Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults

Marcin Wróblewski, M.A.,
Boys Town National Research Hospital, Omaha, NE and Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA

Dawna E. Lewis, Ph.D.,
Boys Town National Research Hospital, Omaha, NE

Daniel L. Valente, Ph.D., and
Boys Town National Research Hospital, Omaha, NE

Patricia G. Stelmachowicz, Ph.D.
Boys Town National Research Hospital, Omaha, NE

Abstract

Objectives—The purpose of this study was to determine how combinations of reverberation and noise, typical of many elementary school classroom environments, affect normal-hearing school-aged children's speech recognition in stationary and amplitude-modulated noise, and to compare their performance to that of normal-hearing young adults. Additionally, the magnitude of release from masking in the modulated noise relative to stationary noise was compared across age in non-reverberant and reverberant listening conditions. Finally, for all noise and reverberation combinations the degree of change in predicted performance at 70% correct was obtained for all age groups using a best-fit cubic polynomial.

Design—Bamford-Kowal-Bench sentences and noise were convolved with binaural room impulse responses representing non-reverberant and reverberant environments to create test materials representative of both audiology clinics and school classroom environments. Speech recognition of 48 school-aged children and 12 adults was measured in speech-shaped and amplitude modulated speech-shaped noise, in the following three virtual listening environments: non-reverberant, reverberant at a 2 m distance, and reverberant at a 6 m distance.

Results—Speech recognition decreased in the reverberant conditions and with decreasing age. Release from masking in modulated relative to stationary noise decreased with age and was reduced by reverberation. In the non-reverberant condition, participants showed similar amounts of masking release across ages. The slopes of performance-intensity functions increased with age, with the exception of the non-reverberant modulated masker condition. The slopes were steeper in the stationary masker conditions, where they also decreased with reverberation and distance. In the presence of a modulated masker, the slopes did not differ between the two reverberant conditions.

Conclusions—The results of this study reveal systematic developmental changes in speech recognition in noisy and reverberant environments for elementary school-aged children. The

Address correspondence to: Marcin Wróblewski, Communication Sciences and Disorders, Wendell Johnson SHC 226C, University of Iowa, Iowa City, IA 52246, Tel: 319-224-0808, Fax: 319-335-8851, marcin-wroblewski@uiowa.edu.

overall pattern suggests that younger children require better acoustic conditions to achieve sentence recognition equivalent to their older peers and adults. Additionally, this is the first study to report a reduction of masking release in children as a result of reverberation. Results support the importance of minimizing noise and reverberation in classrooms, and highlight the need to incorporate noise and reverberation into audiological speech-recognition testing in order to improve predictions of performance in the real world.

Introduction

Communication in the real world rarely occurs in a quiet and anechoic environment. In fact, much of communication and learning occurs in noisy and reverberant environments, and children have more difficulty understanding speech in these conditions, compared to adults (Johnson 2000; Klatte et al. 2010b; Nabelek and Robinson 1982; Neuman et al. 2010). In audiology, word-recognition testing under earphones in quiet is considered the "gold standard" of speech perception, and is often used as a predictor of real-life performance for both normal-hearing and hearing-impaired children (Gelfand 1998; Martin et al. 1994). However, such assessments may not reflect the capacity to understand and learn spoken language in classrooms, where the speech may be affected by reverberation and noise (Bradley 1986; Crandell and Smaldino 2000; Finitzo-Hieber and Tillman 1978; Neuman and Hochberg 1983). The aim of the present study was to assess children's speech recognition in steady-state and modulated noise in less-than-ideal listening conditions (i.e., those that include a combination of reverberation and noise) relative to results obtained in anechoic or near-anechoic environments typical of audiology clinics.

In 1998, the U.S. Access Board joined with the Acoustical Society of America (ASA) in identifying poor classroom acoustics as a potential barrier to education for children, particularly for special populations of children, including those with hearing loss and/or cochlear implants (U.S. Access Board 2011). An adoption of guidelines for classroom design in the United States by the American National Standard Institute (ANSI S12.60-2002) and its recent revision (ANSI S12.60-2009/Part 2 2009; ANSI S12.60-2010/Part 1 2010a) emphasize the importance of classroom acoustics. The ANSI recommended maximum reverberation time (RT defined as the time required for the level of a steady-state sound to decay by 60 dB after it has been turned off) and noise level for mid-sized classrooms ($< 10,000$ ft^3) are 0.6 sec and 35 dBA, respectively. For larger classrooms ($10,000 - 20,000$ ft^3) the allowable RT is 0.7 sec, but the allowable noise level remains unchanged. Despite the need for improved acoustics in schools, research reveals that many current classrooms often exceed the above mentioned standards (Bradley 1986; Bradley and Sato 2008; Knecht et al. 2002; Kodaras 1960).

