## Research Article

# Cognitive and Linguistic Contributions to Masked Speech Recognition in Children

Ryan W. McCreery,[a] (iD) Margaret K. Miller,[b] (iD) Emily Buss,[c] (iD) and Lori J. Leibold[b] (iD)

**Purpose:** The goal of this study was to examine the effects of cognitive and linguistic skills on masked speech recognition for children with normal hearing in three different masking conditions: (a) speech-shaped noise (SSN), (b) amplitude-modulated SSN (AMSSN), and (c) two-talker speech (TTS). We hypothesized that children with better working memory and language skills would have better masked speech recognition than peers with poorer skills in these areas. Selective attention was predicted to affect performance in the TTS masker due to increased cognitive demands from informational masking.

**Method:** A group of 60 children in two age groups (5- to 6-year-olds and 9- to 10-year-olds) with normal hearing completed sentence recognition in SSN, AMSSN, and TTS masker conditions. Speech recognition thresholds for 50% correct were measured. Children also completed standardized measures of language, memory, and executive function.

**Results:** Children's speech recognition was poorer in the TTS relative to the SSN and AMSSN maskers. Older children

had lower speech recognition thresholds than younger children for all masker conditions. Greater language abilities were associated with better sentence recognition for the younger children in all masker conditions, but there was no effect of language for older children. Better working memory and selective attention skills were associated with better masked sentence recognition for both age groups, but only in the TTS masker condition.

**Conclusions:** The decreasing influence of vocabulary on masked speech recognition for older children supports the idea that this relationship depends on an interaction between the language level of the stimuli and the listener's vocabulary. Increased cognitive demands associated with perceptually isolating the target talker and two competing masker talkers with a TTS masker may result in the recruitment of working memory and selective attention skills, effects that were not observed in SSN or AMSSN maskers. Future research should evaluate these effects across a broader range of stimuli or with children who have hearing loss.

The ability to recognize speech in the presence of competing sounds develops throughout childhood and into adolescence. Because many daily listening environments contain numerous competing sound sources (Crukley et al., 2011; Nelson & Soli, 2000), the ability to understand speech in these complex acoustic environments is critical for children to communicate and reach high levels of academic and social functioning. Competing sounds in

a child's listening environments have varying acoustic characteristics, from relatively steady-state noise, such as noise from heating and ventilation systems, to sounds that have fluctuating temporal and spectral characteristics, such as competing talkers. Speech recognition in steady speech-shaped noise (SSN) improves from ages 5 to 10 years (Choi et al., 2008; McCreery et al., 2010; McCreery & Stelmachowicz, 2011; Nittrouer & Boothroyd, 1990; Talarico et al., 2007; Wilson et al., 2010; Yacullo & Hawkins, 1987). Whereas a large number of competing talkers often produce masking effects of similar magnitude to SSN maskers (Carhart et al., 1969), speech recognition in maskers composed of a small number of competing talkers tends to be more challenging for children than for adults (Bonino et al., 2013; Buss et al., 2016; Goldsworthy & Markle, 2019; Hall et al., 2002; Leibold & Buss, 2013). Even though typically developing children reach levels of performance similar to adults by 9–10 years of age for speech recognition in SSN, adultlike levels of performance in a two-talker speech

[a]Audibility, Perception and Cognition Laboratory, Boys Town National Research Hospital, Omaha, NE
[b]Human Auditory Development Laboratory, Boys Town National Research Hospital, Omaha, NE
[c]Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill

Corresponding to Ryan W. McCreery: Ryan.McCreery@boystown.org

(TTS) masker are often not observed until 13 years of age or older (Brown et al., 2010; Corbin et al., 2016). The prolonged developmental trajectory of speech recognition in TTS suggests that different perceptual or cognitive mechanisms play a role in the development of listening in the presence of competing talkers. A better understanding of the cognitive and linguistic skills that support listening in different maskers has important implications for the clinical assessment of hearing using masked speech-recognition tests, as well as the implications of listening and learning in classrooms or other environments where children encounter different sources of competing sound. The goal of this study was to evaluate how school-age children's linguistic and cognitive skills affect masked speech recognition, with the goal of better understanding individual differences and masker effects.

## Developmental Differences in Masked Speech Recognition

One potential mechanism that could account for the protracted development of speech recognition in TTS compared to SSN in children and adolescents is the ability to segregate or selectively attend to target speech while inhibiting attention to speech produced by competing talkers (see Leibold & Buss, 2019, for a review). Masking from SSN produces predominantly energetic masking, where the audibility of the target signal is reduced by the masker because the signals have overlapping representations in the peripheral auditory system (but see work by Stone et al., 2012, on nonenergetic masking effects from SSN maskers). A small number of competing talkers produces energetic masking, but often produce more masking than would be predicted by reductions in audibility. This type of additional interference is known as perceptual or informational masking. Informational masking occurs when the stimulus is adequately represented in the peripheral auditory system, but the listener is unable to perceptually isolate the target from the masker due to stimulus uncertainty and/or perceptual similarity of the target and masker (Brungart, 2001; Durlach et al., 2003; Neff, 1995). Greater perceptual similarity between target and masker reduces the ability to perceptually isolate the target speech from the speech of a small number of competing talkers (Brungart & Simpson, 2002; Buss et al., 2016; also see Leibold, 2017, for a review). The developmental improvement in the ability to recognize speech in a TTS masker is thought to parallel improvements in the abilities to segregate and selectively attend to the stimulus (Corbin et al., 2016; Sobon et al., 2019). The changes in susceptibility to informational masking during childhood and adolescence have often been attributed to immature selective attention (Leibold & Buss, 2019; Wightman & Kistler, 2005), but the influence of selective attention skills on speech recognition with TTS maskers in children has not been directly assessed in previous research.

Children may also be at a disadvantage in a TTS masker because they are developing the ability to use brief temporal fluctuations in a speech masker to "glimpse" the target speech (Hall et al., 2012; Wróblewski et al., 2012). Adults can use portions of the target speech that are audible during temporal fluctuations in the masker to improve their speech recognition in noise compared to a listening condition where the masker is unmodulated (Cooke, 2006; Howard-Jones & Rosen, 1993). The ability to access target speech during fluctuations in the masker to improve speech recognition in noise requires the listener to attend to the target speech during those segments where the target has greater audibility and reconstruct the target speech signal from the components that were audible.

Some studies have shown that improvements in speech recognition associated with the introduction of amplitude modulation to a speech-shaped masker increase as children get older (Buss et al., 2016; Hall et al., 2012, but also see Stuart, 2008). The difference in masked speech recognition between SSN and amplitude-modulated SSN (AMSSN) is often referred to as release from masking or modulation masking release. Both AMSSN and TTS maskers fluctuate in level over time, but AMSSN maskers are less perceptually similar to target speech than a TTS masker. In most studies measuring speech recognition in AMSSN, modulation is imposed via multiplication with a raised sinusoid, resulting in predictable and periodic glimpses of the target. In a TTS masker, the glimpses are more irregular and less predictable, making it more difficult to extract cues that are audible during fluctuations in the masker. As a result, an AMSSN masker generally produces less informational masking than a TTS masker.

