Spectral integration and bandwidth effects on speech recognition in school-aged children and adults

S. Mlot, **E. Buss**, and **J.W. Hall III**

Abstract

Objectives—Previous studies have shown that adult listeners are more adept than child listeners at identifying spectrally-degraded speech. However, the development of the ability to combine speech information from different frequency regions has received little previous attention. The purpose of the present study was to determine the effect of age on the bandwidth necessary to achieve a relatively low criterion level of speech recognition for two frequency bands, then to determine the improvement in speech recognition that resulted when both speech bands were present simultaneously.

Design—Listeners in the present study included normal-hearing children (6–14 years old, n=18) and adults (n=11). In the first stage of testing, sentences were bandpass-filtered around either 500 or 2500 Hz, and the bandwidth of that filter was varied adaptively to determine the width required for approximately 15–25% correct speech recognition. In the second stage of testing, these criterion bandwidths were presented in fixed block trials with either one band or both bands, and percent correct performance was determined.

Results—Results suggest that age is inversely associated with the bandwidth required to achieve a relatively low criterion level of speech recognition for speech bands centered at either 500 Hz or 2500 Hz. However, both adults and children show a similar, large improvement in performance when both bands are presented simultaneously.

Conclusions—Although younger children require more bandwidth to correctly recognize speech filtered around a single frequency, they appear to be relatively adept at integrating frequency-distributed information to recognize a composite stimulus.

INTRODUCTION

Speech recognition is a complex task, and as such it is unsurprising that performance is age-dependent. While speech perception has been shown to improve as children grow older, the age at which performance asymptotes depends to some extent on the specific speech perception task. For example, whereas speech recognition performance with sentences presented in quiet has been shown to reach adult levels by age 7–8, performance with sentences presented in noise does not reach adult levels until ages 11–15 years, depending on the noise condition (e.g., Stuart 2005; Hnath-Chisolm *et al.* 1998; Elliot *et al.* 1979). In all likelihood, this progressive improvement is due to central rather than peripheral factors. Indeed, other studies have shown that children are poorer than adults at a variety of auditory tasks, from backward masking (Hill *et al.* 2004) to amplitude modulation detection (Hall and Grose 1994), and that this relatively poor performance may be due to inefficient central processing as opposed to peripheral encoding.

The idea that developmental improvements in speech perception are related to central rather than peripheral factors is supported by the physiological evidence that infants are born with

fully-functioning cochleae (Eggermont *et al.* 1996) and that brainstem myelination and physiologic function reach maturity by the perinatal period (Moore *et al.* 2007). Central processing, on the other hand, continues to develop well into childhood. For example, electrophysiological testing and cadaveric analysis show that neuronal maturation within the auditory cortex is not complete until age 11 or 12, and that the N1 component of the cortical evoked potential does not become readily apparent until approximately age 9 (Moore *et al.* 2007; Ponton *et al.* 2000). The combination of these physiological and behavioral data suggest that, while the auditory periphery may reach full functionality soon after birth, central processing of complex auditory signals such as speech perception in noise continues to improve throughout childhood.

Many recent studies on speech perception in adults and children have used a model designed to emulate the listening experience of cochlear implant (CI) users. These listeners make use of a processor that divides the speech spectrum into a small number of frequency bands, extracts the amplitude envelope for each of these bands, and uses those envelopes to modulate the stimulus delivered to each of the implanted electrodes. A similar vocoded-speech method has been widely used to simulate CI stimulation in normal-hearing listeners. By this method, the amplitude envelope of each frequency band is multiplied by a sinusoid or narrow-band noise carrier, with the frequency of the carrier being associated with the spectral region of the original speech band (e.g., Shannon *et al.* 2004). In adults, the number of bands needed for normal-hearing listeners to achieve asymptotic levels of performance in speech recognition closely corresponds to the number of electrodes that hearing-impaired listeners fitted with CIs require to achieve a similar level of performance. Studies have shown that normal-hearing adult listeners require four to six spectral channels of speech information for speech recognition in quiet (Shannon *et al.* 2004), while hearing-impaired adult listeners fitted with CIs can achieve this level of performance with the same number of active electrodes (e.g., Fishman *et al.* 1997). Similarly, both of these cohorts require more spectral channels to achieve asymptotic performance when presented with more difficult listening situations, such as speech in background noise or speech in a second language (Shannon *et al.* 2004).

