

Development of speech glimpsing in synchronously and asynchronously modulated noise

Joseph W. Hall III,^{a)} Emily Buss, and John H. Grose

Department of Otolaryngology—Head and Neck Surgery, University of North Carolina at Chapel Hill, 170 Manning Drive, Chapel Hill, North Carolina 27599-7070

(Received 6 August 2013; revised 3 February 2014; accepted 8 April 2014)

This study investigated development of the ability to integrate glimpses of speech in modulated noise. Noise was modulated synchronously across frequency or asynchronously such that when noise below 1300 Hz was “off,” noise above 1300 Hz was “on,” and vice versa. Asynchronous masking was used to examine the ability of listeners to integrate speech glimpses separated across time and frequency. The study used the Word Intelligibility by Picture Identification (WIPI) test and included adults, older children (age 8–10 yr) and younger children (5–7 yr). Results showed poorer masking release for the children than the adults for synchronous modulation but not for asynchronous modulation. It is possible that children can integrate cues relatively well when all intervals provide at least partial speech information (asynchronous modulation) but less well when some intervals provide little or no information (synchronous modulation). Control conditions indicated that children appeared to derive less benefit than adults from speech cues below 1300 Hz. This frequency effect was supported by supplementary conditions where the noise was unmodulated and the speech was low- or high-pass filtered. Possible sources of the developmental frequency effect include differences in frequency weighting, effective speech bandwidth, and the signal-to-noise ratio in the unmodulated noise condition.

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4873518>]

PACS number(s): 43.71.Ft, 43.66.Dc [PBN]

Pages: 3594–3600

I. INTRODUCTION

In environments with fluctuating noise, good speech understanding may depend upon the ability to process fragments of speech that correspond to epochs where the signal-to-noise ratios (SNRs) in given spectral regions are momentarily relatively favorable. This type of listening is often referred to as glimpsing (e.g., Cooke, 2006; Li and Loizou, 2007). The present study investigates developmental effects, comparing performance in steady masking noise with performance in conditions where the masker is temporally amplitude modulated (AM) in various ways. Performance improvement associated with modulation is referred to as masking release.

Only a few developmental studies have compared speech perception in steady versus modulated noise (Stuart, 2005; Stuart *et al.*, 2006; Stuart, 2008; Hall *et al.*, 2012; Wroblewski *et al.*, 2012). Stuart *et al.* (2006) examined speech recognition in steady and modulated noise for monosyllabic words, and Stuart (2008) and Wroblewski *et al.* (2012) examined speech recognition for sentences. None of these studies found a statistically significant difference in masking release between adults and children although Wroblewski *et al.* found reduced masking release for children when reverberation was present. Hall *et al.* (2012), investigating speech recognition for sentences, found that children 4.6–6.9 yr of age had smaller benefit due to noise modulation than adults but no significant difference when comparing children 7.3–11.1 yr of age to adults. One goal of the present investigation was to bring greater

clarity to the question of whether temporal masking release for speech is reduced in children. A feature of the present approach is that the modulation parameters are associated with larger magnitudes of masking release than in previous developmental studies, perhaps creating a larger range over which to observe potential developmental effects.

In addition to examining masking release for temporal modulation, Hall *et al.* (2012) also investigated conditions where the noise was spectrally modulated or both temporally and spectrally modulated. Spectral modulation was achieved by creating four band-stop regions (notches) in the spectrum of the noise masker via digital filtering. Whereas both age groups of children showed adult-like masking release for spectral modulation, masking release in the combined temporal/spectral modulation condition was smaller than for adults, even for the older child group. In the combined temporal/spectral modulation, listeners had access to the entire speech spectrum during temporal modulation minima but only parts of the spectrum during modulation maxima. One possible interpretation of this result is that the ability to process speech glimpses that are distributed across time and frequency develops relatively late. A second goal of the present study was to obtain further information about this class of processing, using the “checkerboard masking” paradigm (Howard-Jones and Rosen, 1993), which was developed specifically to investigate the ability of a listener to integrate speech glimpses that are distributed across time and frequency. In a key condition of this paradigm, the masking noise in different frequency regions is modulated asynchronously, such that when the noise is “on” in one spectral region, the noise is “off” in a complementary spectral region. Comparing performance in this asynchronous modulation

^{a)}Author to whom correspondence should be addressed. Electronic mail: jwh@med.unc.edu

condition to performance in other baseline and control conditions (described in the following section) provides insights about the ability of the listener to integrate speech cues across time and frequency.

