HHS Public Access

Author manuscript

Ear Hear. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Ear Hear. 2016 ; 37(1): 55–63. doi:10.1097/AUD.0000000000000201.

Development of Open-Set Word Recognition in Children: Speech-Shaped Noise and Two-Talker Speech Maskers

Nicole E. Corbin[1], **Angela Yarnell Bonino**[1], **Emily Buss**[2], and **Lori J. Leibold**[1]

[1]Department of Allied Health Sciences, Division of Speech and Hearing Sciences, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC, USA

[2]Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC, USA

Abstract

Objective—The goal of this study was to establish the developmental trajectories for children's open-set recognition of monosyllabic words in each of two maskers: two-talker speech and speech-shaped noise.

Design—Listeners were 56 children (5 to 16 yrs) and 16 adults, all with normal hearing. Thresholds for 50% correct recognition of monosyllabic words were measured in a two-talker speech or a speech-shaped noise masker in the sound field using an open-set task. Target words were presented at a fixed level of 65 dB SPL throughout testing, while the masker level was adapted. A repeated-measures design was used to compare the performance of three age groups of children (5 to 7 yrs, 8 to 12 yrs, and 13 to 16 yrs) and a group of adults. The pattern of age-related changes during childhood was also compared between the two masker conditions.

Results—Listeners in all four age groups performed more poorly in the two-talker speech than the speech-shaped noise masker, but the developmental trajectories differed for the two masker conditions. For the speech-shaped noise masker, children's performance improved with age until about 10 years of age, with little systematic child-adult differences thereafter. In contrast, for the two-talker speech masker, children's thresholds gradually improved between 5 and 13 years of age, followed by an abrupt improvement in performance to adult-like levels. Children's thresholds in the two masker conditions were uncorrelated.

Conclusions—Younger children require a more advantageous signal-to-noise ratio than older children and adults to achieve 50% correct word recognition in both masker conditions. However, children's ability to recognize words appears to take longer to mature and follows a different developmental trajectory for the two-talker speech masker than the speech-shaped noise masker. These findings highlight the importance of considering both age and masker type when evaluating children's masked speech perception abilities.

Name and address for correspondence: Nicole E. Corbin, Department of Allied Health Sciences, 3122 Bondurant Hall, CB #7190, Chapel Hill, NC 27599, USA. Telephone: (919) 843-3672, Fax: (919) 966-0100, nicole_corbin@med.unc.edu.

Conflicts of Interest

No conflicts of interest were declared.

Portions of this article were presented at the 41st Annual Scientific and Technology Meeting of the American Auditory Society, Scottsdale, AZ, March 7, 2014.

INTRODUCTION

The goal of this study was to establish the time course of development for children's susceptibility to speech-on-speech masking for an open-set word recognition task. Children are at a considerable disadvantage relative to adults when listening to speech in the presence of most background sounds (e.g., Elliott et al. 1979; Nittrouer & Boothroyd 1990; Hall et al. 2002). However, the performance gap between children and adults appears to be larger and lasts longer when the background is competing speech compared to when the background is relatively steady-state noise (e.g., Hall et al. 2002). These differences in age effects between masking conditions are of considerable interest to researchers who study auditory development, in part because the auditory processes responsible for masking with speech and noise backgrounds are thought to be different. At least for adults, steady-state noise maskers are often assumed to produce primarily *energetic* masking (Fletcher 1940) due to overlapping excitation patterns on the basilar membrane. In the context of masked speech recognition, this sensory overlap reduces the fidelity with which the target and masker stimuli are represented by the peripheral auditory system, reducing the audibility of the target speech. Speech maskers are believed to be particularly challenging because they produce substantial *informational* masking in addition to energetic masking (e.g., Carhart et al. 1969; Brungart 2001; Freyman et al. 2004; Brungart et al. 2006). Informational masking is thought to reflect a listener's reduced ability to segregate and/or selectively attend to target versus masker speech (e.g., Brungart 2001; Freyman et al. 2004; Brungart et al. 2006). Informational masking effects for speech recognition tasks are most pronounced when the masker is competing speech composed of a small number of talkers (e.g., Carhart et al. 1969; Brungart et al. 2001; Freyman et al. 2004), decreasing as the number of talkers increases. Less informational masking is produced as more talkers are added; presumably because the acoustic waveform of the resulting masker babble is less confusable with target speech, resulting in primarily energetic masking (e.g., Freyman et al. 2004).

It is well established that children require a more advantageous signal-to-noise ratio (SNR) than adults to achieve similar performance for speech recognition in the presence of relatively steady-state noise (e.g., Elliott et al. 1979; Nittrouer & Boothroyd 1990; Hall et al. 2002; Wightman & Kistler 2005; Nishi et al. 2010). Consistent results have been reported by multiple laboratories using different stimuli and measures, with most children achieving adult-like performance by at least 10 years of age (e.g., Elliot et al. 1979; Eisenberg et al. 2000; Nishi et al. 2010; but see McCreery & Stelmachowicz 2011). In contrast to children's performance in relatively steady-state noise, larger and more prolonged age effects have been observed for speech recognition in speech maskers composed of one or two talkers, which are expected to produce substantial informational masking (e.g., Hall et al. 2002; Wightman et al. 2003; Wightman & Kistler 2005; Bonino et al. 2013). For instance, Hall et al. (2002) investigated children's and adults' spondee identification in the presence of either two-talker speech or speech-shaped noise using a four-alternative forced-choice picture-pointing task. Estimates of the SNR at threshold were approximately 7 dB higher for 5- to 10-year-old children than adults in two-talker speech, but only 3 dB higher in speech-shaped noise. Results from subsequent investigations confirmed the finding of a larger performance

gap between children and adults for speech than for relatively steady-state noise maskers (e.g., Wightman et al. 2003; Leibold & Buss 2013).

