HHS Public Access
Author manuscript
J Speech Lang Hear Res. Author manuscript; available in PMC 2015 June 25.

Published in final edited form as:
J Speech Lang Hear Res. 2014 February ; 57(1): 327–337. doi:10.1044/1092-4388(2013/12-0287).

Children's perception of speech produced in a two-talker background

Mallory Baker[1], **Emily Buss**[2], **Adam Jacks**[1], **Crystal Taylor**[1], and **Lori J. Leibold**[1]

[1]Department of Allied Health Sciences, The University of North Carolina at Chapel Hill

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

Abstract

Purpose—This study evaluated the degree to which children benefit from the acoustic modifications made by talkers when they produce speech in noise.

Method—A repeated-measures design compared the speech perception performance of children (5–11 years) and adults in a two-talker masker. Target speech was produced in a two-talker background or in quiet. In experiment 1, recognition with the two target sets was assessed using an adaptive spondee identification procedure. In experiment 2, the benefit of speech produced in a two-talker background was assessed using an open-set, monosyllabic word recognition task at a fixed SNR.

Results—Children performed more poorly than adults, regardless of whether the target speech was produced in quiet or in a two-talker background. A small improvement in the SNR required to identify spondees was observed for both children and adults using speech produced in a two-talker background (experiment 1). Similarly, average open-set word recognition scores were 11 percentage points higher for both age groups using speech produced in a two-talker background compared to quiet (experiment 2).

Conclusions—The results indicate that children can use the acoustic modifications of speech produced in a two-talker background to improve masked speech perception, as previously demonstrated for adults.

INTRODUCTION

Speech recognition in noisy environments can be challenging for listeners of all ages, but the detrimental effects of auditory masking appear to be particularly large during childhood (Elliott et al., 1979; Fallon, Trehub, & Schneider, 2000; Hall, Grose, Buss, & Dev, 2002; Nittrouer & Boothroyd, 1990; Wilson, Farmer, Gandhi, Shelburne, & Weaver, 2010). For example, Fallon et al. (2000) examined sentence recognition in the presence of an eight-talker babble in three groups of children (5, 9, and 11 years) and in a group of adults (19–28 years). All three age groups of children required a more favorable signal-to-noise ratio (SNR) to perform as well as the adults. This child-adult difference was largest for the youngest children tested (5 years of age), who required an additional 5 dB SNR to perform as well as the adults in the multi-talker babble.

The increased difficulties experienced by children in the presence of competing sounds relative to adults do not appear to reflect an immature peripheral auditory system. A basic assumption underlying most contemporary research in auditory development is that the sensory system (i.e., the cochlea and auditory nerve) provides the brain with an accurate representation of sound by at least six months after full-term birth (reviewed by Buss, Hall, & Grose, 2012). Instead, children's relatively poor performance on measures of masked speech recognition is believed to be the result of immature perceptual processing within the central auditory system. One line of evidence supporting this hypothesis comes from studies showing larger age effects with increased stimulus complexity (e.g., Bonino, Leibold, & Buss, 2012; Hall et al., 2002; Wightman, Callahan, Lutfi, Kistler, & Oh, 2003). For example, Hall et al. (2002) compared spondee identification performance between children (5–10 years) and adults in the presence of a speech-shaped noise and a two-talker speech masker. When the masker was presented continuously, children required an additional 3 dB, on average, to perform as well as adults in the speech-shaped noise masker. Children's disadvantage relative to adults increased to 7 dB in the two-talker masker. Originally named "perceptual masking" (Carhart, Tillman, & Greetis, 1969), children's increased difficulties in the two-talker compared to the noise masker may reflect immaturity in how children segregate the target word from the competing background speech and/or how they selectively attend to that word (e.g., Hall et al., 2002; Leibold, 2012). Borrowing terminology from related psychoacoustic studies using multi-tonal stimuli (e.g., Kidd, Mason, Deliwala, Woods, & Colburn, 1994; Oh & Lutfi, 1998), the relatively poor speech recognition performance observed in the presence of maskers comprised of a small number of talkers has been referred to more recently as "informational masking" (e.g., Brungart, Simpson, Ericson, & Scott, 2001; Freyman, Balakrishnan, & Helfer, 2004). In this context, informational masking refers to masking that is produced despite an adequate peripheral representation of the spectral and temporal properties of the target and masker speech stimuli (e.g., Brungart, 2005; Bennett, Billings, Molis, & Leek, 2012). For example, Brungart (2005) stated that "informational masking occurs when the signal and masker are both audible but the listener is unable to disentangle the elements of the target signal from a similar-sounding distractor" (p. 261).

Informational masking is believed to pose a significant problem for children in their everyday listening situations (e.g., Hall et al., 2002; Bonino et al., 2012). In contrast to stimuli used in most laboratory experiments, however, talkers in natural acoustic environments modify their speech productions when competing sounds are present. For both sentences and words, these modifications are collectively referred to as the *Lombard effect* (Lombard, 1911). Several factors have been shown to influence the Lombard effect, including the type and level of the background sounds (e.g., Cooke & Lu, 2010; Dreher & O'Neill, 1957; Summers et al., 1988), the procedures used to obtain speech productions (e.g., Patel & Schell, 2008), and the linguistic content of the speech materials (e.g., Patel & Schell, 2008; Rivers & Rastatter, 1985). Nonetheless, results obtained across multiple laboratories have consistently demonstrated that talkers increase their vocal intensity, word duration, and spectral slope in the presence of competing background sounds (e.g., Cooke & Lu, 2010; Letowski, Frank, & Caravella, 1993; Lombard, 1911; Patel & Schell, 2008; Pittman & Wiley, 2001; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). For example, Patel and Schell (2008) asked 16 adults (8 males and 8 females) to provide spoken instructions to a listener follow animated scenes. These spoken instructions were recorded in quiet and while listening to multi-talker noise at 90 dB SPL, and acoustic analyses were performed on different types of words (e.g., verbs) produced in the two listening environments (i.e., quiet and noise). Compared to words produced in quiet, words produced in noise had a 16-dB increase in peak intensity, a 57-ms increase in duration, and a 55-Hz higher peak fundamental frequency. In addition, the spectral slope was flatter (less negative) for speech produced in noise, reflecting relative increases in high-frequency energy in the speech signal.

The acoustic modifications made by talkers in noisy environments support improvements in masked speech perception for adults, even when speech produced in noise and speech produced in quiet are equated for overall level (e.g., Dreher & O’Neill, 1958; Pittman & Wiley, 2001; Summers et al., 1988). Pitman and Wiley (2002) demonstrated the perceptual benefit of these modifications in a group of 30 adults. The speech stimuli were sentences produced in each of three different conditions: (1) in quiet; (2) in broadband noise presented at an overall level of 80 dB SPL; and (3) in multi-talker babble presented at an overall level of 80 dB SPL. Consistent with previous studies examining speech production, the intensity, duration, and spectral slope of speech produced while listening to either background were significantly different from that produced in quiet. No differences in production were found between the speech produced in the broadband noise and the speech produced in the multi-talker babble. Using target words extracted from each of the three types of sentence productions (produced in quiet, produced in broadband noise, produced in multi-talker babble), speech perception was assessed in the presence of the multi-talker babble at 0, −5, and −10 dB SNR. Speech recognition scores were higher for speech produced in either the broadband noise or the multi-talker babble compared to speech produced in quiet at all three SNRs examined.