Eisenberg *et al.* (2000) tested child listeners using noise-vocoded speech and found that young children require significantly more spectral channels of speech information than both older children and adults. Specifically, children aged 5–7 required eight bands of speech information to reach an asymptotic performance of 94% for sentences presented in quiet, while children aged 10–12 and adults required just six bands to reach a performance of 96% correct. Similar results were obtained using a test measuring the ability to detect contrasts in features of speech such as consonant manner, voicing, and place. These data complement previous observations that speech recognition improves with age through adolescence.

In the paradigm of studies such as Eisenberg *et al.* (2000) that emulate CI stimulation, increasing the number of channels also alters the amount and specificity of spectral information embedded in those channels. Thus, this approach fails to address the question of whether the improved performance of speech recognition with age lies in an improved ability to make use of sparse spectral information, or in an improved ability to integrate speech cues distributed across the frequency spectrum. The primary purpose of the present investigation was to test the ability of children to make use of severely spectrally degraded speech cues and to integrate speech information across frequency bands.

Several past studies of normal-hearing adult listeners have shown that the percent correct associated with the simultaneous presentation of two widely-separated, narrow bands of speech information is super-additive compared to the percent correct values associated with each band separately (e.g., Grant and Braida 1991, Warren *et al.* 1995, Lippmann 1996,

Kasturi *et al.* 2002). This finding of non-linear additivity, while unsurprising, has been taken as evidence for the ability of adult listeners to combine non-redundant, complementary information across frequency (Grant and Braida 1991). The present study investigated this ability in both adults and children using the paradigm of Hall *et al.* (2008). In this paradigm, listeners are presented with sentences filtered into single bands centered at either 500 or 2500 Hz, and the widths of these individual bands are varied adaptively to converge on a bandwidth associated with a percent correct performance of approximately 15–25%. These individually tailored bandwidths are referred to as criterion speech bandwidths (CSBs). Listeners are then presented with their low and high frequency CSBs alone and in combination, and percent correct performance is determined. The relatively low criterion of 15–25% was used in determining CSB in order to ensure that performance would be below the ceiling of 100% correct when the low and high frequency bands are presented in combination.

In the present study, we evaluated the effect of age on a) the width of the low and high frequency CSBs, and b) the ability to integrate speech information across frequency bands. Because previous studies have shown improvements in speech recognition through late childhood, the present study examined performance in children between the ages of 6 and 14, with adults representing asymptotic performance. We hypothesized that there would be an effect of age on the bandwidth necessary to support a criterion level of speech recognition performance, at least up to early adolescence, and that children might not be as adept at integrating speech information across frequency-separated bands.

MATERIALS AND METHODS

A. Listeners

In order to ensure a good representation across the relatively wide age range of the present experiment, six children were recruited in each of three age groups: 6–7 years old (3 male, 3 female), 8–11 years old (6 male), and 12–14 years old (4 male, 2 female). Eleven adults were also recruited, aged 19–35 (5 male, 6 female). With one exception, all listeners were screened to have pure-tone thresholds of 20 dB HL or less bilaterally at octave frequencies from 250 to 8000 Hz (ANSI 1996). One listener in the 12–14 year age group had a 4000-Hz pure-tone threshold of 25 dB HL in the non-test ear.

B. Stimuli

The speech stimuli used in this study were Bamford-Kowal-Bench (BKB) sentences (Bench *et al.* 1979). These materials are organized into 21 lists of 16 sentences, with each sentence consisting of three to five key words. The quantity of sentences available was great enough to ensure that no sentence was repeated for any given listener. Depending on the condition, listeners were presented with speech that had been filtered into a single band centered at either 500 or 2500 Hz, or into two simultaneously present bands centered at 500 and 2500 Hz. Each band was generated by passing the full speech stimulus through a finite impulse response filter with 12-Hz resolution. The width of these bands was either varied adaptively or held constant, as described below. Stimuli were 76 dB SPL prior to filtering; that level was reduced by bandpass filtering.