The factors responsible for children's pronounced and prolonged susceptibility to informational masking are not well understood. The human peripheral auditory system appears to mature prior to 6 months postnatal age (reviewed by Werner 2007). Thus, children's pronounced difficulties on speech-on-speech masking tasks are believed to reflect immature central auditory processes such as sound source determination and selective auditory attention (e.g., Wightman & Kistler 2005; Leibold 2012). Sound source determination refers to the ability of the listener to assign incoming sounds to their respective sources (e.g., Bregman 1990). Auditory selective attention refers to a listener's ability to select an auditory object for further processing while discounting irrelevant auditory information (reviewed by Gomes et al. 2000). The specific age range over which maturation of these central auditory processes occurs has not been firmly established. However, converging evidence from both behavioral and electrophysiological studies indicate that these processes become more refined with increasing age during childhood (e.g., Doyle 1973; Coch et al. 2005; Wightman & Kistler 2005; Bonino et al. 2013; Leibold & Buss 2013). For example, Leibold and Buss (2013) assessed consonant identification in a two-talker speech masker using a 12-alternative, forced-choice procedure. Listeners were 5- to 7-year-olds, 8- to 10-year-olds, 11- to 13-year-olds, and adults. Although child-adult differences in overall identification performance were observed for all three child age groups, the largest child-adult difference was observed for the youngest children (5- to 7-year-olds).

One limitation of previous developmental studies investigating the informational masking of speech is that children older than about 10 years of age are often not included (e.g., Hall et al. 2002; Bonino et al. 2013). Consequently, the time course of development for speech-on-speech recognition has not been fully described. Note, however, that two studies by Wightman and colleagues tested children as old as 16 years of age on conditions in which listeners were asked to attend to a target phrase produced by one talker while ignoring an ipsilateral distractor phrase spoken by a different talker of the same sex (Wightman & Kistler 2005; Wightman et al. 2006). The specific task used in those studies employed the Coordinate Response Measure (CRM) corpus (e.g., Brungart 2001; Brungart et al. 2001). Each CRM sentence has the same structure ("Ready [call sign] go to [color] [number] now"). Two such sentences are played simultaneously, and the target is identifiable by virtue of a unique call sign (e.g, "Baron"). Under these conditions, recognition can be thought of as a pair of forced-choice identifications (color and number). In one study, Wightman and Kistler (2005) tested 4- to 16-year old children and adults using the CRM corpus. Results showed that younger children performed more poorly than older children or adults, and that performance tended to improve with increasing age. Of particular interest to the present study, some children as old as 16 years of age did not achieve adult-like performance.

The primary goal of the present study was to map the developmental trajectory of children's open-set word recognition in the presence of a speech-shaped noise masker, which is expected to produce primarily energetic masking, and a two-talker speech masker, which is expected to produce substantial informational masking. To accomplish this goal, masked

speech recognition performance was assessed in a sample of 56 children ranging in age from 5 to 16 years. Listeners were tested in each of two competing masker conditions: (1) speech-shaped noise and (2) two-talker speech. Consistent with previous studies of masked word recognition (Hall et al. 2002; Bonino et al. 2013), it was predicted that adult-like performance would be observed around 8 to 10 years of age in the speech-shaped noise masker. Based on results from Wightman and Kistler (2005) and Wightman et al. (2006), who assessed speech perception in conditions expected to produce considerable informational masking, we predicted that adult-like performance would be observed in late adolescence for open-set word recognition in the two-talker speech masker.

MATERIALS AND METHODS

Listeners

Fifty-six children (5 to 16 yrs) and 16 adults (18 to 44 yrs) participated in this experiment. Child listeners were distributed approximately uniformly between the lower and upper age limits on a logarithmic scale. The rationale for using the logarithm of age for the purposes of listener recruitment and data analysis was the observation that age-related changes in psychophysical performance occur more rapidly in the younger listeners, decreasing with increasing age during childhood (Mayer & Dobson 1982; Moller & Rollins 2002). Adults were included to provide an estimate of mature performance. Criteria for inclusion were: (1) hearing thresholds less than or equal to 20 dB HL for octave frequencies from 250 Hz to 8000 Hz (ANSI 2010); (2) native speaker of American English; and (3) no known history of chronic ear disease. One 5-year-old child did not complete testing in one session and did not return to finish data collection; complete data were collected in one visit from all other participants. This research was approved by the institutional review board for The University of North Carolina at Chapel Hill.

Stimuli and conditions

Formation and verification of the target word corpus—The target stimuli were monosyllabic words. Since commercial monosyllabic word recognition tests contain a limited number of test items (e.g., 150 words in the Phonetically Balanced Kindergarten Test, Haskins 1949), a larger corpus of words was developed based on children's reading lists for kindergarten and first grade. The corpus was reviewed by five adults to identify homophones and to confirm that young children would likely be familiar with the words. All of the adults were audiology or speech-language pathology graduate students and were native speakers of American English. After omitting homophones and potentially ambiguous and/or unfamiliar words, there were 842 words in the corpus. These 842 words were spoken by a male native speaker of American English and recorded in a double-walled sound booth (IAC). A condenser microphone (AKG-C1000S) was placed approximately six inches from the speaker's mouth using a microphone stand. The single-channel recordings were amplified (TDT, MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (Digital Audio Labs, CARDDELUXE). Each target word was recorded twice, each time with the carrier phrase "say the word" prior to the target word. Audacity sound editing software (v 1.2.6) was used to manually splice target words, which were then saved as individual wav files. The wav files were scaled to equivalent root mean square (RMS) level

and down-sampled at a rate of 24.414 kHz using MATLAB. The first recording of each target word was used unless undesirable sound quality characteristics (e.g., distortion, peak clipping, or irregular speaking rate) were noted.

To verify the sound quality of the recordings, the same five adults who reviewed the preliminary word lists also completed an open-set word recognition task in quiet. The 842 target words were presented sequentially at 60 dB SPL to the right ear through an ER1 Etymotic insert earphone. No listener missed more than a total of 11 words (>98% correct). Based on listener performance and feedback, 18 target words were re-edited either because they were missed by at least one listener or were noted to have undesirable sound qualities. After re-editing was completed, the listening check was repeated by two adults for the full corpus (842 words). Five words were identified and deleted due to clipping. The remaining 837 words were included in the final corpus for the present experiment.

