Effect of response context and masker type on word recognition in school-age children and adults

Emily Buss,^{1,a)} Lori J. Leibold,² and Joseph W. Hall III¹

¹*Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA*

²*Boys Town National Research Hospital, Omaha, Nebraska 68131, USA*

(Received 11 March 2016; revised 13 July 2016; accepted 26 July 2016; published online 12 August 2016)

In adults, masked speech recognition improves with the provision of a closed set of response alternatives. The present study evaluated whether school-age children (5–13 years) benefit to the same extent as adults from a forced-choice context, and whether this effect depends on masker type. Experiment 1 compared masked speech reception thresholds for disyllabic words in either an open-set or a four-alternative forced-choice (4AFC) task. Maskers were speech-shaped noise or two-talker speech. Experiment 2 compared masked speech reception thresholds for monosyllabic words in two 4AFC tasks, one in which the target and foils were phonetically similar and one in which they were dissimilar. Maskers were speech-shaped noise, amplitude-modulated noise, or two-talker speech. For both experiments, it was predicted that children would not benefit from the information provided by the 4AFC context to the same degree as adults, particularly when the masker was complex (two-talker) or when audible speech cues were temporally sparse (modulated-noise). Results indicate that young children do benefit from a 4AFC context to the same extent as adults in speech-shaped noise and amplitude-modulated noise, but the benefit of context increases with listener age for the two-talker speech masker. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4960587>]

[VB]

Pages: 968–977

I. INTRODUCTION

Word recognition can be measured using either an open-set response or a closed-set, forced-choice response. In a typical open-set task, the listener repeats back the target words that he or she hears, and those verbal responses are scored by the tester as correct or incorrect. In a forced-choice task, the listener selects from among a restricted set of allowable response alternatives; these alternatives are commonly presented visually, either orthographically or as a set of pictures. A forced-choice picture-pointing response is frequently used to evaluate speech perception in children, both clinically [e.g., the word intelligibility by picture identification (WIPI) test; Ross and Lerman, 1970] and in laboratory studies (Hall *et al.*, 2002; Litovsky, 2005). In contrast to verbal responses, picture-pointing responses do not rely on accurate speech production on the part of the listener or accurate scoring on the part of the tester. This can be advantageous when testing children with speech production deficits, such as those associated with hearing loss (Blamey *et al.*, 2001), or when the target speech materials are not in the tester's native language (Calandruccio *et al.*, 2014). Interpretation of open- and closed-set word recognition is complicated by the fact that these tasks rely on different cognitive processes (Sommers *et al.*, 1997; Clopper *et al.*, 2006). For example, the familiarity or frequency of a word may affect its recognition in an open-set task but not a closed-set task (Pollack *et al.*, 1959), reflecting the fact that

word retrieval plays a larger role in open-set response contexts (but see also Segui *et al.*, 1982).

For adults, word recognition performance is poorer when assessed using an open-set task than a closed-set forced-choice task, even after accounting for changes in chance performance (Miller *et al.*, 1951). In a forced-choice task, performance depends on the number alternatives and their phonetic similarity to the target: performance tends to be better with a small number of response alternatives (Miller *et al.*, 1951) that are phonetically dissimilar (Pollack *et al.*, 1959; Clopper *et al.*, 2006). Better performance in a closed-set task than an open-set task can be attributed to the information provided by the response set. In an open-set task, the listener must hear the target word clearly enough to differentiate it from other words in his or her vocabulary, with the caveat that listeners use context and linguistic probabilities to narrow the set of likely responses (e.g., word frequency or semantic priming; Pisoni *et al.*, 1985). In contrast to the open-set task, in a forced-choice task the listener must hear just enough of the target word to differentiate it from among the other response alternatives. In the case of a picture-pointing task, capitalizing on the information provided by the response alternatives requires that the listener hold the auditory stimulus in memory while comparing it with the words represented by the pictures in the response set. While performance in a closed-set task tends to be better than that in the comparable open-set task for adults, the closed-set task could nonetheless be cognitively more demanding with respect to working memory.

It is unclear whether school-age children benefit to the same extent as adults from the provision of a closed set of

^{a)}Electronic mail: ebuss@med.unc.edu

response alternatives. Elliott *et al.* (1979) reported that word recognition in quiet was better when evaluated with a four-alternative forced choice (4AFC) than an open-set task for both children and adults; the average benefit associated with the 4AFC context rose from approximately 9 dB in 5-year-olds to approximately 12 dB in adults, although this trend as a function of listener age was not evaluated statistically. One possibility considered in the present study is that children benefit less than adults from a 4AFC context. An age effect, if observed, could be due to a range of factors, including: immature auditory memory or working memory (Hulme *et al.*, 1984; Keller and Cowan, 1994), slower word retrieval of pictorial response alternatives (Winters and Brzoska, 1975; Kail, 1991; Cook and Meyer, 2008), and limited cognitive resources to support the comparison between stimulus memory and response alternatives.

Another consideration in the present study is the effect of masker type on children's ability to benefit from a closed set of response alternatives. A masker may impair word recognition by interfering with peripheral encoding of the target speech or by interfering with the ability to segregate and attend specifically to that target. These two general categories of masking are described as energetic and informational, respectively (e.g., Brungart, 2001; Freyman *et al.*, 2004). Informational masking plays a particularly large role in performance when the target and masker are perceptually similar and the stimulus is unpredictable (Durlach *et al.*, 2003). In the context of speech perception, it is typically assumed that two-talker maskers exert primarily informational masking, and energetic masking plays a larger role in the presence of speech-shaped noise maskers (Brungart *et al.*, 2006).

While masked speech perception tends to be poorer in school-age children than adults for a wide range of stimuli, this child/adult difference tends to be larger when the target speech is masked by two-talker speech than by a speech-shaped noise masker (Hall *et al.*, 2002; Corbin *et al.*, 2016). Despite evidence of greater susceptibility to informational masking in school-age children than adults, both groups benefit from the introduction of segregation cues, including spatial separation of the target and masker (Litovsky, 2005; Yuen and Yuan, 2014), introduction of a sex mismatch between the target and masker (Wightman and Kistler, 2005; Misurelli and Litovsky, 2015), and mismatches in the target and masker language (Calandruccio *et al.*, 2016). Greater susceptibility to informational masking has been interpreted as indicating that adults are better than children at segregating the auditory streams associated with the target and masker talkers, and/or selectively attending to the target.

