



HAL
open science

Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers

Emily Buss, Lori Leibold, Christian Lorenzi

► **To cite this version:**

Emily Buss, Lori Leibold, Christian Lorenzi. Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers. *Journal of the Acoustical Society of America*, 2018, 143 (3), pp.1458-1466. 10.1121/1.5026795 . hal-03898585

HAL Id: hal-03898585

<https://hal.science/hal-03898585>

Submitted on 14 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers

Emily Buss, Lori J. Leibold, and Christian Lorenzi

Citation: [The Journal of the Acoustical Society of America](#) **143**, 1458 (2018); doi: 10.1121/1.5026795

View online: <https://doi.org/10.1121/1.5026795>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/3>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Relationship between speech-evoked neural responses and perception of speech in noise in older adults](#)

[The Journal of the Acoustical Society of America](#) **143**, 1333 (2018); 10.1121/1.5024340

[Effects of age on sensitivity to interaural time differences in envelope and fine structure, individually and in combination](#)

[The Journal of the Acoustical Society of America](#) **143**, 1287 (2018); 10.1121/1.5025845

[Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners](#)

[The Journal of the Acoustical Society of America](#) **143**, 1523 (2018); 10.1121/1.5026788

[Cognitive disruption by noise-vocoded speech stimuli: Effects of spectral variation](#)

[The Journal of the Acoustical Society of America](#) **143**, 1407 (2018); 10.1121/1.5026619

[Speech-material and talker effects in speech band importance](#)

[The Journal of the Acoustical Society of America](#) **143**, 1417 (2018); 10.1121/1.5026787

[Auditory distraction by speech: Sound masking with speech-shaped stationary noise outperforms -5 dB per octave shaped noise](#)

[The Journal of the Acoustical Society of America](#) **143**, EL212 (2018); 10.1121/1.5027765

Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers

Emily Buss,^{1,a)} Lori J. Leibold,² and Christian Lorenzi³

¹*Department of Otolaryngology/Head and Neck Surgery, 170 Manning Drive, University of North Carolina, Chapel Hill, North Carolina 27599, USA*

²*Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131, USA*

³*Département d'études cognitives, Ecole normale supérieure, Paris Sciences et Lettres Research University, Centre National de la Recherche Scientifique, 29 rue d'Ulm, Paris, 75005, France*

(Received 5 October 2017; revised 21 February 2018; accepted 23 February 2018; published online 15 March 2018)

The present study set out to test whether greater susceptibility to modulation masking could be responsible for immature recognition of speech in noise for school-age children. Listeners were normal-hearing four- to ten-year-olds and adults. Target sentences were filtered into 28 adjacent narrow bands (100–7800 Hz), and the masker was either spectrally matched noise bands or tones centered on each of the speech bands. In experiment 1, odd- and even-numbered bands of target-plus-masker were presented to opposite ears. Performance improved with child age in all conditions, but this improvement was larger for the multi-tone than the multi-noise-band masker. This outcome is contrary to the expectation that children are more susceptible than adults to masking produced by inherent modulation of the noise masker. In experiment 2, odd-numbered bands were presented to both ears, with the masker diotic and the target either diotic or binaurally out of phase. The binaural difference cue was particularly beneficial for young children tested in the multi-tone masker, suggesting that development of auditory stream segregation may play a role in the child-adult difference for this condition. Overall, results provide no evidence of greater susceptibility to modulation masking in children than adults. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5026795>

[VB]

Pages: 1458–1466

I. INTRODUCTION

Speech perception is more adversely affected by the presence of masking noise in young children than adults. This effect has been demonstrated for sentence recognition (Elliott, 1979; Stuart, 2008), open-set word recognition (Buss *et al.*, 2016; Corbin *et al.*, 2016), closed-set word recognition (Elliott *et al.*, 1979; Hall *et al.*, 2002; Buss *et al.*, 2016), and phoneme discrimination (Leibold and Buss, 2013). Considering data across studies, speech in noise appears to be adult-like by approximately 10–12 years of age (Stuart, 2008; Corbin *et al.*, 2016). Poorer speech-in-noise recognition for young school-age children compared to adults is often discussed in the context of selective attention, cognition, and linguistic abilities (e.g., McCreery *et al.*, 2017), but it is not clear to what extent these factors explain the observed age effects.

The reduced ability of children to recognize speech in noise may result from a reduced ability to use impoverished speech cues. Compared to adults, children require greater audibility (Scollie, 2008; McCreery and Stelmachowicz, 2011) and greater spectral resolution (Eisenberg *et al.*, 2000) to recognize speech. Age effects in the quantity or quality of cues required to recognize speech may also be responsible for the observation that children tend to benefit less than adults from amplitude modulating a noise masker

(Hall *et al.*, 2012; Buss *et al.*, 2016). Amplitude modulating a noise masker provides the listener with brief glimpses of the target speech at an improved target-to-masker ratio (TMR), resulting in improved intelligibility compared to an unmodulated noise masker. Children's limited ability to benefit from masker modulation could reflect a reduced ability to recognize speech based on brief, temporally sparse cues. This interpretation is corroborated by the finding that speech perception remains poorer in children than adults when the target speech is digitally segregated from the masker, leaving just the epochs of minimally masked speech (Buss *et al.*, 2017). A reduced ability to recognize speech based on sparse cues in children could be related to development of working memory and language ability (McCreery *et al.*, 2017). The lexical restructuring hypothesis proposes that children represent the words they know with increasing detail as their lexicon grows (reviewed by Mainela-Arnold *et al.*, 2008). This shift from word-level to phoneme-level representation of words in memory could confer greater ability to recognize words based on degraded cues.

