Factors affecting the development of speech recognition in steady and modulated noise

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology—Head & Neck Surgery, University of North Carolina at Chapel Hill, 170 Manning Drive, Chapel Hill, North Carolina 27599-7070, USA

(Received 10 December 2015; revised 18 April 2016; accepted 4 May 2016; published online 24 May 2016)

This study used a checkerboard-masking paradigm to investigate the development of the speech reception threshold (SRT) for monosyllabic words in synchronously and asynchronously modulated noise. In asynchronous modulation, masker frequencies below 1300 Hz were gated off when frequencies above 1300 Hz were gated on, and vice versa. The goals of the study were to examine development of the ability to use asynchronous spectro-temporal cues for speech recognition and to assess factors related to speech frequency region and audible speech bandwidth. A speech-shaped noise masker was steady or was modulated synchronously or asynchronously across frequency. Target words were presented to 5–7 year old children or to adults. Overall, children showed higher SRTs and smaller masking release than adults. Consideration of the present results along with previous findings supports the idea that children can have particularly poor masked SRTs when the speech and masker spectra differ substantially, and that this may arise due to children requiring a wider speech bandwidth than adults for speech recognition. The results were also consistent with the idea that children are relatively poor in integrating speech cues when the frequency regions with the best signal-to-noise ratios vary across frequency as a function of time.

© *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4950810]

[EAS] Pages: 2964–2969

I. INTRODUCTION

Many studies have demonstrated that speech recognition is generally better in a temporally modulated masking noise than in steady noise (e.g., Miller and Licklider, 1950; Gustafsson and Arlinger, 1994). Recent investigations have indicated that the speech recognition advantage in modulated noise is smaller in school-age children than in adults (Hall *et al.*, 2012; Hall *et al.*, 2014). One interpretation of this finding is that children may have more difficulty than adults in correctly identifying speech on the basis of fragments or "glimpses" that are available in the envelope minima of modulated masking noises (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Assmann and Summerfield, 2004; Buss *et al.*, 2004; Cooke, 2006; Hall *et al.*, 2008).

A study by Hall *et al.* (2014) examined developmental effects for monosyllable words when the masker envelope minima were synchronous versus asynchronous across frequency, using the checkerboard masking paradigm introduced by Howard-Jones and Rosen (1993). A strength of this paradigm is that the asynchronous modulation condition provides information about the ability of a listener to integrate speech cues when the frequency regions with the best signal-to-noise ratios (SNRs) vary across frequency as a function of time, a situation that occurs in many natural environments. The baseline for all measures of masking release was the speech reception threshold (SRT) in steady noise. Synchronous modulation was achieved by gating a noise masker on and off using a quasi-square wave envelope, with 5-ms ramps smoothing onset and offset, and a frequency of 10 Hz. In asynchronous modulation, the masker spectrum was divided into a low band (low-pass filtered at 1300 Hz) and a high band (high-pass filtered at 1300 Hz), and these two bands were modulated out of phase with respect to each other. The filter roll-off was 36-dB/oct. In this condition, the listener had an opportunity to base speech recognition on asynchronous speech cues, occurring in modulation minima of the alternating low and high spectral regions. Two control conditions were also tested in order to determine whether the listener might simply use only the low-frequency region or only the high-frequency region for speech recognition in the asynchronous condition. In the control conditions, the low-pass masker was on continuously and the high-pass masker was gated on and off, or the high-pass masker was on continuously and the low-pass masker was gated on and off. The results from the main conditions indicated that children 5–10 years of age achieved less masking release than the adults in the synchronous masking condition, but approximately the same masking release as adults in the asynchronous masking condition.

