Research Article

Forward and Backward Masking of Consonants in School-Age Children and Adults

Heather L. Porter,^a Emily R. Spitzer,^b Emily Buss,^c Lori J. Leibold,^a and John H. Grose^c

Purpose: This experiment sought to determine whether children's increased susceptibility to nonsimultaneous masking, particularly backward masking, is evident for speech stimuli.

Method: Five- to 9-year-olds and adults with normal hearing heard nonsense consonant–vowel–consonant targets. In Experiments 1 and 2, those targets were presented between two 250-ms segments of 70-dB-SPL speech-shaped noise, at either –30 dB signal-to-noise ratio (Experiment 1) or at the listener's word recognition threshold (Experiment 2). In Experiment 3, the target was presented in steady speech-shaped noise at listener threshold. For all experiments, percent correct was estimated for initial and final consonants.

Results: In the nonsimultaneous noise conditions, child–adult differences were larger for the final consonant than

the initial consonant whether listeners were tested at –30 dB signal-to-noise ratio (Experiment 1) or at their individual word recognition threshold (Experiment 2). Children were not particularly susceptible to backward masking relative to adults when tested in a steady masker (Experiment 3).

Conclusions: Child–adult differences were greater for backward than forward masking for speech in a nonsimultaneous noise masker, as observed in previous psychophysical studies using tonal stimuli. Children's greater susceptibility to nonsimultaneous masking, and backward masking in particular, could play a role in their limited ability to benefit from masker envelope modulation when recognizing masked speech.

Young school-age children perform more poorly than adults on tasks of masked speech recognition (e.g., McCreery & Stelmachowicz, 2011; Nishi, Lewis, Hoover, Choi, & Stelmachowicz, 2010; Nitttrouer & Boothroyd, 1990). Whereas speech recognition tends to improve for adults when amplitude modulation is applied to a noise masker (e.g., Festen & Plomp, 1990), some data suggest that this benefit may be smaller for school-age children than adults. For example, Buss, Leibold, and Hall (2016) evaluated masked word recognition in 5- to 13-year-olds and found that the benefit associated with amplitude modulating the noise masker increased with child age, with the oldest children deriving the same benefit as adults (see also Hall, Buss, & Grose, 2014). Hall, Buss, Grose, and Roush (2012) showed similar effects for Bamford–Kowal–Bench

sentence recognition in amplitude-modulated or steady speech-shaped noise; there was a difference in the benefit of masker modulation between 4- to 6-year-olds and adults, but not between 7- to 11-year-olds and adults (see also Wróblewski, Lewis, Valente, & Stelmachowicz, 2012). Although the above studies suggest that young children benefit less than adults from masker modulation, a child–adult difference in the ability to benefit from masker modulation is not always observed (Stuart, 2005, 2008; Stuart, Givens, Walker, & Elangovan, 2006).

The ability of young children to benefit from masker amplitude modulation may be influenced by a number of interrelated factors, including reduced ability to perceptually isolate target speech cues coincident with masker envelope minima (e.g., Buss, Leibold, Porter, & Grose, 2017; Hall et al., 2014), susceptibility to informational or “perceptual” masking (e.g., Buss et al., 2017; Leibold, Bonino, & Buss, 2016; Wightman, Kistler, & O'Bryan, 2010), and poor temporal resolution (e.g., Buss, Hall, Porter, & Grose, 2014; Wightman et al., 2010), including increased susceptibility to nonsimultaneous masking (e.g., Buss, He, Grose, & Hall, 2013).

One source of evidence for the development of temporal resolution comes from psychophysical studies indicating that, although thresholds for detecting short-duration tones in noise are higher for school-age children relative

^aCenter for Hearing Research, Boys Town National Research Hospital, Omaha, NE

^bDepartment of Allied Health Sciences, University of North Carolina at Chapel Hill

^cOtolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill

Correspondence to Heather L. Porter: heather.porter@boystown.org

Editor-in-Chief: Frederick (Erick) Gallun

Editor: Jennifer Lentz

Received October 28, 2017

Revision received January 11, 2018

Accepted March 20, 2018

https://doi.org/10.1044/2018_JSLHR-H-17-0403

Disclosure: The authors have declared that no competing interests existed at the time of publication.