Although multiple studies have shown that children require a higher SNR than adults to recognize speech in noise or competing speech (e.g., Hall et al., 2002; Nittrouer & Boothroyd, 1990), these studies have assessed performance using speech produced in quiet. The modifications made by talkers in noise may be particularly beneficial for children as they learn to separate target from background sounds in complex acoustic environments. Although the perceptual consequences of using Lombard speech have not been reported for children, a recent study by Cooke & Garcia Lecumberri (2012) observed that Spanish-speaking adults who were learning English showed a substantial intelligibility benefit for English speech produced in noise compared to English speech produced in quiet. Similar to non-native adults, English-speaking children have less experience listening to the target language. Thus, children may have more difficulty recognizing speech that is embedded in competing sounds when the expected acoustic modifications produced by talkers in noise are absent. Given the larger child-adult differences observed in the presence of complex auditory maskers relative to steady-state noise, the benefit associated with the acoustic features of speech produced in a competing background may be more pronounced for children under conditions expected to produce substantial informational masking.

The present study examined whether the acoustic modifications made by talkers in the presence of two competing streams of speech improve children's perception of target words embedded in the same masker. In Experiment 1, the benefit of speech produced in a two-talker background was compared across children and adults using a closed-set spondee identification task. The rationale for using a closed-set identification task along with highly redundant spondee words (e.g., Wilson & Margolis, 1983) was to reduce the influence of potential child-adult differences in context effects. In Experiment 2, a more challenging task was used to assess the benefit of speech produced in a two-talker background. Specifically, open-set speech recognition was examined for monosyllabic words embedded in a two-talker background at a fixed SNR. As in Experiment 1, speech recognition performance was examined for both children and adults using target words produced in quiet or produced in a two-talker background. Based on results from previous studies (e.g., Hall et al., 2002), the expectation for both experiments was that children would perform more poorly than adults in the presence of a two-talker masker, regardless of whether the speech was produced in quiet or in a two-talker background. It was predicted that both age groups would benefit from the acoustic modifications of speech produced in the presence of a two-talker background compared to speech produced in quiet. However, children were expected to have greater difficulty recognizing speech embedded in competing sounds when the acoustic modifications produced by talkers in noise were absent. Thus, it was hypothesized that child-adult differences in speech recognition performance would be larger for speech produced in quiet than for speech produced in the presence of a two-talker background.

EXPERIMENT 1: BENEFIT OF SPEECH PRODUCED IN A TWO-TALKER BACKGROUND FOR CLOSED-SET SPONDEE IDENTIFICATION

METHOD

A. Listeners—Twenty school-aged children and 18 adults were tested. Listeners were assigned to complete testing in either a two-female-talker or a two-male-talker masker. Criteria for inclusion were: (1) normal hearing sensitivity (<20 dB HL) at all frequencies between 250 and 8000 Hz; (2) no reported risk factors for hearing loss; and (3) native speakers of American English. Listeners were tested individually in a double-walled, sound-treated room (IAC) in a 1-h session. Ten children and 10 adults completed testing in the presence of a masker comprised of two different female talkers (two-female-talker masker). The children in this group ranged in age from 5.6 to 10.4 years (mean = 8.0; standard deviation = 1.4), and the adults ranged in age from 19.9 to 28.9 years (mean = 22.6; standard deviation = 3.0). The second group of 10 children and eight adults completed testing in the presence of a masker comprised of two different male talkers (two-male-talker masker). The children in this group ranged in age from 5.7 to 10.7 years (mean = 8.6; standard deviation = 1.6), and the adults ranged in age from 19.3 to 30.6 years (mean = 23.3; standard deviation = 3.9).

B. Stimuli—Within a group of listeners, the same continuous two-talker stimulus was used as a background to obtain target speech productions and later as a masker in the spondee identification testing. One background/masker was comprised of two streams of meaningful speech produced by two different adult females (two-female-talker). The other background/

masker was comprised of two streams of meaningful speech produced by two different adult males (two-male-talker). The rationale for using two different background/masker stimuli was to evaluate whether children demonstrate a similar pattern of perceptual benefit using speech produced by men and women, as previously shown for adults, (Junqua, 1993). All of the talkers were native speakers of American English between the ages of 20 and 30 years. Each stream of speech was recorded in a sound-treated room (IAC) while the talker read aloud from familiar children's books. Different passages were spoken by the four talkers. The individual speech streams were recorded using a condenser microphone (AKG-C1000S) mounted approximately six inches from the talker's mouth. Productions were amplified (TDT MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (CARDDELUXE). Individual speech streams were manually edited to remove silent pauses greater than 300 ms, resulting in four samples that ranged from 3 to 4 minutes in duration. A 60-minute "seamless" stream was created for each sample by repeating it without discontinuity. The two individual streams from the same-sex talkers were then balanced for overall rms level, mixed, and resampled at a rate of 24.414 kHz using MATLAB.

Target speech productions were obtained from two additional talkers who were native speakers of American English, one adult female (25 years old) and one adult male (40 years old). Following Hall et al. (2002), the target stimuli were 25 visually unambiguous spondee words (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). Each talker recorded the words using a carrier phrase (i.e., "Say the word ________ again") while seated inside the sound-treated booth wearing Sennheiser HD-25 earphones. A condenser microphone (AKG-C1000S) was positioned approximately six inches from the talker's mouth. Both talkers produced the spondee words in two separate target speech conditions: (1) in quiet, and (2) while listening to the same-sex, two-talker speech masker, delivered binaurally through the headphones at 73 dB SPL. Although more robust changes in intensity would be expected using a higher background level, the rationale for using a moderate-level background was to obtain target speech productions at roughly the same level as the masker used in the subsequent perception tasks, while ensuring significant acoustic modifications were made across the two target speech conditions. The female talker produced the 25 spondee words in the presence of the two-female-talker background, and then produced the full corpus of 25 words again in quiet. The male talker first produced the 25 spondee words in quiet, and then in the two-male-talker background. The recording protocol followed Pittman and Wiley (2001), except that talkers were aware that their productions were being recorded. The tester sat outside of the sound booth, wore headphones, and was clearly visible to the talker through a window. The talker was instructed to speak clearly because the observer was unable to discern all of the features of their face. The tester appeared to be recording each production by writing down the target word. Productions were amplified (TDT MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (CARDDELUXE).

The authors reviewed the complete set of recordings for both talkers to ensure each word was intelligible and free of audible distortion. Acoustic analyses were conducted using the Praat speech analysis software (Boersma & Weenink, 2000). Specifically, a custom script was used to auto-identify word boundaries based on intensity thresholds (e.g. Green, Beukelman, & Ball, 2004; Haley, Jacks, de Riesthal, Abou-Khalil, & Roth, in press), with manual verification and adjustment by independent observers when needed. Total word duration was calculated based on the selected word boundaries, in addition to peak vocal level (dB) and spectral slope (dB SPL/kHz). Following the general approach described by Pittman and Wiley (2001), spectral slope was computed by fitting a linear regression line to the levels in each 1/3-octave band for frequencies up to 10 kHz.