Comparisons of the relationship between audibility and speech recognition for different masker types highlight the particular difficulties associated with listening to speech in TTS maskers for children. To examine differences in how children and adults use the information that is audible during masker fluctuations, Sobon et al. (2019) used the Extended Speech Intelligibility Index (Rhebergen et al., 2006) to characterize the amount of audibility required for speech-reception thresholds across different masker conditions for children, young adults, and older adults. Children required greater audibility of the target to reach the same level of performance as young and older adults, but the audibility differences between age groups were larger in a TTS masker than an SSN masker. The psychometric function slopes were also steeper for TTS for children than adults, suggesting that adults were better able to benefit from spectrotemporally sparse audible speech cues than children. Collectively, these results lend support to the idea that the TTS masker was more difficult for children and required greater signal audibility to reach similar levels of performance as adults. However, it remains unclear which cognitive and linguistic abilities might contribute to these differences between children and adults for different masker types.

## Cognitive and Linguistic Contributions to Masked Speech Recognition in Children

When the audibility of a target speech signal is degraded, listeners rely on a combination of their accumulated

knowledge of language and cognitive skills related to language processing to understand the target (Mattys et al., 2012). One theoretical perspective that has emerged to describe how listeners use linguistic and cognitive abilities to understand degraded speech is the Ease of Language Understanding (ELU) model (Rönnberg et al., 2008). The ELU model suggests that listeners are better able to process degraded speech when they have a larger working memory capacity, to store and process incoming speech, and greater accumulated language abilities to support that processing (Zekveld et al., 2013). The ELU model has been applied to a wide range of degraded listening conditions to explain individual differences in speech recognition among adults (see Rönnberg et al., 2013, for a review). The ELU model more recently has been extended to the development of speech recognition in children. For example, individual differences in children's masked speech recognition are partially explained by individual differences in language skills and working memory (MacCutcheon et al., 2019; McCreery et al., 2017, 2019; but also see Magimairaj et al., 2018). The prolonged developmental trajectory of children's speech recognition in a TTS masker compared to SSN maskers (e.g., Corbin et al., 2016) supports the prediction that cognitive and linguistic demands may be particularly crucial for listening in TTS. However, previous studies have not assessed whether the effects of cognitive and linguistic skills on children's masked speech recognition vary depending on the masker type.

For SSN maskers, language abilities are generally associated with better masked speech recognition for children, but the effects of language are not always observed. Phonotactic, lexical, grammatical, and semantic knowledge has been found to support masked speech recognition in some studies (Blamey et al., 2001; Edwards et al., 2004; McCreery & Stelamachowicz, 2011; McCreery et al., 2017; Walley et al., 2003), but not in others (Eisenberg et al., 2000; Magimairaj et al., 2018; Stelmachowicz et al., 2004). Discrepancies across studies could be due to developmental and stimulus factors. Children may not be able to efficiently process linguistic cues, even when they are present in the stimuli. For example, work by Nittrouer and Boothroyd (1990) indicated that children may be less able to benefit from linguistic context than adults, which could limit the observed relationship between language abilities and speech recognition in noise. The ability for children to use linguistic cues may also depend on the masker. Buss et al. (2019) found that children benefited from the addition of semantic context when the masker was SSN, but not when the masker was TTS.

The lexical characteristics of the target stimuli and the age of the children in a study may also affect the degree to which language supports speech recognition in noise. Klein et al. (2017) measured the association between vocabulary size and speech recognition in SSN for monosyllabic words and nonwords for 5- to 12-year-old children. The benefits of larger vocabularies were only observed for recognition of late-acquired words and nonwords. A more recent study by Miller et al. (2019) examined masked speech recognition

for Spanish–English bilingual and English monolingual children in SSN and TTS masking conditions for sentences that included vocabulary that was appropriate for the youngest children in the study. In that study, an association between vocabulary size and masked speech recognition was observed for younger children (5- to 6-year-olds) in both masker conditions but was only observed for older children (9- to 10-year-olds) in the TTS masker. For sentence stimuli, children with better knowledge of sentence-level syntax and grammar also have better masked recognition than children with poorer grammatical knowledge (McCreery et al., 2017). Collectively, the results of these studies suggest that the benefits of language skills for speech recognition are not uniform for school-age children. Rather, the complexity of the target stimuli and masker influence these relationships, such that the strongest associations between language and speech recognition have been observed for younger children, linguistically complex stimuli, and TTS maskers.

The listener's ability to use their accumulated linguistic knowledge to support degraded speech recognition is facilitated by cognitive abilities (Mattys et al., 2012). Cognitive skills associated with language processing have also been examined as predictors of masked speech recognition in children. Domains of cognitive function related to regulation, planning, and actively processing incoming sensory information, known collectively as executive function, are areas that are likely to support masker speech recognition (see a review by Kronenberger, 2019). Executive function includes three core domains: working memory, inhibition of information, and cognitive flexibility (Zelazo et al., 2003). Of these, working memory has received the most attention in previous research as a predictor of masked speech recognition in children, potentially due to established linkages between learning language and working memory (Gathercole & Baddeley, 2014). The ability to store and process incoming auditory information in working memory is a key component of the ELU model (Rönnberg et al. 2013). Working memory may help children to recognize speech in noise, but research pertaining to this association has been mixed. Sullivan et al. (2015) found that school-age children with better working memory skills had better auditory comprehension than peers with poorer working memory. Our previous work revealed that typically developing school-age children with better verbal and visuospatial working memory skills had better speech recognition in SSN for monosyllabic words, semantically anomalous sentences, and sequences of words without syntactic structure or semantic meaning (McCreery et al., 2017). Work by Magimairaj et al. (2018), however, did not find correlations between recognition of Bamford-Kowal-Bench (BKB; Bench et al., 1979) sentences in four-talker babble and any measure of language, selective attention, or working memory in a group of typically developing school-age children. The discrepancies between these studies leave important questions about how a child's cognitive and linguistic abilities might interact with masker characteristics to produce the different patterns observed across studies.

Other aspects of executive function apart from working memory, including selective attention, inhibition of attention for task-irrelevant stimuli, and cognitive flexibility, may also support masked speech recognition in children. In adults, the ability to inhibit attention to irrelevant information and selectively attend to the target has been shown to support segregation of multiple sound sources (Shinn-Cunningham & Best, 2008). Wightman and Kistler (2005) showed that 4- to 16-year-old children had poorer recognition than adults when they were required to selectively attend to target speech in one ear when an SSN masker or competing talker was presented to the opposite ear. The largest differences between children and adults were observed when the contralateral competing signal was a male talker. Studies using broad measures of attention have not found consistent associations between recognition of degraded speech and attention in children. For example, the ability to sustain attention, often known as vigilance, was not associated with speech recognition in noise and reverberation for children with typical hearing or children with hearing loss (McCreery et al., 2019). Similarly, Brännström et al. (2020) did not find a relationship between passage comprehension in multitalker babble and a general measure of executive function that combined multiple components of attention and other aspects of executive function into a single construct. However, neither study included individual measures of the selective attention, inhibition, or cognitive flexibility, all factors known to be important for masked speech recognition in adults (Shinn-Cunningham & Best, 2008). When listening to speech with a small number of masker talkers, attentional control may facilitate selective attention to the target, and working memory facilitates combination of temporally sparse speech information.