C. Procedure

In all stages of testing, the listener sat in a double-walled sound booth and was presented with bandpass-filtered BKB sentences. The listener was instructed to repeat as many words of each sentence as possible and to guess for words that were not intelligible. The experimenter sat in the sound booth with the listener, positioned in front of a computer monitor displaying the listener's current sentence. Errors were recorded as the listener

responded. The experimenter provided no feedback, aside from minimal encouragement to some of the younger listeners. Listeners were offered frequent breaks. All stimuli were presented monaurally over Sennheiser HD 265 earphones. Stimuli were presented to the left ear in all but one case; the right ear was tested in one listener for whom the left ear did not pass the hearing screening.

In the first stage of testing, BKB sentences were bandpass-filtered to frequencies centered at low or high portions of the speech spectrum. The width of these bands was varied adaptively so that the listener's speech recognition scores generally fell within a narrow range of poor performance (15–25%). Thus, each listener achieved a set of criterion bandwidths tailored to his or her individual speech recognition performance. This approach was used to ensure comparable performance in each single-band condition. Individualized bandwidths were used due to previous data indicating reliable differences between listeners in speech recognition performance for fixed-bandwidth conditions (Noordhoek *et al.* 1999, Hall *et al.* 2008), and also in order to evaluate the hypothesis that children would require significantly more bandwidth than adults.

The adaptive procedure was carried out separately for bands centered at either 500 Hz (low band) or 2500 Hz (high band), as described by Hall *et al.* (2008). The bandwidth was changed adaptively by a factor of 1.21. Bandwidth was increased following two sequential sentences in which no key word was reported correctly, and bandwidth was decreased following every sentence in which any key word was reported correctly. Each run was continued through eight reversals in bandwidth adjustment, and the geometric mean of the bandwidths at the last six reversals was taken as the CSB. Three to four runs were conducted for each condition (low band or high band present), and the geometric means of the thresholds for each frequency band were recorded as the low- and high-band CSB, respectively.

In the second stage of testing, the CSBs were used as fixed bands tailored to each individual listener. The listener underwent block trials to determine percent correct speech recognition for three conditions: the low band alone, the high band alone, and the low and high bands together. Percent correct was calculated as the number of key words correctly identified in a block of trials divided by the total number of key words. The two single-band conditions were performed in order to verify that the CSBs determined in the first stage of the study were in fact associated with speech recognition scores of 15–25% correct. The two-band condition was performed to evaluate spectral integration of speech information. For each of these conditions, three runs consisting of 16 BKB sentences were presented. The percent correct value obtained for each condition was then calculated. Results were transformed to rationalized arcsine units (RAUs) prior to statistical analysis to ensure uniform variance across conditions[i] (Studebaker 1985). Such a transformation is important in analyses of percent correct, as data points close to either floor or ceiling (i.e., <20% or >80%) are associated with less variance than those in the middle of the range.

---

[i]Transformation to 'rationalized arcsine units' (RAUs) was accomplished using the equation,

$$RAU = 1.46 * (31.831 * T - 50) + 50,$$

where

$$T = 2 * \arcsin \sqrt{P},$$

and, *P* represents proportion correct.

RESULTS

Results of the first stage of testing are reported as normalized CSB, which is the bandwidth of the CSB divided by the center frequency (either 500 or 2500 Hz). Figure 1 shows the normalized CSB plotted as a function of age for individual child listeners, with results for the low-frequency band in the left panel and those for the high-frequency band in the right panel. Mean data of adult listeners are plotted at the far right of each panel, with error bars indicating +/− one standard deviation (SD) around that mean. The shaded region bounded by dashed lines provides a visual reference for estimating the age at which child performance reaches that of adults. In both panels there is a clear trend for criterion bandwidth to become narrower with increasing listener age. Simple linear regression analyses revealed this relationship to be statistically significant for both the low band (r^2=0.55, p<0.0001) and the high band (r^2=0.61, p<0.0001). Dark lines in Figure 1 show the associated line fits, with the corresponding equations included in the figure background. These lines intersect the $\pm$ 2 SD region around adult thresholds at 12.5 years for the low band and 12.6 years for the high band. The relationship between the CSBs estimated for child listeners was further assessed using a repeated-measures general linear model (GLM) analysis. There were two levels of band (low and high) and the continuous variable of age. This analysis revealed a main effect of age ($F_{1,16}$=27.26, p<0.001, η_p^2=0.63), a main effect of band ($F_{1,16}$=6.33, p<0.05, η_p^2=0.28), and no interaction ($F_{1,16}$=1.45, p=0.25). These results reflect a significant reduction in CSB as a function of child age for both frequency regions.

