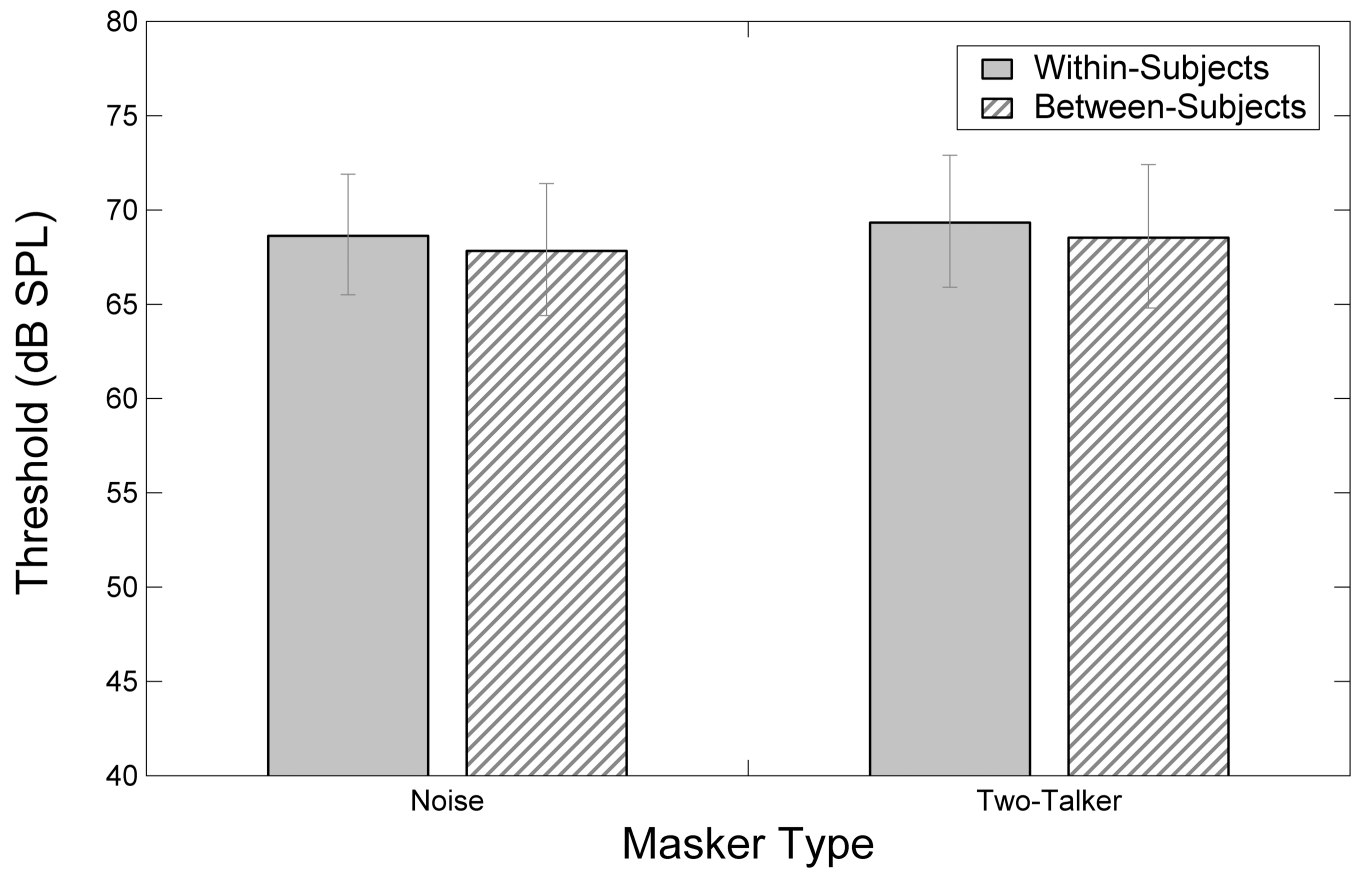
**Figure 3.**
Group average speech detection thresholds (in dB SPL) are shown for seven infants tested in the speech-shaped noise masker (filled bars) and six different infants tested in the two-talker speech masker (open bars). Data for the seven infants tested in both masker conditions are re-plotted from Figure 1. Error bars represent plus or minus one standard error of the mean.

**Table 1**

Average estimates of d′ across listeners are provided for each age group and masker condition. Standard deviations of the mean (SD) and range of d′ estimates across listeners are also provided for each dataset.

| Infants | | | | Children | | | | Adults | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Masker** | **d′** | **SD** | **Range** | **Masker** | **d′** | **SD** | **Range** | **Masker** | **d′** | **SD** | **Range** |
| Shaped Noise | 1.05 | 0.49 | 0.52-1.46 | Shaped Noise | 1.69 | 0.32 | 1.09-2.30 | Shaped Noise | 1.50 | 0.16 | 1.34-1.82 |
| Two-Talker | 0.83 | 0.34 | 0.39-1.22 | Two-Talker | 1.72 | 0.25 | 1.16-2.00 | Two-Talker | 1.81 | 0.19 | 1.47-2.03 |