to adults, this age effect is even larger for nonsimultaneous maskers (Buss, Hall, Grose, & Dev, 1999; Buss et al., 2013; Hartley, Wright, Hogan, & Moore, 2000). For example, Hartley et al. (2000) measured thresholds for a 20-ms tone at 1000 Hz, presented either immediately before or during a 300-ms band of noise (600–1200 Hz). When the tone preceded the masker, children's performance was estimated to be adultlike by 11 years old, but when the tone was simultaneous with the masker, their performance was adultlike by 8 years old. This result indicates a more prolonged development for backward than simultaneous masking. Similarly, Buss et al. (1999) showed significant age-related threshold improvements between 5 and 11 years old for children tasked with detecting a 10-ms tone presented immediately after wideband Gaussian noise with 300-ms duration, a condition associated with forward masking. Interestingly, Buss et al. (2013) found greater child–adult differences for backward masking compared with forward masking for short-duration pure-tone stimuli presented in modulation minima of amplitude-modulated wideband masking noise. These results, indicating development of temporal resolution with child age, have been interpreted in terms of central rather than peripheral limitations. Whereas young school-age children are thought to have mature peripheral auditory encoding, their central auditory processing is not yet adultlike (Moore & Linthicum, 2007).

This study sought to investigate the impact of non-simultaneous masking on children's speech perception. On the basis of tone-in-noise data, one might expect children to experience more nonsimultaneous masking than adults, with a larger child–adult difference for backward than forward masking. However, speech stimuli are longer in duration and more familiar than short-duration tonal stimuli, possibly making speech less prone to nonsimultaneous masking than brief tones. Either outcome would clarify the factors contributing to children's limited ability to benefit from masker envelope fluctuation in previous studies of masked speech perception.

Experiment 1

The purpose of Experiment 1 was to determine whether children experience more nonsimultaneous masking than adults for speech and whether that effect would be greater for backward masking compared with forward masking. Such a result would replicate developmental effects observed for a short-duration pure-tone target presented in fluctuating wideband masking noise (Buss et al., 2013). The approach was to present nonword consonant–vowel–consonant (CVC) targets between two bursts of speech-shaped masking noise.

Stimuli

Target stimuli were CVC nonsense words spoken by a female talker. Nonsense words were used to reduce the impact of language experience on target recognition. The corpus of CVC nonsense words was controlled for phonotactic probability and separated into 27 lists, balanced

for occurrence of initial and final consonants (McCreery & Stelmachowicz, 2011). Each list included 25 nonsense CVCs, each saved to disk in a separate .wav file. On average, recordings were 703 ms in duration (486–936 ms), and duration did not differ significantly across lists, $F(26, 648) = 0.34$, $p = .999$. Target CVC nonsense words were presented between two bursts of noise that were spectrally shaped to match the long-term average speech spectrum. Masker bursts were 250 ms in duration with 5-ms raised-cosine ramps, presented at 70 dB SPL; a new pair of noise samples was generated before each listening interval. The duration of the interval between the two masker bursts was adjusted to match the duration of the target CVC. Figure 1 shows an example stimulus. Target CVCs were presented at a fixed level of –30 dB signal-to-noise ratio (SNR; i.e., 40 dB SPL). Stimuli were generated using custom MATLAB scripts and a real-time processor (RZ6, TDT) and presented using Sennheiser HD 265 linear headphones.

Procedures

Listeners and researchers were seated opposite one another at a table in a sound-attenuated room within the Human Auditory Development Lab at the University of North Carolina at Chapel Hill. Target and masker stimuli were presented to listeners monaurally in the left ear. For the purposes of online scoring, unmasked target CVCs were also played (monaurally) to the researcher, and the target transcription and amplitude spectrum were displayed on the data collection computer, visible only to the researcher.

Stimulus presentation rate was controlled by the researcher. Tracks progressed after initial, middle, and final phonemes were scored correct or incorrect for each trial. Percent correct was determined for each phoneme position. Listeners were presented four lists of 25 nonsense CVCs; list selection was based on a pseudorandom number generator (MATLAB), with the provision that lists were not repeated within a listener. Data were collected in a single session, lasting approximately 15–20 min. Feedback was not provided to listeners during the session.

Figure 1. Example stimulus. The nonsense word /hɛb/ is shown at –10 dB signal-to-noise ratio relative to two 250-ms segments of speech-shaped masking noise.

Audiovisual recording of all sessions was undertaken for offline analysis by trained research assistants at the University of North Carolina at Chapel Hill or Boys Town National Research Hospital.

Listeners

Listeners included 19 children aged 5.1–9.6 years ($M = 7.2$ years, $SD = 1.3$ years) and 20 adults aged 20.4–33.8 years ($M = 24.5$ years, $SD = 3.9$ years). All listeners had hearing sensitivity equal to or better than 20 dB HL at octave frequencies from 250 to 8000 Hz (American National Standards Institute, 2009), and no listeners reported history of chronic ear infections.