II. METHODS

A. Subjects

Three age groups were tested: 5–7 yr (younger child), 8–10 yr (older child), and 18–42 yr (adult). There were 10 subjects in each group, with mean ages of 6.5, 9.2, and 27.3 yr. All listeners had audiometric thresholds better than 20 dB hearing level (HL) (ANSI, 2010) for octave frequencies from 250 to 8000 Hz and no history of otitis media within the past 3 yr.

B. Stimuli

The speech material was adapted from the Word Intelligibility by Picture Identification (WIPI) test (Ross and Lerman, 1970). In this test, the listener hears the phrase “show me,” followed by a target word spoken by a female talker. A set of pictures is shown, and the listener is asked to identify the picture corresponding to the target word. We adapted this test for presentation via computer with pictures displayed on a video screen. In this adaptation, words were presented in a four-alternative-forced choice paradigm where each spoken word was associated with presentation of four pictures in a 2×2 matrix, one of which corresponded to the spoken word. This presentation differed from the traditional WIPI test in that there were three rather than five foil pictures per presentation; pictures without a corresponding audio recording were omitted in the present paradigm. A feature of the WIPI test is that the words displayed on a trial are phonetically similar; in all but one case, words in each set of four shared a common vowel and differed in terms of the initial, or initial and final, consonant. An example target word is “bread” with foils of “bed,” “red,” and “sled.” Thus even though the task is closed set, it requires relatively fine auditory coding because of the phonetic similarity of the target and foils. There were 100 total target words in 25 sets of four. Including the prefix “show me,” recordings were 0.98–1.65 s with a mean duration of 1.19 s.

As in Howard-Jones and Rosen (1993), the masker was pink noise (equal energy per octave band). Depending upon condition, the noise was (1) steady, (2) synchronous AM across the entire noise bandwidth, (3) asynchronous AM such that when the noise below 1300 Hz was on, the noise above 1300 Hz was off, and vice versa, (4) low band AM only with AM below 1300 Hz and steady above this frequency, and (5) high band AM only with AM above 1300 Hz and steady below this frequency. These conditions are summarized schematically in Fig. 1. Masking release was defined as the difference in SRT between the steady condition one of the modulation conditions. As in Howard-Jones and Rosen (1993), the ability to benefit from *asynchronous glimpsing* was taken as the difference between the masking release in the asynchronous modulation condition and the masking release in the better of the conditions where only the low band was modulated or only the high band was modulated.

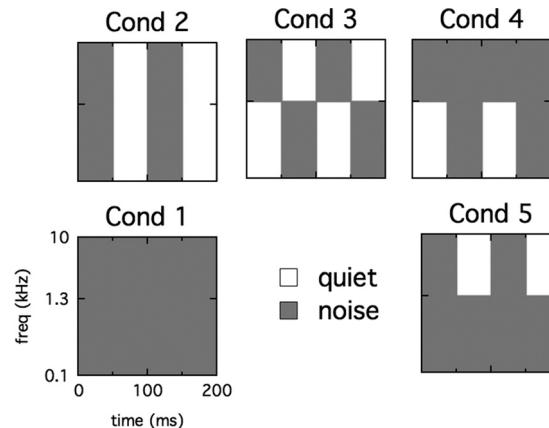


FIG. 1. Spectral/temporal representations of the maskers in the various conditions (frequency on the ordinate and time on the abscissa). The conditions are (1) steady noise, (2) synchronously modulated noise, (3) asynchronously modulated noise, (4) low band noise modulated and high band noise steady, and (5) high band noise modulated and low band noise steady.