Another factor that could limit children's ability to recognize speech in noise is immature auditory stream segregation and/or selective attention, resulting in greater informational masking. The child-adult difference in susceptibility to masking is larger for speech-in-speech recognition than for speech-in-noise recognition (Wightman and Kistler, 2005; Buss *et al.*, 2016; Corbin *et al.*, 2016). The marked age effect for speech-in-speech tasks has been attributed to

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

development in the ability to segregate target speech from a complex background (Wightman and Kistler, 2005). Informational masking can be reduced with the introduction of a binaural difference cue, either by spatially separating sound sources (Freyman *et al.*, 2001) or playing dichotic stimuli over headphones (Kidd *et al.*, 1994; Edmonds and Culling, 2005; Gallun *et al.*, 2005). Adults and children obtain a modest benefit from spatially separating target speech from a noise masker (Murphy *et al.*, 2011; Corbin *et al.*, 2017), suggesting that auditory stream segregation plays at most a minor role in performance for either age group. However, this conclusion is undermined somewhat by data indicating that the ability to fully benefit from the introduction of binaural difference cues develops over childhood (Hall *et al.*, 2004; Cameron and Dillon, 2007; Hall *et al.*, 2007; Yuen and Yuan, 2014), and even contralateral noise can elevate detection thresholds by more than 10 dB in some young children (Wightman *et al.*, 2003).

The present study set out to evaluate another possible factor in the development of speech-in-noise recognition—susceptibility to modulation masking, defined as the detrimental effect of masker amplitude modulation on the ability to detect and discriminate target amplitude modulation. Modulation masking is typically assessed using non-linguistic stimuli, such as amplitude-modulated tones or noise (Bacon and Grantham, 1989; Houtgast, 1989; Strickland and Viemeister, 1996; Sek *et al.*, 2015). Stone and his colleagues have argued that modulation masking limits speech recognition in a noise masker in that random inherent envelope fluctuation of nominally steady noise interferes with the perceptual processing of speech envelope cues (Stone *et al.*, 2011; Stone *et al.*, 2012; Stone and Moore, 2014). In one demonstration of this effect, Stone *et al.* (2012) measured recognition of speech filtered into narrow bands and masked by either spectrally matched narrow bands of noise (the “multi-noise-band” condition) or pure tones (the “multi-tone” condition). Adults experienced more masking for the multi-noise-band masker than the multi-tone masker. The authors argued that performance was limited primarily by energetic masking in the multi-tone condition, and by a combination of energetic and modulation masking in the multi-noise-band condition. By this interpretation, greater susceptibility to masking in the multi-noise-band condition than the multi-tone condition reflects the contributions of modulation masking.

The experimental paradigm of Stone *et al.* (2012) was used in the present study, with the goal of evaluating the role of modulation masking in the child-adult difference observed for speech-in-noise recognition. The first experiment compared sentence recognition for narrow bands of speech masked by either narrow bands of noise or pure tones. If young children are more susceptible to modulation masking than adults, then the effect of child age and the child-adult difference should be larger when performance is evaluated in the multi-noise-band masker than the multi-tone masker. This expectation was not supported: in fact, the developmental effect was *smaller* for the multi-noise-band than the multi-tone masker. The second experiment therefore assessed the possibility of age effects related to selective

auditory attention and auditory stream segregation by evaluating performance with and without a binaural difference cue, which was intended to facilitate segregation of the target and masker.

II. EXPERIMENT 1

The first experiment measured masked speech reception thresholds (SRTs) for sentences in school-age children and adults using stimuli modeled after those of Stone *et al.* (2012). In all conditions, there were 28 maskers consisting of either tones or narrow bands of noise. Odd-numbered maskers were presented to one ear and even-numbered maskers to the other. The rationale for separating neighboring bands across ears was to reduce beating between bands, while retaining the full range of speech information. For children and one group of adults, there were 28 corresponding bands of target speech. A second set of adults was tested using a subset of ten target speech bands distributed across ears. The rationale for testing a subset of adults with a more spectrally sparse target was to approximately match SRTs with those of young children, in the event masker effects depend on the TMR.

A. Methods

Listeners were normal-hearing four- to ten-year-olds ($n=26$) and two groups of normal-hearing adults (18–42 years old), one tested in the same conditions as children ($n=9$, mean of 24 yr) and a second group tested with a spectrally sparser target ($n=11$, mean of 22 yr). All had thresholds of 20 dB hearing level (HL) or lower bilaterally at octave frequencies 250–8000 Hz (ANSI, 2010). All listeners were native speakers of American English, and none reported a history of hearing, speech, or language problems.

The target speech was Bamford-Kowal-Bench sentences (BKB; Bench *et al.*, 1979) produced by a male talker. Individual sentences were spliced from a commercially available recording (Auditec, St. Louis, MO) and saved as wav files, 1.18–2.55 s in duration (mean = 1.81 s). These sentences were filtered into 28 bands between 100 and 7800 Hz, with each band spanning approximately one equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990). This was achieved by constructing a bank of variable length finite impulse response (FIR) filters, with the number of taps selected so that the frequency resolution was three times the passband for each filter. The resulting functions were symmetrically padded with zeros to equalize the number of taps across bands, eliminating temporal asynchronies across frequency. Neighboring filters crossed at -6 dB such that the summed output of this filter bank was perceptually indistinguishable from a stimulus that was bandpass filtered between 100 and 7800 Hz.

The masker was composed of either narrow bands of noise or pure tones. In the case of the multi-noise-band masker, a 5-s sample of speech-shaped noise matching the long-term power spectrum of the BKB sentences was generated. A randomly selected segment of this noise file was selected and filtered prior to each listening interval using the same 28-band filter bank used to filter the target speech. The multi-tone masker consisted of 28 pure tones with frequencies

TABLE I. Center frequencies of bandpass filters in Hz, calculated in ERB units. The top row corresponds to bands presented to the left ear, and the bottom row corresponds to bands presented to the right ear. Stars indicate the bands associated with the target for the second group of adults.

Left	119	209	321	462	639	861	1140	1491	1932	2487	3183	4057	5156	6537
	*			*			*			*			*	
Right	161	261	387	545	743	993	1306	1699	2194	2815	3595	4575	5807	7355
		*			*			*			*			*

corresponding to the center of each noise band, computed in ERB units. Each tone had a new random starting phase on each trial, and its level was scaled to match the root-mean-square (RMS) level of the corresponding noise band. Maskers were 2 s longer than the associated target, and the target was temporally centered in the masker. The masker was ramped on and off with 50-ms raised-cosine ramps.