### The Current Study

The overall goal of this study was to examine masked sentence recognition with three different maskers to determine whether the masker characteristics influence the relationships between masked speech recognition and language, working memory, and selective attention. Based on the ELU model and our previous research, we predicted that children with advanced language and working memory abilities would have better performance in an SSN masker than children with poorer abilities. Groups of younger children (5- to 6-year-olds) and older children (9- to 10-year-olds) were recruited based on our previous research indicating that the effects of language abilities and masker type vary between these two age groups (Klein et al., 2017; Miller et al., 2019). Likewise, we predicted that performance in an AMSSN masker would be predicted by the child's language and working memory abilities because recognizing speech based on sparse glimpses relies on the ability to hold and process speech cues in short-term memory. For a TTS masker, which results in greater informational masking, we predicted that children with better language, working memory, and selective attention would have better recognition than peers with poorer skills in these domains.

This pattern of results would suggest that different maskers may rely on different combinations of cognitive and linguistic skills to support speech recognition.

## Method

### Participants

A total of 60 school-age children participated in this study. Two age groups of 30 children were recruited: 5- to 6-year-olds ($M_{age}$ = 6.1 years) and 9- to 10-year-olds ($M_{age}$ = 9.9 years). Criteria for participation for all children were (a) hearing sensitivity ≤ 20 dB HL between 250 and 8000 Hz (American National Standards Institute, 2004); (b) no incidence of otitis media within 1 month prior to completing the study, as per parent report; (c) no speech or language impairments, as per parent report; and (d) native English speaking. Institutional review board approval was obtained from Boys Town National Research Hospital. All children provided assent (< 7 years) or consent (≥ 7 years) to participate, and all parents provided consent for their children to participate.

### Speech Recognition

#### Stimuli and Conditions

Target stimuli were sentences from the Revised BKB Standard Sentence Test (Bench et al., 1979). This corpus is composed of 21 lists of 16 sentences, developed using a lexicon derived from the speech of young children. Each sentence contains three or four key words, resulting in a total of 50 key words per list. The target sentences were recorded by a female talker who is a native speaker of American English. A sampling rate of 44100 Hz was used for the recording. Each sentence was saved as a .wav file at a reduced sampling rate of 24414 Hz, and recordings were root-mean-squared normalized using MATLAB.

Target sentences were presented in each of three maskers: (a) unmodulated SSN, (b) 10-Hz AMSSN, and (c) TTS. The SSN masker was created taking a Fast-Fourier Transform of the TTS masker, randomizing component phases, and using an Inverse Fast-Fourier Transform to create a noise masker with the same long-term spectrum as the two-talker masker. The AMSSN masker was created by multiplying the SSN by a 10-Hz raised sinusoid. The TTS masker was developed by Calandruccio et al. (2014) and was composed of two separate passages from the story *Jack and the Beanstalk*, each produced by a different female talker. Silent periods longer than 300 ms were reduced to approximately 100 ms using SoundStudio audio software. The recordings were root-mean-squared equalized using Praat, and the long-term average speech spectra were normalized using MATLAB. The two individual streams of speech were then summed to create the TTS masker.

#### Procedure

Participants were tested while sitting on a chair inside a single-walled sound isolating booth. They were instructed to ignore the background noise or speech and repeat aloud

as many words in the sentence that they heard spoken by the target talker. If participants were unsure of what they heard, they were prompted to guess. The tester sat outside the booth and wore headphones connected to a sound card (Avid Fast Track Solo). Participants wore a head-worn microphone (Beta 54, Shure) connected to the same sound card inside the test booth. The tester was able to adjust the volume of the sound card to optimize intelligibility for each participant. Each key word was scored as correct only if it was repeated exactly as presented; substitutions, additions, and deletions were scored as incorrect.

A custom MATLAB script controlled the experiment. All stimuli, including target sentences and the masker, were loaded into a real-time processor (RZ6, TDT) and presented diotically over circumaural headphones (HD25, Sennheiser). The overall level of the combined target and masker was fixed at 60 dB SPL throughout testing, and the signal-to-noise (SNR) was adjusted by adaptively varying both the target and masker level. Each run included two interleaved, one-down, one-up adaptive tracks. For one adaptive track, a correct response was coded whenever the participant repeated one or more key words in the sentence correctly. For the other adaptive track, a correct response was coded whenever the participant repeated all or all but one of the key words in the sentence correctly. The starting SNR for each track was 15 dB in the TTS masker, 5 dB in the SSN masker, and 0 dB in the amplitude-modulated noise masker. The step size for both adaptive tracks was 2 dB. Participants completed a block of 60 trials (2 adaptive tracks × 30 sentences per track) in each masker condition. Condition order was randomized for each participant. The sentence lists used in each condition was also randomized, and participants never heard the same sentence twice. Scoring for each key word on each trial was saved to a computer file, along with the SNR. Two-parameter logit functions were fitted to the data obtained for each listener in each condition, and the speech recognition threshold (SRT) was defined as the SNR corresponding to 50% correct.

## Language and Executive Function Measures

In addition to measures of speech perception in noise, each participant completed measures of language and executive function. The order in which tests were administered varied between subjects. Approximately half of the participants started with measures of speech recognition, and the other half started with language and executive function measures. Most children completed all testing in one session, though some of the younger children needed two sessions to complete testing. In cases when testing was completed over two sessions, Sessions 1 and 2 occurred within 4 weeks of each other.

### Language

Receptive vocabulary was measured with the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4, Form A; Dunn & Dunn, 2007). The PPVT-4 test includes four pictures on each test page, and the child was instructed to point to the picture that best illustrated the word given by the examiner. Receptive vocabulary was measured because children with larger vocabularies often have better masked speech recognition in our previous research and other studies (e.g., Klein et al., 2017; McCreery et al., 2017). Grammar was measured using the Test for Reception of Grammar Version 2 (TROG-2; Bishop, 2003). The TROG-2 is a test for reception of grammar that consists of 80 items, each of which is a four-alternative forced choice. The tester read a sentence aloud for each set of four pictures, and the child was instructed to point to the picture that best illustrated the sentence. A measure of grammar knowledge was included because the ability to parse sentence-level syntax was positively associated with masked speech recognition in our previous research (McCreery et al., 2017).

### Working Memory

Two subtests of the Automated Working Memory Assessment (AWMA; Alloway, 2007), Listening Recall and Odd-One-Out, were used to evaluate memory. Both tests assessed either verbal or visuospatial components of short-term memory and working memory. The test was administered on a laptop computer. During Listening Recall, a series of one to six sentences was played from the testing software via the laptop speaker, and the child was asked to repeat the final word of each sentence in the order they were played. At the end of a block of sentences, the child was asked to recall if each sentence in the block was true or false. During Odd-One-Out, the computer displayed three shapes in a row, and the child was instructed to point out the shape that did not match the other two. The shapes then disappeared from the screen, at which time the child was asked to point to the box that previously contained the odd-one-out shape. If the child correctly identified the position of the odd-one-out, the number of sets of shapes increased with each trial.

### Selective Attention and Cognitive Flexibility

Two subtests from the National Institutes of Health (NIH) Toolbox (Gershon et al., 2013) were used to measure additional components of executive function: The Flanker Inhibitory Control and Attention test, and the Dimensional Change Card Sort (DCCS) test. The Flanker subtest is a test of attention and inhibitory control. The child was instructed to focus on a single stimulus and to inhibit attention to flanking stimuli. The DCCS subtest is a test of attention and cognitive flexibility. The child was instructed to point to pictures on a screen that varied across shape and color. A recorded word indicated whether the child should select the target based on the shape or the color. The child then selected one of two pictures on the touch-screen based on the instructed cue.