Although synchronous and asynchronous masking results were the main focus in the Hall *et al.* (2014) study, interesting developmental differences also occurred in the control conditions. Here, the adults and children achieved about the same masking release when the high band was modulated and the low band was steady, but adults showed much greater masking release than the children when the low band was modulated and the high band was steady. Two explanations were considered for the poor performance of the children in the condition where the low band was

---

[a] Electronic mail: jwh@med.unc.edu

modulated. One was that, in comparison to adults, the children weighted the speech information below 1300 Hz less than the information above 1300 Hz. There has been previous conjecture that children may give preferential weight to higher frequency speech regions (McCreery and Stelmachowicz, 2011), and there is more general evidence that the weighting of different speech cues changes during development (e.g., Nittrouer, 1996). The second explanation concerned the spectra of the masking noise and the speech used in the study. The masker was pink noise, the same masker used in previous checkerboard masking studies (Howard-Jones and Rosen, 1993; Ozmeral *et al.*, 2012). The spectrum of pink noises rolls off at 3 dB/octave, a shallower function than for the speech stimuli, which rolled off at approximately 10 dB/octave above approximately 1000 Hz (see Fig. 1). Hall *et al.* (2014) speculated that the adults may have been able to perform well in the control condition where only the low band was modulated by using only relatively low-frequency speech cues, but that the children may have required additional, higher-frequency cues, which would have necessitated increased signal levels. The relationship between speech level and audible bandwidth in that study was affected by the spectral mismatch between the target words and masking noise. Because the pink noise rolled off more gradually as a function of frequency than the target words, as SNR increased, the low-frequency regions became audible before higher-frequency regions. Audible bandwidth was also dependent on signal level when the bands were filtered and modulated, due to the 36-dB/oct filter roll-off. For example, when the low band was modulated, energy from the unmodulated high band was present in the low-frequency region, particularly near the filter cutoff. The spectral shapes of the speech and masker used in the Hall *et al.* (2014) study are shown in the lower panel of Fig. 1. The idea that the children may have required a broader frequency range than adults to recognize speech is consistent with previous research (Eisenberg *et al.*, 2000; Mlot *et al.*, 2010).

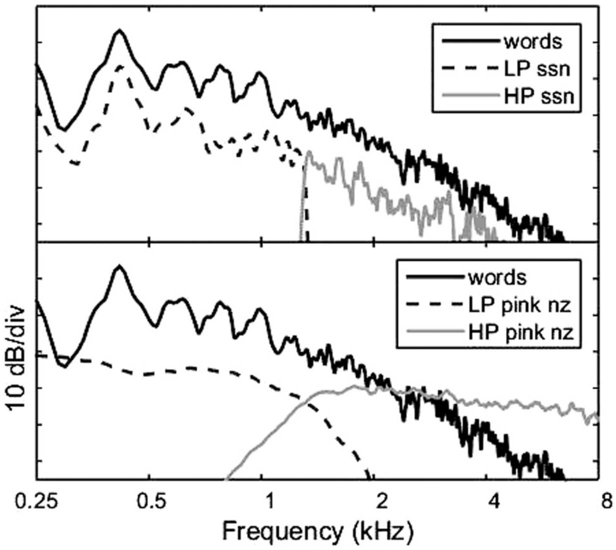

FIG. 1. The top panel shows the spectrum of the target words. Also shown is the spectrum of the speech-shaped noise (ssn) masker along with the filter shape for the low-pass and high-pass skirts associated with the asynchronous modulation condition and the two control conditions. The bottom panel again shows the spectrum of the target words and additionally shows the pink noise masker spectrum used in Hall *et al.* (2014) for comparison.

One goal of the present study was to bring better understanding to the question of whether children are poor with respect to adults when speech cues are relatively low in frequency, or whether developmental differences arise because children require a larger speech bandwidth to achieve the same level of performance. In this study, we again employed the checkerboard-masking paradigm, but with the important difference that the masker had the same spectral shape as the target speech. If children are relatively poor in processing the low-frequency cues of the speech stimuli, the pattern of results in the checkerboard-masking paradigm should be similar to that previously obtained in Hall *et al.* (2014). However, if the results of Hall *et al.* were driven by factors related to audible bandwidth, we would expect the developmental difference between the two control conditions of the checkerboard-masking paradigm to be reduced or eliminated. We would also expect the baseline condition results (the SRT in steady noise) to show a smaller developmental difference than the 6-dB difference found in Hall *et al.* (2014) for children 5–7 years of age. This is because, with pink noise, the higher frequency components of the speech would be less audible than lower frequency components. If children are less able than adults to achieve threshold on the basis of restricted signal bandwidths associated with relatively low-frequency speech in pink noise, they may show less disadvantage in speech-shaped noise where the speech signal should be more consistently audible over a wider range of frequencies.