Filtering was accomplished by passing the masker through a third order Butterworth filter twice, once forward and once backward, to cancel non-linear phase effects. The transition frequency of 1300 Hz was chosen because pilot listening in adults indicated this frequency resulted in approximately equal performance in conditions 4 and 5, as defined in the preceding text. The modulation had a frequency of 10 Hz and an approximately square-wave shape (5-ms raised cosines were used to smooth the modulation transitions). The level of the masking noise was fixed at 68 dB sound pressure level (SPL) before amplitude modulation was applied. Each masker presentation was 2 s in duration, and the signal was temporally centered in the masker. A new sample of pink noise was generated for each trial.

C. Procedure

An adaptive threshold procedure was used to estimate SRTs. At the outset of each track, the 25 sets of words were arranged in random order. On each trial, a word from the associated set was randomly selected to serve as the target. Once the previously determined random order was exhausted, a new random order was determined with the proviso that the first word set in the new order could not be the same as the last word set in the previous order. For each trial, a word was played and four pictures were presented on a video display. The listener was asked to touch the picture corresponding to the auditory stimulus and to guess when unsure. Following a correct response, the stimulus level was decreased, and following an incorrect response, it was increased. These signal level adjustments were made in steps of 4 dB prior to the second track reversal and in steps of 2 dB thereafter. The threshold track was stopped after 10 reversals in tracking direction, and the threshold estimate was taken as the average of the last 8 reversal values. Stimuli were presented to the right ear via a Sennheiser HD 265 headset. Listeners were seated in a double-wall sound booth. Thresholds were blocked by condition, and the order of conditions was selected pseudo-randomly for each listener to control for possible order effects.

TABLE I. SRTs (dB SPL) for the five conditions in the main experiment. The standard error of the mean appears in parentheses below each mean.

Group	Steady noise	Sync AM	Async AM	Low AM Only	High AM Only
Younger children (5–7 yr)	72.0 (0.4)	56.7 (0.8)	60.5 (1.0)	70.1 (1.1)	64.0 (0.6)
Older children (8–10 yr)	70.3 (0.7)	52.9 (0.8)	57.7 (0.5)	65.9 (0.7)	61.5 (0.4)
Adults	65.5 (0.6)	45.4 (0.8)	51.8 (0.8)	57.7 (1.5)	57.5 (0.4)

III. RESULTS AND SUPPLEMENTARY CONDITIONS

Table I shows average SRTs for the different masking conditions for the three age groups. Because the SRT in steady noise was the baseline for the measures of masking release, an initial analysis of variance (ANOVA) was performed to assess the effect of age group on this measure. This analysis showed a significant effect of age group ($F_{2,27} = 34.4$; $p < 0.001$; $\eta_p^2 = 0.72$). Preplanned comparisons indicated improved SRT in steady noise with increasing age for the younger versus older children ($p = 0.04$), and the older children versus adults ($p < 0.001$).

Figure 2 shows the masking release for the four modulated noise conditions investigated. As in Howard-Jones and Rosen (1993), the SRT in steady noise was used as the reference for calculating masking release. As can be seen in this figure, the effect of age on masking release appears to vary markedly across conditions. A repeated measures ANOVA performed on the masking release measures indicated a significant effect of modulation condition ($F_{3,81} = 205.1$; $p < 0.001$; $\eta_p^2 = 0.88$), a significant effect of age group ($F_{2,27} = 5.2$; $p = 0.012$; $\eta_p^2 = 0.28$), and a significant interaction between modulation condition and age group ($F_{6,81} = 4.2$; $p < 0.001$; $\eta_p^2 = 0.24$). Because the interaction