## Statistical Analyses

All statistical analyses and data visualization were completed using R Statistical Software (R Core Team,

2018). Data visualization was completed using the ggplot2 (Wickham & Chang, 2016) and sjPlot (Lüdecke & Schwemmer, 2018) packages for R. Descriptive statistics were calculated for each predictor and outcome measure. A repeated-measures analysis of variance was used to examine differences in speech recognition across age group and masker type. Two-tailed tests were used for all $t$ tests and correlations. For language and cognitive measures, standard scores were used to compare children in the experiment to the normative sample for each test, whereas raw scores were used to represent each construct in statistical models, with age in months as a covariate. Pearson correlations were calculated to examine the pattern of relationship between predictors and outcomes. Linear regression models were used to test the effects of linguistic and cognitive predictors on masked speech recognition in the three masker conditions. Variables and predictors were mean centered prior to regression analyses to simplify the interpretation of raw regression coefficients. The normality of each linear regression model's residuals was assessed to identify potential violations of statistical assumptions. All possible interaction terms were assessed for each model. To adjust for inflation of Type I error rate in analyses with multiple comparisons, the False Discovery Rate–adjusted (Benjamini & Hochberg, 1995) $p$ values are reported.

## Results

Figure 1 shows the SRT for each subject for the three masker conditions by age group. A repeated-measures analysis of variance with age group and masker type was conducted to determine the pattern of speech recognition performance across age and condition. The main effects of masker, $F(2, 174) = 1026.6$, $p < .001$, $\eta_p^2 = .92$, and age group, $F(1, 174) = 96.15$, $p < .001$, $\eta_p^2 = .35$, were significant. Children performed best in the AMSSN masker ($M = -12.1$ dB, $SD = 2.9$ dB), poorer in the SSN masker ($M = -6.1$ dB, $SD = 1.2$ dB), and poorest in the TTS masker ($M = 1.82$ dB, $SD = 1.9$ dB). The two-way interaction between age group and masker type, $F(2, 174) = 7.14$, $p = .001$, $\eta_p^2 = .08$, was significant. Post hoc testing using Tukey's honestly significant difference indicated that the mean differences between masker, age group, and the Masker × Age Group interactions were all significant. Older children had lower SRTs relative to younger children for all three maskers. The significant Masker × Age interaction reflects larger differences between the two age groups for the AMSSN masker than the SSN and TTS conditions. The average difference in SRT between older and younger children was 1.4 dB in the SSN masker ($p < .001$), 3.7 dB in the AMSSN ($p < .001$), and 2.1 dB in the TTS masker ($p < .001$).

Table 1 shows individual standard scores for measures of language and executive function. Differences between groups on each predictor variable were assessed using independent-samples $t$ tests. The older children had higher scores on the Odd-One-Out subtest of the AWMA and on the DCCS task of the NIH Toolbox, but there were no differences in standard scores between age groups on other predictors. Table 2 shows the Pearson correlations between SRT in each masker condition and each of the predictor variables. There was a pattern of medium to large correlations between each of the predictor variables and SRT for each masking condition with the exception of the DCCS task, which was not significantly correlated with SRT or any other predictor variable.

To examine the factors that predicted individual differences in speech recognition across the three masker conditions, one linear regression model was tested for each of the three masker conditions. All measures were included in the models, except those that (a) did not correlate with speech recognition in any masking condition or (b) had high correlations with another predictor variable that represented the same cognitive or linguistic construct, to avoid collinearity between predictors. On this basis, the DCCS task was excluded from further analysis because of a lack of correlation with speech recognition in any masking condition. The Odd-One-Out subtest of the AWMA was excluded to avoid collinearity between predictors due to the high correlation with the Listening Recall subtest, which also represents working memory. The Test of Reception of Grammar was omitted to avoid collinearity because of high correlations with both PPVT and measures of working memory. The remaining cognitive and linguistic measures were included in the regression models to assess individual differences in SRT for each masker.

Table 3 shows the linear regression model results for each masker type. For the SSN and AMSSN, age and PPVT were negatively associated with SRT. Each model also had a significant Age × PPVT interaction. Increasing age and PPVT scores were associated with lower SRT for SSN and AMSSN, but the significant interaction indicated that the relationship between PPVT and SRT decreased with age. For the TTS masker, significant effects of age, PPVT, and the Age × PPVT interaction were observed, but higher NIH Flanker Test and AWMA Listening Recall were also associated with lower SRT. Figure 2 shows the relationship between SRT and PPVT by age group. Pearson correlations between PPVT and SRT for each age group and masker were calculated and are shown in Table 4. Strong correlations ($r = -.66$ to $-.74$) between vocabulary size and SRT were observed for younger children, whereas the correlations between vocabulary size and SRT for the older children were weak ($-.19$ to $.14$) and not statistically significant. This pattern of results demonstrates that the effect of vocabulary on SRT was large for younger children and not significant for older children. Table 5 shows the linear regression model for modulation masking release, which was calculated as the difference between the SRT for the SSN and the AMSSN. For modulation masking release, PPVT was the only significant predictor, indicating that children with higher vocabulary scores had greater mod>ulation masking release than children with lower vocabulary scores. Although vocabulary was a significant predictor of performance in both the SSN and AMSSN

**Figure 1.** The speech recognition threshold (SRT) in dB as a function of age group (5- to 6-year-olds = gold; 9- to 10-year-olds = blue) and masker type, indicated at the top of the figure. The filled circles represent individual data points for each age group and condition. The box plots represent the 25th–75th percentiles and vertical lines extend to the 5th and 95th percentiles. The horizontal line bisecting each boxplot represents the median for each condition and age group.
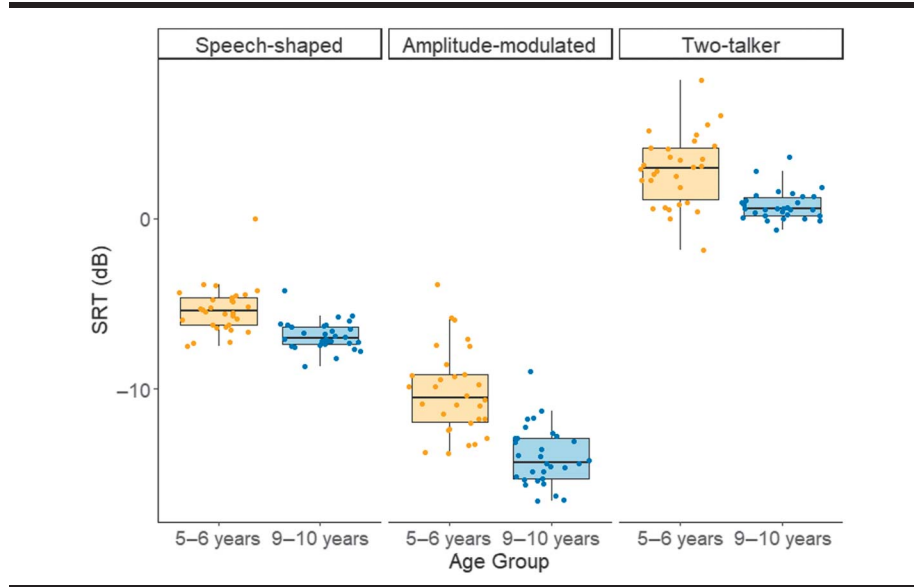


masker conditions, that association was larger in the SSN masker condition.