Limiting the response alternatives to a small number of words could be particularly helpful in a two-talker speech masker by reducing stimulus uncertainty. On the other hand, the cognitive resources required to accurately compare the response alternatives and the auditory memory of the stimulus may be higher in the two-talker speech than the speech-shaped noise masker. If listeners benefit from the closed response set by comparing those alternatives to all speech cues available in the stimulus, this comparison would include three streams of speech in the two-talker masker (one target plus two masker talkers), but only one in the speech-shaped

noise masker (one target talker). The increased number of speech streams in the two-talker masker condition could require more time and cognitive resources. Pupilometry data from adults corroborate the idea that speech-in-speech recognition is cognitively more demanding than speech in noise (Koelewijn *et al.*, 2012a; Zekveld *et al.*, 2014). If children have more limited cognitive resources or more rapid memory decay than adults, then their ability to benefit from provision of a closed response set could be more limited relative to adults in the two-talker masker. Based on these considerations, one prediction is that whereas adults may benefit from a closed set context in a two-talker masker, due to reduced stimulus uncertainty regarding the target, young children may benefit less or not at all, due to limited memory capacity or more rapid decay of auditory memory.

Two experiments were performed to evaluate children's ability to benefit from the information provided by a closed set of response alternatives in a two-talker speech masker or a speech-shaped noise masker. The first experiment compared word recognition in the two maskers for an open-set and a closed-set 4AFC paradigm. The second experiment compared performance in two 4AFC tasks, which differed with respect to the phonetic similarity between words in the four-alternative response set. This experiment also evaluated performance in an amplitude-modulated noise masker, to better understand the role of informational masking vs masker fluctuation in the effects of the two-talker speech masker. In both cases the goal was to better understand development of the ability to benefit from the information provided by a closed response set.

II. EXPERIMENT 1

A. Methods

1. Listeners

Listeners were 5- to 13-year olds ($n = 16$) and adults (18–33 years, $n = 10$). All listeners were native speakers of American English, were free from known language deficits, and had normal hearing bilaterally, defined as thresholds of 20 dB hearing level (HL) or less at octave frequencies between 250 and 8000 Hz (ANSI, 2010).

2. Stimuli

Target stimuli were 30 disyllabic words judged to be familiar to children as young as 5 yr old (Storkel and Hoover, 2010). Pictures illustrating each word were custom drawn with colored pencil. Listeners were familiarized with these illustrations and the words they represented prior to the experiment. Targets were produced by a female talker, with a mean F0 of 203 Hz. The two-talker masker was a 2.8-min sample of two women reading different passages from *Jack and the Beanstalk*. The masker talkers had mean F0s of 170 and 208 Hz. All three talkers, the target talker and the two masker talkers, were young adults who were native speakers of American English. The two masker recordings were equalized in level, and pauses longer than 300 ms were edited out, as described by Calandruccio *et al.* (2014). The speech-shaped noise masker was a 90-s sample of noise with

the same long-term power spectrum as the two-talker masker. These test materials were all used previously by Calandruccio *et al.* (2014) to evaluate word recognition in young school-age children.

3. Procedures

The experiment was controlled through a custom MATLAB script. Stimuli were loaded into a real-time processor (RZ6, TDT) and presented diotically over headphones (HD-25, Sennheiser). In the 4AFC conditions, pictures associated with the four response alternatives were viewed on a computer monitor, positioned in front of the listener. A target word was randomly selected to serve as the target on each trial, and three foils were randomly selected from the remaining 29 stimuli. Each trial began with the visual presentation of a matrix of four pictures displayed in black and white, with the target appearing in a randomly selected quadrant. After a delay of 1 s, the target word was presented. The pictures then transitioned from black and white to color, letting the listener know that it was time to indicate a response by pointing to the picture that most closely matched the target word. In the open-set response conditions, a randomly selected word was presented and the listener was asked to verbally repeat back that word. The monitor in front of the listener was turned off during open-set testing, and the listening interval was not indicated apart from presentation of the target word. The experimenter scored the response as correct or incorrect. Other aspects of the open-set task were the same as the 4AFC task. Listeners were encouraged to guess when unsure.

Masked speech reception thresholds (SRTs) were estimated by adaptively adjusting the signal level following a 2-down, 1-up stepping rule, which estimates the level associated with 71% correct (Levitt, 1971). The masker played continuously at 60 dB sound pressure level (SPL) throughout a threshold estimation track. Signal level adjustments were made in steps of 4 dB prior to the second track reversal, and in steps of 2 dB thereafter. Each track continued until eight reversals had been obtained, and the final threshold estimate was the mean signal level at the last six reversals. All listeners completed testing with both the two-talker speech and speech-shaped noise maskers, and in both 4AFC closed-set and open-set tasks, for a total of four conditions. At least two threshold estimates were obtained for each condition. A third threshold was obtained if the first two differed by 5 dB or more, and the mean of all thresholds obtained is reported below. In speech-shaped noise, a third estimate was required for one child (closed set); in the two-talker masker, a third estimate was required for three children (one open set, two closed set) and six adults (two open set, four closed set). Testing order was randomized across the four testing conditions.

Two approaches were used to analyze developmental trends. First, the association between thresholds and the natural log of child age in years was evaluated. The rationale for using a log transform of age is the observation that development tends to progress more rapidly in younger listeners, decelerating with increasing age. While adult thresholds were collected, they were not included in analyses evaluating the association between performance and listener age, on

the assumption that development is at asymptote within the adult age group. Instead, thresholds from adults were used to estimate mature performance and an associated 95% confidence interval. The second approach to evaluating development was to compare line fits to child data with the adult estimate of mature performance.

B. Results and discussion

The SRTs for individual child listeners are shown in Fig. 1 as a function of age. Mean adult thresholds are shown at the far right, with error bars representing the 95% confidence interval around the mean. The top panel shows results for the speech-shaped noise masker, and the bottom panel shows results for the two-talker speech masker. Symbol shading reflects the response type, either open set (unfilled circles) or 4AFC (filled circles). This figure illustrates a trend for better performance with increasing listener age for both maskers and both tasks. Linear regression was used to evaluate the association between thresholds and the natural log of child age in years. For the speech-shaped noise masker, the trend for better thresholds in older children approached but failed to reach significance for either the open-set ($\beta = -1.85$, $t_{14} = -2.09$, $p = 0.056$) or the 4AFC ($\beta = -1.71$, $t_{14} = -1.55$, $p = 0.144$) task. For the two-talker masker, however, improvements in performance with increasing child age were significant for both the open-set ($\beta = -4.52$, $t_{14} = -2.59$,

FIG. 1. The SRTs of individual child listeners are plotted as function of age. Adult means are shown at the right, with error bars indicating the 95% confidence interval. Symbol shading indicates the listener's task, which was either open set (open circles) or 4AFC (filled circles). Results for the speech-shaped noise masker are shown in the top panel, and those for the two-talker speech masker are shown in the bottom panel. Associations between SRT and child age are indicated with lines; solid lines indicate significant correlations, and dotted lines indicate non-significant trends.