Subsequent to the acoustic analyses, tokens were resampled at a rate of 24.414 kHz using MATLAB, and then normalized to an equal rms level. The selection and presentation of stimuli for the perceptual testing was controlled using custom software (MATLAB). The target spondee words and the masker were mixed (TDT SM3), sent to a headphone buffer (TDT HB6), and presented to the listener through Sennheiser HD-25 headphones.

C. Spondee identification procedure and conditions—Listeners were tested while seated in the sound booth. Prior to testing, each listener completed a familiarization phase in quiet, in which they listened to and identified each of the pictured spondees presented on a laminated board. This familiarization phase was completed with ease by all listeners.

Following the familiarization phase, separate testing conditions were completed for the two target speech conditions (produced in quiet, produced in a two-talker background). The task was a four-alternative, forced-choice spondee identification procedure using a picture-pointing response. Listeners held a 7-inch, touchscreen monitor (MIMO) during testing. One of the 25 spondee words was randomly selected as the target on each trial. Three additional spondees were randomly selected without replacement from the remaining set of 24, and these served as foils. Approximately 20-ms prior to presenting the selected target spondee, images associated with the target and three foils appeared on the touchscreen monitor. Each picture was randomly assigned to appear in one quadrant of the monitor. After stimulus presentation, listeners indicated their response by touching the image corresponding to the target spondee on the touchscreen monitor. Visual feedback was provided on the monitor after each response by flashing the target picture in isolation.

Target spondees were presented at a fixed level of 50 dB SPL (average rms) throughout testing. The masker level was adapted using a 2-up, 1-down rule (Levitt, 1971) to obtain an estimate of the level corresponding to 70.7% correct spondee identification. The starting level for the masker was approximately 10 dB below the expected threshold for each masker condition. An initial step size of 4 dB was reduced to 2 dB after the first two reversals. Runs were stopped after eight reversals. The masker threshold was estimated by computing the average of the masker level at the final six reversals. At least two runs were completed for each target spondee condition. A third run was obtained if the first two estimates differed by more than 4 dB. Testing order for the two target spondee conditions was counterbalanced across blocks of testing.

RESULTS

A. Acoustic analysis of target spondee productions—Table 1 summarizes the mean, minimum, and maximum peak levels, word durations, and spectral slopes for both talkers recorded in quiet and in the presence of their sex-matched, two-talker background. For the female talker, introduction of the two-female-talker masker was associated with a 5 dB increase in average peak level, a 58 ms (7.0%) increase in average word duration, and a +0.3 dB/kHz change in average spectral slope (indicating an increase in high-frequency energy). For the male talker, introduction of the two-male-talker masker was associated with a 2 dB increase in average peak level, a 55 ms (7.3%) increase in average word duration, and a +0.3 dB/kHz change in average spectral slope.

Two sets of planned comparisons (female talker, male talker) were conducted using two-tailed, paired t-tests to examine differences in peak level, word duration, and spectral slope across the spondees produced in quiet in the two-talker backgrounds. Bonferroni adjusted alpha levels of 0.017 per test (0.05/3) were used. For both the female and male speech productions, these planned comparisons confirmed a significant increase in peak level, word duration, and spectral slope for spondees produced in the corresponding two-talker background compared to spondees produced in quiet ($p < 0.017$ for all paired t-tests)

B. Spondee identification results—Figure 1 shows masker thresholds for children and adults tested in the two-female-talker masker (left panel) and for listeners tested in the two-male-talker masker (right panel). The open boxes show the range of performance spanning from the 25th to the 75th percentile for the words produced in quiet. Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines. Higher thresholds indicate greater sensitivity. The shaded boxes show the same range of performance for the words produced in the two-female-talker (left panel) or the two-male-talker (right panel) masker.

Masker thresholds for children assigned to complete testing using the female speech stimuli ranged from 53.7 to 69.2 dB SPL (mean = 60.2) for spondees produced by the female talker in quiet and from 53.7 to 76.5 dB SPL (mean = 64.2) for spondees produced by the female talker in the two-female-talker background (Figure 1, left panel). A significant correlation was observed between masker threshold and age for spondees produced by the female talker in the two-female-talker background ($r = 0.64$; $p = 0.023$). This correlation was positive, indicating that the SNR required for 70.7% correct spondee identification performance decreased for children as age increased. No significant correlation was observed between masker thresholds and age for spondees produced by the female talker in quiet ($p = 0.131$). Corresponding masker thresholds for adults ranged from 71.8 to 79.8 dB SPL (mean = 77.2) for spondees produced in quiet and from 76.5 to 82.7 dB SPL (mean = 79.4) for spondees produced in the two-female-talker background.

The right panel of Figure 1 presents masker thresholds for listeners tested in the two-male-talker masker. Children's thresholds ranged from 54.9 to 63 dB SPL (mean = 57.8) for spondees produced by the male talker in quiet and from 55.3 to 64.5 dB SPL (mean = 58.9) for spondees produced in the two-male-talker background. Consistent with the data obtained for children tested using the spondees produced by the female talker, there was a significant correlation between masker threshold and age for spondees produced by the male talker in the two-male-talker background ($r = 0.64$; $p = 0.048$), but not for spondees produced by the male talker in quiet ($p = 0.222$). Masker thresholds for adults ranged from 63.8 to 75.5 dB

SPL (mean = 70.6) and from 68.8 to 70.7 dB SPL (mean = 72.7) for spondees produced in quiet and in the two-male-talker background, respectively.

A three-way repeated-measures analysis-of-variance (ANOVA) on masker threshold confirmed the trends observed in Figure 1. An alpha level criterion of 0.05 was used. The analysis included the within-subjects factor of target speech (produced in quiet, produced in a two-talker background), as well as the between-subjects factors of age group (children, adults) and speech sex (female speech, male speech). The analysis revealed a significant main effect of target speech [$F(1,34) = 16.7$; $p < 0.0001$; $\eta_p^2 = 0.33$], indicating better performance using spondees produced in the two-talker backgrounds compared to spondees produced in quiet. The main effect of age group was also significant [$F(1,34) = 9556.3$; $p < 0.0001$; $\eta_p^2 = 0.99$], providing evidence of greater informational masking for children than for adults in the two-talker masker. A significant main effect of speech sex was also observed [$F(1,34) = 14.33$; $p < 0.001$; $\eta_p^2 = 0.30$], indicating higher masker thresholds (better performance) for listeners tested using the female speech compared to listeners tested using the male speech. None of the interaction terms were significant.

Figure 2 shows the difference in thresholds observed for individual listeners for spondees produced in a two-talker background and spondees produced in quiet. Data for listeners tested with the female stimuli are shown in the top panel, and data for listeners tested with the male stimuli are shown in the bottom panel. Filled circles show difference scores for individual children, and open squares show difference scores for individual adults. Data at or below the dashed horizontal line indicate no benefit of using spondees produced in the corresponding two-talker masker. For listeners tested with the female stimuli, 5 of 10 children and 5 of 10 adults showed a benefit (positive difference score) of 2 dB of greater. Difference scores ranged from −2.7 to 11.5 dB for children (mean = 4.0) and from −0.7 to 4.2 dB for adults (mean = 2.1). For listeners tested with the male stimuli, 4 of 10 children and 3 of 8 adults showed a benefit of 2 dB or greater. Difference scores ranged from −3.5 to 5.4 dB (mean = 1.1 dB) for children and from −1.3 to 8.7 dB (mean = 2.1) for adults.