Data Analysis

Before analysis, percent correct scores were converted to rationalized arcsine units (RAUs) to normalize variance (Studebaker, 1985). The logarithm of age was used for evaluating age effects within the child group,¹ based on the expectation of rapid development in early childhood that slows with maturity (Mayer & Dobson, 1982; Moller & Rollins, 2002). The initial consonant, vowel, and final consonant in the nonsense CVC targets are referred to as C1, V, and C2, respectively, from this point forward. Online and offline scoring was found to be in good agreement (88%–92%), and data from online scoring were used for data analysis.² One-tailed Pearson correlations were used to test for significant positive associations between child age and recognition of C1 and C2. A repeated-measures analysis of variance was used to test for main effects and interactions of the within-group factor of phoneme position (C1, C2) and the between-group factor of listener age (child, adult). Vowel recognition performance was excluded from these analyses because the primary intent of this study was to examine forward and backward masking on initial and final consonant recognition and because performance tended to be at or near ceiling for vowel recognition.

¹The pattern of significance was not dependent on the units used to represent age (linear or log scale).

²Offline scoring was conducted based on audio and video recordings. The scorer also had access to both the orthographic (Klattese) and audio recordings of the original stimuli. On the basis of this information, each phoneme was scored as correct or incorrect. In cases where a phoneme was scored as incorrect, a transcription of the phoneme produced by the listener was attempted. Agreement between online and offline scoring reported in the text is the percentage of trials for which both scorers indicated a correct or an incorrect response. For data sets with more than one offline rater, agreement between offline scores was comparable with the agreement between online and offline scores. This observation supports the conclusion that offline scores were not more reliable than online scores. Although correct/incorrect scoring was relatively consistent across raters, there was only 50% agreement across offline raters in the transcriptions made for incorrect responses. This variability rules out an error analysis of listener responses.

Results and Discussion

Figure 2 shows phoneme recognition scores for individual child listeners, plotted as a function of age, and mean phoneme recognition for adult listeners in RAU. The horizontal gray bar indicates the 95% confidence interval for adult data. Significant and nonsignificant associations between child age and performance are shown by the solid and dashed lines, respectively. Data are plotted separately for C1 (left panel) and C2 (right panel). Table 1 lists mean and standard deviation in RAU for children and adults for C1, V, and C2 as well as the difference between performance for backward and forward masking conditions (i.e., C2–C1).

Several patterns are evident in these data. First, phoneme recognition was positively associated with child age for C1 ($r = .48$, $p = .019$), and there was a nonsignificant association with age for C2 ($r = .364$, $p = .063$). Performance was better for adults than children and better for C2 than C1. The difference between scores for children and adults was larger for C2 than C1. The significance of these observations was confirmed with a repeated-measures analysis of variance. This analysis indicated a significant main effect of group, $F(1, 37) = 61.62$, $p < .001$; a significant main effect of phoneme position, $F(1, 37) = 95.35$, $p < .001$; and a significant interaction, $F(1, 37) = 9.61$, $p = .004$.

The finding of poorer performance for children than adults is consistent with published data showing that school-age children are more susceptible to masking than adults for speech targets presented in dynamic maskers, such as two-talker speech (e.g., Corbin, Bonino, Buss, & Leibold, 2016; Leibold & Buss, 2013) or amplitude-modulated speech-shaped noise (e.g., Hall et al., 2014; Stuart, 2008; Wróblewski et al., 2012), results that have been attributed to informational masking.

Figure 2. Phoneme recognition scores for nonsense consonant–vowel–consonant (CVC) targets played in between 250-ms bursts of speech-shaped masking noise at –30 dB signal-to-noise ratio (open circles). Performance in rationalized arcsine units (RAUs) is shown for individual children, plotted as a function of age (log scale), and mean phoneme recognition for adult listeners is shown at the right of each panel. The gray bar shows the 95% confidence interval for adult data. Significant and nonsignificant associations between child age and performance are shown by the solid and dashed lines, respectively. Data are plotted separately for C1 (left panel) and C2 (right panel).

Table 1. Mean and standard deviation C1, V, and C2 phoneme recognition in rationalized arcsine units (RAUs) for child and adult listeners in each of the three experiments.