Masker formation and stimuli presentation—Performance was assessed in each of two continuous masker conditions: (1) two-talker speech or (2) speech-shaped noise. Following Bonino et al. (2013), the two-talker speech masker was composed of recordings of two males, each reading separate passages from a popular series of fantasy novels written for children ages 8 to 12 years. Each of the individual masker streams was manually edited to shorten silent pauses, ensuring no silent pauses greater than 300 ms. The rationale for this editing was to reduce opportunities for “dip listening” (e.g., Gustafsson & Arlinger 1994; Cooke 2006) after mixing of the individual streams. The duration of one edited sample was 4 min and 17 s. The duration of the other sample was 7 min 47 s. Each stream ended with a complete sentence. The two streams were scaled to equal RMS level. A continuous 20-minute masker was created by concatenating copies of each speech sample head-to-tail, and then combining the two streams. The speech-shaped noise masker was created by extracting the long-term spectral envelope of the two-talker speech masker, and then shaping Gaussian noise with the extracted spectral envelope. The speech-shaped noise was also 20 minutes in duration. Both maskers were stored as wav files.

The selection and presentation of stimuli were controlled through custom software (MATLAB). The target speech tokens were mixed with either the two-talker speech or speech-shaped noise masker (TDT, RZ6), amplified (Applied Research and Technology SLA-4), and presented through a loudspeaker (JBL, Professional Control 1).

Procedure

Performance was assessed using an open-set word recognition procedure. Listeners were tested while seated approximately 3 ft from the loudspeaker in the sound field of a 7 × 7 ft, single-walled, sound-treated booth. The height and position of the listener’s chair was adjusted so that stimuli would be presented at approximately 0-degree azimuth and 0-degree elevation. Listeners wore an FM transmitter (Sennheiser, ew 100 G3) with a wireless lapel microphone during testing. The microphone was attached to the listener’s shirt within 6 inches of his/her mouth. The signal inside the booth was delivered to an examiner seated outside the booth via an FM receiver coupled to high-quality headphones (Sennheiser, HD25). This approach optimized the signal-to-noise ratio (SNR) for the primary tester, who

also watched the listener's face through a double-paned window throughout testing. In addition to the primary tester, an assistant sat inside the booth during testing, positioned behind the listener and in front of a computer monitor and keyboard. Exceptions include seven listeners (two 9-year-olds, an 11-year-old, two 13-year-olds, four 14-year-olds, four 15-year-olds, and four adults) who completed testing with only the primary tester.

Listeners were instructed to ignore the continuous background sounds and repeat the target words. Guessing was encouraged. Listeners were instructed to say "I don't know" if they thought a token was presented but could not make out what the word was. The primary tester was seated outside the booth, monitored the listeners' productions through the FM system, and entered their responses via a keyboard. This information was immediately provided to the assistant seated inside the booth via a second computer monitor located behind the listener. The assistant was prompted to agree or disagree with the primary tester's entry. If the assistant disagreed, she was prompted to enter her coding of the listener's response. Because the primary tester outside the booth had a more advantageous SNR than the assistant and could see the listener's face, the primary coder made the final decision as to what was ultimately recorded as the listener's response. Disagreements between the tester and the assistant were uncommon. The maximum response window for each trial was 5 seconds following the end of the target -presentation. If the listener did not respond within this window, the tester coded the trial as incorrect.

Target words were presented at a fixed level of 65 dB SPL throughout testing. One target word was randomly selected from the corpus on each trial. Target words were selected without replacement within a run, but could be re-selected by the computer software in subsequent runs. Therefore, a word could be presented to a listener multiple times within a given test session; however, recall that the full corpus contained 837 words, so repetition was uncommon. For example, if a listener heard 100 total words (5 runs with 20 words per run), the probability of a word being repeated is approximately 4%. The masker level was adapted using a 1-up, 1-down rule (Levitt, 1971) to obtain an estimate of the level corresponding to 50% correct on the psychometric function. The starting level for the masker was approximately 10 dB below the expected threshold for each masker condition, adjusted for individual listeners. An initial step size of 4 dB was reduced to 2 dB after the first two reversals. Masker level was adjusted following the primary tester's final input of the listener's verbal responses and at least 300 ms before the next trial was initiated. Runs were terminated after eight reversals. The masker threshold was estimated by computing the average of the masker level at the final six reversals. Listeners completed two runs in each masker (speech-shaped noise and two-talker speech). A third run was completed if the first two estimates differed by 5 dB or more. Thresholds for the two masker conditions were collected in random interleaved order. Seven listeners required three runs in the two-talker speech masker (two 5- to 7-year-olds; one 8- to 12-year-olds; two 13- to 16-year-olds; and two adults), and 15 listeners required three runs in the speech-shaped noise masker (four 5- to 7-year-olds; six 8- to 12-year-olds; two 13- to 16-year-olds; and three adults). Two children (one and 8- to 12-year-old and one 13- to 16-year-old) and one adult completed three runs for both listening conditions. For listeners completing more than two runs for a given listening condition, the two runs with the greatest agreement were used to compute

threshold. Listeners completed an entire testing session in less than one hour and took breaks as needed.

RESULTS

Group data

The data were first analyzed by comparing performance across four age groups of listeners: 5 to 7 years (n=19); 8 to 12 years (n=19); 13 to 16 years (n=18); and 18–44 years (n=16). These age groups were selected to sample the age range over which many auditory skills are thought to develop. Inclusion of 5- to 7-year-olds was based on data showing greater speech-on-speech masking for children greater than 7 years of age than for older children (e.g., Leibold & Buss 2013). Inclusion of 8- to 12-year-olds was based on observations of immature behavior persisting into adolescence for masked speech perception (e.g., Wightman et al. 2010). Inclusion of teenagers (13–16 yrs) was aimed at capturing the full range of developmental effects. Recall that speech recognition performance for children older than 13 years of age has not been previously assessed in a two-talker speech masker. Figure 1 shows the average SNR at threshold for each age group. Error bars represent ± 1 standard error of the mean (SEM). Results for the speech-shaped noise masker are presented to the left, and results for the two-talker speech masker are presented to the right. Higher thresholds indicate poorer performance than lower thresholds.

Younger children required a more advantageous SNR than older children and adults to achieve 50% correct recognition of target words for both masker conditions. However age-related changes appear to be more pronounced for the two-talker speech compared with the speech-shaped noise masker. Group average thresholds in each masker condition are summarized in Table 1. All four age groups of listeners performed more poorly in the two-talker speech than the speech-shape noise masker. As shown in Table 1, group average thresholds in the two-talker speech masker were about 3 to 6 dB higher compared to thresholds in the speech-shaped noise masker.