Overall, these data indicate that children required a more advantageous SNR than adults to achieve 70.7% spondee identification accuracy in either a two-female-talker or a two-male-talker masker, but that average performance was better for both age groups using speech produced in a two-talker background compared to speech produced in quiet. Although the average benefit of using speech produced in a two-talker background was statistically significant, this benefit was modest (range = 1 to 4 dB). Moreover, an examination of the individual data shows that about half of the subjects tested did not show an improvement in performance with speech produced in a two-talker background. Recall that spondee words were selected because of their high redundancy, and a four-alternative, closed-set identification task was used. Thus, a second experiment was conducted to evaluate the possible role of stimulus and task difficulty.

EXPERIMENT 2: BENEFIT OF SPEECH PRODUCED IN A TWO-TALKER BACKGROUND FOR OPEN-SET MONOSYLLABIC WORD RECOGNITION

METHOD

A. Listeners—Listeners were 16 school-aged children and 18 adults. The children ranged in age from 5.1 to 10.3 years (mean = 8.0; standard deviation = 1.8), and the adults ranged in age from 19.0 to 28.7 years (mean = 22.5; standard deviation = 3.0). All listeners had normal hearing sensitivity (<20 dB HL) at frequencies between 250 and 8000 Hz, reported no risk factors for hearing loss, and were native speakers of American English. Listeners were tested individually in a single-walled booth in a 1-h session.

Stimuli: The target stimuli were Phonetically-Balanced Kindergarten words (PBK; Haskins 1949). Two different male talkers produced the words from List 1 and List 2 of the PBK word corpus. Each list consists of 50 monosyllabic words that were selected based on the spoken vocabulary of kindergarteners. Sanderson-Leepa and Rintelman (1976) demonstrated that typically-developing 5- to 7-year-old children can achieve 96 to 98% correct recognition of these words in quiet. The first male talker produced the words in quiet, and then produced the words again while listening to the two-male-talker background described in experiment 1. The second male talker initially produced the words in the background, and then in quiet. The method for obtaining the target speech productions was identical to that described for experiment 1.

The complete set of 400 tokens (100 words X 2 conditions X 2 talkers) was manually cropped to create separate files for each token (Audacity), and reviewed by the authors to ensure that each word was intelligible and free from distortion. As outlined in experiment 1, acoustic analyses were performed using Praat (Boersma & Weenink, 2000). These analyses were verified and manually adjusted (if necessary) by two independent observers. The tokens were then scaled to have equal rms level, and exported as WAV files with a 44.1-kHz sampling rate. During testing, target words were presented at 60 dB HL for all listeners and conditions.

The masker used for all listeners was the continuous two-male-talker masker described above in Experiment 1. Different masker levels were used to test children and adults. The masker level was 50 dB HL for children (+10 dB SNR) and 55 dB HL for adults (+5 dB SNR). The rationale for using different masker levels was to roughly equate performance across children and adults, and to avoid floor and ceiling effects. The selection of these masker levels was based on pilot data collected in the laboratory prior to testing and on previous work using similar stimuli (Bonino et al., 2012).

The target PBK words and the two-male-talker masker were recorded onto different compact disks. The recordings were routed to independent channels of a two-channel audiometer (Grason-Stadler GSI 61; Eden Prairie, MN). A 1000-Hz calibration tone was used to calibrate both channels of speech prior to each testing session. Stimuli were presented to both ears using TDH-50P headphones.

Word recognition procedure and conditions: Listeners were tested while seated inside the booth. The task was to repeat back the target PBK words. Eight children and nine adults were assigned to complete testing using target words produced by the first male talker. Eight children and nine adults were tested using target words produced by the second male talker. Each listener completed one list of 50 words for each target speech condition (produced in quiet, produced in the two-male-talker background). The order of presentation of target speech conditions and word lists was counterbalanced across listeners within each age group. Guessing was encouraged. A tester sat inside the booth with the listener throughout testing and scored responses in real time. The listener's spoken responses were also were recorded throughout testing and scored offline by a second tester. The first five words from each list were considered practice, and the remaining 45 words were used to compute a percent correct score for each target speech condition. The average correlation coefficient across the live and offline coders was computed using the data from nine children and nine adults. The average correlation between coders was 0.94 for children (range = 0.86 – 0.99) and 0.96 for adults (range = 0.93 – 0.99). Thus, scores from the real-time coder were used in all subsequent analyses.

RESULTS

A. Acoustic analysis of target PBK word productions—Table 2 presents the mean, minimum, and maximum peak vocal levels, word durations, and spectral slopes for PBK words produced by each talker in quiet and while listening to the two-male-talker background. For PBK words produced by the first male talker, average peak level increased by 4 dB, average word duration increased by 142 ms (26.5%), and average spectral slope changed by +0.3 dB SPL/kHz in the two-male-talker background compared to words produced in quiet. For the second male talker, average peak level increased by 2 dB, word duration increased by 56 ms (12.3%), and slope changed by +0.8 dB SPL/kHz for words produced in the background compared to words produced in quiet.

A separate set of planned comparisons was conducted for each of the two male talkers using two-tailed, paired t-tests. Bonferroni adjusted alpha levels of 0.017 per test (0.05/3) were used. The results of the planned comparisons confirmed a significant increase in peak level, word duration, and spectral slope for PBK words produced in the two-male-talker background compared to PBK words produced in quiet ($p < 0.001$ for all paired t-tests).

B. Word recognition results—Figure 3 shows percent correct scores (45/50 words per condition) for children and adults (45 words/condition) tested using PBK words produced by the first target talker (left panel) and for children and adults tested using words produced by the second target talker (right panel). The open boxes show the range of performance spanning from the 25th to the 75th percentile for the words produced in quiet. Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines. The shaded boxes show the same range of performance for the words produced in the two-male-talker background. Recall that children were tested at a +10 dB SNR, and adults were tested at a +5 dB SNR.

For children tested using PBK words produced by the first talker, average word recognition score improved from 76% (range = 64 to 84) for words produced in quiet to 89% (range = 82 to 96) for words produced in the two-male-talker background. For children testing using PBK words produced by the second talker, average word recognition score improved from 64% (range = 42 to 74) for words produced in quiet to 74% (range = 44 to 90) for words produced in the two-male-talker background. Similar to the pattern of results observed for children, the average word recognition score for adults tested using PBK words produced by the first talker improved from 76% (range = 71 to 87) using words produced in quiet to 87% (range = 82 to 96) using words produced in the two-male-talker background. For adults tested using the PBK words produced by the second talker, average word recognition score improved from 71% (range = 52 to 82) for words produced in quiet to 83% (range = 64 to 91) for words produced in the two-male-talker background.