Odd-numbered target and masker bands were presented to the left ear and even-numbered bands to the right. In the primary conditions, all 28 bands were presented to the listener for both the target and masker. In the second group of adults, all masker bands were presented, but only ten target bands were presented (five to each ear); the center frequencies of those target bands are indicated with stars in Table I.

Stimuli were generated in MATLAB (Natick, MA) played out of a real-time processor (RP2, TDT, Alachua, FL), routed to a headphone buffer (HB7, TDT), and presented over headphones (Sennheiser, HD 265, Old Lyme, CT). The stimulus was 70 dB sound pressure level (SPL) overall, and the TMR was adjusted by manipulating both the target and masker level; since most thresholds were <0 dB TMR, this is approximately equivalent to fixing the masker level at 70 dB SPL and adjusting the target level. Performance was evaluated using two interleaved one-down, one-up adaptive tracks. These tracks differed with respect to the rule defining a correct response: one track required the listener to report one or more words correctly, while the other required the listener to report all or all but one word correctly. The TMR was adjusted in steps of 2 dB, and each track contained 30 trials. Scoring for each keyword on each trial was saved to disk, along with the TMR. Each listener completed 2 such blocks in each condition for 120 trials per condition. Two-parameter logit functions were fitted to the proportion of keywords correct, weighted by the number of observations at each level, for each listener in each condition, and the SRT was defined as the 50% correct point based on these fits. One potential advantage of this method over a single adaptive track is that both threshold and slope can be estimated. However, preliminary analysis of slope estimates in the current dataset revealed no systematic effects of condition or age, so estimates of slope are not considered below.

Data were collected in random interleaved order, and listeners did not hear any sentence more than once. Most listeners completed testing in a single 1-h visit to the laboratory. Effects of age were evaluated with respect to the natural log of child age in years, based on the expectation of decelerating effects of development with increasing age (e.g., Mayer and Dobson, 1982).¹ Data were analyzed using linear regression and mixed models (Pinheiro *et al.*, 2016; R Core Team, 2016) with subject as a random factor and a significance criterion of $p < 0.05$. For analysis of child data,

age was centered on $\log(7)$ because the mean age of child listeners was approximately 7 years of age. Parameters reported include the slope (β), standard error (SE) around the slope estimate, and Pearson correlation (r).

B. Results

Results are shown in Fig. 1. The SRTs for child listeners are plotted on the left as a function of age, and mean SRTs for adults are shown at the right of the panel. Boxplots reflect the distribution of adult data in each condition. Results for adults tested with all 28 target bands are plotted separately from those tested with a subset of 10 target bands.

There was an association between the log of child age and SRT for both the multi-noise-band masker ($r = -0.61$, $\beta = -4.19$, $p < 0.001$) and the multi-tone masker ($r = -0.63$, $\beta = -6.50$, $p = 0.001$). While age accounted for ~40% of the variance across estimates in each condition, there were reliable individual differences that appeared to be unrelated to listener age. There was a significant correlation between SRTs in the multi-noise-band and multi-tone maskers after controlling for age ($r = 0.85$, $p < 0.001$). The effect of child age appears to differ for the two maskers; SRTs improve more rapidly with increasing age for the multi-tone than the multi-noise-band masker. The significance of this observation was confirmed with a linear mixed model; masker and the log of listener age were fixed factors, and subject was a random factor. There was a significant effect of masker

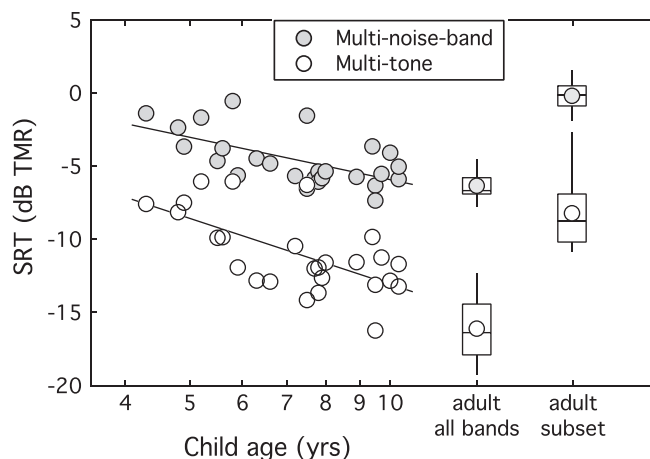


FIG. 1. SRTs for the masker composed of narrow bands of noise (filled circles) or tones (open circles). Results for individual child listeners are plotted as a function of age, with solid black lines indicating linear fits to data as a function of the log of child age. Group mean SRTs are shown for adults. Results are plotted separately for the two groups of adults: those tested in the primary conditions with all 28 bands of target speech, and those tested in more difficult conditions with a subset of 10 target speech bands. Boxes span the 25th–75th percentiles, horizontal lines indicate the median, and vertical lines span the 10th–90th percentiles.

($\beta = -6.32$, $SE = 0.23$, $p < 0.001$) and age ($\beta = -4.19$, $SE = 1.39$, $p = 0.006$), and a significant interaction between age and masker ($\beta = -2.32$, $SE = 0.90$, $p = 0.017$). The interaction reflects the larger developmental effect for the multi-tone masker than the multi-noise-band masker. This developmental effect is also evident when comparing data from children and adults tested with all 28 bands. Based on line fits to the child data, thresholds fall within the 95% confidence interval around the adult mean by 10.0 yr for the multi-noise-band masker, and by 13.2 yr for the multi-tone masker.