Figures 2 and 3 illustrate the relationship between speech recognition performance and age for each of the stimulus conditions in the second stage of testing. The low band-alone and high band-alone data are shown in Figure 2, and the combined band data are shown in Figure 3. Results are shown in RAUs, units that are analogous to percent correct within the range of 10–90%, with increasing deviation from percent correct as values approach ceiling and floor. As in Figure 1, data for individual child listeners are plotted as a function of age, and mean adult data are plotted as a single point to the right of each panel, with error bars indicating +/− one standard deviation around that mean. In contrast to the bandwidth data, results of the second stage of testing do not show any clear dependence on listener age, either within the child group or across groups.

Speech recognition scores represented in RAUs were submitted to repeated-measures ANOVA. There were three levels of band (low, high, both) and two groups (children, adults). This analysis resulted in a main effect of band ($F_{2,54}$=571.78, p<0.0001, η_p^2=0.95), reflecting the better performance for the condition including both the low and high speech bands presented together. The main effect of band was further quantified using a Helmert contrast, wherein results collapsed across the two single-band conditions were contrasted with the two-band results. Performance with both bands present was significantly different from that of the two single-band conditions ($F_{1,27}$=943.73, p<0.0001, η_p^2=0.97), but the two single-band conditions did not differ significantly from each other ($F_{1,27}$=0.64, p=0.43). This latter result verifies that the determination of CSB in the first stage of testing was successful, in that the individual CSBs were associated with a similar percent correct performance at all ages tested. There also was no main effect of group ($F_{1,27}$=1.16, p=0.29), indicating that there was not a significant difference in speech recognition abilities between children and adults when using CSBs. A Helmert contrast was used to assess spectral integration in child as compared to adult data. In this analysis, the interaction between group and number of bands (one vs. two) failed to reach significance ($F_{1,27}$=3.01, p=0.09). Most importantly for the hypothesis of the present study, mean performance for the two-band condition was not better for the adults than children. In fact, the mean percent correct in the two-band condition was better for the children (84.4) than for the adults (74.4), though this difference was not significant.

The analyses of adult versus child data from the second stage of testing demonstrated no significant difference across age groups, but it was also of interest to determine whether performance varied as a function of age within the child group. To answer this question, regression analyses were undertaken on percent correct (in RAUs) as a function of child age. In no case was this association significant, using an experiment-wise α>0.05 and Bonferroni correction for multiple tests. It was hypothesized at the outset of the study that younger children might perform more poorly than older children in the condition where both bands were present; in this condition there was no evidence of an association between child age and performance (r=−0.02, p=0.94). These results support the conclusion of no association between child age and percent correct in the second stage of testing, and in particular no evidence of an impaired ability to integrate speech information across the 500- and 2500-Hz frequency regions.

DISCUSSION

A. Comparison of adult data with previous results

The adult data of the present study can be compared most directly to those of Hall *et al.* (2008), where similar methods were used to derive CSBs for low and high frequencies and to evaluate the effects of band combination. One finding of the present study was that the normalized CSB was significantly smaller for the band centered on 2500 Hz than for the band centered on 500 Hz. This relationship was also found in Hall *et al.*, and is consistent with published estimates of frequency band importance. For example, when CID sentences are assessed for band importance, the third octave band centered on 500 Hz is associated with approximately 5.2% of total speech information, and the band at 2500 Hz with 9.3% (ANSI 1997). The greater importance of the higher frequency region may lie in the fact that it carries information essential for determining consonant place, which is more critical in enabling the listener to discriminate among words than the vowel voicing information carried in the low-frequency regions (e.g., Kasturi *et al.* 2002).

One minor difference between the results of the present study and those of Hall *et al.* (2008) is that the CSBs were slightly narrower in the present study. The mean adult CSB at 500 Hz was 0.34 in the present study versus 0.41 in the previous study; the mean CSB at 2500 Hz was 0.26 in the present study versus 0.35 in the previous study. However, these mean values cannot be directly compared due to differences in stimulus presentation level between the two studies. Specifically, Hall *et al.* (2008) applied a boost in level to the high-frequency band in order to ensure audibility in hearing-impaired listeners and reduce effects of upward spread of masking among hearing-impaired listeners. Since the present study did not test hearing impaired listeners, such a level boost was not applied to the high-frequency band. There also exists the possibility that the difference between studies represents an effect related to age among *adult* listeners, as the listeners in the previous study were approximately a decade older than the adults in the present study. Although such an age effect might seem unlikely, Grose *et al.* (2006) have shown that auditory processing deficits can begin in middle age, and it is possible that similar effects could extend to the kind of speech processing investigated here.