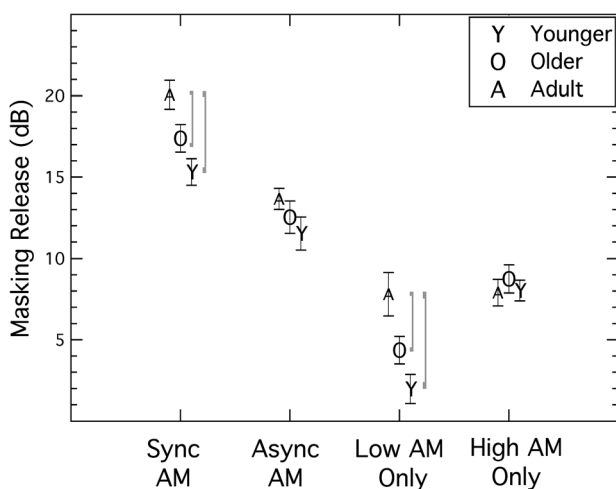


FIG. 2. Masking release (dB) for noise that was synchronously modulated, asynchronously modulated, modulated below 1300 Hz and steady above 1300 Hz, and steady below 1300 Hz and modulated above 1300 Hz. Data are shown for the adults (A), older children (O), and younger children (Y). Brackets identify developmental differences that were statistically significant (see text). Error bars show plus and minus 1 standard error of the mean.

was significant, care should be taken in interpreting the main effects of modulation condition and age. Simple effects testing (Kirk, 1968) was performed to reveal the sources of the significant interaction.

First, we consider results by modulation condition: For the synchronous AM condition, the two child groups did not differ significantly from each other ($p = 0.097$), but the adults had significantly larger masking release than both the younger children ($p < 0.001$) and the older children ($p = 0.036$).

For the asynchronous AM condition, there was no significant difference between any of the age groups.

For the low band AM only condition (high band steady), the two child groups did not differ significantly from each other ($p = 0.118$), but the adults had significantly larger masking release than both the younger children ($p < 0.001$) and the older children ($p = 0.028$).

For the high band AM only condition (low band steady), there was no significant difference between any age group.

Next, we consider results by age group: For adults, there was no significant difference between the conditions where only the low band was modulated and the condition where only the high band was modulated ($p = 0.931$), but all the results in the remaining masking release conditions differed from one another ($p < 0.001$).

For both groups of children, the results in all masking release conditions differed significantly from one another ($p < 0.001$).

As in Howard-Jones and Rosen (1993), *asynchronous glimpsing* was defined as the difference between the masking release in the asynchronous modulation condition and those in the better of the conditions where only the low band was modulated or only the high band was modulated. This comparison was intended to determine whether performance in the asynchronous modulation condition could be accounted for by listening to only one of the asynchronous bands. A repeated measures ANOVA showed that the asynchronous condition was associated with better performance than the better of the single-band-AM conditions ($F_{1,27} = 74.0$; $p < 0.001$; $\eta_p^2 = 0.73$), no effect of age group ($F_{2,27} = 1.4$; $p = 0.26$; $\eta_p^2 = 0.10$), and no interaction between age group and stimulus condition ($F_{2,27} = 0.06$; $p = 0.94$; $\eta_p^2 = 0.005$). These results are consistent with an interpretation of similar asynchronous glimpsing across the age groups.

Because the developmental difference between the two control conditions (low band AM only and high band AM only) was not anticipated, we decided to collect a set of additional data to test the generality of this frequency-region effect and also provide a better foundation to consider the results of the study as a whole. The masker was the same pink noise used as the steady masker in the main experiment, and the target stimuli were again WIPI words presented in a four-alternative-forced choice paradigm. The masker was always steady in these conditions. There were three conditions: Unfiltered speech, low-pass filtered speech, and high-pass filtered speech (1300-Hz transition frequency). Note that the unfiltered speech condition replicates the reference condition of the main experiment. The masker level and the threshold estimation procedure were the same as in the main experiment. Nine new adults (19–23 yr of age with a mean

of 22.1 yr) and seven new children (5–7 yr of age with a mean of 6.1 yr) participated. All listeners had audiometric thresholds better than 20 dB HL (ANSI, 2010) for octave frequencies from 250 to 8000 Hz and no history of otitis media within the past 3 yr. The results are shown in Table II, where it can be seen that the children were poorer than adults in all conditions, but particularly for the low-pass speech condition. A repeated measures ANOVA showed significant effects of condition ($F_{2,28} = 129.4$; $p < 0.001$; $\eta^2_p = 0.90$) and group ($F_{1,14} = 100.66$; $p < 0.001$; $\eta^2_p = 0.88$), and a significant interaction ($F_{2,28} = 12.8$; $p < 0.001$; $\eta^2_p = 0.48$). The interaction was due to the fact that the children were particularly poor, in comparison to adults, in the low-pass speech condition. Thus the developmental frequency region effect that occurred in the modulated noise control conditions of the main experiment was also seen in the supplementary conditions assessing filtered speech recognition in steady noise. That is, children were less able to recognize speech that was restricted to the low frequencies.