## Discussion

The primary goal of this study was to examine the effects of masker type on the relationship between masked speech recognition, language, and executive function for two age groups of typically developing children. For SSN and AMSSN maskers, we predicted that children with better language and working memory skills would have better masked speech recognition based on previous research in children (e.g., McCreery et al., 2017) and adults (Rönnberg et al., 2013). We expected that the effects of language on masked speech recognition might be largest for the youngest age group, as in previous studies (Klein et al., 2017; Miller

et al. 2019). Finally, we predicted that speech recognition in a TTS masker would require better language, working memory, and selective attention skills, consistent with the previous literature on cognitive demands that occur when the listener must perceptually isolate the target talker from a small number of competing talkers (Corbin et al., 2016; Sobon et al., 2019). The findings of this study supported some of these predictions, but not others. The prediction that stronger language, working memory, and selective attention skills would be associated with better speech recognition was confirmed, but only for the TTS masker. Vocabulary ability was associated with masked speech recognition for all three masking conditions, but only for the 5- to 6-year-old age group. Younger children with better receptive vocabulary skills, measured by the PPVT, had better masked speech recognition across all conditions

**Table 1.** Standard scores for measures of language, working memory, and executive function.

| Variable | Mean, 5- to 6-year-olds | Mean, 9- to 10-year-olds | Difference |
|---|---|---|---|
| Age | 6.1 years | 9.9 years | *t* = 23.94, *p* < .0001 |
| PPVT | 119.1 | 115.8 | *t* = 0.439, *p* = .66 |
| TROG | 108.4 | 103.3 | *t* = 1.08, *p* = .29 |
| LR | 115.6 | 115.2 | *t* = 0.09, *p* = .92 |
| OOO | 111 | 122.5 | *t* = 3.44, *p* = .001 |
| Flanker | 110.8 | 113.4 | *t* = 0.65. *p* = .515 |
| DCCS | 105.6 | 127.7 | *t* = 4.19, *p* < .0001 |

*Note.* Difference with *p* < .05 are bolded. PPVT = Peabody Picture Vocabulary Test; TROG = Test of Reception of Grammar; LR = Listening Recall subtest of the Automated Working Memory Assessment; OOO = Odd-One-Out subtest of the Automated Working Memory Assessment; Flanker = NIH Toolbox Flanker task; DCCS = NIH Toolbox Dimensional Change Card Sort Task.

**Table 2.** Pearson correlations between predictor variables.

| Variable | SSNR50 | AM50 | TTN50 | Age | PPVT | TROG | LR | OOO | Flanker | DCCS |
|---|---|---|---|---|---|---|---|---|---|---|
| SSN50 | 1 | | | | | | | | | |
| AM50 | .74* | 1 | | | | | | | | |
| TTN50 | .75* | .75* | 1 | | | | | | | |
| Age | −.65* | −.74* | −.65* | 1 | | | | | | |
| PPVT | −.65* | −.74* | −.66* | .88* | 1 | | | | | |
| TROG | −.59* | −.57* | −.54* | .69* | .79* | 1 | | | | |
| LR | −.56* | −.60* | −.61* | .77* | .83* | .74* | 1 | | | |
| OOO | −.57* | −.60* | −.61* | .87* | .78* | .69* | .79* | 1 | | |
| Flanker | .51* | .52* | −.44* | −.88* | −.72* | −.59* | −.65* | −.76* | 1 | |
| DCCS | .12 | .08 | .22 | −.13 | −.13 | .04 | −.04 | −.06 | .09 | 1 |

*Note.* SSN50 = Target-to-masker ratio for 50% correct (TMR50) for speech-shaped noise; AM50 = TMR50 for amplitude-modulated noise; TTN50 = TMR50 for two-talker masker; PPVT = Peabody Picture Vocabulary Test; TROG = Test of Reception of Grammar; LR = Listening Recall subtest of the Automated Working Memory Assessment; OOO = Odd-One-Out subtest of the Automated Working Memory Assessment; Flanker = NIH Toolbox Flanker task; DCCS = NIH Toolbox Dimensional Change Card Sort Task.

*$p < .05$ after adjustment using False Discovery Rate (Benjamini & Hochberg, 1995).

relative to peers with poorer receptive vocabulary skills, but the relationship between receptive vocabulary and masked speech recognition was not significant for the older children. These results suggest that the relationship between language, executive function, and masked speech recognition depends on the type of masker and on age.

## Linguistic and Cognitive Predictors Across Masking Conditions

The main novel finding of this study was that different masking conditions may require children to use different combinations of cognitive and linguistic skills to recognize speech. The different cognitive mechanisms that

**Table 3.** Linear regression models for speech recognition threshold by masker type.

| Speech-shaped noise | $R^2 = .37$, $F(5, 54) = 7.87$, $p < .0001$ | | |
|---|---|---|---|
| **Factor** | **Raw coefficient** | **t value** | **p value** |
| Intercept | 0 | 1.36 | .18 |
| **Age** | **−1.21** | **−2.88** | **< .01** |
| **PPVT** | **−0.58** | **−2.27** | **.03** |
| Flanker | −0.02 | −0.64 | .53 |
| Listening Recall | −0.01 | −0.29 | .77 |
| **Age × PPVT** | **0.006** | **2.17** | **.04** |
| **Amplitude-modulated speech-shaped noise** | $R^2 = .59$, $F(5, 54) = 18.08$, $p < .0001$ | | |
| **Factor** | **Raw coefficient** | **t value** | **p value** |
| Intercept | 0 | 1.96 | .05 |
| **Age** | **−2.32** | **−3.05** | **< .01** |
| **PPVT** | **−0.14** | **−2.95** | **< .01** |
| Flanker | −0.01 | −0.20 | .84 |
| Listening Recall | 0.01 | 0.26 | .79 |
| **Age × PPVT** | **0.01** | **2.31** | **.02** |
| **Two-talker** | $R^2 = .65$, $F(5, 54) = 22.59$, $p < .0001$ | | |
| **Factor** | **Raw coefficient** | **t value** | **p value** |
| **Intercept** | **0** | **7.69** | **< .01** |
| **Age** | **−2.85** | **−6.04** | **< .01** |
| **PPVT** | **−0.13** | **−4.81** | **< .01** |
| **Flanker** | **−0.07** | **−2.04** | **.04** |
| **Listening Recall** | **−0.09** | **−2.05** | **.04** |
| **Age × PPVT** | **0.02** | **4.95** | **< .01** |

*Note.* Coefficient for the intercept for each model is zero because all variables were mean centered. Significant factors ($p < .05$) in each model are bolded. PPVT = Peabody Picture Vocabulary Test; Flanker = NIH Toolbox Flanker task.