One question of interest is whether the relatively modest effect of masker type observed for young children is due to immature listening abilities, or whether it is a consequence of the fact that young children perform more poorly than older children and adults. Listening at a higher TMR might tend to reduce the importance of masker type due to the decreasing perceptual salience of the masker. Reducing the number of target speech bands from 28 to 10 increased SRTs for adult listeners by 6.1 dB in the multi-noise-band masker and by 7.9 dB in the multi-tone masker. Based on line fits to child data, SRTs for adults in these conditions were higher than those of even the youngest children for the multi-noise-band masker and comparable to those of five-year-olds for the multi-tone masker. With a subset of speech bands, adults' performance was 8.0 dB better for the multi-tone than the multi-noise-band masker. This effect is smaller than the 9.8-dB masker effect seen in the adult data for the full complement of 28 speech bands ($t_{10} = -3.12$, $p = 0.011$), but it is larger than the 5.5-dB effect associated with five-year-olds ($t_{10} = 4.50$, $p = 0.001$). In other words, the poorer overall performance of young children can account for about 1.8 dB of the child-adult difference in the masker effect for the 28-band data, but it leaves 2.5 dB unaccounted for. This observation supports the conclusion that the more pronounced age effects in the multi-tone condition than the multi-noise-band condition reflects a true developmental effect.

C. Discussion

The goal of this experiment was to evaluate the role of modulation masking in the child-adult difference observed for speech-in-noise recognition. For adults, recognition of speech filtered into ERB-wide bands is poorer when the masker is composed of spectrally matched bands of noise compared to pure tones at the center frequencies of each band. Stone *et al.* (2012) explained this masker effect in terms of modulation masking; the inherent amplitude modulation of the multi-noise-band masker was argued to interfere with envelope processing of temporal envelope cues associated with the target speech. If children are more susceptible to modulation masking than adults, then the child-adult difference should be larger for the multi-noise-band masker than the multi-tone masker. This expectation was not borne out in the data, however. Age effects were *smaller* for the multi-noise-band masker than the multi-tone masker. The results of experiment 1 therefore fail to provide evidence

that modulation masking contributes to the child-adult difference for speech-in-noise perception.

One question of interest is how the selection of stimuli in the present experiment bears on the effects of age and masker type that were observed. The BKB corpus was developed for use with young hearing-impaired children (Bench and Bamford, 1979; Bench *et al.*, 1979). Clinically, this corpus is recommended for use in children as young as 5–6 years of age (e.g., BKB-SIN, Etymotic Research Inc., Elk Grove Village, IL), but three children tested in experiment 1 were four-year-olds. While hearing-impaired children often have smaller vocabularies than normal-hearing children (Moeller *et al.*, 2007; Halliday *et al.*, 2017), it is possible that some words in the BKB corpus could be unfamiliar to a normal-hearing four-year-old. Filtering and dichotic presentation of bands could introduce further challenges to speech recognition, particularly in younger listeners. However, the primary outcome of interest in experiment 1 was the effect of masker type, and there is no reason to believe that challenges related to language ability and signal processing differ for the two masker conditions. It is therefore unlikely that the choice of stimuli was entirely responsible for the larger effects of child age in the multi-tone than the multi-noise-band masker, although task demands could play a role in the magnitude of this effect.

It is unclear how to explain the finding of a larger child-adult difference for the multi-tone masker than the multi-noise-band masker, but one possibility is related to the availability of spectrally sparse cues in the multi-tone masker condition. While the multi-noise-band masker provides relatively uniform masking of the target bands, the multi-tone masker provides more masking at the spectral center than the edges of the target speech bands. The optimal listening strategy would therefore be to make use of speech cues near the edges of the speech bands via off-frequency listening. This strategy could fail in young listeners for two reasons: inability to recognize speech based on spectrally sparse cues, or inability to segregate and selectively attend to those sparse cues.

While adults are relatively good at making use of speech cues that are restricted in frequency, children are not. For example, Tarr and Nittrouer (2013) measured masked vowel identification thresholds in five-year-olds, eight-year-olds, and adults. Target stimuli were synthesized vowels /I/ and /ε/ that were either limited to F_1 (peaks at 375 and 625 Hz, respectively), or also included F_2 and F_3 components, which were identical across the two vowels (2200 and 2900 Hz). The child-adult difference was reduced by inclusion of the high-frequency components despite the fact that they did not provide additional distinguishing information. This result is broadly consistent with the suggestion that children may process speech in larger linguistic units than adults (e.g., words rather than phonemes; reviewed by Mainela-Arnold *et al.*, 2008). It is also consistent with the observation that children require a wider bandwidth than adults to recognize speech (Mlot *et al.*, 2010; McCreery and Stelmachowicz, 2011).

The ability to benefit from improved TMR in some auditory channels for speech in a multi-tone masker also relies on the listener's ability to listen selectively in frequency.

Despite adult-like peripheral frequency selectivity, children are more susceptible to off-frequency masking than adults (Leibold and Neff, 2011; Leibold and Buss, 2016). Interestingly, a recent study by Leibold and Buss (2016) showed that children's increased susceptibility to off-frequency masking was largely limited to cases where the noise masker was gated on during each listening interval. Playing the masker continuously reduced masking for all listeners. This finding was interpreted in terms of auditory stream segregation; whereas synchronous target and masker onsets may interfere with a child's ability to segregate the target from the masker, playing the masker continuously allows the child to form separate auditory streams and attend selectively to the target frequency. In the context of the present experiment, it is possible that children were unable to make use of speech cues in channels offset from those dominated by the tone masker due to immature segregation abilities.

The second experiment evaluated the possibility that the larger child-adult difference in SRTs for the multi-tone than the multi-noise-band masker could be due to maturation in the ability to segregate and selectively attend to speech cues available to adults in the multi-tone masker. Binaural difference cues were used to facilitate segregation of the target and masker. If children's particularly poor performance in the multi-tone masker is due to poor segregation, then they should derive a particularly large benefit from the binaural cue for that condition.