IV. DISCUSSION

As in previous findings (e.g., Elliott *et al.*, 1979; Stuart, 2005; McCreery and Stelmachowicz, 2011), our results in the baseline steady noise condition indicated that children require a higher SNR than adults at the speech reception threshold (see Table I). The main goals of the present study were to investigate age effects on masking release with synchronous and asynchronous temporal modulation. These effects are considered first. The pattern of results obtained in the main and supplementary conditions dealing with possible developmental effects related to speech frequency region will then be considered.

A. Developmental effects for synchronously modulated noise

The present results showed that both groups of children obtained smaller masking release than the adults in the condition where the noise modulation was synchronous across frequency. This result is in contrast to the masking release findings of Stuart *et al.* (2006) for monosyllabic words. The Stuart *et al.* study measured percent correct as a function of SNR, comparing performance in adults and in children 4–5 yr of age. They found that although the performance of the children was poorer than that of the adults in both steady and modulated noise, masking release was not significantly different between the age groups. One possible reason for the difference in outcome between the present study and that

TABLE II. SRTs (dB SPL) for supplementary conditions examining low-pass, high-pass, and full spectrum speech in steady noise. The standard error of the mean appears in parentheses below each mean.

Group	Low-pass speech	High-pass speech	Full spectrum speech
Children (5–7 yrs)	83.0 (1.7)	87.1 (1.0)	72.4 (1.1)
Adults	68.8 (1.0)	81.3 (0.6)	65.8 (0.4)

of Stuart *et al.* is related to the nature of the monosyllabic word stimuli used in the two studies. Stuart *et al.* used monosyllabic words from NU-CHIPS word lists for children (Elliott and Katz, 1980). Although both the NU-CHIPS test and the WIPI test used here involve the recognition of monosyllabic words in a closed set context using four pictures, the WIPI test is more difficult: In the NU-CHIPS test, the four words on a trial are distinctively different, whereas in the WIPI test, the four words are phonetically similar to one another (see Sec. II). It is possible that good performance on the WIPI task, where the phonetic distinctions between the target and foil words are relatively subtle, depends more heavily upon a relatively long history of speech/language experience, thus putting children at a greater disadvantage. It is also possible that the phonetic similarity between the target and foil words increases the cognitive load when holding the perceived word in memory while viewing the four pictures displayed on the monitor, a situation that would again favor more mature processing capacity. Another possible reason for the difference in outcome between the studies is that Stuart *et al.* used a randomly varying modulation frequency with an average frequency that was higher than that used in the present study. This could have limited the overall amount of masking release and therefore the range over which developmental differences could have been observed.

The present developmental differences in masking release for synchronously modulated noise also differ in detail from the results of Hall *et al.* (2012) for sentence material. The Hall *et al.* study found that children 5–7 yr of age had smaller masking release than adults but that children 8–10 yr of age did not differ significantly from adults. The present study found that both child age groups had smaller masking release than adults. It is possible that the difference between studies may be related to the difference in speech material. Another possibility is that the difference is related to a greater sensitivity in observing a developmental effect due to the relatively large masking release associated with the present methods: Adults achieved masking release of approximately 20 dB with the 10-Hz square-wave modulation used here, contrasted with masking release of approximately 5 dB with the 10-Hz sinusoidal modulation used in Hall *et al.* (2012). As noted by Hall *et al.* (2012), one possible interpretation of the observed developmental effect in synchronously modulated noise is that children are relatively poor at piecing together a speech signal from cues that are distributed across separated temporal intervals.