**Figure 2.** The speech recognition threshold (SRT) in dB as a function of receptive vocabulary (Peabody Picture Vocabulary Test raw score) for younger (gold circles) and older (blue triangles) children for each masking condition. The solid lines represent the linear relationship between SRT and receptive vocabulary for younger (gold) and older (blue) children in each masking condition.
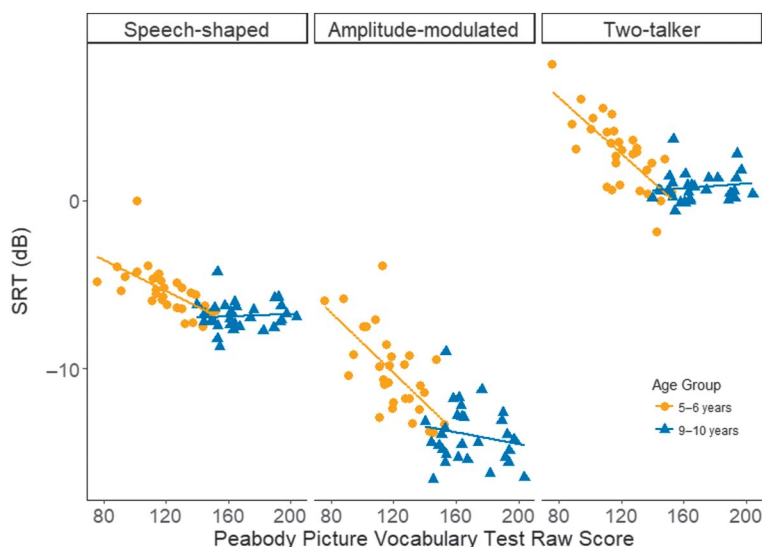


are apparent across different masking conditions could explain different maturational trajectories for different maskers that have been observed in previous studies of masked speech recognition in children. We examined the effects of individual differences in language and specific domains of executive function related to working memory and selective attention on masked speech recognition for different masking conditions. The finding that children's speech recognition in a TTS masker depends on language, working memory, and selective attention supports previous research indicating a protracted developmental trajectory for speech recognition in TTS (Buss et al., 2017, 2019; Corbin et al., 2016). In contrast, speech recognition in the SSN and AMSSN was only associated with receptive vocabulary abilities, and the relationship between vocabulary and masked speech recognition was only observed for the 5- to 6-year-old children. The additional influence of working memory and selective attention for the TTS condition is consistent with the suggestion that maturation of selective attention is critical for children's ability to recognize speech when the competing signal is speech (e.g., Allen & Wightman, 1995). These results are also broadly consistent

with the results of Sobon et al. (2019). That study showed that school-age children required more audible target cues than adults to recognize speech in noise, but the child/adult difference was even larger in a TTS masker. Relatively modest effects of age in the SSN masker could reflect maturation of language. More pronounced effects of age in the TTS masker could reflect additional contributions of the maturation of selective attention and working memory. Collectively, these results are all consistent with the idea that children with better selective attention skills may be more able to segregate the target talker from the competing talkers or inhibit the competing talkers as irrelevant information.

Working memory abilities were associated with masked speech recognition in the TTS masker, but not for the SSN or AMSSN. Research with children (McCreery et al., 2017; Sullivan et al., 2015) and adults (Akeroyd, 2008; Souza & Sirow, 2014) supports the main idea of the ELU model, that working memory skills help the listener to store and interpret degraded acoustic input (see Rönnberg et al., 2013, for a review). Working memory is predicted to help listeners in conditions that are particularly cognitively demanding, which may explain the differences in the influence of working

**Table 4.** Pearson correlations between Peabody Picture Vocabulary Test score and speech recognition threshold by age group and masking condition.

| Masking condition | 5- to 6-year-old children | 9- to 10-year-old children |
|---|---|---|
| Speech-shaped noise masker | −.69 | −.19 |
| Amplitude-modulated speech-shaped noise masker | −.66 | .08 |
| Two-talker masker | −.74 | .14 |

*Note.* Significant correlations (*p* < .05) in each model are bolded.

**Table 5.** Linear regression model for the difference in speech recognition thresholds for speech-shaped noise and amplitude-modulated speech-shaped noise.

| | $R^2 = .39$, $F(3, 56) = 13.29$, $p < .0001$ | | |
|---|---|---|---|
| Factor | Raw coefficient | t value | p value |
| Intercept | 0 | 0.03 | .98 |
| Age | −0.67 | −0.91 | .36 |
| **PPVT** | **−0.42** | **−2.78** | **< .01** |
| Flanker | −0.03 | −0.42 | .62 |
| Listening Recall | 0.04 | 0.66 | .51 |
| Age × PPVT | 0.02 | 1.2 | .23 |

*Note.* Significant correlations ($p < .05$) are bolded. PPVT = Peabody Picture Vocabulary Test; Flanker = NIH Toolbox Flanker task.

memory observed for different masker conditions in the current data. For example, previous studies that have demonstrated a relationship between children's speech recognition in SSN and working memory included conditions that were difficult in terms of the target level of performance (McCreery et al., 2017) or the use of a comprehension task (Sullivan et al., 2015). In the Magimairaj et al. (2018) study that did not find an association between working memory and speech recognition in four-talker babble, the stimuli were simple sentences based on the lexicon of the youngest children in the study. Children had higher levels of recognition than were targeted in previous studies (~67% on average compared to 50% in McCreery et al., 2017). The current results, which suggest that working memory was only associated with performance in TTS, are consistent with the idea that the task demands, including the stimuli and masker characteristics, could moderate the degree to which children rely on working memory to recognize masked speech. Both TTS and AMSSN offer the listener opportunities to glimpse the target during fluctuations in the masker. Reconstructing the audible components of the target during masked speech recognition was expected to rely on working memory. A key difference between the TTS and AMSSN maskers is that the TTS is perceptually similar to the target speech, which makes perceptually isolating the target and the masker more challenging. The lack of an association between working memory and masking release suggests that individual differences in working memory may not be related to glimpsing in an AMSSN. Rather, children with better working memory skills may have advantages when the glimpses are less predictable and the similarity of the target and masker require them to perceptually isolate the target talker from the TTS masker.

The prediction that children with better working memory skills have larger release from masking (i.e., difference in SRTs for SSN and AMSSN masker conditions) was not supported by the data. Some data have indicated child/adult differences in the ability to benefit from masker modulation (Buss et al., 2016; Hall et al., 2012), although this is not always observed (Stuart, 2008). Working memory was expected to be associated with the listeners' ability to store and process glimpses of speech that were audible during

fluctuations in the AMSSN. However, language was the only significant predictor of the SRT difference between the two maskers. Children with larger vocabularies had larger release from masking than children with smaller vocabularies. It is possible that the speech recognition task or the use of periodic modulation in the absence of informational masking placed minimal cognitive demands on the listener, such that working memory was not a limiting factor in this protocol.

Another key finding was that, across all maskers, younger children with higher language abilities had better masked speech recognition than peers with poorer language skills, but language did not predict masked speech recognition in older children. This finding is consistent with several previous studies (Blamey et al., 2001; Klein et al., 2017; McCreery et al., 2017; Miller et al., 2019). Klein et al. (2017) found that the association between language and speech recognition in SSN was only observed for older children when the stimuli included later acquired vocabulary words. The stimuli in that study were BKB sentences, which were constructed using words within the lexicon of children as young as 5 years of age. Stimuli with lexical and semantic content that reflects the language skills of the youngest age group may not require older children to apply their advanced linguistic knowledge in these areas as much as younger children. Speech recognition tasks are often selected to be developmentally appropriate for the youngest children in a study, but stimuli developed for younger children may not place the same language processing demands on older children. This may explain why some previous studies that included mixed groups of younger and older children did not find an effect of language on speech recognition in degraded conditions (Eisenberg et al., 2000; Magimairaj et al., 2018; Stelmachowicz et al., 2004). A related explanation could account for the failure to observe a correlation between masked sentence recognition and sentence-level grammatical knowledge as measured by the TROG. The BKB sentences have relatively simple syntactic structure, such that grammatical knowledge may not have been a limiting factor in even the youngest subjects. In contrast to our previous research (McCreery et al., 2017), sentence-level grammatical knowledge was not associated with masked speech recognition in the present data set.