A second goal of this study was to examine development of the ability to integrate speech cues when the frequency regions with the best SNRs vary as a function of time. Hall *et al.* (2012) investigated speech recognition in a speech-shaped noise that was both temporally and spectrally modulated. The spectral modulation was imposed via notches placed in several frequency regions of the noise masker. In this condition, listeners had access to the whole spectrum of speech during temporal minima but only parts of the spectrum during temporal envelope maxima. Children 4.6 to 6.9 years of age performed more poorly than adults in this condition. It was noted that one possible interpretation of this result is that the ability to process fragments of the speech stimulus that are distributed across both time and frequency is relatively poor in young children. The pink-noise checkerboard masking results of Hall *et al.* (2014) did not show an apparent deficit in the ability of children to integrate asynchronous speech cues. A goal of the present study was to examine this question further using a speech-shaped noise masker.

II. METHODS

A. Subjects

Eight children 5–7 years of age and eight adults were tested. The mean ages for the two groups were 6.0 and 27.2 years, respectively. Listeners had audiometric thresholds that were 20 dB hearing level (ANSI, 2010) or better for octave frequencies from 250 to 8,000 Hz. Adults and children had similar audiograms, with the largest difference between groups being 3.9 dB at 250 Hz. The listeners had no history of otitis media within the previous three years.

B. Stimuli

The speech target stimuli, monosyllabic words spoken by a female talker, were adapted from the Word Intelligibility by Picture Identification, or WIPI, test (Ross and Lerman, 1970). Each presentation of a target word was preceded by the words "show me," spoken by the same female talker. The total duration of the carrier phrase plus target word ranged from 0.98 to 1.65 s, with a mean of 1.18 s. For each trial, the listener's task was to select a picture corresponding to the target word among three foil pictures in a 2×2 matrix displayed on a video monitor. A challenging aspect of this test is that the foils correspond to words that are phonetically similar to the target word, with the task usually depending upon cues related to the initial or final consonants. In all, there were 100 target words contained in 25 sets of four words.

In contrast to previous checkerboard masking studies, the masker was a speech-shaped noise whose spectral shape was derived from the 100 WIPI target words. The masking noise had a level of 68 dB sound pressure level (SPL) before amplitude modulation and filtering were applied. There were three main masking noise conditions: a steady masker; a masker that was synchronously modulated; and a masker that was asynchronously modulated, with the noise above 1300 Hz modulated out of phase with the noise below 1300 Hz. There were also two control conditions: a masker where the noise below 1300 Hz was modulated and the noise above 1300 Hz was steady; and a masker where the noise below 1300 Hz was steady and the noise above 1300 Hz was modulated. Noise filtering was accomplished by converting the stimulus into the frequency domain, reducing component magnitudes to zero outside the passband, and converting the stimulus back into the time domain. The abrupt filter skirts were intended to maximize the SNR associated with the band-stop regions. The spectral shapes of the speech and noise used in this study are shown in the top panel of Fig. 1. Modulation, when present, had a frequency of 10 Hz. This modulation rate was chosen in part because it gives rise to a reliable masking release for adults in all of the checkerboard masking conditions, and is therefore appropriate for the investigation of possible developmental differences in masking release. The modulation had a 50% duty cycle and rapid, 5-ms raised cosine transitions. Masker presentations were 2 s in duration, with the target word temporally centered in the masker.

After Howard-Jones and Rosen (1993), a measure of asynchronous glimpsing was also derived by subtracting the SRT in the asynchronous masking condition from the SRT in the better of the two control conditions.