III. EXPERIMENT 2

The goal of the second experiment was to test the hypothesis that the larger child-adult difference for the multi-tone masker than the multi-noise-band masker is due to immaturity in the ability to segregate and attend to cues in auditory channels associated with the edges of the target bands. The approach taken here was to facilitate segregation of the target and masker by manipulating interaural target phase. Presenting a diotic masker with target speech that is out of phase at the two ears results in better performance than when both the target and masker are diotic (Levitt and Rabiner, 1967; Wilson *et al.*, 1982; Johansson and Arlinger, 2002; Goverts and Houtgast, 2010). This benefit is referred to as the binaural intelligibility level difference (BILD; Levitt and Rabiner, 1967). If children obtain a larger BILD than adults for the multi-tone masker, this could indicate development of the ability to segregate the speech from the multi-tone masker and selectively attend to target cues present at the spectral edges of the speech bands.

It is unclear how the BILD might be affected by modulation masking associated with the multi-noise-band masker, but previous data using non-speech stimuli suggest that segregation cues, including binaural difference cues, can reduce modulation masking. For example, Grantham and Bacon (1991) reported detection thresholds for target modulation imposed on a wideband noise carrier in the presence of masker modulation imposed on a second wideband noise carrier. Stimuli were presented either monaurally or dichotically, with the carrier bands and masker modulation in phase across ears and target modulation out of phase across ears.

Thresholds for detecting target modulation were better with than without the binaural difference cue (Grantham and Bacon, 1991). Modulation masking can also be observed for stimuli distributed across frequency. Under these conditions, stimulus manipulations designed to segregate stimulus energy associated with the target and masker, such as asynchronous onset, reduce or eliminate masking (e.g., Oxenham and Dau, 2001). These observations support the expectation that greater susceptibility to modulation masking in young children could lead to greater BILD for the multi-noise-band than the multi-tone masker, although this was not expected to occur given the more pronounced age effects for the multi-tone masker in experiment 1.

A. Methods

Listeners were normal-hearing five- to ten-year-olds ($n = 18$) and two groups of normal-hearing adults (19–35 years old), one group tested in the same conditions as children ($n = 10$, mean of 26 yr) and a second group tested with a subset of stimulus bands ($n = 11$, mean of 24 yr). Inclusion and exclusion criteria were the same as for experiment 1. Most listeners were tested in a single 1-h session.

Stimuli and test procedures closely followed those of experiment 1 with the following exceptions. For children and the first group of adults, stimuli were restricted to the 14 odd-numbered bands, indicated in the top row of Table I. The second group of adults was tested with the 14 odd-numbered masker bands and a subset of 5 target bands, indicated with stars in the top row of Table I. Stimulus bands were presented to both ears in two binaural conditions. The masker was always diotic (Mo), and the target was either diotic (To) or out of phase in the two ears ($T\pi$). Each listener completed a single block of 60 trials in each of 4 conditions: 2 maskers (multi-tone and multi-noise-band) \times 2 stimulus phase configurations (MoTo and Mo $T\pi$). Conditions were completed in random order. Stimuli were generated in MATLAB, played out of a real-time processor (RZ6, TDT), and presented over headphones (Sennheiser, HD 25). As in experiment 1, psychometric functions were fitted to estimate the SRT associated with 50% correct for each listener and condition.

B. Results

Figure 2 shows results plotted separately for the multi-noise-band masker (left column) and the multi-tone masker (right column). Results for individual children are plotted as a function of age. The mean of adult data appears at the far right of each panel, and boxplots indicate the distribution of individual results. The top row shows SRTs, where symbol shape indicates stimulus phase, as defined in the legend. The bottom row indicates the magnitude of the BILD in dB, computed as the difference in SRTs in the two stimulus phase conditions (MoTo–Mo $T\pi$). Due to experimenter error, one child (5.4 years old) did not provide data in the Mo $T\pi$ condition for the multi-noise-band masker.

The SRT improved with child age for all four conditions, but the effect of age depended on both masker type and stimulus phase. For the multi-noise-band masker, the

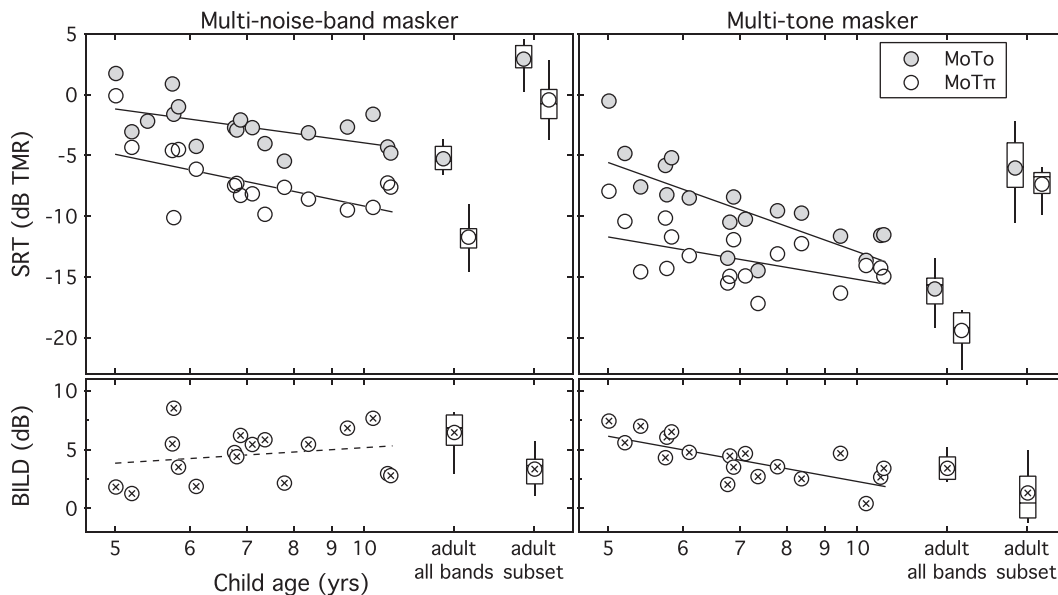


FIG. 2. Results obtained in the MoTo and MoT π conditions. The top row of panels shows SRTs for the multi-noise-band masker (left) and the multi-tone masker (right). Symbol shape indicates the stimulus phase condition, which was either MoTo (filled circles) or MoT π (open circles). The bottom row of panels shows the BILD, computed as the difference between SRTs in the MoTo and MoT π conditions for individual listeners. Solid lines indicate a significant correlation with child age, and dashed lines indicate non-significant trends. As in Fig. 1, data are plotted as a function of child age, and adult means are shown at the right of each panel. Boxes span the 25th–75th percentiles, horizontal lines indicate the median, and vertical lines span the 10th–90th percentiles.