## Clinical Implications

In addition to extending our scientific knowledge of the underlying mechanisms responsible for the development of masked speech recognition in school-age children, results of this study have important implications for clinical and educational practices. Masked speech recognition is often used by audiologists as a functional assessment of listening in children. The SSN maskers that are typically used to assess speech recognition clinically may not impose the same cognitive demands that children encounter when listening in real-world environments, which often include competing speech. The development of clinical speech recognition protocols for children has been almost solely focused

on developing stimuli that represent the continuum of linguistic complexity from nonwords to words to sentences (Uhler et al., 2017). Different masker types may also be an important parameter for audiologists to consider when developing and selecting clinical speech recognition materials. A multitalker babble composed of four or more individual talkers often produces similar masking effects to SSN maskers (Carhart et al. 1975), so the use of a TTS may be more indicative of the challenges experienced in everyday listening than maskers comprising a larger number of talkers (Hillock-Dunn et al., 2015). Efforts are underway to develop clinical assessment tools for audiologists to reliably assess masked speech recognition with TTS maskers (Calandruccio et al., 2014; Miller et al., 2019).

### Limitations and Future Research Directions

In interpreting the results of this study, there are several limitations that should be considered. The population for the study was limited to typically developing children with no history of listening problems, hearing loss, and/or other developmental concerns. For the most part, the children who participated in this study do not experience the same significant communication problems as children with additional developmental concerns. Children with hearing loss, developmental language disorders, or other developmental conditions that affect language or cognition may be at additional risk for difficulty understanding speech in complex, multisource listening environments. Data from children with hearing loss support the hypothesis that cognitive and linguistic skills are important for supporting masked speech recognition (Blamey et al., 2001; McCreery et al., 2015). Future research should examine the cognitive and linguistic mechanisms of energetic and informational masking in children with hearing loss or with developmental conditions that affect language and cognitive abilities, as these children may be at even greater risk for difficulty in complex listening environments than typically developing peers with normal hearing.

## Conclusions

In summary, we examined speech recognition for a large group of typically developing children across three masker conditions to examine the influence of language, working memory, and selective attention on masked speech recognition. In a TTS masker, children with better language, working memory, and selective attention had better masked speech recognition, whereas language was the only significant predictor of masked speech recognition in the SSN and AMSSN. The effect of receptive vocabulary for speech recognition across all three masker conditions decreased as children increased in age, suggesting that the effects of language on speech recognition in degraded conditions may depend on the stimuli and the age of the children. The difference between the SSN and the AMSSN was used to estimate modulation masking release. Children with better language abilities had larger modulation masking

release than peers with poorer language abilities. These results suggest that different linguistic and cognitive mechanisms influence masked speech recognition, depending on the cognitive demands of the task, characteristics of the masker, and age of the child.

## References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*(Suppl. 2), S53–S71. https://doi.org/10.1080/14992020802301142

Allen, P., & Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech, Language, and Hearing Research, 38*(2), 503–511. https://doi.org/10.1044/jshr.3802.503

Alloway, T. P. (2007). *Automated Working Memory Assessment.* Pearson Assessment.

American National Standards Institute. (2004). *Methods for manual pure-tone threshold audiometry (ANSI S3.21–2004).*

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bishop, D. V. M. (2003). *Test for Reception of Grammar Version 2.* Pearson Assessment.

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., & Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech, Language, and Hearing Research, 44*(2), 264–285. https://doi.org/10.1044/1092-4388(2001/022)

Bonino, A. Y., Leibold, L. J., & Buss, E. (2013). Release From perceptual masking for children and adults: Benefit of a carrier phrase. *Ear and Hearing, 34*(1), 3–14. https://doi.org/10.1097/AUD.0b013e31825e2841

Brännström, K. J., von Lochow, H., Lyberg Åhlander, V., & Sahlén, B. (2020). Passage comprehension performance in children with cochlear implants and/or hearing aids: The effects of voice quality and multi-talker babble noise in relation to executive function. *Logopedics Phoniatrics Vocology, 45*(1), 15–23. https://doi.org/10.1080/14015439.2019.1587501

Brown, D. K., Cameron, S., Martin, J. S., Watson, C., & Dillon, H. (2010). The North American Listening in Spatialized Noise—Sentences Test (NA LiSN-S): Normative data and test-retest

reliability studies for adolescents and young adults. *Journal of the American Academy of Audiology, 21*(10), 629–641. https://doi.org/10.3766/jaaa.21.10.3

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *The Journal of the Acoustical Society of America, 112*(6), 2985–2995. https://doi.org/10.1121/1.1512703

Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. *Ear and Hearing, 40*(5), 1117–1126. https://doi.org/10.1097/AUD.0000000000000692

Buss, E., Leibold, L. J., & Hall, J. W., III. (2016). Effect of response context and masker type on word recognition in school-age children and adults. *The Journal of the Acoustical Society of America, 140*(2), 968–977. https://doi.org/10.1121/1.4960587

Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America, 141*(4), 2650–2660. https://doi.org/10.1121/1.4979936

Calandruccio, L., Gomez, B., Buss, E., & Leibold, L. J. (2014). Development and preliminary evaluation of a pediatric Spanish-English speech perception task. *American Journal of Audiology, 23*(2), 158–172. https://doi.org/10.1044/2014_AJA-13-0055

Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *The Journal of the Acoustical Society of America, 58*(Suppl. 1), S35. https://doi.org/10.1121/1.2002082

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America, 45*(3), 694–703. https://doi.org/10.1121/1.1911445

Choi, S., Lotto, A., Lewis, D., Hoover, B., & Stelmachowicz, P. (2008). Attentional modulation of word recognition by children in a dual-task paradigm. *Journal of Speech, Language, and Hearing Research, 51*(4), 1042–1054. https://doi.org/10.1044/1092-4388(2008/076)

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing, 37*(1), 55–63. https://doi.org/10.1097/AUD.0000000000000201

Crukley, J., Scollie, S., & Parsa, V. (2011). An exploration of non-quiet listening at school. *Journal of Educational Audiology, 17*(1), 23–35.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Pearson Assessments. https://doi.org/10.1037/t15144-000

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G., Jr. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America, 114*(1), 368–379. https://doi.org/10.1121/1.1577562

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in non-word repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107*(5), 2704–2710. https://doi.org/10.1121/1.428656

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology, 80*(11 Suppl. 3), S2–S6. https://doi.org/10.1212/WNL.0b013e3182872e5f

Goldsworthy, R. L., & Markle, K. L. (2019). Pediatric hearing loss and speech recognition in quiet and in different types of background noise. *Journal of Speech, Language, and Hearing Research, 62*(3), 758–767. https://doi.org/10.1044/2018_JSLHR-H-17-0389

Hall, J. W., III., Buss, E., Grose, J. H., & Roush, P. A. (2012). Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation. *Ear and Hearing, 33*(3), 340–348. https://doi.org/10.1097/AUD.0b013e31823fa4c3

Hall, J. W., III., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing, 23*(2), 159–165. https://doi.org/10.1097/00003446-200204000-00008

Hillock-Dunn, A., Taylor, C., Buss, E., & Leibold, L. J. (2015). Assessing speech perception in children with hearing loss: What conventional clinical tools may miss. *Ear and Hearing, 36*(2), e57. https://doi.org/10.1097/AUD.0000000000000110

Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acta Acustica United With Acustica, 78*(5), 258–272.