Condition		RAU, <i>M</i> (<i>SD</i>)			
		C1	V	C2	C2–C1
Experiment 1	Child	43.5 (10.5)	95.4 (8.5)	51.3 (12.3)	7.8 (7.9)
	Adult	64.4 (9.9)	110.3 (3.9)	79.5 (8.8)	15.1 (6.8)
Experiment 2	Child	73.4 (5.6)	108.9 (5.7)	79.0 (5.6)	5.6 (5.2)
	Adult	71.4 (5.4)	112.2 (2.5)	84.1 (6.7)	12.7 (5.8)
Experiment 3	Child	66.8 (5.2)	93.2 (10.5)	58.8 (9.4)	–8.0 (9.7)
	Adult	68.7 (8.6)	99.5 (7.9)	59.4 (2.5)	–9.3 (8.9)

Note. Experiment 1 presented the signal between two masker bursts at –30 dB signal-to-noise ratio. Experiment 2 presented the signal between two masker bursts at the listener's word recognition threshold. Experiment 3 presented the signal in 45-dB-SPL steady masking noise at the listener's C2 recognition threshold. The relative impact of backward versus forward masking is estimated as the difference scores for C2 and C1.

Data from this study can also be characterized in terms of the relative magnitude of forward and backward masking. Whereas children's C2 scores were 7.8 RAUs higher than their C1 scores, that difference was 15.1 RAUs for adults. Results from this study indicate that the increased susceptibility to backward masking in children observed for tones and noise (e.g., Buss et al., 2013; Hartley et al., 2000) is also observed for nonsense CVCs presented in a gap between bursts of speech-shaped masking noise. In adults, backward and forward masking are reduced by the introduction of a contralateral cue in some conditions. This result has been interpreted as indicating that central auditory system effects, such as informational masking, contribute to estimates of threshold for these conditions (Elliott, 1962; Moore & Glasberg, 1982; Puleo & Pastore, 1980). Whereas informational masking is thought to be the primary influence in backward masking, forward masking is thought to include peripheral auditory system effects in addition to informational masking effects (Moore & Glasberg, 1982; Puleo & Pastore, 1980). Considering these studies together with results showing that informational masking impacts children to a greater degree than adults (e.g., Leibold et al., 2016; Wightman & Kistler, 2005), it is reasonable to consider that larger child–adult differences for backward than forward masking for tones and speech may be related to children's greater susceptibility to informational masking.

Experiment 2

Experiment 1 showed larger child–adult differences for backward than forward masking in the context of a speech-in-noise task, a result previously observed in the context of tone detection in a fluctuating noise masker (i.e., Buss et al., 2013). Although this result is consistent with greater susceptibility to backward masking in children, it is also possible that differences in the psychometric

function slope for recognizing C1 and C2 with a non-simultaneous masker could introduce a child–adult difference when the two groups are tested at the same SNR. The goal of Experiment 2 was to determine whether the greater child–adult difference for backward masking can be eliminated by testing listeners at an SNR that is equally challenging for the two groups. This was accomplished by evaluating performance at each individual listener's word recognition threshold.

Method

All stimuli and procedures were identical to Experiment 1, with the exception that Experiment 2 presented nonsense CVC stimuli at each listener's threshold for recognizing the entire CVC target. As for Experiment 1, the masker was presented at 70 dB SPL. Threshold for nonsense CVC targets was estimated using a 1-down, 1-up rule to adjust target presentation level (Levitt, 1971), which estimates the level associated with 50% correct on the psychometric function. A response was judged to be correct if all three phonemes were reported back correctly. The initial signal level was 50 dB SPL. The step size was 4 dB before the second reversal and 2 dB thereafter. Each track continued for 12 reversals, and the final estimate of threshold was the mean of the last 10 reversals. Four threshold estimates were averaged to calculate the final threshold. Listeners who participated in Experiment 2 were identical to those in Experiment 1. All data were collected in a single session, lasting approximately 30–45 min.

Results and Discussion

Figure 3 shows thresholds for individual child listeners and the average adult threshold (dB SNR) for

Figure 3. Word recognition thresholds for nonsense consonant–vowel–consonant (CVC) targets presented in between 250-ms bursts of speech-shaped noise presented at 70 dB SPL. Thresholds for individual children are plotted as a function of age (log scale), and the average adult threshold is shown at the right of the panel. Thresholds are plotted in dB signal-to-noise ratio (SNR) (left ordinate) and dB SPL (right ordinate). A significant association between child age and performance is shown by the solid line. The gray bar shows the 95% confidence interval for adult data.

nonsense CVC targets presented between 250-ms bursts of speech-shaped noise presented at 70 dB SPL. The standard deviation across the four threshold estimates was 3.6 dB on average for adults and 3.5 dB on average for children. The gray bar in Figure 3 shows the 95% confidence interval for the adult data. A significant association between child age and performance is shown as a solid line. These data illustrate a significant, negative, one-tailed Pearson correlation between child age and threshold ($r = -.47$, $p = .022$). This result is consistent with studies indicating improvement with age for speech recognition in steady masking noise (e.g., Hall, Grose, Buss, & Dev, 2002; Nishi et al., 2010; Wróblewski et al., 2012).