B. Developmental effects for asynchronously modulated noise

The asynchronously modulated noise condition was run to test the possibility that there are specific developmental effects in the ability to process glimpses of speech that are distributed across time and frequency. The results of the asynchronously modulated noise condition were not consistent with such a developmental effect. Furthermore, all groups showed significant asynchronous glimpsing (better performance in the asynchronously modulated noise condition than in the low-only or high-only condition). This developmental

result might be viewed as puzzling with respect to the synchronously modulated noise finding. That is, if children have difficulty piecing together temporally separated glimpses of speech in synchronously modulated noise, why was a similar, or even poorer, outcome not observed for glimpses that are separated in time and in frequency? One possible explanation for the lack of a developmental effect in the asynchronously modulated noise condition is related to the fact that the adult masking release was relatively larger for the synchronously modulated noise condition (see Fig. 2), allowing a larger range over which to observe a developmental deficit. However, such an argument is undercut by the finding that the children *did* demonstrate a developmental deficit for the low-only modulation condition, where the adult masking release was even smaller than for the asynchronously modulated noise condition (see Fig. 2).

Another possible explanation for the lack of a developmental effect in the asynchronous modulation condition is related to the nature of the speech cues in the synchronous versus asynchronous modulation conditions (Buss *et al.*, 2009). In the synchronous condition, the listener has good access to the entire speech spectrum in some intervals, but very poor access in adjacent intervals. In the asynchronous condition, although no intervals provide access to the full speech spectrum, all intervals provide access to at least part of the spectrum. It is possible that, compared to adults, children are less negatively affected when all intervals provide at least partial speech information than when some of the intervals provide little or no information.

C. Frequency effects in the control conditions of the main experiment and filtered speech conditions of the supplementary conditions

Whereas the adults showed approximately equal performance in the low-only and high-only conditions of the main experiment, the children showed better performance in the high-only condition than the low-only condition. Thus the performance of the children was more adult-like when the masker modulation enhanced the audibility of the higher speech frequencies. The results of the supplementary conditions, where the masker was steady and the speech was either low-pass or high-pass filtered, also indicated that the performance of the children was more adult-like when listeners had access to speech cues in the higher spectral region (see Table II).

One possible interpretation of the developmental differences for spectral regions found in the main and supplementary conditions is that, in contrast to adults, the children weight the information above 1300 Hz relatively more than the lower frequency information. Such an effect is, at least generally, consistent with previous findings indicating that the use of particular speech cues can continue to develop over time in school-age children. For example, Nittrouer and colleagues showed that in making judgments about syllable-initial fricatives, children appeared to weight the dynamic vowel formant transition portion of the stimuli more than the static noise-simulated consonant when compared to adults (e.g., Nittrouer, 1996). McCreery and Stelmachowicz (2011)

have suggested that, relative to adults, children may give relatively more weight to higher speech frequencies. Although McCreery and Stelmachowicz did not find such an effect for nonsense words, they speculated that it might be more likely to find such an effect for linguistically meaningful stimuli, where adults could use experience and cognitive skills to support performance. The present results for WIPI words in noise are consistent with that supposition. It may be relevant to point out that Mlot *et al.* (2010) did not show a developmental speech frequency effect in a study on sentence recognition. That study examined sentence recognition in quiet, determining the speech bandwidth required for approximately 20% correct recognition for bands centered on either 500 or 2500 Hz. The results indicated that children required a greater bandwidth than adults, but the effect was similar at both frequencies. It is possible that developmental frequency effects for speech perception are complex and may depend upon a number of factors, including the speech material and the specific frequency regions investigated.

D. Effects of SNR on audible bandwidth

As noted in the preceding text, some of the present results suggest that developmental differences for WIPI word recognition are greater when the listener has access to low- rather than high-frequency speech cues. This could have implications not only for the control conditions of the main experiment but also for the results of the other experimental conditions. While this effect could be due to general differences in the weighting of speech cues by children and adults, it could also be affected by the spectral characteristics of the particular stimuli used in the present experiment.