Percent correct performance was converted to rationalized arcsine units (RAUs) prior to statistical analyses to prevent bias due to non-uniformity of (Studebaker, 1985). The RAU scores were analyzed separately for the group of listeners tested using PBK words produced by the first target talker and the group of listeners tested using PBK words produced by the second target talker, using a pair of repeated-measures ANOVAs with the within-subjects factors of target speech (produced in quiet, produced in the two-male-talker background) and the between-subjects factor of age group (children, adults). For children and adults tested using words produced by the first male talker, a significant main effect was found for target speech [$F(1,15) = 82.4$; $p < 0.0001$; $\eta_p^2 = 0.85$], indicating better performance using target PBK words produced in the two-male-talker background compared to target PBK words produced in quiet. Since the use of different SNRs was designed to equate performance across children and adults, the absence of a main effect for age group was expected [$F(1,15) = 0.5$; $p = 0.513$]. There was no target speech X age group interaction [$F(1,15) = 0.2$; $p = 0.670$], indicating a comparable benefit of using speech produced in the two-talker background for children and adults. The same pattern of statistical results was observed for the data obtained from children and adults using words produced by the second male talker. Performance was better using PBK words produced in the two-male-talker background compared to PBK words produced in quiet, indicated by the significant main effect for target speech [$F(1,15) = 27.4$; $p < 0.0001$; $\eta_p^2 = 0.65$]. Neither the main effect of age group [$F(1,15) = 0.2$; $p = 0.680$] nor the target speech X age group interaction were significant [$F(1,15) = 3.0$; $p = 0.106$].

Figure 4 shows the difference in percent correct scores for individual listeners, plotted as a function of age. The open symbols show the advantage of using PBK words produced in noise for children (squares) and adults (circles) tested using words produced by the first target talker. The shaded symbols show this advantage for children (squares) and adults (circles) tested using words produced by the second target talker. Data at or below the dashed horizontal line indicate no benefit of using PBK words produced in the two-male-talker background. For the first group of listeners tested using PBK words from Talker 1, difference scores ranged from 2 to 22 percentage points (mean = 12 percentage points) across individual children, and from 7 to 18 percentage points across individual adults (mean = 11 percentage points. For the second group of listeners tested using PBK words from Talker 2, difference scores ranged from −5 to 21 percentage points (mean = 10 percentage points) across individual children, and from 5 to 24 percentage points across individual adults (mean = 12 percentage points. Positive difference scores indicate improved performance using words produced in the two-male-talker background relative to words produced in quiet. Although individual differences in benefit scores were observed, the individual data are consistent with the trends observed in the group data. Thirteen (out of 18) children received a benefit of at least five percentage points from the speech produced in the two-male-talker background, and all but two children received a positive benefit. For the two children who did not benefit, one child (8.6 years) performed similarly across the two target speech conditions, and the other child (8 years) showed a decrease of five percentage points in performance. All 18 adults received a benefit of at least five percentage points.

DISCUSSION

The main goal of this study was to evaluate the degree to which children benefit from the acoustic modifications made by talkers when they produce speech in noise. This goal was accomplished by comparing performance across conditions using speech produced in a two-talker background and speech produced in quiet. Closed-set spondee identification was assessed in experiment 1 using an adaptive procedure, and open-set monosyllabic word recognition was measured at a fixed SNR in experiment 2. The results from both experiments indicate that children can use the acoustic modifications of speech produced in a two-talker background to improve their perception of speech in the presence of the same masker, as has previously been demonstrated for non-native (Cooke & Garcia Lecumberri, 2012) and native English-speaking adults (e.g., Dreher & O'Neill, 1958; Pittman & Wiley, 2001; Summers et al., 1988). The present results do not support the hypothesis that child-adult differences in speech perception are larger for speech produced in quiet than for speech produced in a two-talker background. Although children were more susceptible to informational masking than adults, a similar benefit was observed for both age groups.

Acoustic modifications to speech produced in a two-talker background

A significant increase in peak vocal level, an increase in word duration, and a flattening of spectral slope were found in the speech produced by all four target talkers in the two-talker backgrounds compared to the same words produced in quiet. The increase in word duration and the flattening of the spectral slope are in general agreement with the pattern of acoustic modifications reported in previous investigations (e.g., Cooke & Lu, 2010; Letowski et al., 1993; Lombard, 1911; Patel & Schell, 2008; Pittman & Wiley, 2001; Summers et al., 1988). For example, Pittman and Wiley (2001) reported a 12.6% increase (65 ms) in average word duration for speech produced in a six-talker babble presented at 80 dB SPL compared to speech produced in quiet. In comparison, word duration in the present study increased by 7.0% (58 ms) for the female talker and 7.3% (55 ms) for the male talker comparing the spondee productions in experiment 1. Productions increased by 26.5% (142 ms) for the first male talker and by 12.3% (56 ms) for the second male talker comparing PBK productions in experiment 2. Also in agreement with Pittman and Wiley (2001), changes in the spectral slope were observed for all four talkers in the present study for speech produced in the two-talker backgrounds relative to quiet. These slope changes reflect a flattening of the spectral slope of the target speech in the two-talker backgrounds, indicating an increase in the contribution of high-frequency energy relative to speech produced in quiet. Substantial variability in the magnitude of intensity-related changes associated with speech produced in noise has been reported across previous studies (e.g., Patel & Schell, 2008; Pittman & Wiley, 2001; Summers et al., 1988). Nonetheless, the level increases observed in the present study were smaller than typically observed (reviewed by Patel and Schell 2008). For example, Patel and Schell (2008) reported average increases of 6.9 and 15.7 dB in speech produced in a multi-talker babble at 60 dB SPL and 90 dB SPL, respectively. In contrast, peak intensity increased by only 2–5 dB across the four talkers in the present study. Differences in the target and masker stimuli used across studies, the procedures used to obtain speech productions, and individual talker variability complicate efforts to determine the basis for this discrepancy. However, a likely contributor to the reduced intensity effects found in the present study is the relatively low level of the two-talker background (73 dB SPL) used to obtain speech productions. As discussed in the introduction, intensity effects become larger with increasing background level (e.g., Patel & Schell, 2008). Recall, however, that overall word level was equated across target speech conditions prior to the perception tasks.

Child-adult differences in susceptibility to informational masking

As expected based on previous work using similar stimuli (Bonino et al., 2012; Hall et al., 2002), the results from this study demonstrated pronounced informational masking effects for children compared to adults in the presence of a two-talker masker. These child-adult differences in susceptibility to informational masking were observed for target speech produced in quiet, and also for target speech produced in the two-talker backgrounds. In experiment 1, an adaptive paradigm was employed to estimate the masker level required to achieve 70.7% correct spondee identification performance. Adults showed an average advantage over children ranging from 13–17 dB SNR across the four testing conditions. These child-adult differences are larger than observed by Hall et al. (2002), who reported that 5- to 10-year-old children required an average advantage of 7 dB SNR compared to adults to achieve 79.7 % correct identification on the same spondee task used in the present study. Differences in methodology make it difficult to determine the basis for the larger child-adult differences observed in the present study relative to Hall et al. (2002). In that study, the SNR corresponding to 79.4% correct spondee identification was estimated in a two-male-talker masker by varying the level of the target spondees. The masker was presented at a fixed level of 70 dB SPL throughout testing. In contrast, spondees were presented at a fixed level of 50 dB SPL in the present study, and masker level was varied to estimate 70.7% correct spondee identification. One likely explanation for the larger age effect observed in the present study relative to Hall et al. (2002) is that different two-talker masker samples and/or target-masker combinations vary in masking effectiveness. This possibility is supported by the significant difference in thresholds observed between listeners tested in the female- and male-talker conditions of experiment 1 and the two male-talker conditions of experiment 2. Variability in masking across masker samples comprised of a small number of talkers has also been reported in the adult literature (e.g., Freyman et al., 2004). Additional studies are needed to quantify the variability in masking effects across different two-talker masker samples as it pertains to speech perception in children.