The other speech perception effect measured in this study was the ability to combine information from frequency-separated bands. The results of several studies of speech perception in adult listeners have shown that although relatively narrow bands of speech result in a very low level of speech recognition performance, the percent correct obtained when multiple narrow bands are presented simultaneously often greatly exceeds the sum of percent correct associated with the individual bands (e.g., Grant and Braida 1991, Warren *et al.* 1995, Lippmann 1996, Kasturi *et al.* 2002). The present results replicate this finding. They also agree closely with the results of Hall *et al.* (2008), where individual bands were

associated with approximately 15–25% correct, and both bands together were generally associated with performance of better than 70% correct.

B. Developmental data

Results from the first stage of testing clearly indicated that the CSB varied as a function of listener age for both low- and high-frequency bands. Specifically, younger children needed more bandwidth than either older children or adults to achieve a speech recognition performance of approximately 15–25% correct (Figure 1). For both the low and high bands, the width of the CSB narrowed with increasing age through approximately 12.5 years of age, at which point the magnitude appeared to approach that of adults. These data are consistent with the results of previous studies, such as Eisenberg *et al.* (2000), which used noise-vocoded speech to show that children require more spectral bands of speech information to achieve maximal speech recognition performance until the age of 10–12, when the performance of children equaled that of adults. There are important stimulus differences between the vocoded speech method and the paradigm of the present experiment; notably, the stimuli used here preserved temporal fine structure and spectral detail within the stimulus passband, whereas both temporal and spectral features are degraded in the vocoded speech method. Despite these differences, results from the first stage of testing can be interpreted as corroborating the broad conclusion of the Eisenberg study -- namely, that children require more spectral detail than adults to identify speech at a low criterion level of performance.

Although the present and past results indicate that children require greater spectral detail than adults to accurately recognize speech, the underlying reasons for this are not yet clear. The hypothesis examined in the present study is related to the ability of listeners to combine information across auditory frequency channels. Two results from the second stage of testing are consistent with an interpretation that the ability to combine across-frequency speech information is similar for all age groups tested: mean performance for each group was approximately 80–85% correct when presented with low- and high-frequency bands simultaneously and, while some listeners performed outside of this range, performance was not significantly correlated with age among the child listeners.

Another potential possible explanation for the bandwidth effect found in the first stage of testing could be a differential ability to make use of sentence context in children and adults. For example, adults might make better use of contextual cues to identify key words that otherwise would have been difficult to recognize. This type of context effect would not have affected calculation of CSB, since the bandwidth estimation procedure did not discriminate between correctly reporting one keyword versus more than one keyword in a sentence. If contextual cues were more effective for adults than children, then identifying one word correctly might increase the probability of correctly identifying other words in the sentence, particularly for adult listeners. Such a group difference in conditional probabilities would not be reflected in CSB estimates, but it *would* have affected the percent correct upon which the adaptive procedure converged. If, indeed, the adaptive procedure converged upon a different percent correct for the children versus adults, an effect of age would be expected in the second stage of testing where percent correct was determined at the CSBs estimated in stage 1. The finding of no significant difference in percent correct between adults and children when the bands were presented individually undermines an interpretation that the adaptive strategy converged upon different levels of percent correct for the different age groups. To consider this question further, we performed a post hoc analysis of the number of keywords correctly recognized in each sentence for single band conditions in the second stage of testing. This analysis showed nearly identical error patterns between age groups; adults and children correctly recognized all words in only 6% of sentences and missed all words in 60% and 56% of sentences, respectively. This result indicates that the conditional

probabilities of correctly reporting words in context were consistent across groups, casting doubt on the idea that differential ability to make use of sentence context is responsible for the age effects observed.