Average child and adult C1, V, and C2 performance in RAU and the difference between performance for backward and forward masking conditions (i.e., C2–C1) for the current experiment are included in Table 1. Despite tracking threshold to 50% correct during the adaptive portion of this experiment, speech recognition performance for individual phonemes is above 50% correct. However, recall that threshold was measured adaptively based on correct recognition of all three phonemes. In this scenario, it is possible for an incorrect response to contain one or two correctly identified phonemes. In contrast to Experiment 1, children did not perform more poorly than adults overall. A significant main effect of group was not observed, confirming that the strategy to equate performance across individuals by testing at SNRs corresponding to 50% recognition worked reasonably well, $F(1, 37) = 0.93$, $p = .340$. Similarly, there was no correlation between log of child age and consonant recognition scores (C1: $r = -.16$, $p = .256$; C2: $r = -.04$, $p = .435$).

As observed for nonsense CVC targets presented at –30 dB SNR (Experiment 1), backward masking had less impact on speech recognition than forward masking, particularly for adults. This was confirmed by a significant main effect of phoneme position, $F(1, 37) = 106.62$, $p < .001$, and a significant Group $\times$ Phoneme interaction, $F(1, 37) = 15.94$, $p < .001$. The difference between scores for C1 and C2 was 5.6 RAUs for children and 12.7 RAUs for adults. Another way to characterize this interaction is by comparing scores between groups. For C1, scores were roughly similar for children and adults on average, $t(37) = -1.10$, $p = .279$. For C2, scores were 5.1 RAUs worse (lower) on average for children than adults, $t(37) = 2.46$, $p = .019$.

The current experiment was designed to assess speech recognition in nonstationary masking noise at listener threshold to evaluate the role of overall performance in the relative effects of forward and backward masking. Findings are consistent with the results of Experiment 1, showing that children are more susceptible to backward masking than adults even when nonsense CVC stimuli were presented at each listener's 50% threshold for CVC targets.

One assumption up to this point is that the larger child–adult difference for C2 than C1 is due to children's greater susceptibility to backward masking and not to poorer masked recognition of C2 relative to C1 generally.

Although past studies have shown that steady masking noise has greater detrimental effects on speech recognition for children than adults for a variety of speech stimuli (e.g., Hall et al., 2002; Nishi et al., 2010; Wróblewski et al., 2012), the impact of steady noise on initial and final phoneme recognition in children remains unknown. It is possible that developmental effects related to auditory working memory or the ability to segregate and maintain auditory streams (Sussman, Wong, Horváth, Winkler, & Wang, 2007; Vuontela, Steenari, & Koivisto, 2003) could result in poorer C2 than C1 performance in children than adults, regardless of the nonstationary nature of the masking noise.

Experiment 3

The purpose of Experiment 3 was to determine whether steady masking noise has a comparable effect on C1 and C2 phoneme recognition for children and adults for nonsense CVC targets. For adults, the consonant at the beginning of a word is often less susceptible to masking than a consonant later in the word (e.g., Dubno & Levitt, 1981; Woods, Yund, Herron, & Cruadhlaioich, 2010); however, the initial consonant advantage is not consistently present for all consonants and may be influenced by vowel context (Redford & Diehl, 1999). An initial consonant advantage would not necessarily impact the interpretation of data from Experiments 1 and 2, unless the effects were different for children and adults.

Method

Procedures and stimuli were identical to Experiment 2, with the exception that Experiment 3 measured each listener's 50% threshold for C2 recognition in steady speech-shaped masking noise presented at 45 dB SPL. This masker level was chosen to approximately match thresholds to those obtained with a nonstationary 70-dB-SPL masker in Experiment 2. Masked phoneme recognition was then measured at each listener's 50% threshold for C2. To verify that 45 dB SPL was an adequate masker level to elevate thresholds relative to presentation in quiet, 50% threshold for C2 recognition in quiet was also measured for adult listeners. Data were collected in a single session, lasting approximately 30–45 min for children and approximately 1 hr for adults.

Experiment 3 was undertaken after completion of Experiments 1 and 2. As such, child and adult listener groups included in Experiment 3 differed from those included in the first two experiments. Listeners in Experiment 3 included 13 children aged 5.3–9.4 years ($M = 6.7$ years, $SD = 1.1$ years) and 10 adults aged 19.8–41.6 years ($M = 27.2$ years, $SD = 6.7$ years). As for previous experiments, all listeners had normal hearing sensitivity defined as thresholds equal to or better than 20 dB HL at octave frequencies from 500 to 8000 Hz (American National Standards Institute, 2009).