Figure 3 shows the average spectrum of the target words used in this study. Low- and high-pass filtered pink noise spectra are also shown to represent the approximate frequency regions of these two bands. Consider first the steady noise masker of the main experiment and the effect of raising the level of the speech with respect to the level of the masker. Because of the differing spectral shapes of the signal and pink noise masker, the lower speech frequencies will become audible before the higher regions. As the level of the speech increases further, the listener will gain access to higher and higher speech frequencies. If an adult listener is

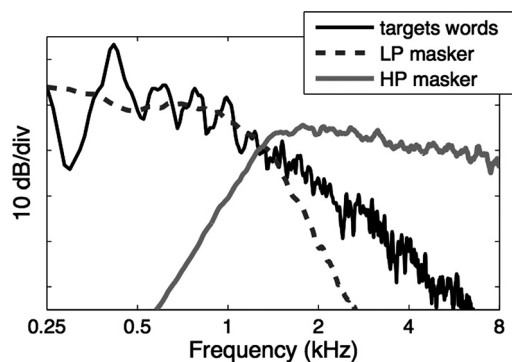


FIG. 3. Spectra of the target words, low-pass noise, and high-pass noise. Note that the filtered speech used in the supplementary conditions had the same roll-off at the 1300-Hz transition frequency as indicated here for the filtered noise.

able to “get by” on relatively low-frequency speech information, a relatively low masked SRT will be obtained. In contrast, a child who may not do well with limited low-frequency information will not obtain SRT until a higher SNR, where additional higher-frequency speech cues become available. Because of the difference in spectral shape between the signal and masker, this consideration applies to all of the conditions but may have been particularly relevant in the low-frequency band AM condition. Although the speech was broadband in this case, higher frequencies would be masked by the continuous high-pass noise. Furthermore, although the nominal frequency cutoff in this condition is 1300 Hz, speech frequencies below the cutoff are subject to masking by energy in the filter skirts of the high-pass noise. The restricted access to high speech frequencies could have made this condition especially difficult for the younger listeners. The filter skirt issue may also have been important for the low-pass speech supplementary condition. Here the speech information in the skirts of the filter (above the nominal cut-off of 1300 Hz) may be particularly helpful for children but available only at relatively high SNRs. In these two conditions, the developmental difference could be due in part to the differences in spectra of the speech and pink noise masker and in part to masking in the transition region between the low and high band.

The idea that children may not have been able to attain criterion performance until higher SNRs where higher-frequency speech cues became available is broadly consistent with previous research showing that children require a greater bandwidth than adults for criterion speech performance (Eisenberg *et al.*, 2000; Mlot *et al.*, 2010). While increasing the SNR would also increase the audible bandwidth in the high band AM and high-pass speech conditions, inclusion of low-frequency cues could be less relevant to the discrimination of WIPI words, given that all but one set of words share a vowel and differ in terms of consonants. This raises the possibility that a different pattern of results might be obtained for stimuli relying on predominantly low-frequency cues.

E. Relation between the present results and effects related to SNR

Recent investigations have provided evidence that reduced masking release for speech in temporally modulated noise is expected in cases where listeners require a relatively high SNR in steady noise for criterion performance (Bernstein and Grant, 2009; Bernstein and Brungart, 2011). This interpretation is consistent with findings from a previous filtered speech study by Oxenham and Simonson (2009), indicating that normal-hearing listeners showed relatively small benefit from masker fluctuation for filtered speech conditions where the SNR at SRT was relatively high in a steady noise baseline condition. It has been suggested that the effect of SNR on masking release may be related to the non-uniform distribution of speech cues as a function of intensity (Freyman *et al.*, 2008; Bernstein and Grant, 2009; Bernstein and Brungart, 2011). Smits and Festen (2013) have also reported data consistent with the idea that decreases in masking release are predictable from the SNR required in steady

noise. Although this line of reasoning has generally been associated with effects related to hearing impairment, it is relevant in other situations where listener groups differ in the SNR for the baseline steady noise condition, such as found in studies comparing adults and children (e.g., Stuart, 2005; Stuart *et al.*, 2006; Stuart, 2008; Hall *et al.*, 2012; Wroblewski *et al.*, 2012). While the present experiment was not designed to examine masking release effects related to SNR, it is possible that an SNR effect made some contribution in conditions where masking release was reduced in children. However, it is worth noting that there were two outcomes where the difference in SNR between children and adults was similar between modulated and unmodulated noise. One was the asynchronous modulation masking condition, where the three age groups did not differ in masking release even though the SRT in steady noise indicated a developmental difference of approximately 6 dB. The other condition that showed no developmental effect for masking release was the high-only modulated noise condition. A tentative conclusion is that although masking release in children may be influenced by factors related to the SNR in the baseline (steady) condition, other developmental considerations related to the processing of spectrally and temporally distributed speech and to the speech spectrum may have dominated the pattern of results obtained in the present study.