The effect of age on CSB may also have been related to the well-known effect of age on attention and motivation. This is particularly a concern when the task is adaptive as opposed to fixed, as several investigators have argued that attention may often be correlated with signal level (Schneider and Trehub, 1992; Viemeister and Schlauch, 1992; Allen and Wightman, 1994). In the first stage of testing, during the course of the adaptive track it is likely that the listener would be presented with several trials in a row in which the signal would be incomprehensible. These very difficult trials might lead to a lapse in attention or motivation in child listeners. In such cases, the bandwidth might need to be widened several steps above the "true" threshold in order for the listener to become re-engaged in the task.

Two pieces of evidence argue against the interpretation that lapses in attention or motivation elevated the CSB estimated in children compared to adults. First, the second stage of testing showed that, when presented with the low and high bands alone, children performed similarly to adults in terms of percent correct. If inattention were a problem for children in the adaptive task, then the criterion bandwidth tracked in the first stage of testing would have been elevated, resulting in a higher-than-expected percent correct when child listeners were presented with that criterion bandwidth as a fixed band. As has been previously discussed, however, there was no significant difference between the performance of children and adults in the single band conditions of the second stage of testing. The second piece of evidence is related to the variability of performance within the adaptive tracks. Variable attention within an adaptive track should result in heightened track variability. This was examined via a post-hoc analysis of the variability in each listener's performance among the trials of the adaptive tracks. This analysis was performed by computing the standard deviation of the log of the bandwidth at the final six reversals of each track; the resulting values of standard deviation were comparable for child and adult groups at the median (0.23 and 0.22), as well as the 10th (0.14 and 0.16) and 90th (0.31 and 0.34) percentiles. The finding of nearly identical track variability for child and adult listeners undermines an interpretation that the group differences in CSB were strongly related to group differences in attention. Although the above analyses fail to show evidence for a developmental effect based on attention, we cannot rule out the possibility that attention in children could have been reduced in conditions associated with low percent correct (in both the adaptive and fixed conditions), and that this could have influenced performance.

Another possible account of the increased bandwidth required for speech recognition in child listeners is related to speech and language experience. While the sentences used in this study employed vocabulary that is easily identifiable by children, it may be the case that since adults are more experienced in using and identifying speech, they are better able to correctly identify spectrally-degraded speech stimuli. The task of recognizing a degraded speech signal can be regarded as a process of comparing ambiguous acoustic information with a set of likely utterances, where that set of utterances is constrained by context cues and general linguistic knowledge. The experience with speech processing that comes with age, particularly in conditions where the signal-to-noise ratio is poor, presumably would be beneficial to the process of sifting through candidate words and choosing a response that best matches the ambiguous information that is available. This possibility was not fully examined in this study, but may partially account for the obtained results.

In summary, the results of the present study are convergent with other data showing that children require more spectro-temporal detail than adults to accurately recognize speech, and they provide new evidence demonstrating that children and adults are equally able to

integrate widely separated spectral information. Because there are important differences between the stimuli used here and those used with cochlear implants, it is not possible to draw definitive conclusions about the implications of the present results for the ability of children to utilize the kinds of speech cues provided with cochlear implant use. However, the results of the present study suggest that limitations in speech perception by such children may be unlikely to arise from a fundamental deficit in the ability to combine speech information across frequency channels. The present results may also be relevant to the broader issue of speech perception by children under conditions where the frequency regions containing speech information are limited and spectrally separated. This could occur in the research lab where, for example, speech is masked by a noise with multiple spectral notches, or in natural environments where dynamic variations in background noise create spectral gaps that are separated in frequency but occur roughly synchronously in time. The present results would suggest that children might not be limited by a deficit in the ability to combine across-frequency cues under such circumstances.

CONCLUSIONS

1. The percent correct for two frequency-separated speech bands was substantially better than the sum of percent correct for the bands presented separately. This effect was similar for adults and children. Furthermore, the ability to combine spectral information from these frequency-separated bands did not vary with age in the range of 6–14 years.

2. Compared to adults, children required a wider band of a given frequency region to achieve a criterion speech recognition performance. Children attained performance comparable to that of adults in this task by approximately 12.5 years of age.

3. Both children and adults required a wider normalized criterion speech bandwidth at the low frequency (500 Hz) tested in this study in comparison to the high frequency (2500 Hz).

References

American National Standards Institute. American National Standards Specification for Audiometers (ANSI S3-1996). New York: ANSI; 1996.