Results and Discussion

Figure 4 shows child listener thresholds and mean adult threshold in dB SNR for nonsense CVC targets in 45-dB-SPL steady speech-shaped masking noise. The gray bar shows the 95% confidence interval for adult data. A nonsignificant association between child age and performance is shown by the dashed line. For adults, the mean threshold for C2 recognition in 45-dB-SPL steady speech-shaped masking noise was 44.5 dB SPL (shown as -0.5 dB SNR; Figure 4), whereas the mean threshold in quiet was 32.6 dB SPL (not shown). This difference in threshold between quiet and masked conditions indicates that the masker level chosen for use in Experiment 3 was adequate to elevate thresholds, $F(1, 9) = 732.13$, $p < .001$.

Thresholds for child listeners tended to decrease with age, although a one-tailed Pearson correlation failed to reach statistical significance ($r = -.39$, $p = .095$). Unlike the data for modulated noise (shown in Figure 3), some older children performed as well as adults in steady masking noise. A significant correlation was expected based on age effects in published data for speech recognition in steady masking noise (e.g., Hall et al., 2002; Nishi et al., 2010; Wróblewski et al., 2012); however, variability in our listener data set, particularly for children approximately 6 years old, may have obscured age effects.

The bottom row of Table 1 lists mean and standard deviation in RAU for C1, V, and C2 as well as the difference between performance for backward and forward masking conditions (i.e., C2–C1) for child and adult listeners tested in steady speech-shaped masking noise at 45 dB SPL. Despite presenting targets at each listener's 50% threshold for C2, percent correct for C2 is approximately 59% for both adults and children. This better-than-expected performance could reflect bias associated with relatively short adaptive threshold estimation tracks that start above threshold, or with effects of practice, which have been shown to be particularly evident for backward masking in some conditions (e.g., Buss et al., 1999; Wright,

Fitzgerald, Johnston, & Reid, 1999; Wright, Johnston, & Reid, 1996). In contrast to Experiments 1 and 2, scores were better (higher) for C1 than C2, and there was no evidence of a difference between performance of children and adults. There was a significant main effect of phoneme position, $F(1, 21) = 19.30$, $p < .001$, but no effect of group, $F(1, 21) = 0.290$, $p = .596$, and no Group $\times$ Phoneme Position interaction, $F(1, 21) = 0.115$, $p = .738$. Better recognition for C1 than C2 is consistent with past research with adults (e.g., Dubno & Levitt, 1981; Redford & Diehl, 1999; Woods et al., 2010). The present data indicate a similar result for children.

Results of Experiment 3 indicate that the larger child–adult difference in backward than forward masking observed in previous experiments cannot be explained in terms of differential susceptibility to masking of C1 and C2 in general. Steady masking noise had similar effects on child and adult phoneme perception for initial and final consonants in nonsense CVC targets presented at threshold SNR. These data support the assumption that child–adult differences for consonant perception observed under conditions of nonsimultaneous masking reflect differences in susceptibility to forward and backward masking.

General Discussion

This study evaluated nonsense speech recognition for nonsense CVC targets presented between two bursts of speech-shaped noise. Children were more susceptible to nonsimultaneous masking than adults, and this effect was larger for backward than forward masking. Results from the current experiment are analogous to those found by Buss et al. (2013). That study reported a larger child–adult difference for backward masking than forward masking for brief pure-tone bursts presented at different positions in the modulation cycle of a fluctuating noise masker. Although direct comparison of data between Buss et al. (2013) and this study is hampered by differences in methodology, a qualitative comparison suggests broadly parallel results. In one set of conditions, Buss et al. (2013) reported detection thresholds for a brief 1000-Hz tone presented either before or after a broadband masker burst; the child–adult difference for 5- to 6.5-year-olds was 6.1 dB for forward masking and 9.8 dB for backward masking. This study showed comparable results for CVC nonsense words, in that the child–adult difference was smaller for C1 than C2. Comparing performance of adults with that of 5- to 6.5-year-olds, estimated based on line fits to the data (Figure 2), the child–adult difference for C1 was 25–30 RAUs, and the child–adult difference for C2 was 32–37 RAUs.