$p = 0.021$) and 4AFC ($\beta = -9.17$, $t_{14} = -5.97$, $p < 0.001$) task. The association between SRT and child age is illustrated by the line fits shown in Fig. 1; solid lines indicate significant correlations, and dotted lines indicate non-significant trends. For reference, adult thresholds for these target words presented in quiet are approximately 20 dB SPL for an open-set task.¹

Across listeners, thresholds tended to be lower for the 4AFC than the open-set task. In adults this difference was significant for both speech-shaped noise (mean = 3.9 dB, $t_9 = 7.55$, $p < 0.001$) and two-talker speech (mean = 7.5 dB, $t_9 = 6.25$, $p < 0.001$); response type had a significantly larger effect on adult listeners' SRTs in the two-talker masker than the speech-shaped noise ($F_{1,9} = 7.90$, $p = 0.020$, $\eta_p^2 = 0.47$). In children, the benefit associated with the 4AFC response was relatively consistent in speech-shaped noise, but increased with age for the two-talker speech masker. A repeated-measures analysis of covariance (rmANCOVA) was performed to evaluate the significance of this observation, with the natural log of child age in years as a continuous variable. There were two levels of masker (speech-shaped noise, two-talker speech) and two levels of response (open set, 4AFC). This analysis resulted in a significant three-way interaction between masker, response, and child age ($F_{1,14} = 7.19$, $p = 0.018$, $\eta_p^2 = 0.34$). This result is consistent with an effect of age on the ability to benefit from the 4AFC context in the two-talker speech masker, but not the speech-shaped noise. To confirm that interpretation, an additional rmANCOVA was performed for each of the two maskers. For the speech-shaped noise masker there was a significant effect of response ($F_{1,14} = 4.88$, $p = 0.044$, $\eta_p^2 = 0.26$), a non-significant trend for an effect of age ($F_{1,14} = 3.76$, $p = 0.073$, $\eta_p^2 = 0.21$), and no response-by-age interaction ($F_{1,14} = 0.03$, $p = 0.860$, $\eta_p^2 < 0.01$). For the two-talker masker there was a non-significant trend for an effect of response ($F_{1,14} = 3.42$, $p = 0.086$, $\eta_p^2 = 0.20$), a significant effect of age ($F_{1,14} = 22.85$, $p < 0.001$, $\eta_p^2 = 0.62$), and a significant response-by-age interaction ($F_{1,14} = 8.28$, $p = 0.012$, $\eta_p^2 = 0.37$). The finding of a significant response-by-age interaction for the two-talker but not the speech-shaped noise masker confirms the observation that child age has little effect on the ability to benefit from the 4AFC context in speech-shaped noise, but that younger children benefit less than older children from a 4AFC context in the two-talker masker.

The second set of analyses compared SRTs of children and adults. For the speech-shaped noise masker, there is substantial overlap between SRTs of child and adult listeners. Four child thresholds fell within the 95% confidence interval of the adult data for both the open-set (9.7, 11.0, 11.2, and 13.4 years) and the closed-set (7.1, 10.3, 11.2, and 13.5 years) response conditions. The line fit to child data predicts mean performance within the 95% confidence interval of adult data by 13.4 years of age for the open set condition. For the 4AFC condition, the line fit indicates immature performance even in the oldest child listeners tested. For the two-talker speech masker there was very little overlap between SRTs of child and adult listeners. Only one child threshold fell within the 95% confidence interval of the adult mean; that was in the open-set condition (11.0 years). Line

fits to child data indicate that mean performance falls well outside the 95% confidence interval of adult data across the age range of children tested here.

One factor affecting SRTs for the open-set and 4AFC closed-set tasks is the fact that chance performance differs in the two conditions. In the absence of auditory information, performance converges on 25% correct in the 4AFC, but chance is nearly zero in an open-set task. The difference in chance performance across tasks is most evident at or near the lower asymptote of the psychometric function, with little or no effect at or near the upper asymptote. The magnitude of the effect at 71% correct depends on the slope of the psychometric function. To evaluate the benefit associated with higher chance performance in the 4AFC task, a logit function was fitted to the trial-by-trial data of adults for the open-set task in each of the two maskers. Given the limited (30-item) word list, it is possible that even in the absence of explicit alternatives listeners could treat the task as a 30-alternative forced choice rather than a truly open set, particularly with increasing stimulus familiarity. Because the adaptive track started above threshold and honed in on 71% correct, the lower asymptote could not be evaluated empirically. A lower asymptote of 0% was assumed here, as this represents the most conservative baseline for estimating the effect of introducing a 4AFC context.

Psychometric function fits to adults' open-set data accounted for 89% of the variance in the speech-shaped noise data, and 62% of the data in the two-talker speech noise data. The effect of better performance at chance in a 4AFC can be modeled by compressing the psychometric function based on open-set data by a factor of 0.75 and raising it by 0.25. Using this procedure, adults' thresholds in the open- and closed-set tasks are predicted to differ by 1.5 dB in the speech-shaped noise and by 4.3 dB in the two-talker speech. These values were compared to the mean effects observed in adult listeners with a pair of two-tailed t-tests. The benefit associated with the 4AFC context was greater than expected based on changes in chance performance for both the speech-shaped noise ($t_9 = 4.62$, $p = 0.001$) and the two-talker speech masker ($t_9 = 2.67$, $p = 0.025$). This analysis was not repeated on child data, as the individual differences associated with age pose a problem for fitting group psychometric functions. However, the finding of comparable benefits of context across listener age groups in the speech-shaped noise is consistent with the idea that listeners in all age groups used the response set of the 4AFC task to improve their performance beyond that expected based on increases in chance performance.