American National Standards Institute. American National Standards Methods for Calculation of Speech Intelligibility Index (ANSI S3.5-1997). New York: ANSI; 1997.

Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br J Audiol. 1979; 13:108–112. [PubMed: 486816]

Eggermont JJ, Brown DK, Ponton CW, et al. Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation. Ear Hear. 1996; 17:386–394. [PubMed: 8909886]

Eisenberg LS, Shannon RV, Schaefer Martinez A, et al. Speech recognition with reduced spectral cues as a function of age. J Acoust Soc Am. 2000; 107:2704–2710. [PubMed: 10830392]

Elliott LL. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. J Acoust Soc Am. 1979; 66:651–653. [PubMed: 489836]

Fishman K, Shannon RV, Slattery WA. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J Speech Lang Hear Res. 1997; 40:1201–1215. [PubMed: 9328890]

Grant KW, Braida LD, Renn RJ. Single band amplitude envelope cues as an aid to speechreading. Q J Exp Psychol. 1991; 43A:621–645.

Grose JH, Hall JW, Buss E. Temporal processing deficits in the pre-senescent auditory system. J Acoust Soc Am. 2006; 119(4):2305–2315. [PubMed: 16642844]

Hall JW, Buss E, Grose JH. Spectral integration of speech bands in normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2008; 124:1105–1115. [PubMed: 18681600]

Hall JW, Grose JH. Development of temporal resolution in children as measured by the temporal modulation transfer function. J Acoust Soc Am. 1994; 96:150–154. [PubMed: 7598757]

Hill PR, Hartley DE, Glasberg BR, Moore BC, Moore DR. Auditory processing efficiency and temporal resolution in children and adults. J Speech Lang Hear Res. 2004; 47:1022–1029. [PubMed: 15603460]

Hnath-Chisolm TE, Laipply E, Boothroyd A. Age-related changes on a children's test of sensory-level speech perception capacity. J Speech Lang Hear Res. 1998; 41:94–106. [PubMed: 9493737]

Kasturi K, Loizou PC, Dorman M, et al. The intelligibility of speech with 'holes' in the spectrum. J Acoust Soc Am. 2002; 112:1102–1111. [PubMed: 12243158]

Lippmann R. Accurate consonant perception without mid-frequency speech energy. IEEE Trans Speech Audio Proc. 1996; 4:66–69.

Moore JK, Linthicum FH Jr. The human auditory system: A timeline of development. Int J Audiol. 2007; 46:460–478. [PubMed: 17828663]

Noordhoek IM, Houtgast T, Festen JM. Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I Normal-hearing listeners. J Acoust Soc Am. 1999; 107:1685–1696. [PubMed: 10738821]

Ponton CW, Eggermont JJ, Kwong B, et al. Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. Clin Neurophysiol. 2000; 111:220–236. [PubMed: 10680557]

Shannon RV, Fu QJ, Galvin J. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. Acta Otolaryngol Suppl. 2004; 552:50–54. [PubMed: 15219048]

Stuart A. Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. Ear Hear. 2005; 26:78–88. [PubMed: 15692306]

Studebaker GA. A 'rationalized' arcsine transform. J Speech Hear Res. 1985; 28:455–462. [PubMed: 4046587]