Consistent with the closed-set spondee identification results from experiment 1, the open-set PBK recognition results from experiment 2 indicate that children are more susceptible to informational masking than adults in the presence of a two-talker masker. Children required an additional +5 dB SNR relative to adults to achieve comparable performance. Recent results reported by Bonino et al. (2012) are consistent with the idea that the pronounced child-adult differences in spondee identification in a two-talker masker, first reported by Hall et al. (2002), extend to monosyllabic word recognition. Bonino et al. (2012) measured open-set recognition of PBK words in two groups of children (5–7 and 8–10 years) and a group of young adults in the presence of the same two-male-talker masker used in the present study. All three age groups were tested at a +10 dB SNR. Both groups of children performed more poorly than the group of adults, with average percent correct scores of 48% for 5- to 7-year-olds, 62% for 8- to 10-year-olds, and 82% for adults.

Benefit of speech produced in noise

The main finding of this study is that children were able to take advantage of the acoustic modifications associated with speech produced in a two-talker background to improve their word recognition performance, as has previously been demonstrated for adults (e.g., Dreher & O’Neill, 1958; Lu & Cooke, 2008; Pittman & Wiley, 2001; Summers et al., 1988). Although modest, a significant improvement in the SNR required to identify spondees in a either a two-male-talker or a two-female-talker masker was observed for both children and adults using speech produced in a two-talker background compared to speech produced in quiet (experiment 1). Similarly, open-set word recognition scores at a fixed SNR for both age groups were higher using speech produced in a two-male-talker masker compared to speech produced in quiet (experiment 2). Word recognition scores were an average of 11 percentage points higher for both children and adults using produced in a two-talker masker compared to the same words produced in quiet.

It was hypothesized that child-adult differences in speech perception would be larger for speech produced in quiet than for speech produced in a two-talker background. This hypothesis was based on previous work demonstrating larger child-adult differences in speech recognition in a two-talker masker than in speech-shaped noise (e.g., Bonino et al., 2012; Hall et al., 2002). These previous data indicate that the potential for improvement in masked speech perception is higher for children than for adults, particularly when the masker is comprised of two competing streams of speech. However, the present results do not support the a priori hypothesis. Despite age effects in susceptibility to informational masking, children and adults experienced a similar benefit associated with speech produced in a two-talker background. For example, the average child-adult difference in threshold observed in experiment 1 was 14.9 dB for spondees produced in quiet and 14.5 dB for spondees produced in either a two-male-talker or a two-female talker background.

The basis for children’s increased susceptibility to informational masking relative to adults’ is unknown. Similar findings have been reported in the developmental literature for some studies examining informational masking release using both speech (e.g., Bonino et al., 2012; Litovsky, 2005) and non-speech (e.g., Leibold & Neff, 2007) stimuli. For example, Bonino et al. (2012) recently examined whether a carrier phrase (i.e., “say the word”) improves children's word recognition performance in the presence of a two-talker masker. Listeners were two age groups of children (5–7 and 8–10 years) and a group of young adults. Both age groups of children performed more poorly than the adults in the two-talker masker when the carrier phrase was absent. Consistent with the present findings, however, a similar benefit of providing a carrier phrase was observed across the three age groups. Thus, children remained more susceptible to informational masking than adults despite the provision of the carrier phrase. In contrast to the present results, other studies of informational masking release have observed a smaller release from masking for children than for adults (e.g., Hall et al., 2002; Wightman et al., 2003). It appears that children's ability to benefit from cues that reduce informational masking in adult listeners depends on the particular cue that is introduced (e.g., Hall et al., 2002).

Conclusions and Implications

Although the sources of children's increased susceptibility to informational masking relative to adults' in the present study are not fully understood, the present data clearly support the idea that children are likely to have considerable difficulty hearing speech and other important sounds in complex acoustic environments. The results also indicate that children, like adults, can use the acoustic modifications made by talkers in noise to improve their perception of masked speech. From a theoretical perspective, these findings provide new information about the types of acoustic cues that help the developing child separate target speech from a complex speech masker. From a clinical perspective, these results highlight the importance of considering the influence of selecting the appropriate stimuli for measures aimed at assessing hearing in noise. Children, like adults, are likely to show better performance when tested with target speech produced in noise than target speech produced in quiet. Further, results with speech produced in noise are likely to be more representative of real-world performance.

Acknowledgments

This work was supported from the National Institute of Deafness and Other Communication Disorders (DC011038). Portions of these results were presented to the American Auditory Society Annual Meeting in Scottsdale, AZ in March 2012. We are grateful to the members of the Human Auditory Development Laboratory for their assistance with data collection.

References

Bennett KO, Billings CJ, Molis MR, Leek MR. Neural encoding and perception of speech signals in informational masking. Ear Hear. 2012; 33(2):231–238. [PubMed: 22367094]

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. 2000. http://www.praat.org

Bonino AY, Leibold LJ, Buss E. Release from perceptual masking for children and adults: benefit of a carrier Phrase. Ear Hear. 2012; 34(1):3–14. [PubMed: 22836239]

Brungart, DS. Informational and energetic masking effects in multitalker speech perception. In: Divenyi, P., editor. Speech separation in humans and machines. New York: Springer; 2005. p. 261-267.

Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am. 2001; 110(5 Pt 1):2527–2538. [PubMed: 11757942]

Buss, E.; Hall, JW.; Grose, JH. Development of Auditory Coding as Reflected in Psychophysical Performance. In: Werner, LA.; Fay, RR.; Popper, AN., editors. Human Auditory Development. New York, NY: Springer; 2012. p. 107-136.

Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am. 1969; 45(3):694–703. [PubMed: 5776931]

Cooke M, Garcia Lecumberri ML. The intelligibility of Lombard speech for non-native listeners. J Acoust Soc Am. 2012; 132(2):1120–1129. [PubMed: 22894231]

Cooke M, Lu Y. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. J Acoust Soc Am. 2010; 128(4):2059–2069. [PubMed: 20968376]

Dreher JJ, O'Neill JJ. Effects of ambient noise on speaker intelligibility of words and phrases. Laryngoscope. 1958; 68(3):539–548. [PubMed: 13551103]

Elliott LL, Connors S, Kille E, Levin S, Ball K, Katz D. Children's understanding of monosyllabic nouns in quiet and in noise. J Acoust Soc Am. 1979; 66(1):12–21. [PubMed: 489827]

Fallon M, Trehub SE, Schneider BA. Children's perception of speech in multitalker babble. J Acoust Soc Am. 2000; 108(6):3023–3029. [PubMed: 11144594]

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(5 Pt 1):2246–2256. [PubMed: 15139635]

Green JR, Beukelman DR, Ball LJ. Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. J Med Speech Lang Pathol. 2004; 12(4):149–154. [PubMed: 20628555]

Haley KL, Jacks A, de Riesthal M, Abou-Khalil R, Roth HL. Toward a quantitative basis for assessment and diagnosis of apraxia of speech. J Speech Lang Hear Res. (in press).

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

Haskins, H. Unpublished Master's thesis. Northwestern University; Evanston, Illinois: 1949. A phonetically balanced test of speech discrimination for children.