Whereas children and adults benefitted from the 4AFC context to a comparable degree in the speech-shaped noise masker, the results of experiment 1 are consistent with the idea that children are less adept than adults at using the information available in a forced-choice context when the masker is two-talker speech. One possible explanation is that failure to segregate the target from the two-talker masker increases the number of speech sounds to evaluate with respect to the response set, and the child/adult difference could be increased under these greater task demands. There are other factors that could have produced an age-by-response interaction,

however. First, the response interface of the 4AFC provided a temporal cue that was not present in the open-set task. We cannot rule out the possibility that this temporal cue is responsible for the results obtained, as opposed to provision of four response alternatives. Second, inherent fluctuation of the two-talker masker could have had a different effect on children and adults. Amplitude modulating a masker tends to improve masked SRTs for speech (Miller and Licklider, 1950), but this benefit is sometimes smaller in children than adults (Hall *et al.*, 2012). A reduced benefit of masker modulation in younger children could reflect differences in the ability to piece together sparse “glimpses” of speech, coincident with masker modulation minima (see also Elliott *et al.*, 1979). This raises the possibility children’s reduced ability to benefit from a closed response set in the two-talker masker could be related not to informational masking, but rather to greater cognitive demands associated with understanding speech associated with temporally sparse speech cues audible during modulation minima of the two-talker speech masker. This possibility receives some support from studies of hearing-impaired adults, which show that cognitive capacity and working memory is more predictive of speech reception in modulated noise than steady noise (reviewed by Ronnberg *et al.*, 2010).

1. Experiment 2

Experiment 1 showed that young children benefit as much as adults from the provision of a 4AFC response context when the masker is speech-shaped noise, but that this benefit increases with child age when the masker is two-talker speech. While this result is consistent with differential effects of cognitive development in conditions dominated by energetic and informational masking, respectively, we could not rule out a contribution of temporal cuing and/or envelope fluctuation in the two-talker masker. Experiment 2 was designed to clarify the role of these factors.

Experiment 2 compared performance for two 4AFC conditions, which differed with respect to the phonetic information provided by the response alternatives. In one condition the response alternatives were all phonetically similar, and in the other condition the response alternatives were phonetically dissimilar. When the response set is composed of dissimilar words, the correct response can be determined based on minimal phonetic information; in contrast, the similar response set requires that the listener identify one of a smaller set of unique phonemes. Based on the results of experiment 1, we expected a relatively constant effect of response condition across age in the speech-shaped noise, but an increase in the effect of response condition with age in the two-talker masker. Because these two conditions both used a 4AFC, different results could not be attributed to differences in chance performance or to differences in temporal cuing. In addition to the speech-shaped noise and two-talker speech maskers, experiment 2 also evaluated performance in an amplitude modulated noise masker. If the age effects observed for the two-talker speech masker in experiment 1 are due to immaturity in the ability to recognize speech based on temporally sparse glimpses, then an interaction

between listener age and response set should also be observed for the modulated masker.

C. Methods

1. Listeners

Listeners were 5- to 13-year olds ($n = 22$) and adults (19–29 years, $n = 9$). All were native speakers of American English, were free from known language deficits, and had normal hearing bilaterally, defined as thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz (ANSI, 2010). Only one of these listeners had previously participated in experiment 1; this listener was 7.5 years old, and participation in the two protocols was separated by 1.5 years.

2. Stimuli

Target words were 100 monosyllables from the WIPI test (Ross and Lerman, 1970), recorded by a female with a mean F0 of 209 Hz (Auditec, St. Louis). These recordings include the carrier phrase “show me” prior to the target word. The two-talker and speech-shaped noise maskers were the same as used in experiment 1. In the modulated-noise condition, the speech-shaped noise was amplitude-modulated via multiplication with a 10-Hz raised sinusoid. All three maskers were played continuously, without perceptible discontinuities at the beginning and end of the stimulus array.

In the original form of the WIPI test there are 25 cards each with six illustrations, four of which correspond to the target words in each of four lists. Two of the pictures on each card are foils that never correspond to an auditory stimulus, and those non-target pictures were omitted from the present experiment. The remaining four words illustrated on each card are phonetically similar (e.g., “arm,” “car,” “barn,” and “star”). In addition to the phonetically similar word sets adapted from the original form of the WIPI, the present study also evaluated performance for the same words organized into response sets that are phonetically dissimilar. This was achieved by randomly scrambling the order of words within each list, such that all 25 four-word sets had a unique vowel (e.g., “arm,” “meat,” “spring,” and “black”). Table I shows all 25 response sets for the phonetically similar and phonetically dissimilar conditions. Notice that all but one of the phonetically similar word sets share a vowel in common, the exception being the set comprising “ball,” “bowl,” “bell,” and “bow”. All the phonetically dissimilar word sets contained words with unique vowels, but there were instances of word sets in which one or more words shared a consonant in common (e.g., “pan,” “door,” “sun,” and “green”).

Phonetic similarity within the response set was manipulated in order to evaluate the extent to which children and adults use the phonetic information provided by the response set to optimize masked word recognition. In the present study, phonetically similar and dissimilar response sets were defined primarily with respect to vowels. Vowels tend to be correctly identified at a lower SNR than consonants in both noise and speech-based maskers (Varnet *et al.*, 2012). Since vowels differ for each word in the phonetically dissimilar

TABLE I. Four-alternative response sets, which are either phonetically similar (left column) or phonetically dissimilar (right column). Words in the phonetically similar response sets share a vowel in common in all but one case (marked with an asterisk). Words in the phonetically dissimilar response sets have unique vowels.

Phonetically similar	Phonetically dissimilar
school, broom, moon, spoon	school, fan, corn, gum
ball, bowl, bell, bow*	ball, ship, nest, bee
smoke, coat, coke, goat	smoke, barn, feet, ring
floor, door, corn, horn	floor, thumb, queen, bow
fox, socks, box, blocks	fox, church, thread, spoon
hat, flag, bag, black	hat, pie, jail, frog
pan, fan, can, man	pan, door, sun, green
bread, red, thread, bed	bread, broom, can, horn
neck, desk, nest, dress	neck, tea, snake, bug
stair, bear, chair, pear	stair, clown, coke, star
eye, pie, fly, tie	eye, nail, saw, blocks
knee, tea, key, bee	knee, cake, cup, fish
street, meat, feet, teeth	street, string, bag, tie
wing, string, spring, ring	wing, flag, fly, tail
mouse, clown, crown, mouth	mouse, coat, car, teeth
shirt, church, dirt, skirt	shirt, red, moon, man
gun, thumb, sun, gum	gun, seal, bell, lip
bus, rug, cup, bug	bus, stick, chair, mouth
train, cake, snake, plane	train, rug, dish, pear
arm, barn, car, star	arm, meat, spring, black
chick, stick, dish, fish	chick, bear, crown, goat
crib, ship, bib, lip	crib, desk, key, plane
wheel, seal, queen, green	wheel, bowl, bib, dress
straw, dog, saw, frog	straw, socks, dirt, bed
pail, nail, jail, tail	pail, dog, box, skirt

response sets, listeners can select the correct response based solely on an audible vowel, even if they are unable to identify any of the consonants. In contrast, phonetically similar words cannot be discriminated based solely on the vowel; correct word recognition of a target in the context of phonetically similar foils requires that one or more of the consonants is audible. Because consonants are typically recognized at a higher SNR than vowels, thresholds obtained with phonetically similar response sets were expected to be higher than those obtained with phonetically dissimilar response sets. We know that adults' forced-choice performance improves when the response alternatives are phonetically dissimilar (Pollack *et al.*, 1959), presumably because they can select from among the alternatives based on partial auditory information about the target word. It is unclear whether children are able to use this strategy, or whether they benefit to the same extent as adults.