change in SRT with age was similar for MoTo ($r = -0.54$, $\beta = -3.95$, $p = 0.021$) and MoT π ($r = -0.59$, $\beta = -6.03$, $p = 0.013$) conditions. But for the multi-tone masker, the effect of age was larger for MoTo ($r = -0.73$, $\beta = -10.34$, $p = 0.001$) than MoT π ($r = -0.53$, $\beta = -4.93$, $p = 0.024$). The SRT improvement with increasing age was compared across conditions with a linear mixed model; masker type, stimulus phase, and the log of listener age were fixed factors, and subject was a random factor. This model resulted in a significant three-way interaction between masker type, stimulus phase, and child age ($\beta = 7.18$, $SE = 2.45$, $p = 0.006$), indicating that effects of age and stimulus phase differed for the two maskers. Evaluating SRTs for just the multi-noise-band masker, there was an effect of age ($\beta = -3.94$, $SE = 1.80$, $p = 0.044$) and stimulus phase ($\beta = -4.44$, $SE = 0.52$, $p < 0.001$), and no interaction between age and stimulus phase ($\beta = -2.02$, $SE = 2.15$, $p = 0.362$). Evaluating SRTs for just the multi-tone masker, there was an effect of age ($\beta = -10.34$, $SE = 2.20$, $p < 0.001$) and stimulus phase ($\beta = -4.31$, $SE = 0.30$, $p < 0.001$), and a significant interaction between age and stimulus phase ($\beta = 5.41$, $SE = 1.24$, $p = 0.001$). This interaction reflects a larger age effect in the MoTo condition than MoT π condition for the multi-tone masker. A model including just the MoTo data resulted in an interaction between age and masker type ($\beta = -6.40$, $SE = 2.25$, $p = 0.012$), replicating the results of experiment 1. A model including just the MoT π data found no interaction between child age and masker type ($\beta = 0.56$, $SE = 1.50$, $p = 0.714$), consistent with the interpretation that the binaural difference cue eliminates the greater masking that young children experienced for the multi-tone masker relative to the multi-noise-band masker.

As in experiment 1, the effect of masker for the MoTo condition was smaller for children than for adults tested in

the 14-band condition; the difference between SRTs in the multi-noise-band and multi-tone masker was 4.4 dB for five-year-olds (based on line fits), and 10.7 dB in adults. This result is similar to the values of 5.5 and 9.8 dB observed in experiment 1, where the 28 bands were presented dichotically (odd-numbered bands to the left ear and even-numbered bands to the right ear). In contrast, for the MoT π condition, the difference in SRTs was 6.8 dB for five-year-olds and 7.7 dB for adults; a single-sample t -test indicates a non-significant trend for the masker effect in adults to exceed 6.8 dB ($t_9 = 2.26$, $p = 0.050$). In other words, the binaural difference cue markedly reduced the child-adult difference in the difference between SRTs in the multi-noise-band and multi-tone conditions.

Attention now turns to adults tested with a subset of five target bands. Reducing the number of bands from 14 to 5 increased SRTs by 8.2 dB (MoTo) and 11.3 dB (MoT π) in the multi-noise-band masker, and by 10.0 dB (MoTo) and 12.0 dB (MoT π) in the multi-tone masker. The resulting SRTs were comparable to, or slightly higher than, those of the youngest children. Of particular interest here, the BILD was smaller for adults tested in the 5-band than the 14-band conditions. This reduction in the BILD was 3.1 dB for the multi-noise-band masker ($t_{19} = 4.43$, $p < 0.001$) and 2.1 dB for the multi-tone masker ($t_{19} = 2.91$, $p = 0.009$). In contrast, the BILD was larger for the younger, poorer-performing children than the older children and adults. This result lends support for the idea that the larger BILD for young children in the multi-tone masker condition reflects development and is not simply a consequence of the higher TMRs at threshold. Results obtained for adults in the five-band conditions also provide evidence of an interaction between masker type and age with diotic stimuli, which is eliminated by the introduction of the binaural difference cue. For the MoTo

condition, the difference between SRTs in the multi-noise-band and multi-tone masker was 4.4 dB for five-year-olds (based on line fits) and 8.9 dB for adults ($t_{10}=8.41$, $p<0.001$). For the MoT π condition, that difference was 6.8 dB for five-year-olds and 6.9 dB for adults ($t_{10}=0.34$, $p=0.745$).

C. Discussion

Results of experiment 2 replicated the more pronounced age effects observed for speech recognition in a multi-tone masker compared to a multi-noise-band masker in the absence of a binaural difference cue (MoTo), as observed in experiment 1. If modulation masking played a dominant role in children's poor masked speech perception, then the opposite pattern of results would be expected: a larger child-adult difference in the multi-noise-band masker due to inherent amplitude modulation of the noise bands.

Introducing a binaural difference cue (MoT π) improved performance for all listeners. In the multi-noise-band condition, the mean BILD of 6.5 dB for adults is broadly consistent with previously reported BILD values of 5–8 dB for unfiltered speech (Levitt and Rabiner, 1967; Wilson *et al.*, 1982; Johansson and Arlinger, 2002; Goverts and Houtgast, 2010). There was a non-significant trend for the BILD observed in a multi-noise-band masker to increase with child age. For a tone-in-noise detection task, children and adults benefit from binaural difference cues to a comparable degree for wide masker bandwidths, but children benefit less than adults for narrowband maskers (Hall and Grose, 1990; Grose *et al.*, 1997). This result has been interpreted as reflecting development of temporal resolution of binaural cues (Hall *et al.*, 2004; Hall *et al.*, 2007), the effects of which are more pronounced for narrowband stimuli due to perceptual prominent inherent fluctuation of narrow bands. If stimulus fluctuation limited the values of BILD for the multi-noise-band masker obtained for young children in the present study that could complicate interpretation of BMLD data as a means of differentiating modulation masking from other effects. However, adults tested with a subset of 5 target speech bands also obtained a smaller BILD than those tested with the full complement of 14 bands. This outcome is consistent with the conclusion that the higher TMR at threshold for young children may account for their smaller BILD in the multi-noise-band condition.