Junqua JC. The Lombard reflex and its role on human listeners and automatic speech recognizers. J Acoust Soc Am. 1993; 93:510–524. [PubMed: 8423266]

Kidd G Jr, Mason CR, Deliwala PS, Woods WS, Colburn HS. Reducing informational masking by sound segregation. J Acoust Soc Am. 1994; 95(6):3475–3480. [PubMed: 8046139]

Leibold, LJ. Development of Auditory Scene Analysis and Auditory Attention. In: Werner, LA.; Fay, RR.; Popper, AN., editors. Human Auditory Development. New York, NY: Springer; 2012. p. 137-162.

Leibold LJ, Neff DL. Effects of masker-spectral variability and masker fringes in children and adults. J Acoust Soc Am. 2007; 121(6):3666–3676. [PubMed: 17552718]

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

Letowski T, Frank T, Caravella J. Acoustical properties of speech produced in noise presented through supra-aural earphones. Ear Hear. 1993; 14(5):332–338. [PubMed: 8224576]

Levitt H. Transformed up-down methods in psychoacoustics. J Acoustic Soc Am. 1971; 49(2 Pt 2): 467–477.

Lombard E. Le signe de l'elevation de la voix (The sign of the rise in the voice). Ann Maladiers Oreille, Larynx, Nex, Pharynx. 1911; 37:101–119.

Lu Y, Cooke M. Speech production modifications produced by competing talkers, babble, and stationary noise. J Acoust Soc Am. 2008; 124(5):3261–3275.10.1121/1.2990705 [PubMed: 19045809]

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

Oh EL, Lutfi RA. Nonmonotonicity of informational masking. J Acoust Soc Am. 1998; 104(6):3489–3499. [PubMed: 9857508]

Patel R, Schell KW. The influence of linguistic content on the Lombard effect. J Speech Lang Hear Res. 2008; 51(1):209–220. [PubMed: 18230867]

Pittman AL, Wiley TL. Recognition of speech produced in noise. J Speech Lang Hear Res. 2001; 44(3):487–496. [PubMed: 11407555]

Rivers C, Rastatter MP. The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. J Aud Res. 1985; 25:37–45. [PubMed: 3836994]

Sanderson-Leepa ME, Rintelmann WF. Articulation functions and test-retest performance of normal-hearing children on three speech discrimination tests: WIPI, PBK-50, and NV Auditory Test No. 6. J Speech Hear Disord. 1976; 41(4):503–519. [PubMed: 994481]

Studebaker GA. A “rationalized” arcsine transform. J Speech Hear Res. 1985; 28(3):455–462. [PubMed: 4046587]

Summers WV, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA. Effects of noise on speech production: acoustic and perceptual analyses. J Acoust Soc Am. 1988; 84(3):917–928. [PubMed: 3183209]

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

Wilson RH, Farmer NM, Gandhi A, Shelburne E, Weaver J. Normative data for the Words-in-Noise Test for 6- to 12-year-old children. J Speech Lang Hear Res. 2010; 53(5):1111–1121. [PubMed: 20699343]

Wilson, RH.; Margolis, RH. Measurement of the Auditory Thresholds for Speech Stimuli. In: Konkle, D.; Rintelmann, W., editors. Principles of Speech Audiometry. Baltimore, MD: University Park Press; 1983. p. 79-126.

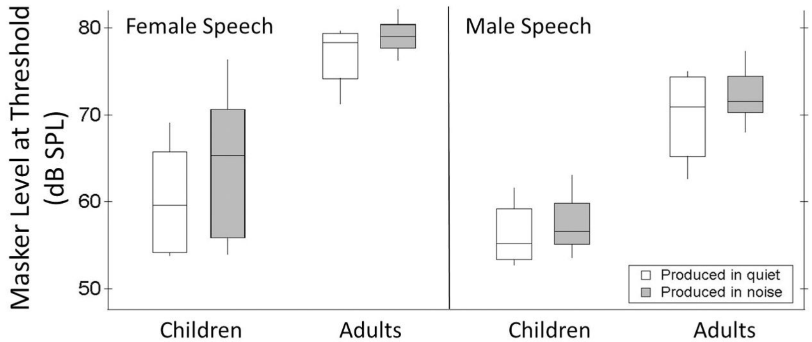

Figure 1.

The masker level corresponding to 70.7% correct spondee identification is shown for children and adults tested using the female target spondees in the presence of the two-female-talker masker (left), and for children and adults testing using the male target spondees in the presence of the two-male-talker masker (right). The open rectangles span the 25th to 75th percentiles for speech produced in quiet. The shaded rectangles show this range for speech produced in the corresponding two-talker background. The horizontal line inside each box shows the median. The vertical bars show the 10th and 90th percentiles.