C. Procedure

Listeners sat in a double-walled sound booth, and SRTs were measured using a four-alternative, forced-choice, adaptive procedure. On each trial, the listener was presented with a randomly selected target word, presented to the right ear via a Sennheiser HD 265 earphone. After the word was presented, four pictures were presented on a video monitor and the listener touched the one corresponding to the perceived sound. All listeners were instructed to guess when unsure. Each correct response resulted in a decrease in signal level and each incorrect response resulted in an increase in signal level. The level adjustments were 4 dB prior to the second track reversal and 2 dB thereafter. After ten reversals in tracking direction, threshold was estimated as the average of the final eight reversals. Thresholds were blocked by condition with an order that was selected pseudo-randomly for each listener.

III. RESULTS

SRTs for the conditions run in this experiment are shown in Table I, and the derived masking release values are shown in Fig. 2. The reference for all values of masking release was the SRT in steady noise.

A. Steady noise reference condition

A t-test was performed to determine whether the adults and children differed for the steady noise condition, with effect size reported as Cohen's *d*. Children had significantly poorer thresholds than adults ($t = 5.1$; $df = 14$; $p < 0.001$; $d = 2.55$). The magnitude of the developmental effect was 2.9 dB.

B. Masking release conditions

A repeated measures analysis of variance (RMANOVA) was performed on the masking release values in the two main experimental conditions, synchronously and asynchronously modulated masking noise. Effect size is reported as partial eta squared (η^2). The RMANOVA showed a significant effect of condition ($F1,14 = 60.71,14$; $p < 0.001$; $\eta^2 = 0.83$), with the synchronous modulation showing larger masking release. The adults showed greater masking release than the children ($F1,14 = 59.42$; $p < 0.001$; $\eta^2 = 0.81$). The interaction between condition and group was not significant ($F1,14 = 1.65$; $p = 0.220$; $\eta^2 = 0.10$). The average masking release in synchronously modulated noise was 21.3 dB for the adults and 12.9 dB for the children. The average masking release in asynchronously modulated noise was 15.2 for the adults and 8.6 dB for the children.

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. Thresholds are shown for the steady noise, synchronously modulated noise (sync AM), asynchronously modulated noise (async AM), modulation of only the low band (low AM only), and modulation of only the high band (high AM only).

Group	Steady noise	Sync AM	Async AM	Low AM only	High AM only
Children	71.0	58.1	62.4	66.5	67.3
	(0.4)	(0.9)	(0.8)	(0.4)	(0.5)
Adults	68.1	46.8	52.9	60.0	62.1
	(0.4)	(1.1)	(1.1)	(1.3)	(1.0)

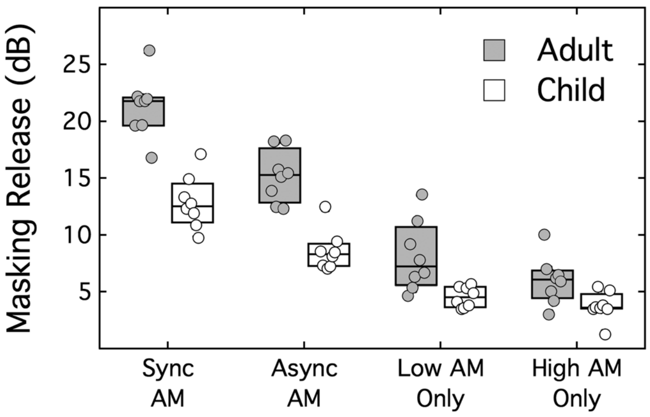

FIG. 2. Box plots showing masking release (dB) for conditions where the masker was synchronously modulated, asynchronously modulated, modulated below 1300 Hz and steady above 1300 Hz, and steady below 1300 Hz and modulated above 1300 Hz. Data are shown for both the adults (gray) and children (white). Horizontal lines indicate the median, boxes indicate the 25th-to-75th percentile range, and circles indicate the individual data.