3. Procedures

Procedures closely resembled those of experiment 1. Masked SRTs were measured using a 2-down 1-up tracking rule, which estimates the signal level associated with 71% correct. Masker level was 60 dB SPL, and signal level was adjusted in steps of 4 dB until the second track reversal, and in steps of 2 dB thereafter. A track continued until ten reversals were obtained, and the final threshold estimate was the mean signal level at the last eight reversals. Two thresholds

were obtained in each condition, with a third collected in cases where the first two estimates differed by 5 dB or more; the final threshold estimate was the mean of all estimates obtained. A third estimate was required in approximately 15% of cases, irrespective of listener age or condition. Thresholds in the six conditions (3 maskers $\times$ 2 response sets) were obtained in random interleaved order. Adults completed data collection in a single visit to the lab, whereas children required two visits.

D. Results and discussion

Figure 2 shows SRTs for individual child listeners as a function of age; adult means are shown at the far right, with error bars indicating the 95% confidence interval. Each panel shows results for a different masker, and symbol shading reflects the 4AFC response set, in which the target and foils were either phonetically similar (open circles) or phonetically dissimilar (filled circles). There was a trend for better

FIG. 2. The SRTs of individual child listeners are plotted as function of age. Adult means are shown at the right, with error bars indicating the 95% confidence interval. Symbol shading indicates the phonetic similarity between words in the 4AFC response set, which was either similar (open circles) or dissimilar (filled circles). Results for the three masker conditions are shown in separate panels; speech-shaped noise (top panel), two-talker speech (middle panel), and modulated noise (bottom panel). Associations between SRT and child age are indicated with lines; solid lines indicate significant correlations, and dotted lines indicate non-significant trends.

performance in older than younger children for all conditions, although this trend did not reach significance in all cases. When responses were obtained using phonetically similar response sets, the association between SRT and the log of child age failed to reach significance for the two-talker speech masker ($\beta = -3.13$, $t_{20} = -1.71$, $p = 0.104$), but it was significant for the speech-shaped noise masker ($\beta = -3.88$, $t_{20} = -2.38$, $p = 0.028$) and the modulated-noise masker ($\beta = -7.44$, $t_{20} = -3.48$, $p = 0.002$). When responses were obtained using phonetically dissimilar response sets, the correlation between SRT and the log of child age was significant for the speech-shaped noise masker ($\beta = -5.98$, $t_{20} = -4.05$, $p = 0.001$), two-talker speech masker ($\beta = -12.06$, $t_{20} = -6.30$, $p < 0.001$), and the modulated-noise masker ($\beta = -10.64$, $t_{20} = -4.29$, $p < 0.001$). For reference, adult thresholds obtained in quiet with the phonetically similar response set are approximately 28 dB SPL.²

The first analysis evaluated the effects of child age in the speech-shaped noise and the two-talker masker. As expected, thresholds tended to be lower for the phonetically dissimilar response set than the phonetically similar response set. For adult data, the effect of response set was significant for both the speech-shaped noise (mean = 5.6 dB, $t_8 = 6.06$, $p < 0.001$) and the two-talker speech masker (mean = 9.4 dB, $t_8 = 9.65$, $p < 0.001$); the effect of response set was significantly larger for the two-talker speech masker than the speech-shaped noise masker ($F_{2,16} = 7.76$, $p = 0.004$, $\eta_p^2 = 0.49$). For child data, the benefit associated with a dissimilar response set was relatively consistent for the speech-shaped noise and grew with age for the two-talker speech masker. A rmANCOVA with two levels of masker (speech-shaped noise, two-talker speech) and two levels of response set (similar, dissimilar) was performed to confirm a significant three-way interaction between age, response set, and masker ($F_{1,20} = 5.11$, $p = 0.035$, $\eta_p^2 = 0.20$). A follow-up rmANOVA was conducted for each masker alone to better understand this three-way interaction. For the speech-shaped noise, there was a main effect of age ($F_{1,20} = 18.26$, $p < 0.001$, $\eta_p^2 = 0.48$), but no main effect of response set ($F_{1,20} = 0.02$, $p = 0.889$, $\eta_p^2 < 0.01$), and no age-by-set interaction ($F_{1,20} = 1.02$, $p = 0.325$, $\eta_p^2 = 0.05$). For the two-talker masker, there was a significant effect of age ($F_{1,20} = 23.08$, $p < 0.001$, $\eta_p^2 = 0.54$), a significant main effect of response set ($F_{1,20} = 11.16$, $p = 0.003$, $\eta_p^2 = 0.36$), and a significant age-by-set interaction ($F_{1,20} = 19.60$, $p < 0.001$, $\eta_p^2 = 0.50$). This result confirms the observation that phonetic similarity of between the target and foils had a relatively consistent effect across age for the speech-shaped noise masker, but that the effect of phonetic similarity grew with listener age for the two-talker speech masker.

There was substantial overlap between child and adult SRTs in the speech-shaped noise. Seven child thresholds fell within the 95% confidence interval of adult data for both the similar response set and the dissimilar response set; in both cases all but one of these good-performing children were 10 years of age and older. Based on line fits to the child data, mean child performance fell within the 95% confidence interval of adult data by 12.2–12.8 years of age for both speech-shaped noise conditions. In contrast to the speech-shaped noise data, there was only modest overlap

between child and adult SRTs in the two-talker speech masker. Two child thresholds fell within the 95% confidence interval of adult data for both the similar response set (10.2 and 11.8 years) and the dissimilar response set (12.2 and 13.9 years). Based on line fits to child data, mean child performance remained above the 95% confidence interval around the adult mean over the age range tested here.