In contrast to results obtained with the multi-noise-band masker, the BILD observed for the multi-tone masker was larger for young children than older children and adults. This cannot be explained in terms of young children's poor performance overall, as adults tested with a subset of 5 bands had a *smaller* BILD for the multi-tone masker than adults tested with the full complement of 14 bands. One interpretation of the large BILD obtained for young children with the multi-tone masker is that those listeners had greater difficulties in the MoTo condition with auditory stream segregation and/or selective attention to the target speech compared to adults. Immature ability to segregate and selectively attend to a speech target has been proposed to account for the pronounced and prolonged child-adult difference for speech

recognition in a speech masker (Wightman and Kistler, 2005; Corbin *et al.*, 2016). This result is often described in terms of the perceptual similarity between a speech target and speech masker interfering with young children's ability to weight the appropriate acoustic information. Leibold *et al.* (2016) recently argued that infants may experience similar difficulties when listening to speech in noise. That is, the perceptual features that allow adults to easily segregate speech from noise may be learned. The present results suggest that the perceptual differences between target speech and the multi-tone masker may support relatively good segregation in adults but pose a greater challenge for young children.

Whereas the multi-tone masker proved to be particularly challenging for young children in the absence of binaural difference cues, the age-by-masker interaction was eliminated with the introduction of a binaural difference cue. This result suggests that once young children were able to segregate target speech from the multi-tone masker, their speech recognition was no more immature than in the multi-noise-band masker. This observation suggests that segregation may play a dominant role in the age-by-masker interaction observed *without* a binaural difference cue. While young children may be less adept than adults at recognizing speech based on sparse cues (Buss *et al.*, 2017), this factor appears not to explain young children's particular susceptibility to masking with the multi-tone masker.

Young children's limited ability to segregate the target from the multi-tone masker is broadly consistent with off-frequency masking in a tone detection paradigm. Leibold and Buss (2016) measured detection thresholds for a 2-kHz pure-tone signal in the presence of an off-frequency masker. In one set of conditions the masker was a band of noise, filtered between 4 and 10 kHz. When this masker was gated on during each listening interval, it raised thresholds by 10.9 dB in four- to six-year-olds and by 2.0 dB in adults, but when it played continuously masking dropped to 2.9 and 0.2 dB, respectively. Children's pronounced susceptibility to masking with the gated masker is consistent with the idea that synchronous target and masker onsets interfere with the child's ability to segregate the target from the masker and listen selectively at the target frequency. It is possible that speech recognition in the multi-tone masker represents an example of children's reduced ability to listen in a frequency-selective manner, particularly for gated maskers. This possibility receives some support from the observation that the present study used gated maskers and showed a stronger age effect for the multi-tone masker than the multi-noise-band masker, whereas Hall *et al.* (2012) used continuous maskers and showed similar child-adult differences for spectrally modulated and unmodulated noise maskers.

IV. CONCLUSIONS

The present results provide no evidence of greater susceptibility to modulation masking in young children than adults for a speech-in-noise recognition task. On the contrary, children's SRTs were elevated more for the multi-tone masker, which lacks marked inherent amplitude modulation,

than the multi-noise-band masker. Young children had a larger BILD than adults for the multi-tone masker, but not for the multi-noise-band masker. This result suggests that young children's greater susceptibility to masking with the multi-tone masker could be due to immature auditory segregation and/or selective attention to the target in the presence of a spectrally sparse masker. This result undermines comparison of speech masked by narrow bands of noise vs tones as a means of assessing susceptibility to modulation masking, particularly in listeners with limited abilities to segregate auditory streams or listen selectively in frequency. Although the present results do not provide evidence of deleterious developmental effects of modulation masking, they do not rule out the possibility that such effects might occur in other paradigms.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) Grant No. R01 DC000397 (E.B.). We thank Joseph W. Hall III for helpful comments on the manuscript, and for serving as inspiration for the study generally.

¹The decision to represent age in log of years is not critical to the results. Repeating all analyses with age represented in years did not change the pattern of significance.