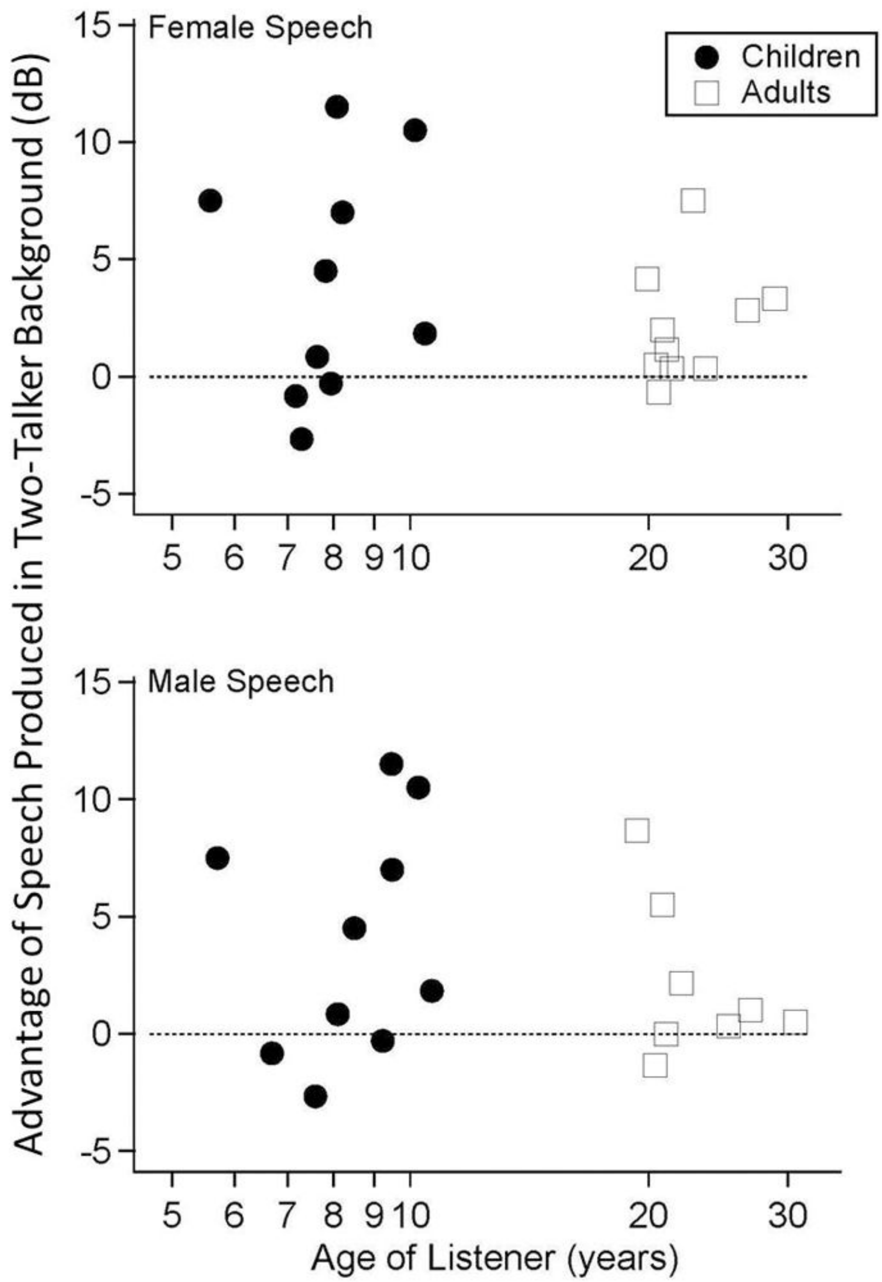

Figure 2.
The difference between masker thresholds for spondee words produced in a two-talker background and the same spondee words produced in quiet are shown for individual children (filled circles) and adults (open squares). Data are ordered by age, with data for listeners tested with the female-speech stimuli shown in the top panel, and data for listeners tested with the male-speech stimuli shown in the bottom panel. Data above the dotted horizontal line indicate a benefit of speech produced in the two-talker masker compared to speech produced in quiet.

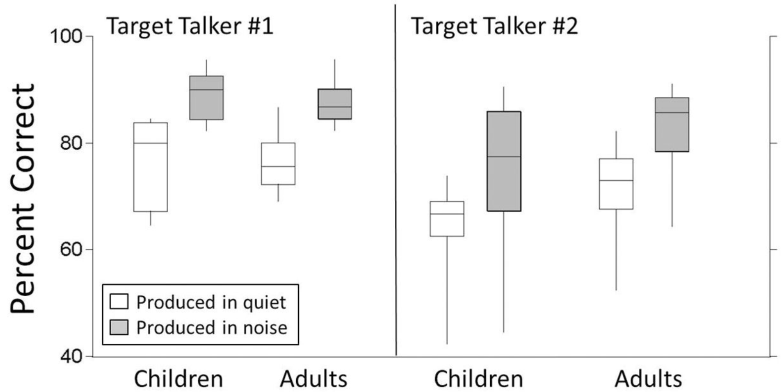

Figure 3.

Percent correct performance is shown for children and adults for PBK words produced by the first male target talker (left panel) and by the second male target talker (right panel). The open rectangles span the 25th to 75th percentiles for speech produced in quiet. The shaded rectangles show this range for speech produced in the corresponding two-talker background. The horizontal line inside each box shows the median. The vertical bars show the 10th and 90th percentiles. Note that children were tested at +10 dB SNR and adults were tested at +5 dB SNR.

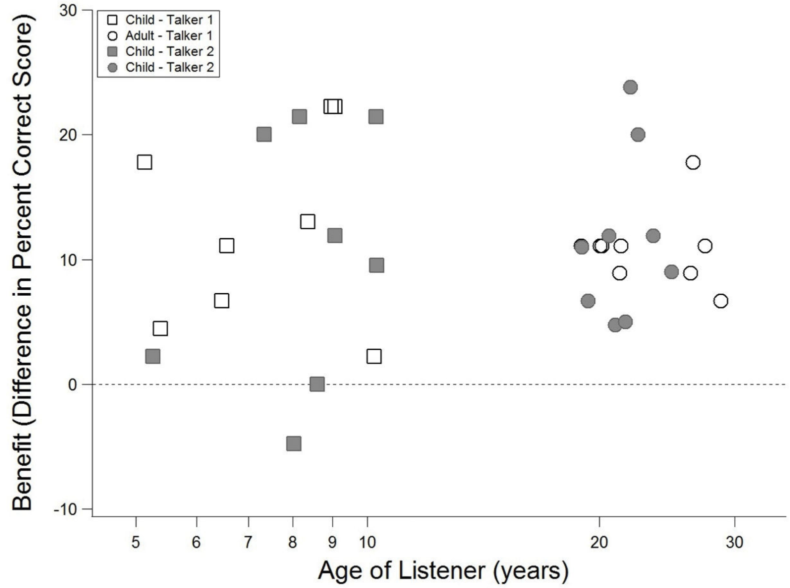

Figure 4.

The difference between percent correct scores for PBK words produced in the two-male-talker background and the same words produced in quiet are shown for individual children (squares) and adults (circles). Data are ordered by age, with differences for listeners tested using words produced by the first male target talker shown by the open symbols, and data for listeners tested using words produced by the second male target talker shown by the shaded symbols. Difference scores above the dotted horizontal line indicate a benefit of speech produced in the two-talker masker compared to speech produced in quiet.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary of the mean, minimum, and maximum peak level (in dB SPL), word duration (in ms), and spectral slope (in dB SPL/kHz) for both talkers' spondee word productions in quiet and in the presence of a two-talker background.

Spondee Words - Female Talker			
Condition	**Peak Level (dB SPL)**	**Duration (ms)**	**Spectral Slope (dB/kHz)**
Produced in quiet	Mean = 75 Min = 71 Max = 80	Mean = 824 Min = 659 Max = 1090	Mean = −4.8 Min = −7.4 Max = −2.4
Produced in two-female- talker background	Mean = 80 Min = 75 Max = 84	Mean = 882 Min = 741 Max = 1084	Mean = −4.5 Min = −7.1 Max = −1.9
Spondee Words - Male Talker			
Condition	**Peak Level (dB SPL)**	**Duration (ms)**	**Spectral Slope (dB/kHz)**
Produced in quiet	Mean = 76 Min = 72 Max = 79	Mean = 753 Min = 602 Max = 909	Mean = −4.9 Min = −7.9 Max = −2.1
Produced in two-male- talker background	Mean = 78 Min = 72 Max = 81	Mean = 808 Min = 641 Max = 1000	Mean = −4.6 Min = −7.4 Max = −1.2

Table 2

Summary of the mean, minimum, and maximum peak level (in dB SPL), word duration (in ms), and spectral slope (in dB SPL/kHz) for both male talkers' productions of PBK words (List 1 and List 2) in quiet and in the presence of a two-male-talker background.

PBK Words - Male Talker #1			
Condition	**Peak Level (dB SPL)**	**Duration (ms)**	**Spectral Slope (dB/kHz)**
Produced in quiet	Mean = 76 Min = 69 Max = 88	Mean = 536 Min = 237 Max = 875	Mean = −4.5 Min = −8.3 Max = −1.1
Produced in two-male-talker background	Mean = 80 Min = 73 Max = 87	Mean = 678 Min = 397 Max = 1027	Mean = −4.2 Min = −7.4 Max = −1.3
PBK Words - Male Talker #2			
Condition	**Peak Level (dB SPL)**	**Duration (ms)**	**Spectral Slope (dB/kHz)**
Produced in quiet	Mean = 74 Min = 69 Max = 80	Mean = 456 Min = 243 Max = 732	Mean = −5.4 Min = −8.7 Max = −1.8
Produced in two-male-talker background	Mean = 76 Min = 72 Max = 84	Mean = 512 Min = 271 Max = 810	Mean = −4.6 Min = −8.0 Max = −1.3