A repeated-measures analysis of variance (ANOVA) was conducted to evaluate the statistical significance of the trends observed in Figure 1. This analysis included the within-subjects factor of Masker (speech-shaped noise and two-talker speech) and the between-subjects factor of Age Group (5- to 7-year-olds, 8- to 12-year-olds, 13- to 16-year-olds, and adults). There was a significant main effect of Masker [$F(1, 68) = 125.24$, $p < 0.001$, $\eta^2_p = 0.65$], a significant main effect of Age Group [$F(3, 68) = 34.47$, $p < 0.001$, $\eta^2_p = 0.60$], and a significant interaction of Masker × Age Group [$F(3, 68) = 2.93$, $p = 0.04$, $\eta^2_p = 0.12$]. The significant Masker × Age Group interaction was examined with follow-up pairwise comparisons on thresholds across Age Group for each Masker condition, with Bonferroni adjustments for multiple comparisons. The *p*-values associated with this analysis are shown in Table 2. In speech-shaped noise, thresholds for the two youngest groups of children (5 to 7 yrs and 8 to 12 yrs) were significantly higher than those of the oldest group of children (13 to 16 yrs) and adults. There was no difference in thresholds between 5- to 7-year-olds and 8- to 12-year-olds or between 13- to 16-year-olds and adults. In the two-talker speech masker, thresholds for 5- to 7-year-olds were significantly higher than for the older three groups of listeners, including 8- to 12-year-olds. In addition, thresholds for 8- to 12-year-olds were

significantly higher than for 13- to 16-year-olds and adults. There was no difference in thresholds between the oldest group of children (13 to 16 yrs) and adults.

Individual data and developmental trajectories

In addition to the analysis of the group data, the developmental trajectories for the two masker conditions were compared. Data for individual listeners for the speech-shaped noise masker are presented in Figure 2, plotted as a function of listener age. Open circles and filled squares indicate thresholds for children and adults, respectively. Consistent with the group data shown in the left panel of Figure 1, there was a trend for better performance with increasing age during childhood. Although between-subjects differences were large, just under half of the thresholds for children younger than 8 years of age fell within the range of thresholds observed for adults. In contrast, all but five children (out of 37 total children) 8 years of age or older had thresholds within the range of those observed for adults. Results of a correlational analysis conducted on the data of individual children for the speech-shaped noise masker indicated a significant negative correlation between the logarithm of age and SNR at threshold ($r = -0.55$; $p = 0.001$).

Figure 3 shows individual thresholds for the two-talker speech masker condition, using the same format as in Figure 2. The individual data are in agreement with the group data summarized in the right panel of Figure 1. The highest thresholds were observed for children younger than 14 years of age; similar performance was observed for children older than 14 years of age and adults. For example, thresholds for 41/42 children younger than 14 years of age were positive, ranging from 0.8 to 12.8 dB SNR (mean = 3.9). The lone exception is a child who was 13.9 years of age with a threshold of −0.7 dB SNR. In contrast, only four children older than 14 years of age and three adults had positive thresholds for the two-talker speech masker condition. Estimates of thresholds ranged from −4.3 to 1.7 dB SNR (mean = −0.9) for children older than 14 years of age and from −4.0 to 2.2 dB SNR (mean = −1.3) for adults. As with the data for the speech-shaped noise masker, there was a significant negative correlation between the logarithm of age and thresholds for children in the two-talker speech masker ($r = -0.68$; $p < 0.001$).

Performance for both masker conditions was significantly correlated with child age, but the pattern of age-related changes appeared to differ between the two masker conditions. For the speech-shaped noise masker, a gradual improvement in performance was observed with increasing age, with children over 10 years of age performing within the range of adults. However, considerable overlap in the data was evident between listeners of all ages. For the two-talker speech masker, performance remained consistently poor for children between 5 and 13 years of age. An apparent “break point” in the developmental trajectory was observed between 13 and 14 years of age, with thresholds for children over 14 years of age being indistinguishable from those of adults. To determine the extent to which performance in the speech-shaped noise could predict performance in the two-talker speech masker on either side of the observed age break point, we conducted separate correlations between thresholds obtained in both maskers for 5- to 13-year-olds and 14- to 16-year-olds. The correlations between thresholds in the speech-shaped noise and two-talker speech maskers were not significant for either child age range [5- to 13-year-olds ($r = 0.19$; $p = 0.24$), and

14- to 16-year-olds ($r = -0.40$; $p = 0.16$)]. That is, children who performed more poorly in the two-talker speech masker were not necessarily the same children who performed more poorly in the speech-shaped noise masker.

DISCUSSION

The goal of this study was to compare the time course of development for children's susceptibility to speech-on-speech masking for open-set word recognition in two-talker speech or speech-shaped noise. The present results are in line with previous data (e.g., Hall et al. 2002; Wightman & Kistler 2005; Bonino et al. 2013) showing that children's ability to recognize speech takes longer to develop in the presence of maskers composed of 1 to 2 competing talkers than relatively steady-state noise. Furthermore, extending the age range of children to include teenagers provides new data that suggest differences in the pattern of age-related changes between the two masker conditions.

Age-related changes in speech-shaped noise

Consistent with previous investigations of speech recognition in the presence of Gaussian or speech-shaped noise (e.g., Elliott et al. 1979; Nittrouer & Boothroyd 1990), we observed age-related improvements in the SNR required to achieve 50% correct speech recognition in speech-shaped noise. Specifically, children's open-set performance linearly increased with age until about 10 years. A significant negative correlation was found between children's performance in the speech-shaped noise and the logarithm of child age. Note, however, the substantial degree of overlap in performance between children and adults; some children younger than 10 years of age performed as well as adults, and some adults had thresholds similar to those of young children.