- ANSI (2010). ANSI S3.6-2010, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Bacon, S. P., and Grantham, D. W. (1989). "Modulation masking: Effects of modulation frequency, depth, and phase," *J. Acoust. Soc. Am.* **85**, 2575–2580.
- Bench, R. J., and Bamford, J. (1979). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic Press, London).
- Bench, J., Kowal, A., and Bamford, J. (1979). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," *Br. J. Audiol.* **13**, 108–112.
- Buss, E., Leibold, L. J., and Hall, J. W., III (2016). "Effect of response context and masker type on word recognition in school-age children and adults," *J. Acoust. Soc. Am.* **140**, 968–977.
- Buss, E., Leibold, L. J., Porter, H. L., and Grose, J. H. (2017). "Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing," *J. Acoust. Soc. Am.* **141**, 2650–2660.
- Cameron, S., and Dillon, H. (2007). "Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)," *Ear Hear.* **28**, 196–211.
- 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.
- Corbin, N. E., Buss, E., and Leibold, L. J. (2017). "Spatial release from masking in children: Effects of simulated hearing loss," *Ear Hear.* **38**, 223–235.
- Edmonds, B. A., and Culling, J. F. (2005). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," *J. Acoust. Soc. Am.* **117**, 3069–3078.
- 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.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**, 2112–2122.
- Gallun, F. J., Mason, C. R., and Kidd, G., Jr. (2005). "Binaural release from informational masking in a speech identification task," *J. Acoust. Soc. Am.* **118**, 1614–1625.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Goverts, S. T., and Houtgast, T. (2010). "The binaural intelligibility level difference in hearing-impaired listeners: The role of supra-threshold deficits," *J. Acoust. Soc. Am.* **127**, 3073–3084.
- Grantham, D. W., and Bacon, S. P. (1991). "Binaural modulation masking," *J. Acoust. Soc. Am.* **89**, 1340–1349.
- Grose, J. H., Hall, J. W., III, and Dev, M. B. (1997). "MLD in children: Effects of signal and masker bandwidths," *J. Speech Lang. Hear. Res.* **40**, 955–959.
- Hall, J. W., III, Buss, E., and Grose, J. H. (2007). "The binaural temporal window in adults and children," *J. Acoust. Soc. Am.* **121**, 401–410.
- Hall, J. W., Buss, E., Grose, J. H., and Dev, M. B. (2004). "Developmental effects in the masking-level difference," *J. Speech Lang. Hear. Res.* **47**, 13–20.
- 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., III, and Grose, J. H. (1990). "The masking-level difference in children," *J. Am. Acad. Audiol.* **1**, 81–88.
- 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.
- Halliday, L. F., Tuomainen, O., and Rosen, S. (2017). "Language development and impairment in children with mild to moderate sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **60**, 1551–1567.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection," *J. Acoust. Soc. Am.* **85**, 1676–1680.
- Johansson, M. S., and Arlinger, S. D. (2002). "Binaural masking level difference for speech signals in noise," *Int. J. Audiol.* **41**, 279–284.
- Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). "Reducing informational masking by sound segregation," *J. Acoust. Soc. Am.* **95**, 3475–3480.
- Leibold, L. J., and Buss, E. (2013). "Children's identification of consonants in a speech-shaped noise or a two-talker masker," *J. Speech Lang. Hear. Res.* **56**, 1144–1155.
- Leibold, L. J., and Buss, E. (2016). "Factors responsible for remote-frequency masking in children and adults," *J. Acoust. Soc. Am.* **140**, 4367–4377.
- Leibold, L. J., and Neff, D. L. (2011). "Masking by a remote-frequency noise band in children and adults," *Ear Hear.* **32**, 663–666.
- Leibold, L. J., Yarnell Bonino, A., and Buss, E. (2016). "Masked speech perception thresholds in infants, children, and adults," *Ear Hear.* **37**, 345–353.
- Levitt, H., and Rabiner, L. R. (1967). "Binaural release from masking for speech and gain in intelligibility," *J. Acoust. Soc. Am.* **42**, 601–608.
- Mainela-Arnold, E., Evans, J. L., and Coody, J. A. (2008). "Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task," *J. Speech Lang. Hear. Res.* **51**, 381–393.
- Mayer, D. L., and Dobson, V. (1982). "Visual acuity development in infants and young children, as assessed by operant preferential looking," *Vision Res.* **22**, 1141–1151.
- McCreery, R. W., Spratford, M., Kirby, B., and Brennan, M. (2017). "Individual differences in language and working memory affect children's speech recognition in noise," *Int. J. Audiol.* **56**, 306–315.
- McCreery, R. W., and Stelmachowicz, P. G. (2011). "Audibility-based predictions of speech recognition for children and adults with normal hearing," *J. Acoust. Soc. Am.* **130**, 4070–4081.
- Mlot, S., Buss, E., and Hall, J. W. (2010). "Spectral integration and bandwidth effects on speech recognition in school-aged children and adults," *Ear Hear.* **31**, 56–62.
- Moeller, M. P., Tomblin, J. B., Yoshinaga-Itano, C., Connor, C. M., and Jerger, S. (2007). "Current state of knowledge: Language and literacy of children with hearing impairment," *Ear Hear.* **28**, 740–753.
- Murphy, J., Summerfield, A. Q., O'Donoghue, G. M., and Moore, D. R. (2011). "Spatial hearing of normally hearing and cochlear implanted children," *Int. J. Pediatr. Otorhinolaryngol.* **75**, 489–494.
- Oxenham, A. J., and Dau, T. (2001). "Modulation detection interference: Effects of concurrent and sequential streaming," *J. Acoust. Soc. Am.* **110**, 402–408.

- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2016). *nlme: Linear and Nonlinear Mixed Effects Models* (R Foundation for Statistical Computing, Vienna, Austria).
- R Core Team (2016). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
- Scollie, S. D. (2008). "Children's speech recognition scores: The Speech Intelligibility Index and proficiency factors for age and hearing level," *Ear Hear.* **29**, 543–556.
- Sek, A., Baer, T., Crinnion, W., Springgay, A., and Moore, B. C. (2015). "Modulation masking within and across carriers for subjects with normal and impaired hearing," *J. Acoust. Soc. Am.* **138**, 1143–1153.
- Stone, M. A., Fullgrabe, C., Mackinnon, R. C., and Moore, B. C. J. (2011). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," *J. Acoust. Soc. Am.* **130**, 2874–2881.
- Stone, M. A., Fullgrabe, C., and Moore, B. C. J. (2012). "Notionally steady background noise acts primarily as a modulation masker of speech," *J. Acoust. Soc. Am.* **132**, 317–326.
- Stone, M. A., and Moore, B. C. J. (2014). "On the near non-existence of 'pure' energetic masking release for speech," *J. Acoust. Soc. Am.* **135**, 1967–1977.
- Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 3638–3646.
- 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.
- Tarr, E., and Nittrouer, S. (2013). "Explaining coherence in coherence masking protection for adults and children," *J. Acoust. Soc. Am.* **133**, 4218–4231.
- Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (2003). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," *J. Acoust. Soc. Am.* **113**, 3297–3305.
- 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., Hopkins, J. L., Mance, C. M., and Novak, R. E. (1982). "Detection and recognition masking-level differences for the individual CID W-1 spondaic words," *J. Speech Hear. Res.* **25**, 235–242.
- 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.