The next set of analyses evaluated performance as a function of child age for the modulated noise masker. As for the other masker conditions, there was a trend for thresholds in the modulated noise masker to be higher when the response set was phonetically similar to the target than when it was dissimilar. For adults, this difference was significant (mean = 8.8 dB, $t_8 = 10.04$, $p < 0.001$); this difference was significantly larger than the 3.9-dB effect observed for speech-shaped noise ($t_8 = 5.60$, $p = 0.001$). As evident in Fig. 2, there was a comparable effect of response set in child listeners, which did not appear to change with listener age. A rmANCOVA was performed to confirm this observation, with two levels of response set (similar, dissimilar) and child age (log of years) of as a continuous variable. There was a significant main effect of age ($F_{1,20} = 29.72$, $p < 0.001$, $\eta_p^2 = 0.60$), but no main effect of response set ($F_{1,20} = 0.12$, $p = 0.729$, $\eta_p^2 = 0.01$) and no age-by-set interaction ($F_{1,20} = 0.98$, $p = 0.334$, $\eta_p^2 = 0.05$). There was only modest overlap between child and adult SRTs. Two child thresholds fell within the 95% confidence interval for both the similar response set (11.8 and 13.4 years) and the dissimilar (11.3 and 11.7 years) response set conditions. Based on line fits, mean child performance fell within the 95% confidence interval of adult data by 13.0 years of age for the dissimilar response set; for the similar response set performance remained above the adult range throughout the range of ages tested here.

The inclusion of the modulated-noise masker in the present experiment was intended to evaluate the ability to benefit from information provided by a closed response set when the stimulus provides temporally sparse glimpses of speech. While masker modulation is often associated with improved speech perception in adults, the magnitude of this benefit differs widely across stimuli and conditions (e.g., Buss *et al.*, 2009). Some data indicate a child/adult difference in the ability to benefit from masker modulation (Hall *et al.*, 2012), although this is not always observed (Stuart *et al.*, 2006; Stuart, 2008). It is therefore of interest to evaluate the benefit associated with masker modulation across listener age. The effect of modulation was computed as the threshold in speech-shaped noise minus the threshold in modulated noise, a metric sometimes described as modulation masking release. Figure 3 shows the modulation masking release as a function of listener age. Symbol shape reflects type of response set, which contained words that were phonetically similar (open circles) or phonetically dissimilar (filled circles). Masking release tended to increase with increasing child age. The correlation between masking release and the log of child age was significant for the dissimilar response set ($\beta = 4.66$, $t_{20} = 2.56$, $p = 0.018$) but not for the similar response set ($\beta = 3.56$, $t_{20} = 1.48$, $p = 0.153$).

FIG. 3. Modulation masking release, computed as the difference between SRTs in the speech-shaped noise and modulated noise conditions, is plotted as function of child listener age. Adult means are shown at the right of the panel, with error bars indicating the 95% confidence interval. Symbol fill indicates the phonetic similarity between words in the 4AFC response set, which was either similar (open circles) or dissimilar (filled circles).

Overall masking release was larger for the phonetically dissimilar than the phonetically similar response sets. For adult data this difference was significant (mean = 3.2 dB, $t_8 = 2.67$, $p = 0.028$). This result was not predicted, but it could be related to the lower baseline (speech-shaped noise) thresholds for the dissimilar response condition (Bernstein and Grant, 2009; Smits and Festen, 2013). As illustrated in Fig. 3, masking release tended to increase with age, but the effect of response set was relatively consistent. The significance of this observation was evaluated with a rmANCOVA, with age in log of years as a continuous variable and two levels of response set (similar, dissimilar). This analysis resulted in a main effect of age ($F_{1,20} = 6.70$, $p = 0.018$, $\eta_p^2 = 0.25$), no main effect of response set ($F_{1,20} = 0.09$, $p = 0.769$, $\eta_p^2 < 0.01$) and no age-by-set interaction ($F_{1,20} = 0.15$, $p = 0.701$, $\eta_p^2 = 0.01$). The main effect of age corroborates the idea that children are less adept than adults at using the temporally sparse cues available in a modulated masker. The lack of an interaction between response set and age indicates that children's ability to benefit from a phonetically dissimilar response set is not significantly affected by masker modulation. It therefore is unlikely children's reduced benefit from the phonetically dissimilar response set in the two-talker masker is due to envelope modulations.

III. GENERAL DISCUSSION

Previous studies have shown that masked SRTs in adults are lower when the task is a closed-set forced choice than when it is an open-set free response, particularly if the closed response set contains a small number of phonetically dissimilar alternatives (Miller *et al.*, 1951; Pollack *et al.*, 1959; Clopper *et al.*, 2006). The present study was designed to evaluate whether school-age children benefit to the same extent as adults from the information provided by a closed response set when listening to masked words, and whether this ability depends on masker type. The first experiment demonstrated that, on average, word recognition in 5- to 13-year-olds benefitted from a 4AFC closed-set context for both maskers. Whereas children and adults benefitted from the 4AFC context to a similar degree for the speech-shaped

noise masker, the benefit conferred by the 4AFC increased with child age for a two-talker masker. This general finding was corroborated by the second experiment, which showed that the effect of phonetic similarity among words in the 4AFC response set impacted performance of 5- to 13-year-olds and adults similarly for the speech-shaped noise, but that response set had less impact on younger listeners' performance in the two-talker masker. The second experiment also evaluated performance in a modulated-noise masker. Whereas phonetic similarity had a larger effect on thresholds in the modulated-noise than the speech-shaped noise masker, this effect did not depend on listener age. The uniform effect of response set across age in the modulated-noise masker is inconsistent with the idea that envelope fluctuation of the two-talker speech masker is responsible for the age-dependent use of context in that masker.

The primary finding of the present study is that school-age children seem to make use of closed-set response context to a similar extent as adults when the masker is noise or modulated noise, but they use context less effectively than adults when the masker is two-talker speech. One possible explanation for this finding is that the informational masking associated with a two-talker masker taxes the cognitive resources of the child listeners. In adults, the cognitive resources necessary for good masked speech perception depend on the type of masker: listening in modulated-noise requires more cognitive effort than steady noise in some conditions, but the two-talker speech is even more cognitively challenging (Koelewijn *et al.*, 2012b; Zekveld *et al.*, 2014). The cognitive demands exerted by the masker could limit children's ability to capitalize on response context.