A RMANOVA was also carried out on masking release in the two control conditions, where either the low band was steady and the high band was modulated, or the low band was modulated and the high band was steady. This analysis did not show a significant effect of condition ($F1,14 = 3.87$; $p = 0.069$; $\eta^2 = 0.22$). However, the effect of group was significant, with adults showing larger masking release than children ($F1,14 = 18.70$; $p = 0.001$; $\eta^2 = 0.57$). The interaction between condition and group was not significant ($F1,14 = 0.77$; $p = 0.400$; $\eta^2 = 0.05$). The average masking release with the low band modulated was 8.1 dB for the adults and 4.5 dB for the children. The average masking release with the high band modulated noise was 6.0 dB for the adults and 3.7 dB for the children.

A t-test was performed to determine whether the magnitude of asynchronous glimpsing differed between the two groups. Asynchronous glimpsing was defined as the difference between the threshold in the asynchronous modulation condition and the threshold in the better of the two control conditions. Note that using the better of the two control conditions is a conservative measure of asynchronous glimpsing, as it introduces a statistical bias in the direction of underestimating asynchronous glimpsing. The t-test ($t = 2.52$; $df = 14$; $p = 0.025$; $d = 1.26$) indicated that the magnitude of asynchronous glimpsing was larger in adults (6.4 dB) than in the children (3.8 dB).

IV. DISCUSSION

A. Issues related to speech bandwidth and speech frequency region

Results from the checkerboard masking study of Hall *et al.* (2014) indicated that children performed more poorly than adults when low-frequency speech cues were preferentially available, but performed more similarly to adults when high-frequency speech cues were preferentially available. One purpose of the present study was to examine whether such effects were likely to be due to developmental differences in the ability to process speech information from different frequency regions, or were instead driven by speech bandwidth effects that occurred because the spectral roll-off of the masking noise was shallower than that of the speech signal. It has been previously shown that children require a wider speech bandwidth than adults to obtain comparable speech identification performance (Eisenberg *et al.*, 2000; Mlot *et al.*, 2010).

The present results provide clear indications that some developmental differences found by Hall *et al.* (2014) in the checkerboard masking paradigm were likely to have arisen due to spectral differences between the speech and masker rather than to a developmental effect related to frequency-specific speech cues. One indication of this can be seen in the results of the control conditions. In the previous study, the adults showed similar performance for the two control conditions. In contrast, children's thresholds were higher when just the low band was modulated than when just the high band was modulated. The average SRTs for the two control conditions differed by 6.1 dB for the 5- to 7-year-old children but by only 0.2 dB for the adults. One explanation for the previous finding that children had a much higher SRT than the adults in the condition where the lower masker frequencies were modulated is based on the spectral slopes of the filtered noise and the fact that the higher frequencies of the speech signal fell off more steeply than the higher frequencies of the pink noise masker. In order to provide the children with sufficient speech bandwidth to attain threshold, the speech level would need to be increased to bring the higher speech frequencies into audibility. This explanation is consistent with the present findings in the control conditions. Specifically, for the present stimuli where the signal and masker were spectrally matched, the children showed much more similar performance in the two control conditions (66.5 dB SPL for the low band modulated and 67.3 dB SPL for the high band modulated). The present findings are consistent with an explanation in terms of the children requiring a larger speech bandwidth than adults, but do not support an interpretation that the children were specifically poor in processing lower-frequency speech information.

Further support for a developmental effect in terms of speech bandwidth is evident in the results of the baseline, steady noise condition. In Hall *et al.* (2014), where the high-frequency roll-off was steeper for the speech than for the pink noise masker, the SRT for the children was approximately 6 dB higher than for the adults. This large difference could be due to the fact that the children required a relatively great increase in signal level in order bring the higher frequencies to a level that could contribute to the total audible speech bandwidth. In support of this interpretation, the results of the present study, where the speech and masker had the same spectral shape, the SRT of the children was only about 3 dB higher than for the adults.