The increased speech recognition difficulties observed on average for younger children in speech-shaped noise relative to older children and adults reflect immature central auditory processing rather than immature peripheral resolution (reviewed by Buss et al. 2012). Several investigators have proposed that this additional masking reflects an immaturity in the ability to recognize degraded speech due to limited acoustic and/or linguistic experience (e.g., Eisenberg et al. 2000; Mlot et al. 2012). In support of this idea, findings from several studies have shown that children require greater spectral detail than adults to achieve the same speech recognition performance under degraded stimulus conditions (e.g., Eisenberg et al. 2000; Mlot et al. 2012). For instance, Eisenberg et al. (2000) tested adults and children (5- to 7-year-olds and 10- to 12-year-olds) on speech recognition tasks using 4-, 6-, 8-, 16- and 32-band noise-vocoded speech. For sentence recognition, 5- to 7-year-olds required more spectral bands than the older children and adults to reach a performance asymptote of >90% in quiet. Eisenberg et al. (2000) posited that 5- to 7-year-old children's limited acoustic and linguistic knowledge, and limited experience restricts their ability to recognize speech when stimuli are spectrally degraded.

Age-related changes in two-talker speech

Children's word recognition performance in the two-talker speech masker took longer to mature than their performance in the speech-shaped noise. Moreover, there were distinct

differences in the pattern of age-related changes during childhood for the two masker conditions. As evident in Figure 3, almost all children younger than about 14 years of age performed more poorly than adults in the two-talker speech. While children's performance in the speech-shaped noise improved approximately linearly with increasing age, a more gradual improvement was observed from ages 5 to 13 years in the two-talker speech, followed by a sharp improvement to adult levels between 13 and 14 years of age.

Differences in stimuli, procedures, and the age range of children tested limit direct comparisons between the present and previous data. Few studies of speech-on-speech recognition have included children older than 13 years of age. Exceptions are studies carried out by Wightman and colleagues, who used the CRM corpus (Wightman & Kistler 2005; Wightman et al. 2006; Wightman et al. 2010). In those studies, structured CRM sentences were time-aligned with a single stream of competing speech. Recall that in the present study, monosyllabic words were presented in a continuous two-talker speech masker, and the task was open-set word recognition. It is possible that the longer stimuli and more predictable format of the CRM sentences facilitated auditory streaming to a greater extent than the monosyllabic words and open-set recognition task used in the present study. Despite substantial differences in stimuli and procedures, the present results are generally consistent with those reported by Wightman and colleagues, in that immature performance was observed into adolescence. Note, however, that Wightman and colleagues (2006) did not observe a dramatic change in performance for children between 13 and 14 years of age for the single-talker masker used in that study.

Supplemental data: Deviation from 0 dB SNR and performance in two-talker speech

A striking feature of the data collected in the two-talker speech masker is the marked improvement in thresholds between 13 and 14 years of age. As evident in Figure 3, thresholds for listeners younger than 14 years were all greater than 0 dB SNR. In sharp contrast, thresholds were lower than 0 dB SNR for 8 out of 11 children older than 14 years of age and 11 out of 12 adults. Previous experiments on adults using the CRM corpus have demonstrated a plateau or a non-monotonicity in thresholds between 0 and −8 dB SNR (e.g., Brungart et al. 2001). Wightman and Kistler (2005) observed a similar performance plateau for children older than about 6 years of age. Energetic masking is expected to increase as the SNR decreases below 0 dB. Deviations from the expected steady improvement with increasing SNR has been interpreted in terms of informational masking and the target/masker level differences present below 0 dB SNR, which could be used to segregate the less intense target speech from the more intense masker speech (e.g., Brungart 2001). If word recognition is particularly difficult at 0 dB SNR, this could account for the abrupt change in performance in the two-talker speech masker as a function of increasing listener age. For example, consider how the adaptive tracks would be affected if young children's responses were consistently incorrect for SNRs near 0 dB, but older children and adults occasionally responded correctly at 0 dB SNR. For young children, an adaptive track could approach 0 dB SNR, but an incorrect response would prevent it from dropping into the range of negative SNRs. For older children and adults, even infrequent correct responses at 0 dB SNR would allow the adaptive track to fall below 0; once the SNR became slightly negative,

recognition could become easier, such that the remainder of the track would be more likely to stay low.

To evaluate the potential role of target/masker level differences on performance in the present study, supplemental data were collected from five additional adults (19 to 38 yrs). These adults completed testing in a single visit lasting about two hours. The same stimuli and procedure were used as in the main experiment, except that listeners were tested in 50-trial blocks using a fixed SNR between −12 and 14 dB. The level of the target words was 65 dB SPL, and the masker level was adjusted to obtain the desired SNR. Each listener was tested at a minimum of 9 different SNRs.

The supplemental data are shown in Figure 4. Circles indicate average percent-correct performance as a function of SNR in dB, and symbols size indicates the number of observations contributing to each estimate. The solid curve shows a standard logistic function that was fitted to the pooled data using a constrained maximum-likelihood procedure (Wichmann & Hill 2001). The data and fitted psychometric function are inconsistent with the idea that the abrupt improvement in performance between 13 and 14 years of age reflects a particular difficulty associated with speech recognition near 0 dB SNR. No evidence of a plateau in performance was observed. Instead, speech recognition performance improved monotonically with increasing SNR. This trend was observed in the pooled data shown in Figure 4, and in each of the five individual psychometric functions. This finding is inconsistent with the idea that the abrupt improvement in performance between 13 and 14 years of age is due to the lack of a level cue around 0 dB SNR.

Possible explanations for the rapid improvement in speech-on-speech recognition observed between 13 and 14 years of age

It is possible that the onset of puberty (Ponton et al. 2000) and/or the development and maturation of executive function in late adolescence (reviewed by Crone 2009) contributed to the rapid improvement in performance observed for children ages 13 to 14 years of age in the two-talker speech masker. Changes in cortical evoked potentials obtained from children ages 12 years and older (Ponton et al. 2000) are consistent with maturation of cortical axons through adulthood, and have been associated with the improvement of speech perception in the presence of reverberation and noise around 15 years of age (reviewed by Eggermont & Moore 2012). It is therefore possible that the listening conditions from the current study tapped into functional correlates of neural maturation between 13 and 14 years of age. Similarly, factors involved in complex auditory processing that fall under the umbrella of executive function, such as attention and working memory, do not become adult-like until late adolescence (e.g., reviewed by Gomes et al. 2000; Pisoni 2000; reviewed by Crone 2009). Masked speech recognition may rely on mature executive function to a greater extent in two-talker speech than in speech-shaped noise due to the perceptual similarity between the target and masker speech. Increased perceptual similarity appears to interfere with auditory stream segregation, thus placing a greater processing load on the cognitive system (e.g., Zekveld et al. 2013). A greater load on the auditory and cognitive systems could require the listener to allocate more processing resources (e.g., sustained attention and working memory) to achieve accurate performance in the two-talker speech masker (e.g.,

reviewed by Rönnberg et al. 2010). Future research is needed to examine the potential relationship between children's performance on speech-on-speech recognition tasks and measures of executive function.