Warren RM, Riener KR, Bashford JA Jr, et al. Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. Percept Psychophys. 1995; 57:175–182. [PubMed: 7885815]
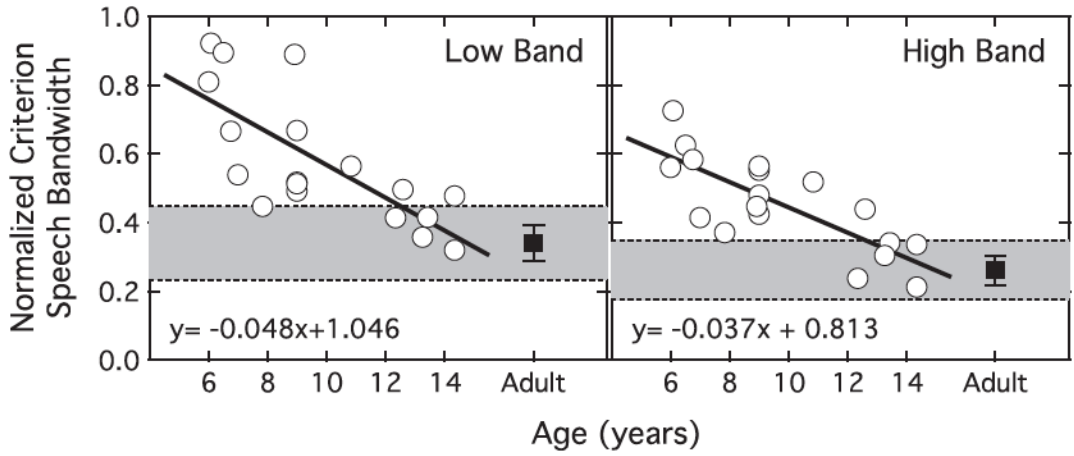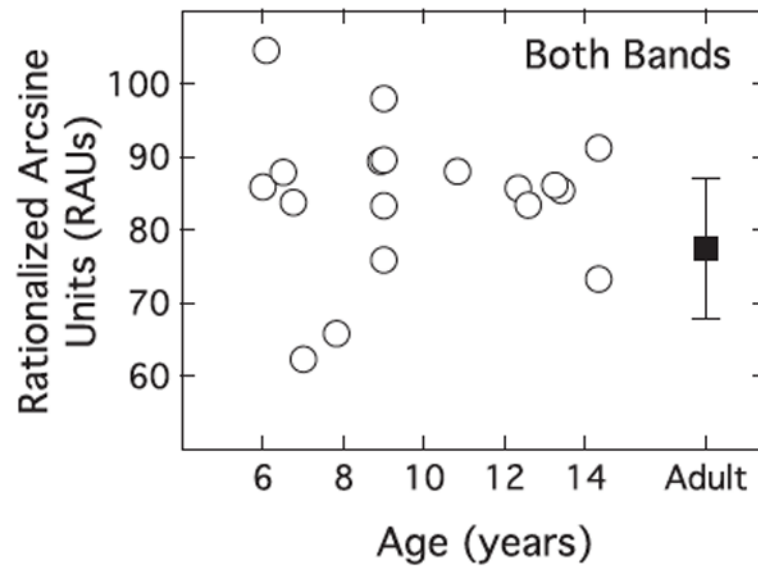
Figure 1.
Normalized criterion speech bandwidth (CSB) as a function of age for individual child listeners between the ages of 6 and 14 years. Results for the low-frequency band are in the left panel, and results for the high-frequency band are in the right panel. The mean normalized CSB for adult listeners is plotted as a filled square on the right side of each panel, with error bars demarcating one standard deviation above and below this mean. The shaded region indicates ± 2 standard deviations around the adult mean. The dark line in each panel characterizes the relationship between bandwidth and child age, calculated using a least squared method; correlations between CSB and age are r=−0.74 and r=−0.78 for the low- and high-frequency bands, respectively.
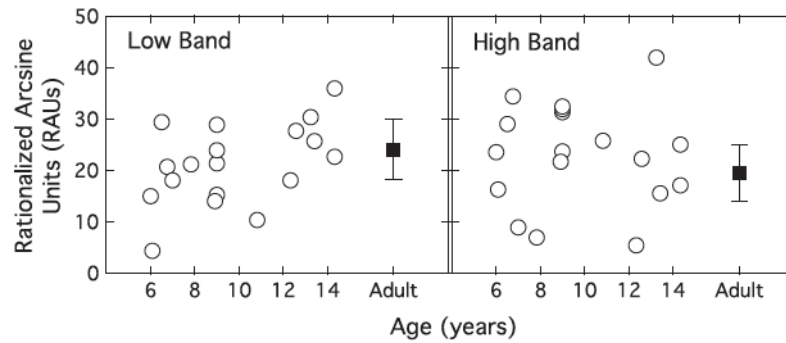
Figure 2.
Speech recognition in rationalized arcsine units (RAUs) for the low and high bands alone as a function of age for individual child listeners between the ages of 6 and 14 years. Results for the low-frequency band alone are in the left panel and results for the high-frequency band alone are in the right panel. The mean of RAU data for adult listeners is plotted as a filled square on the right side of both panels, with error bars demarcating one standard deviation above and below this mean.

Figure 3.
Speech recognition in RAUs for both bands presented together. The mean of RAU data for adult listeners is plotted as a filled square on the right side of the panel, with error bars demarcating one standard deviation above and below this mean.