One conclusion that can be drawn in comparing the present work to the previous study of Hall *et al.* (2014) is that developmental differences in masked speech perception can be quite different, depending upon the relative spectra of the speech and masker stimuli. Furthermore, these differences may arise from the fact that children require a larger speech bandwidth than adults to achieve a comparable level of performance (Eisenberg *et al.*, 2000; Mlot *et al.*, 2010). This larger bandwidth requirement possibly reflects a more general need for greater redundancy in the speech signal on the part of children. In line with this, there are indications in the present and previous studies (e.g., Hall *et al.*, 2012) that normal-hearing children can show less benefit than adults from synchronous temporal modulation of the masker, a manipulation that improves audibility of *temporally* sparse segments of the speech signal. As children's linguistic experience increases, tolerance for reduced speech redundancy likely also increases (e.g., Eisenberg *et al.*, 2000; McCreery and Stelmachowicz, 2011; Mlot *et al.*, 2010; Stelmachowicz *et al.*, 2004).

A general expectation arising from developmental speech bandwidth considerations is that developmental differences will be accentuated when the masker and speech spectra differ from each other. This is because in such cases the audible bandwidth broadens with increases in the SNR. In cases where the noise spectrum differs between the speech and the masker, the adults can be expected to take good advantage of parts of the speech spectrum that rise to audibility, but, because children require greater bandwidth, they will require additional signal level to bring that extra bandwidth into audibility. Although many developmental studies of masked speech recognition use masking noise that is filtered to have the same spectral shape as the test speech material it is not uncommon for the masker to have a different spectral shape (Stuart *et al.*, 2006; Gustafson and Pittman, 2011; Nittrouer *et al.*, 2013; Hall *et al.*, 2014).

B. Masking release

Another issue of interest in comparing the present results to those of Hall *et al.* (2014), where a pink noise masker was used, is the magnitude of the masking release for synchronous masking. Typically, experimental features that result in high SNR in the steady noise baseline condition are associated with relatively small masking release in the synchronous modulation condition. Such experimental features include listener group type, such as normal hearing vs hearing impairment (e.g., Wilson and Carhart, 1969; Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Peters *et al.*, 1998; George *et al.*, 2006), or stimulus characteristics such as unfiltered or filtered speech presented to normal-hearing listeners (Oxenham and Simonson, 2009). This association between a high SRT in steady noise and low magnitude of masking release is *not* followed uniformly when comparing our previous results for a pink noise masker (Hall *et al.*, 2014) to the current results for a speech-shaped masker. In the pink noise study, the children had a baseline SRT 6 dB higher than for the adults and a masking release that was 4.8 dB lower than for the adults (15.3 dB for children and 20.1 dB for adults). In the present speech-shaped noise study, the children had a baseline SRT that was only 3 dB higher than for the adults and a masking release that was 8.8 dB lower than for the adults (12.9 dB for children and 21.3 dB for adults). Thus the children showed more adult-like masking release magnitude in the pink noise, where their baseline, steady noise SRT was relatively high. This might suggest that the stimulus features causing the poor baseline threshold for children in pink noise had a less deleterious effect in the modulated noise condition. This could occur because the pink noise baseline condition is particularly difficult for children due to their requirement of a relatively large speech bandwidth at threshold.

C. Asynchronous glimpsing

In the present study, asynchronous glimpsing (the difference between the better of the two control condition SRTs and the SRT for the asynchronous modulation condition) was significantly greater for the adults (6.4 dB) than for the children (3.8 dB). This is in contrast to our previous study (Hall *et al.*, 2014), where the children and adults did not differ significantly in asynchronous glimpsing. This prompted us to re-examine some aspects of our previous data. In the previous dataset, the better of the two control conditions was the same for every child: SRTs for children were uniformly lower when just the high band was modulated than when just the low band was modulated. In contrast, the adults were closely divided in terms of the control condition associated with the better SRT. Since adults' thresholds were very similar in the two control conditions, the differences were likely dominated by measurement variability. Under these conditions, selecting the better of the two SRTs on an individual-by-individual basis would tend to underestimate true performance in the control condition, which would in turn underestimate asynchronous glimpsing in adults. This would tend to reduce the power in examining the possibility of superior asynchronous glimpsing for the adults. We performed a reanalysis of the data from Hall *et al.* (2014), using the condition where the high band was modulated not only as the control condition for the children, but also for the adults. This resulted in a significant difference in asynchronous glimpsing ($t = 2.4$; $df = 18$; $p = 0.027$; $d = 1.14$), with adults showing asynchronous glimpsing of 5.8 dB versus 3.5 dB for children, similar to the present study. As expected, the adults showed a similar magnitude of asynchronous glimpsing (5.9 dB) when the low band modulated condition was used as the control. Overall, these analyses are consistent with the idea that the developmental effects on asynchronous glimpsing were similar in the present dataset and the published study of Hall *et al.* (2014).