V. CONCLUSIONS

- (1) The children tested in this study showed reduced masking release for WIPI words in synchronously modulated pink noise relative to adults, but there was no age effect in masking release for asynchronously modulated noise. One possible interpretation of these results is that children are poorer than adults in processing speech glimpses that are temporally separated by noise intervals that contain little or no speech information (synchronous modulation), but that they have less difficulty when all temporal intervals provide at least partial speech information (asynchronous modulation).
- (2) Results from modulated noise control conditions and steady noise supplementary conditions were consistent with an interpretation that children are poorer than adults in processing the WIPI words when provided with cues below 1300 Hz, but they were more adult-like when provided with cues above 1300 Hz. We note the 1300-Hz cutoff used here does not represent a sharp division because the skirts of the filters used to generate the two bands overlapped considerably.
- (3) In the present experiment, the audible speech bandwidth increased as the SNR increased. It is therefore possible that part of the reason that the children had higher SRTs than adults is related to past findings indicating that children require larger speech bandwidth than adults for criterion speech performance. This effect could have been larger in conditions associated with predominantly low-frequency cues due to the fact that children would gain access to higher-frequency regions of the WIPI words only at relatively high SNRs.

ACKNOWLEDGMENTS

Patricia Stelmachowicz and Ryan McCreery provided helpful comments on an early version of this manuscript. This work was supported by Grant No. NIH NIDCD R01DC00397.

- ANSI (2010). S3.6, *Specification for Audiometers* (American National Standards Institute, New York).
- Bernstein, J. G., and Brungart, D. S. (2011). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio," *J. Acoust. Soc. Am.* **130**, 473–488.
- Bernstein, J. G., and Grant, K. W. (2009). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **125**, 3358–3372.
- Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W. (2009). "Masking release for words in amplitude-modulated noise as a function of modulation rate and task," *J. Acoust. Soc. Am.* **126**, 269–280.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- 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., 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.
- Elliott, L. L., and Katz, D. (1980). "Development of a new children's test of speech discrimination," Technical Manual (St. Louis, MO).
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2008). "Spatial release from masking with noise-vocoded speech," *J. Acoust. Soc. Am.* **124**, 1627–1637.
- 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.
- Howard-Jones, P. A., and Rosen, S. (1993). "Uncomodulated glimpsing in 'checkerboard' noise," *J. Acoust. Soc. Am.* **93**, 2915–2922.
- Kirk, R. E. (1968). *Experimental Design: Procedures for the Behavioral Sciences* (Wadsworth, Belmont, CA), pp. 179–182.
- Li, N., and Loizou, P. C. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165–1172.
- 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.
- Nittrouer, S. (1996). "The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM," *J. Speech Hear. Res.* **39**, 1059–1070.
- Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," *J. Acoust. Soc. Am.* **125**, 457–468.
- Ross, M., and Lerman, J. (1970). "A picture identification test for hearing-impaired children," *J. Speech Hear. Res.* **13**, 44–53.
- Smits, C., and Festen, J. M. (2013). "The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners. II. Fluctuating noise," *J. Acoust. Soc. Am.* **133**, 3004–3015.
- Stuart, A. (2005). "Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise," *Ear Hear.* **26**, 78–88.
- Stuart, A. (2008). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children," *J. Am. Acad. Audiol.* **19**, 135–146.
- Stuart, A., Givens, G. D., Walker, L. J., and Elangovan, S. (2006). "Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise," *J. Acoust. Soc. Am.* **119**, 1946–1949.
- Wroblewski, M., Lewis, D. E., Valente, D. L., and Stelmachowicz, P. G. (2012). "Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults," *Ear Hear.* **33**, 731–744.