Klein, K. E., Walker, E. A., Kirby, B., & McCreery, R. W. (2017). Vocabulary facilitates speech perception in children with hearing aids. *Journal of Speech, Language, and Hearing Research, 60*(8), 2281–2296. https://doi.org/10.1044/2017_JSLHR-H-16-0086

Kronenberger, W. G. (2019). Executive functioning and language development in children with cochlear implants. *Cochlear Implants International, 20*(Suppl. 1), 2–5.

Leibold, L. J. (2017). Speech perception in complex acoustic environments: Developmental effects. *Journal of Speech, Language, and Hearing Research, 60*(10), 3001–3008. https://doi.org/10.1044/2017_JSLHR-H-17-0070

Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech-shaped noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research, 56*(4), 1144–1155. https://doi.org/10.1044/1092-4388(2012/12-0011)

Leibold, L. J., & Buss, E. (2019). Masked speech recognition in school-age children. *Frontiers in Psychology, 10*, 1981–1981. https://doi.org/10.3389/fpsyg.2019.01981

Lüdecke, D., & Schwemmer, C. (2018). *sjPlot: Data visualization for statistics in social science* (Version 2.6.2). https://www.rdocumentation.org/packages/sjPlot/versions/2.8.4

MacCutcheon, D., Pausch, F., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., Fel, J., & Ljung, R. (2019). The contribution of individual differences in memory span and language ability to spatial release from masking in young children. *Journal of Speech, Language, and Hearing Research, 62*(10), 3741–3751. https://doi.org/10.1044/2019_JSLHR-S-19-0012

Magimairaj, B. M., Nagaraj, N. K., & Benafield, N. J. (2018). Children's speech perception in noise: Evidence for dissociation from language and working memory. *Journal of Speech,*

*Language, and Hearing Research, 61*(5), 1294–1305. https://doi.org/10.1044/2018_JSLHR-H-17-0312

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McCreery, R. W., Ito, R., Spratford, M., Lewis, D., Hoover, B., & Stelmachowicz, P. G. (2010). Performance-intensity functions for normal-hearing adults and children using computer-aided speech perception assessment. *Ear and Hearing, 31*(1), 95–101. https://doi.org/10.1097/AUD.0b013e3181bc7702

McCreery, R. W., Spratford, M., Kirby, B., & Brennan, M. (2017). Individual differences in language and working memory affect children's speech recognition in noise. *International journal of audiology, 56*(5), 306–315. https://doi.org/10.1080/14992027.2016.1266703

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America, 130*(6), 4070–4081. https://doi.org/10.1121/1.3658476

McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., & Brennan, M. (2019). Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss. *Frontiers in Neuroscience, 13.* https://doi.org/10.3389/fnins.2019.01093

McCreery, R. W., Walker, E. A., Spratford, M., Oleson, J., Bentler, R., Holte, L., & Roush, P. (2015). Speech Recognition and Parent Ratings From Auditory Development Questionnaires in Children Who Are Hard of Hearing. *Ear and Hearing, 36,* 60S–75S. https://doi.org/10.1097/AUD.0000000000000213

Miller, M. K., Calandruccio, L., Buss, E., McCreery, R. W., Oleson, J., Rodriguez, B., & Leibold, L. J. (2019). Masked English speech recognition performance in younger and older Spanish–English bilingual and English monolingual-children. *Journal of Speech, Language, and Hearing Research, 62*(12), 4578–4591. https://doi.org/10.1044/2019_JSLHR-19-00059

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *The Journal of the Acoustical Society of America, 98*(4), 1909–1920. https://doi.org/10.1121/1.414458

Nelson, P. B., & Soli, S. (2000). Acoustical barriers to learning: Children at risk in every classroom. *Language, Speech, and Hearing Services in Schools, 31*(4), 356–361. https://doi.org/10.1044/0161-1461.3104.356

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 87*(6), 2705–2715. https://doi.org/10.1121/1.399061

R Core Team. (2018). *R: A language and environment for statistical computing* (Version 3.4.4). R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2006). Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *The Journal of the Acoustical Society of America, 120*(6), 3988–3997. https://doi.org/10.1121/1.2358008

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7,* 31. https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*(Suppl. 2), S99–S105. https://doi.org/10.1080/14992020802301167

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12*(4), 283–299. https://doi.org/10.1177/1084713808325306

Sobon, K. A., Taleb, N. M., Buss, E., Grose, J. H., & Calandruccio, L. (2019). Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging. *The Journal of the Acoustical Society of America, 145*(4), EL284–EL290. https://doi.org/10.1121/1.5097377

Souza, P. E., & Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing aids. *American Journal of Audiology, 23*(4), 394–401. https://doi.org/10.1044/2014_AJA-14-0006

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., & Moeller, M. P. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. *Archives of Otolaryngology–Head & Neck Surgery, 130*(5), 556–562. https://doi.org/10.1001/archotol.130.5.556

Stone, M. A., Füllgrabe, C., & Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America, 132*(1), 317–326. https://doi.org/10.1121/1.4725766

Stuart, A. (2008). Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. *Journal of the American Academy of Audiology, 19*(2), 135–146. https://doi.org/10.3766/jaaa.19.2.4

Sullivan, J. R., Carrano, C., & Osman, H. (2015). Working memory and speech recognition performance in noise: Implications for classroom accommodations. *Journal of Communication Disorders, Deaf Studies and Hearing Aids, 3,* 136–141. https://doi.org/10.4172/2375-4427.1000136

Talarico, M., Abdilla, G., Aliferis, M., Balazic, I., Giaprakis, I., Stefanakis, T., Foenander, K., Grayden, D. B., & Paolini, A. G. (2007). Effect of age and cognition on childhood speech in noise perception abilities. *Audiology and Neurotology, 12*(1), 13–19. https://doi.org/10.1159/000096153

Uhler, K., Warner-Czyz, A., Gifford, R., & PMSTB Working Group. (2017). Pediatric Minimum Speech Test Battery. *Journal of the American Academy of Audiology, 28*(3), 232–247. https://doi.org/10.3766/jaaa.15123

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16*(1–2), 5–20. https://doi.org/10.1023/A:1021789804977

Wickham, H., & Chang, W. (2016). *ggplot2: Create elegant data visualisations using the grammar of graphics* (Version 2.2.1). https://cran.r-project.org/web/packages/ggplot2/index.html

Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America, 118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Wilson, R. H., Farmer, N. M., Gandhi, A., Shelburne, E., & Weaver, J. (2010). Normative data for the words-in-noise test for 6- to 12-year-old children. *Journal of Speech, Language, and Hearing Research, 53*(5), 1111–1121. https://doi.org/10.1044/1092-4388(2010/09-0270)

Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and

young adults. *Ear and Hearing, 33*(6), 731–744. https://doi.org/10.1097/AUD.0b013e31825aecad

Yacullo, W. S., & Hawkins, D. B. (1987). Speech recognition in noise and reverberation by school-age children. *Audiology, 26*(4), 235–246. https://doi.org/10.3109/00206098709081552

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 134*(3), 2225–2234. https://doi.org/10.1121/1.4817926

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., & Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3), vii–137. https://doi.org/10.1111/j.0037-976x.2003.00260.x