The effect of masker type on children's ability to benefit from a 4AFC, closed set contrasts somewhat with the results of Elliott *et al.* (1979). That study evaluated monosyllabic word recognition in 5- to 10-year-olds and adults. The task was either open-set word recognition in quiet or a 4AFC, picture-pointing task in one of three masker conditions: quiet, speech-shaped noise, and 12-talker babble. In quiet, performance improved with listener age, with a benefit from the 4AFC context observed across listener age. Masked performance in the 4AFC task also tended to improve with age. Comparing 5-year-olds to adults, the mean child/adult difference was 3.4 dB in the speech-shaped noise masker and 5.5 dB for the 12-talker babble. Based on line fits to data in experiments 1 and 2 of the present study, 4AFC thresholds for speech-shaped noise masker differed by 3.2–6.4 dB between 5-year-olds and adults. In contrast, the child/adult difference for a 4AFC in two-talker speech differed depending on the listening conditions, with values of 16.6 dB (experiment 1), 16.9 dB (experiment 2, phonetically dissimilar conditions), and 7.2 dB (experiment 2, phonetically similar response condition). This suggests that the modest (5.5-dB) child/adult difference observed by Elliott *et al.* (1979) for 4AFC thresholds in the 12-talker babble could be due either to the phonetic similarity of the response alternatives or to the degree to which the 12-talker babble produced informational masking. Given the description of stimulus development provided by Elliott *et al.* (1979), it seems likely that the target words were comparable to those tested in experiment 1

of the present study with respect to phonetic similarity. However, it has been argued that informational masking is highly dependent on the number of talkers comprising the masker. For example, Freyman *et al.* (2004) showed that informational masking for sentence recognition peaked at two talkers and dropped off sharply as the number of talkers increases to ten. Based on this observation, the Elliott *et al.* (1979) finding of similar age effects for a 4AFC word recognition in noise and speech maskers may have been due to relatively modest informational masking associated with the 12-talker babble.

One question of interest is the extent to which the present results generalize to other speech tasks, such as sentence recognition. Under normal listening conditions, speech perception often occurs in a rich linguistic context, including pragmatic, semantic and syntactic information. In that light, open-set word recognition is not typical of everyday communication. However, the context provided by a 4AFC is fundamentally different from the linguistic context of a sentence. For example, word retrieval based on illustrations does not rely on a listener's ability to predict syntactic or semantic structures in an ongoing speech stream. For sentence materials, school-age children benefit from predictable sentence context in multiple situations, including noise-masked sentence recognition (Nittrouer and Boothroyd, 1990; Fallon *et al.*, 2002), sentence recognition near threshold in quiet (Stelmachowicz *et al.*, 2000), and detection of mispronounced words in running speech (Cole and Perfetti, 1980). Whereas adults and children appear to benefit to comparable degrees from lexical and syntactic context, children tend to benefit less than adults from semantic context (Elliott, 1979; Nittrouer and Boothroyd, 1990; Eisenberg *et al.*, 2000).

Results of the present study indicate that the benefit conferred by a 4AFC context increases between 5 and 13 years of age when the masker is composed of a small number of talkers. One question for future research is whether masker type affects the use of sentence-level context differently for children and adults. While inconclusive, the results of Elliott (1979) are consistent with such an effect. That study tested 9- to 17-year-olds using the SPIN test (Kalikow *et al.*, 1977), which comprises sentences with either low-predictability or high-predictability, presented in a babble masker. Performance tended to improve with age in all conditions, with the 17-year-olds performing like adults. Of particular interest here, the 11- to 13-year-olds performed more poorly than older children only in the high-predictability condition. Elliott (1979) speculated that the babble masker may have interfered with the 11- to 13-year-olds' ability to make use of sentence context. In conjunction with the present results, we would predict a smaller child/adult difference in the use of sentence context for a speech-shaped noise than a two-talker speech masker.

The present results have practical implications for the clinical assessment of masked perception in children. Whereas masked speech perception has traditionally been evaluated in noise, there is a growing consensus that it is also important to characterize speech perception under more complex listening conditions, such as in the presence of a speech masker. For children with hearing loss, speech recognition evaluated in a two-talker masker is a better predictor

of real-world listening difficulties, as assessed by parental report, than performance in noise (Hillock-Dunn *et al.*, 2015). A number of speech perception tests appropriate for use in children include speech maskers, some evaluating masked word recognition with a forced-choice response (Litovsky, 2005; Calandruccio *et al.*, 2014), and others evaluating masked word or sentence recognition with an open-set response (Cameron and Dillon, 2007; Wilson *et al.*, 2010; Holder *et al.*, 2015). Based on the results of the present study, it is possible that interpretation of these test results may be affected differently by the interaction between age, response context, and informational masking. This complex interplay could be particularly important to consider when response set is manipulated to adjust task difficulty (Eisenberg *et al.*, 2006; Wang *et al.*, 2008).

ACKNOWLEDGMENTS

This work was supported by a grant from NIH NIDCD R01-DC014460 (E.B.). Mary Flaherty provided helpful comments on this work.

¹Thresholds were measured in quiet with five adults using an open-set response. Other than the absence of a masker, the methods were the same as those described for experiment 1.

²Thresholds were measured in quiet with five adults using a 4AFC with similar response alternatives. Other than the absence of a masker the methods were the same as those described for experiment 2.

- ANSI (2010). ANSI S3.6-2010, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Bernstein, J. G., and Grant, K. W. (2009). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **125**, 3358–3372.
- Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., and Tooher, R. (2001). "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," *J. Speech Lang. Hear. Res.* **44**, 264–285.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.* **120**, 4007–4018.
- Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W. (2009). "Masking release for words in amplitude modulated noise as a function of modulation rate and task," *J. Acoust. Soc. Am.* **126**, 269–280.
- Calandruccio, L., Gomez, B., Buss, E., and Leibold, L. J. (2014). "Development and preliminary evaluation of a pediatric Spanish-English speech perception task," *Am. J. Audiol.* **23**, 158–172.
- Calandruccio, L., Leibold, L. J., and Buss, E. (2016). "Linguistic masking release in school-age children and adults," *Am. J. Audiol.* **25**, 34–40.
- Cameron, S., and Dillon, H. (2007). "Development of the listening in spatialized noise-sentences test (LISN-S)," *Ear Hear.* **28**, 196–211.
- Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (2006). "Effects of open-set and closed-set task demands on spoken word recognition," *J. Am. Acad. Audiol.* **17**, 331–349.
- Cole, R. A., and Perfetti, C. A. (1980). "Listening for mispronunciations in a children's story—The use of context by children and adults," *J. Verb. Learn. Verb. Behav.* **19**, 297–315.
- Cook, A. E., and Meyer, A. S. (2008). "Capacity demands of phoneme selection in word production: New evidence from dual-task experiments," *J. Exp. Psychol. Learn. Mem. Cog.* **34**, 886–899.
- Corbin, N., Bonino, A. Y., Buss, E., and Leibold, L. J. (2016). "Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers," *Ear Hear.* **37**, 55–63.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (2003). "Informational masking:

- Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," *J. Acoust. Soc. Am.* **114**, 368–379.
- Eisenberg, L. S., Johnson, K. C., Martinez, A. S., Cokely, C. G., Tobey, E. A., Quittner, A. L., Fink, N. E., Wang, N. Y., and Niparko, J. K. (2006). "Speech recognition at 1-year follow-up in the childhood development after cochlear implantation study: Methods and preliminary findings," *Audiol. Neurotol.* **11**, 259–268.
- Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (2000). "Speech recognition with reduced spectral cues as a function of age," *J. Acoust. Soc. Am.* **107**, 2704–2710.
- Elliott, L. L. (1979). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," *J. Acoust. Soc. Am.* **66**, 651–653.
- Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (1979). "Children's understanding of monosyllabic nouns in quiet and in noise," *J. Acoust. Soc. Am.* **66**, 12–21.
- Fallon, M., Trehub, S. E., and Schneider, B. A. (2002). "Children's use of semantic cues in degraded listening environments," *J. Acoust. Soc. Am.* **111**, 2242–2249.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," *J. Acoust. Soc. Am.* **115**, 2246–2256.
- Hall, J. W., Buss, E., Grose, J. H., and Roush, P. A. (2012). "Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation," *Ear Hear.* **33**, 340–348.
- Hall, J. W., Grose, J. H., Buss, E., and Dev, M. B. (2002). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," *Ear Hear.* **23**, 159–165.
- Hillock-Dunn, A., Taylor, C., Buss, E., and Leibold, L. J. (2015). "Assessing speech perception in children with hearing loss: What conventional clinical tools may miss," *Ear Hear.* **36**, e57–e60.
- Holder, J. T., Sheffield, S. W., and Gifford, R. H. (2015). "Speech understanding in children with normal hearing: Sound field normative data for BabyBio, BKB-SIN, and QuickSIN," *Otol. Neurotol.* **37**, e50–e55.
- Hulme, C., Thomson, N., Muir, C., and Lawrence, A. (1984). "Speech rate and the development of short-term-memory span," *J. Exp. Child Psychol.* **38**, 241–253.
- Kail, R. (1991). "Processing time declines exponentially during childhood and adolescence," *Dev. Psychol.* **27**, 259–266.
- Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351.
- Keller, T. A., and Cowan, N. (1994). "Developmental increase in the duration of memory for tone pitch," *Dev. Psychol.* **30**, 855–863.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012a). "Pupil dilation uncovers extra listening effort in the presence of a single-talker masker," *Ear Hear.* **33**, 291–300.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., Ronnberg, J., and Kramer, S. E. (2012b). "Processing load induced by informational masking is related to linguistic abilities," *Int. J. Otolaryngol.* **2012**, 865731.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Litovsky, R. Y. (2005). "Speech intelligibility and spatial release from masking in young children," *J. Acoust. Soc. Am.* **117**, 3091–3099.
- Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials," *J. Exp. Psychol.* **41**, 329–335.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Misurelli, S. M., and Litovsky, R. Y. (2015). "Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity," *J. Acoust. Soc. Am.* **138**, 319–331.
- Nittrouer, S., and Boothroyd, A. (1990). "Context effects in phoneme and word recognition by young children and older adults," *J. Acoust. Soc. Am.* **87**, 2705–2715.
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. (1985). "Speech perception, word recognition and the structure of the lexicon," *Speech Commun.* **4**, 75–95.
- Pollack, I., Rubenstein, H., and Decker, L. (1959). "Intelligibility of known and unknown message sets," *J. Acoust. Soc. Am.* **31**, 273–279.
- Ronnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (2010). "When cognition kicks in: Working memory and speech understanding in noise," *Noise Health* **12**, 263–269.
- Ross, M., and Lerman, J. (1970). "A picture identification test for hearing-impaired children," *J. Speech Hear. Res.* **13**, 44–53.
- Segui, J., Mehler, J., Frauenfelder, U., and Morton, J. (1982). "The word frequency effect and lexical access," *Neuropsychol.* **20**, 615–627.
- Smits, C., and Festen, J. M. (2013). "The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise," *J. Acoust. Soc. Am.* **133**, 3004–3015.
- Sommers, M. S., Kirk, K. I., and Pisoni, D. B. (1997). "Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. 1. The effects of response format," *Ear Hear.* **18**, 89–99.
- Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W. L., and Pittman, A. L. (2000). "The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children," *J. Speech Lang. Hear. Res.* **43**, 902–914.
- Storkel, H. L., and Hoover, J. R. (2010). "An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English," *Behav. Res. Methods* **42**, 497–506.
- Stuart, A. (2008). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children," *J. Am. Acad. Audiol.* **19**, 135–146.
- Stuart, A., Givens, G. D., Walker, L. J., and Elangovan, S. (2006). "Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise," *J. Acoust. Soc. Am.* **119**, 1946–1949.
- Varnet, L., Meyer, J., Hoen, M., and Meurier, F. (2012). "Phoneme resistance during speech-in-speech comprehension," in *13th Annual Conference of the International Speech Communication Association*, Portland, Oregon.
- Wang, N. Y., Eisenberg, L. S., Johnson, K. C., Fink, N. E., Tobey, E. A., Quittner, A. L., Niparko, J. K., and Team, C. D. I. (2008). "Tracking development of speech recognition: Longitudinal data from hierarchical assessments in the Childhood Development after Cochlear Implantation Study," *Otol. Neurotol.* **29**, 240–245.
- Wightman, F. L., and Kistler, D. J. (2005). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," *J. Acoust. Soc. Am.* **118**, 3164–3176.
- Wilson, R. H., Farmer, N. M., Gandhi, A., Shelburne, E., and Weaver, J. (2010). "Normative data for the Words-in-Noise Test for 6- to 12-year-old children," *J. Speech Lang. Hear. Res.* **53**, 1111–1121.
- Winters, J. J., Jr., and Brzoska, M. A. (1975). "Development of lexicon in normal and retarded persons," *Psychol. Rep.* **37**, 391–402.
- Yuen, K. C. P., and Yuan, M. (2014). "Development of spatial release from masking in mandarin-speaking children with normal hearing," *J. Speech Lang. Hear. Res.* **57**, 2005–2023.
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., and Kramer, S. E. (2014). "The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load," *Neuroimage* **101**, 76–86.