Implications and future directions

The present results add to the growing body of evidence that hearing and understanding speech in complex acoustic environments remains a significant challenge for children throughout most of childhood. These findings have significant implications for children in classrooms, which often contain multiple sources of competing sounds. It has long been recognized that classrooms typically contain multiple sources of relatively steady-state noise, such as heating and air conditioning systems, computer fans, and fish tanks (e.g., Knecht et al. 2002). This noise often results in poor classroom acoustics, with many classrooms failing to meet national standards for unoccupied classrooms (reviewed by Picard & Bradley 2001; ANSI/ASA S12.60-2002). In addition to relatively steady-state sounds, there is a growing recognition that most occupied classrooms contain competing speech, often produced by multiple people talking at the same time (e.g., Sato & Bradley 2008). Multi-talker environments may occur when children talk out of turn, or when the teacher uses instructional strategies that divide students into smaller groups (e.g., center-based or cooperative learning). The present results add to the mounting evidence that listening to target speech in the presence of competing talkers poses a significant challenge during childhood. To date, the influence of competing speech has generally not been captured in the classroom acoustics literature.

Clinically, children's speech perception in noise is usually assessed using relatively steady-state backgrounds. However, estimates of performance obtained in noise maskers may not be predictive of the challenges faced by children in real-world environments (e.g., Carhart 1965; Hillock-Dunn et al. 2014). Furthermore, the present results suggest that there are marked differences between the development of speech perception in the presence of relatively steady-state noise and speech maskers. The finding that performance in the two maskers was not correlated is remarkable, and may indicate that these listening conditions are tapping into very difference auditory skills. It is critical that such factors are considered in the assessment and interpretation of clinical speech perception measures. The recent development of clinical measures assessing speech perception in the presence of a small number of acoustically dissimilar talkers (e.g., PRESTO; Gilbert, Tamati, & Pisoni 2013) reflects the rising acknowledgment of this problem. It is important to consider adopting a more diverse battery of speech perception tests in the audiology clinic that incorporates assessments of performance in the presence of both energetic and informational masking. The present results may provide normative data for the development of potential clinical measures that will capture the real-world difficulties children, including those who are hard of hearing, encounter.

Acknowledgments

Source of Funding: This work was supported by the National Institute of Deafness and Other Communication Disorders (NIDCD, R01 DC011038).

We are grateful to the members of the Human Auditory Development Laboratory for their assistance with data collection. We thank Eric Sanders and members of the Human Auditory Development Lab for their help in developing the monosyllabic word corpus, particularly Jack Hitchens and Chas Phillip.

References

American National Standards Institute. Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools. New York: American National Standards Institute; 2002. ANSI/ASA S12.60-2002

Bonino AY, Leibold LJ, Buss E. Release from perceptual masking for children and adults: benefit of a carrier phrase. Ear Hear. 2013; 34:3–14. [PubMed: 22836239]

Bregman, AS. Auditory scene analysis. Cambridge, MA: MIT Press; 1990.

Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoustic Soc Am. 2001; 109:1101–1109.

Brungart DS, Chang PS, Simpson BD, et al. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J Acoust Soc Am. 2006; 120:4007–4018. [PubMed: 17225427]

Brungart DS, Simpson BD, Ericson MA, et al. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am. 2001; 110:2527–2538. [PubMed: 11757942]

Buss, E.; Hall, JW., III; Grose, JH. Development of auditory coding as reflected in psychophysical performance. In: Werner, L.; Fay, RR.; Popper, AN., editors. Human Auditory Development. New York, NY: Springer; 2012. p. 107-136.

Carhart R. Problems in the measurment of speech discrimination. Arch Otolaryng. 1965; 82:253–260. [PubMed: 14327024]

Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am. 1969; 45:694–703. [PubMed: 5776931]

Coch D, Sanders LD, Neville HJ. An event-related potential study of selective auditory attention in children and adults. J Cogn Neurosci. 2005; 17:605–622. [PubMed: 15829081]

Cooke M. A glimpsing model of speech perception in noise. J Acoust Soc Am. 2006; 119:1562–73. [PubMed: 16583901]

Crone EA. Executive functions in adolescence: inferences from brain and behavior. Developmental Science. 2009; 12:825–830. [PubMed: 19840037]

Doyle AB. Listening to distraction: a developmental study of selective attention. J Exp Child Psychol. 1973; 15:100–115. [PubMed: 4706960]

Eggermont, JJ.; Moore, JK. Morphological and functional development of the auditory nervous system. In: Werner, L.; Fay, RR.; Popper, AN., editors. Human Auditory Development. New York, NY: Springer; 2012. p. 61-106.

Eisenberg LS, Shannon RV, Martinez AS, Wygonsky J, Boothroyd A. Speech recognition with reduced spectral cues as a function of age. J Acoustic Soc Am. 2000; 107:2704–2710.

Elliott LL, Connors S, Kille E, et al. Children's understanding of monosyllabic nouns in quiet and noise. J Acoust Soc Am. 1979; 66:12–21. [PubMed: 489827]

Fletcher H. Auditory patterns. Reviews of Modern Physics. 1940; 12:47–65.

Freyman RL, Balakrishnan U, Helfer KS. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J Acoustic Soc Am. 2004; 115:2246–2256.