Another previous study (Hall *et al.*, 2012) used an entirely different method to explore auditory development of the ability to benefit from speech cues in the context of a spectro-temporally modulated masker. That study used a paradigm introduced by Peters *et al.* (1998) where a noise masker was either steady, spectrally modulated, temporally modulated at 10 Hz, or both temporally and spectrally modulated. Children were particularly poor in the condition where both spectral and temporal modulations were applied to the noise. For this noise, the listeners had access to the entire spectrum of the speech during temporal envelope minima, and had access to frequency-separated regions associated with masker *spectral* minima during the temporal envelope maxima. One conclusion of that study was that children might experience difficulty in integrating speech cues when the frequency regions with the best SNRs vary across frequency as a function of time. The present results on asynchronous glimpsing are consistent with this idea.

V. CONCLUSIONS

Consideration of the present results, along with the results of Hall *et al*. (2014), supports the following conclusions regarding differences between children 5–7 years of age and adults.

(1) Age effects in masked speech recognition depend on the spectral match between the speech signal and noise masker, presumably due to greater requirements with respect to audible bandwidth in younger listeners.
(2) When the masker is a speech-shaped noise, the frequency region associated with masker modulation has a comparable effect on performance of children and adults.
(3) Children are less able than adults to benefit from masker modulation whether that modulation is synchronous or asynchronous across frequency.
(4) The present study found that children demonstrated smaller asynchronous glimpsing than adults. This limit in the ability to integrate speech cues that differ in spectral location as a function of time is consistent with previous results using a different paradigm where masking noise was both temporally and spectrally modulated (Hall *et al*., 2012).

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, Grant NIDCD: R01 DC000397 (J.W.H.). We are grateful for helpful comments from the Associate Editor, Elizabeth Strickland, and two anonymous reviewers.

ANSI (**2010**). S3.6-2010, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).

Assmann, P. F., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer Verlag, New York).

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

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.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Gustafson, S. J., and Pittman, A. L. (**2011**). "Sentence perception in listening conditions having similar speech intelligibility indices," Int. J. Audiol. **50**, 34–40.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.

Hall, J. W., Buss, E., and Grose, J. H. (**2008**). "The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency," J. Acoust. Soc. Am. **123**, 955–962.

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, Buss, E., and Grose, J. H. (**2014**). "Development of speech glimpsing in synchronously and asynchronously modulated noise," J. Acoust. Soc. Am. **135**, 3594–3600.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.

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.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

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.

Nittrouer, S., Caldwell-Tarr, A., Tarr, E., Lowenstein, J. H., Rice, C., and Moberly, A. C. (**2013**). "Improving speech-in-noise recognition for children with hearing loss: Potential effects of language abilities, binaural summation, and head shadow," Int. J. Audiol. **52**, 513–525.

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.

Ozmeral, E. J., Buss, E., and Hall, J. W. (**2012**). "Asynchronous glimpsing of speech: Spread of masking and task set-size," J. Acoust. Soc. Am. **132**, 1152–1164.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Ross, M., and Lerman, J. (**1970**). "A picture identification test for hearing-impaired children," J. Speech Hear. Res. **13**, 44–53.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., and Moeller M. P. (**2004**). "The importance of high-frequency audibility in the speech and language development of children with hearing loss," Arch. Otolaryngol. Head Neck Surg. **130**, 556–562.

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.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.

Wilson, R. H., and Carhart, R. (**1969**). "Influence of pulsed masking on the threshold for spondees," J. Acoust. Soc. Am. **46**, 998–1010.