Despite the similarity in results, there are important differences between tone-in-noise and speech-in-noise paradigms for studying nonsimultaneous masking. One salient difference is the task: detection in the case of tone-in-noise and recognition in the case of speech-in-noise. Another difference is stimulus duration; whereas tonal stimuli can be quite brief (e.g., 5 ms), speech stimuli are

Figure 4. Recognition thresholds for C2 in stationary speech-shaped noise presented at 45 dB SPL. Child listener thresholds are plotted as a function of age (log scale), and the average adult threshold is shown at the right of the panel. Thresholds are plotted in dB signal-to-noise ratio (SNR; left ordinate) and dB SPL (right ordinate). The gray bar shows the 95% confidence interval for adult data. A nonsignificant association between child age and performance is shown by the dashed line.

necessarily longer, and phoneme duration can be difficult to define precisely. When speech stimuli are spliced out of a long recording, there is some ambiguity regarding when the recording reaches the noise floor, and the point at which the level drops below audibility depends on presentation level. Speech recordings in this study were spliced by hand, and there was some variability across stimuli with respect to the delay between masker bursts and stimulus features responsible for consonant recognition. Although these differences were consistent across children and adults, they could theoretically impact the results reported here.

Up to this point, scores for C1 and C2 have been interpreted as reflecting forward and backward masking associated with the noise bursts, respectively. Effects of backward masking on recognition of C1, or forward masking on recognition of C2, are assumed to be minimal. The results of Buss et al. (2013) provide some support for that assumption. In that study, the temporal windows fitted to 1000-Hz thresholds presented between two broadband masker bursts had shorter time constants for adults than children, but the mean equivalent rectangular duration was less than 50 ms for all age groups. Recall that masker bursts before and after the stimulus in this study were separated by 486–936 ms. Although the exact temporal distribution of cues for C1 and C2 is not known, it is quite likely that the dominant cues to C1 were temporally resolved from the trailing masker burst, and cues associated with C2 were temporally resolved from the leading masker burst. One caveat is that the temporal window duration of < 50 ms measured for detection of a brief 1000-Hz tone may not be accurate for the stimulus conditions used here. It is also theoretically possible that the vowel contributed additional forward and backward masking effects on C2 and C1, respectively. However, this effect is likely to be minimal compared with masking associated by the noise masker, given the relatively low level of the vowel compared with the noise bursts.

Backward masking is thought to be influenced by central auditory system effects, like informational masking. This interpretation is consistent with a reduction in backward masking for dichotic conditions (Elliott, 1962) or with the addition of a temporally synchronous contralateral listening cue (e.g., pure tone, band-limited noise, or temporal gap in an otherwise constant noise masker; Puleo & Pastore, 1980). Although the ability to perceptually segregate a target and a masker is often discussed in the context of speech targets in speech maskers, Leibold et al. (2016) hypothesized that the ability to segregate speech targets and noise maskers may also develop with age. That study found a child–adult difference for speech presented in a two-talker masker but not for speech in noise, a result consistent with greater informational masking for the speech-in-speech task in children. However, infants performed more poorly than adults for both masker types, with a similar infant–adult difference in the two conditions. On the basis of this result, Leibold et al. hypothesized that segregation may be a limiting factor in infants for both

two-talker speech and noise maskers. It is unclear when segregation of speech and steady noise becomes adultlike, but it is likely that masker modulation is a relevant factor in this process.

In summary, results from this study indicate that the increased susceptibility to backward masking in children observed for tones in amplitude-modulated noise generalizes to speech presented between bursts of speech-shaped noise. Larger child–adult differences for the second consonant (C2) than the first consonant (C1) were evident whether listeners were tested at –30 dB SNR or at an SNR based on their individual word recognition thresholds. Children and adults had similar patterns of C1 and C2 recognition in steady masking noise, suggesting that the larger child–adult differences in backward than forward masking are specific to nonstationary masking conditions. These findings are consistent with the idea that immature temporal resolution could play a role in masked speech recognition with fluctuating maskers.

Acknowledgments

The work was supported by the National Institute on Deafness and Other Communication Disorders (R01 DC011038, granted to Lori J. Leibold, and F32 DC014209, granted to Heather L. Porter). The authors are grateful to Margaret Miller, Heidi Lang, and the members of the Psychoacoustics Laboratories and the Human Auditory Development Lab at the University of North Carolina at Chapel Hill for their assistance with data processing.