Gilbert JL, Tamati TN, Pisoni DB. Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. J Am Acad Audiol. 2013; 24:26–36. [PubMed: 23231814]

Gomes H, Molholm S, Christodoulou C, et al. The Development of Auditory Attention in Children. Frontiers in Bioscience. 2000; 5:d108–d120. [PubMed: 10702373]

Gustafsson HA, Arlinger SD. Masking of speech by amplitude modulated-noise. J Acoustic Soc Am. 1994; 95:518–529.

Hall JW 3rd, Grose JH, Buss E, et al. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear. 2002; 23:159–165. [PubMed: 11951851]

Haskins, H. Unpublished master's thesis. Northwestern University; Evanston, IL: 1949. A phonetically balanced test of speech discrimination for children.

Hillock-Dunn A, Taylor CN, Buss E, et al. Assessing speech perception in children with hearing loss: What conventional tools may miss. Ear Hear. 2014 Epub Ahead of Print.

Knecht HA, Nelson PB, Whitelaw GM, Feth LL. Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements. American Journal of Audiology. 2002; 11:65–71. [PubMed: 12691216]

Leibold, LJ. Development of auditory scene analysis and auditory attention. In: Werner, LA.; Fay, RR.; Popper, AN., editors. Human Auditory Development. New York, NY: Springer; 2012. p. 137-162.

Leibold LJ, Buss E. Children's identification of consonants in a speech-shaped noise or a two-talker masker. J Speech Lang Hear Res. 2013; 56:1144–1155. [PubMed: 23785181]

Levitt H. Transformed up-down methods in psychoacoustics. J Acoustic Soc Am. 1971; 49(Suppl 2): 467.

Mayer DL, Dobson V. Visual acuity development in infants and young children, as assessed by operant preferential looking. Vis Res. 1982; 22:1141–1151. [PubMed: 7147725]

McCreery RW, Stelmachowicz PG. Audibility-based predictions of speech recognition for children and adults with normal hearing. J Acoustic Soc Am. 2011; 130:4070–4081.

Mlot S, Buss E, Hall JW 3rd. Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. Ear Hear. 2012; 31:56–62. [PubMed: 19816182]

Moller AR, Rollins PR. The non-classical auditory pathways are involved in hearing in children but not in adults. Neuroscience Letters. 2002; 319:41–44. [PubMed: 11814649]

Nishi K, Lewis DE, Hoover BM, et al. Children's recognition of American English consonants in noise. J Acoust Soc Am. 2010; 127:3177–3188. [PubMed: 21117766]

Nittrouer S, Boothroyd A. Context effects in phoneme and word recognition by young children and older adults. J Acoust Soc Am. 1990; 87:2705–2715. [PubMed: 2373804]

Picard M, Bradley JS. Revisiting speech interference in classrooms. Audiology. 2001; 40:221–244. [PubMed: 11688542]

Pisoni DB. Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception. Ear Hear. 2000; 21:70–78. [PubMed: 10708075]

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

Rönnberg J, Rudner M, Lunner T, et al. When cognition kicks in: Working memory and speech understanding in noise. Noise & Health. 2010; 12:263–269. [PubMed: 20871181]

Sato H, Bradley JS. Evaluation of acoustical conditions for speech communication in working elementary school classrooms. J Acoust Soc Am. 2008; 123:2064–2077. [PubMed: 18397014]

Werner LA. Issues in human auditory development. Journal of communication disorders. 2007; 40(4): 275–283. [PubMed: 17420028]

Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys. 2001; 63:1293–1313. [PubMed: 11800458]

Wightman FL, Callahan MR, Lutfi R, et al. Children's detection of pure-tone signals: Informational masking with contralateral maskers. J Acoustic Soc Am. 2003; 113:3297–3305.

Wightman FL, Kistler DJ. Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. J Acoust Soc Am. 2005; 118:3164–3176. [PubMed: 16334898]

Wightman FL, Kistler DJ, Brungart DS. Informational masking of speech in children: Auditory-visual integration. J Acoust Soc Am. 2006; 119:3940–3949. [PubMed: 16838537]

Wightman FL, Kistler DJ, O'Bryan A. Individual differences and age effects in dichotic informational masking paradigms. J Acoust Soc Am. 2010; 128:270–279. [PubMed: 20649222]

Zekveld AA, Rudner M, Johnsrude IS, et al. The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. J Acoust Soc Am. 2013; 134:2225–2234. [PubMed: 23967952]

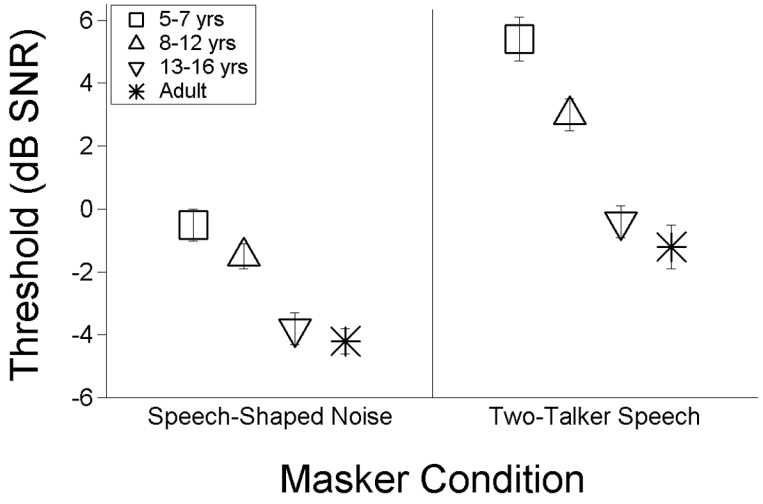

Figure 1.

Group average thresholds (in dB SNR) required to reach 50% correct word recognition are shown for each of the five age groups. Results are shown separately for data collected in the speech-shaped noise (left panel) and the two-talker speech (right panel) maskers. Error bars represent ± one standard error of the mean.

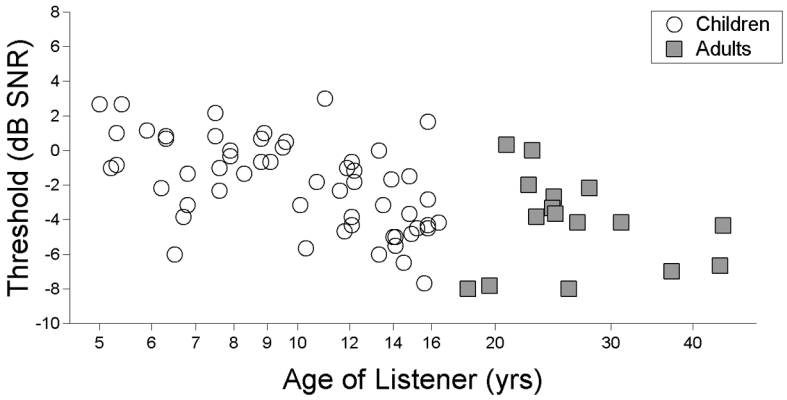

Figure 2.

Individual thresholds (in dB SNR) for word recognition in the speech-shaped noise masker. Data are plotted as a function of listener age on a logarithmic scale.

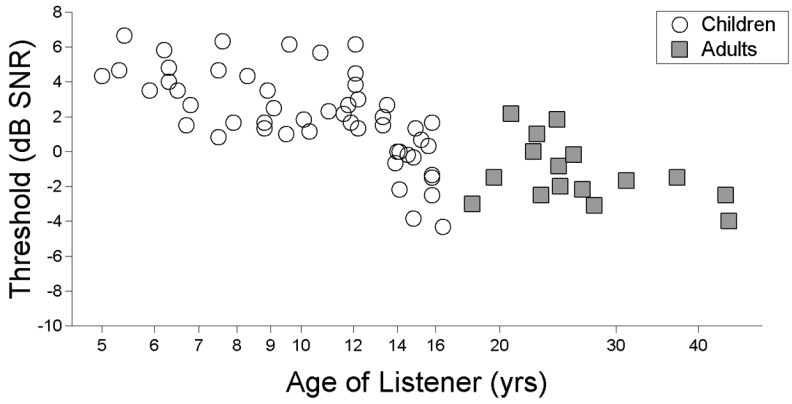

Figure 3.

Individual thresholds (in dB SNR) for word recognition in the two-talker speech masker. Data are plotted as a function of listener age on a logarithmic scale.

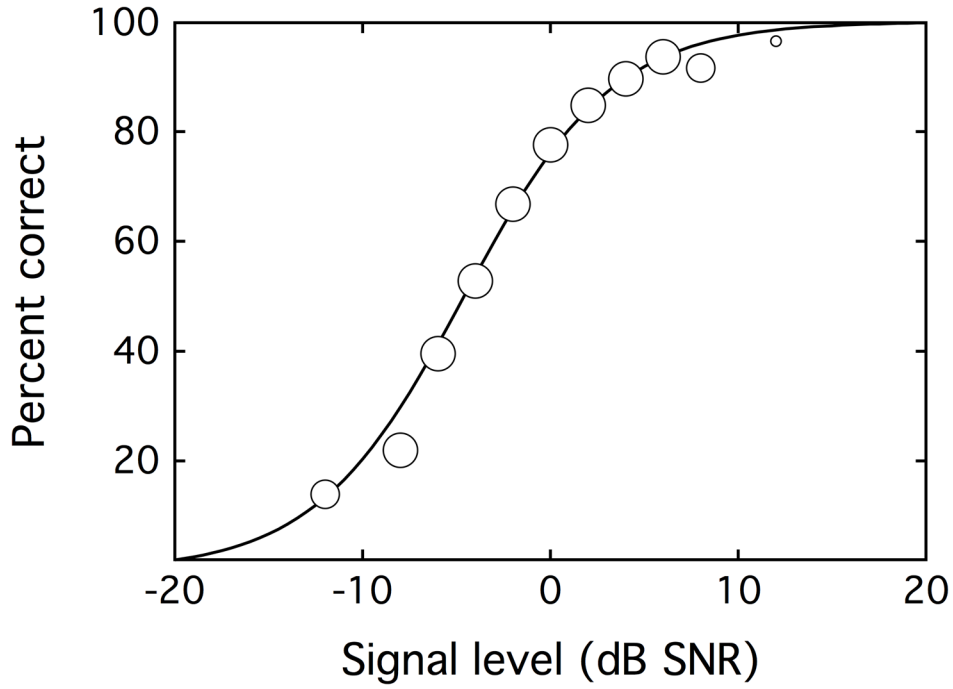

Figure 4.

Supplemental data showing percent correct word recognition in the two-talker speech masker, as a function of SNR in dB. All five adults provided data at 2-dB intervals between −8 and 8 dB SNR, and a subset of listeners provided data above and below this range. Symbol size reflects the number of listeners providing data at each SNR. The solid line shows the logit fit to the data.

Table 1

Group average thresholds in dB SNR are provided for the four listener age groups in the two-talker speech and speech-shaped noise maskers. Standard deviation estimates are shown in parentheses. The average difference in threshold (in dB) between the two masker conditions is shown for each age group in the rightmost column.

	Speech-Shaped Noise	Two-Talker Speech	Two-Talker minus Noise
5–7 yrs	−0.5 (2.3)	5.4 (3.5)	5.9
8–12 yrs	−1.5 (2.2)	3.0 (1.7)	4.5
13–16 yrs	−3.8 (2.3)	−0.4 (2.0)	3.4
Adults	−4.2 (2.6)	−1.3 (1.8)	2.9

Table 2

Results (p-values) of pairwise comparisons (adjusted Bonferroni) are shown for both masker conditions. Significant effects are indicated by an asterisk.

Speech-Shaped Noise Masker				**Two-Talker Speech Masker**			
	8–12 yrs	**13–16 yrs**	**Adults**		**8–12 yrs**	**13–16 yrs**	**Adults**
5–7 yrs	$p = 1.00$	$\mathbf{p < 0.001}$*	$\mathbf{p < 0.001}$*	5–7 yrs	$\mathbf{p < 0.05}$*	$\mathbf{p < 0.001}$*	$\mathbf{p < 0.001}$*
8–12 yrs	-----	$\mathbf{p < 0.05}$*	$\mathbf{p < 0.01}$*	8–12 yrs	-----	$\mathbf{p < 0.001}$*	$\mathbf{p < 0.001}$*
13–16 yrs	-----	-----	$p = 1.00$	13–16 yrs	-----	-----	$p = 1.00$