References

- American National Standards Institute. (2009). *ANSI/ASA S3.21-2004 (R2009) methods manual pure-tone threshold audiometry*. New York, NY: Author.
- Buss, E., Hall, J. W., Grose, J. H., & Dev, M. B. (1999). Development of adult-like performance in backward, simultaneous, and forward masking. *Journal of Speech, Language, and Hearing Research*, 42(4), 844–849.
- Buss, E., Hall, J. W., Porter, H., & Grose, J. H. (2014). Gap detection in school-age children and adults: Effects of inherent envelope modulation and the availability of cues across frequency. *Journal of Speech, Language, and Hearing Research*, 57, 1098–1107. https://doi.org/10.1044/2014_JSLHR-H-13-0132
- Buss, E., He, S., Grose, J. H., & Hall, J. W. (2013). The monaural temporal window based on masking period pattern data in school-aged children and adults. *The Journal of the Acoustical Society of America*, 133(3), 1586–1597.
- Buss, E., Leibold, L. J., & Hall, J. W. (2016). Effect of response context and masker type on word recognition in school-age children and adults. *The Journal of the Acoustical Society of America*, 140(2), 968–977.
- Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America*, 141(4), 2650–2660.
- Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children:

- Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*, 37(1), 55–63.
- Dubno, J. R., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *The Journal of the Acoustical Society of America*, 69(1), 249–261.
- Elliott, L. L. (1962). Backward masking: Monotic and dichotic conditions. *The Journal of the Acoustical Society of America*, 34(8), 1108–1115.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725–1736.
- Hall, J. W., Buss, E., & Grose, J. H. (2014). Development of speech glimpsing in synchronously and asynchronously modulated noise. *The Journal of the Acoustical Society of America*, 135(6), 3594–3600.
- Hall, J. W., Buss, E., Grose, J. H., & Roush, P. A. (2012). Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation. *Ear and Hearing*, 33(3), 340–348.
- Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, 23(2), 159–165.
- Hartley, D. E. H., Wright, B. A., Hogan, S. C., & Moore, D. R. (2000). Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. *Journal of Speech, Language, and Hearing Research*, 43(6), 1402–1415.
- Leibold, L. J., Bonino, A. Y., & Buss, E. (2016). Masked speech perception thresholds in infants, children, and adults. *Ear and Hearing*, 37(3), 345–353.
- Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech-shaped noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 56(4), 1144–1155.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2), 467–477.
- Mayer, D. L., & Dobson, V. (1982). Visual acuity development in infants and young children, as assessed by operant preferential looking. *Vision Research*, 22(9), 1141–1151.
- McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America*, 130(6), 4070–4081.
- Moller, A. R., & Rollins, P. R. (2002). The non-classical auditory pathways are involved in hearing in children but not in adults. *Neuroscience Letters*, 319(1), 41–44.
- Moore, B., & Glasberg, B. R. (1982). Contralateral and ipsilateral cueing in forward masking. *The Journal of the Acoustical Society of America*, 71(4), 942–945.
- Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460–478.
- Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., & Stelmachowicz, P. G. (2010). Children's recognition of American English consonants in noise. *The Journal of the Acoustical Society of America*, 127(5), 3177–3188.
- Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, 87(6), 2705–2715.
- Puleo, J. S., & Pastore, R. E. (1980). Contralateral cueing effects in backward masking. *The Journal of the Acoustical Society of America*, 67(3), 947–951.
- Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *The Journal of Laryngology and Otology*, 106(3), 1555–1565.
- Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and Hearing*, 26(1), 78–88.
- Stuart, A. (2008). Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. *Journal of the American Academy of Audiology*, 19(2), 135–146.
- Stuart, A., Givens, G. D., Walker, L. J., & Elangovan, S. (2006). Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise. *The Journal of the Acoustical Society of America*, 119(4), 1946–1949.
- Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, 28(3), 455–462.
- Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, 225, 117–127.
- Vuontela, V., Steenari, M., & Koivisto, J. (2003). Audiospatial and visuospatial working memory in 6–13 year old school children. *Learning & Memory*, 10, 74–81.
- Wightman, F., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, 118(5), 3164–3176.
- Wightman, F., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *The Journal of the Acoustical Society of America*, 128(1), 270–279. <https://doi.org/10.1121/1.3436536>
- Woods, D. L., Yund, E. W., Herron, T. J., & Cruadhlaoich, M. A. U. (2010). Consonant identification in consonant–vowel–consonant syllables in speech-spectrum noise. *The Journal of the Acoustical Society of America*, 127(3), 1609–1623.
- Wright, B. A., Fitzgerald, M. B., Johnston, P. A., & Reid, M. D. (1999). Influence of condition variation on learning in auditory backward masking. *The Journal of the Acoustical Society of America*, 105(2), 1237.
- Wright, B. A., Johnston, P. A., & Reid, M. D. (1996). Learning and generalization in auditory backward masking. *The Journal of the Acoustical Society of America*, 100(4), 2818.
- Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear and Hearing*, 33(6), 731–744.