The ANSI guidelines are relevant to the design of classrooms and should be helpful in promoting better classroom acoustics. However, these guidelines do not address concerns about the noise generated in occupied classrooms at times when schools are in session. Noise generated by students, heating ventilation and air-conditioning (HVAC) systems, and instructional equipment also can degrade speech recognition in the classroom. Numerous studies have reported that the noise levels in occupied classrooms generally range between 50 and 65 dBA, and that signal-to-noise ratios (SNR) range from −7 to +5 dB (Crandell and Smaldino 2000; Jamieson et al. 2004; Picard and Bradley 2001). The highest noise levels (65–75 dBA) occurred in the classrooms of the youngest children (Picard and Bradley 2001).

The purpose of children's listening in a classroom is different from that of adults engaged in a conversation. In adult conversations, participants often navigate within limited context and can utilize context-dependent (top-down) processing to fill in missing information when the

acoustic environment is less than optimal. When learning new material in unfamiliar context, children are often required to rely more on context-independent (bottom-up) processing. Evidence suggests that children gradually improve their top-down language processing skills well into adolescence (Bitan et al. 2009). Since children do not reach adult-like speech recognition in noisy and reverberant environments until 14 – 15 years of age (Johnson 2000), it is important to develop and implement ecologically valid audiological test materials, especially for the assessment of speech recognition of children.

In a seminal study, Finitzo-Hieber and Tillman (1978) examined the effects of RT and noise on monosyllabic word recognition of normal-hearing and hearing-impaired children (8 – 13 yrs). The authors demonstrated that increased reverberation (RT = 0.4, 1.2 s) and decreased SNR (quiet, +12 and 0 dB) resulted in poorer performance. The combined effects of noise and reverberation were greater than would be predicted from the decrease in performance due to either type of signal degradation alone. This discrepancy was even greater for hearing-impaired children. However, Finitzo-Hieber and Tillman may have overestimated the effects of noise and reverberation due to the fact that all testing was performed monaurally. To address the influence of monaural-only stimulus presentation, Neuman and Hochberg (1983) assessed the developmental aspects of speech recognition (phoneme identification) in anechoic condition and a reverberant field (RT = 0.4, 0.6 s) in monaural and binaural listening conditions in 5 –13 year old children. An increase in reverberation alone negatively impacted speech recognition, even for older children, but performance did improve with age. For the youngest group, binaural hearing at a RT of 0.6 sec improved phoneme identification relative to the monaural condition. This apparent increase in performance with age in reverberant but not anechoic conditions suggests that reverberation may impede the ability for young normal-hearing children to understand speech in classrooms.

Bradley and Sato (2008) assessed children's ability to identify a closed set of monosyllabic rhyming words as a function of age (6, 8, and 11 year olds), noise (−15 to +30 dB SNR) and reverberation (RT = 0.3 to 0.7 s) in their own classrooms. Forty-one Canadian classrooms (most of which were in compliance with the highest RT recommended by the 2002 ANSI standard) were surveyed. The findings indicated that younger children had more trouble understanding acoustically compromised speech than older children; all children performed significantly poorer in noisier classrooms; and younger children's performance decreased more sharply with increasing noise levels. Estimates of the SNR needed for near-optimal performance decreased as a function of age. The scatter in speech intelligibility scores suggested that many students would have more difficulty understanding speech in noisy and reverberant conditions than the mean data would imply. Yang and Bradley (2009) assessed the effects of reverberation (RT = 0.3, 0.6, 0.9 and 1.2 s) on the recognition of monosyllabic words by 6–11 year old children in the presence of a steady-state noise. The results revealed that younger children performed more poorly than their older peers, and that performance was better at shorter RTs. The benefit of adding early-arriving reflections to the direct sound was confirmed, suggesting that children sitting closer to the teacher (i.e., in or near the direct field of speech signal) receive a more robust signal than those sitting farther away (i.e., in the reverberant field of the original sound).

A recent report by Neuman et al. (2010) examined the combined effects of reverberation (RT = 0.3, 0.6 and 0.8 s) and noise in young children (6 – 12 yrs) using the Bamford-Kowal-Bench Speech in Noise (BKB-SIN) test (Etymotic Research 2005). A virtual classroom approach was used, with reverberation incorporated into the testing materials. Results of the study revealed trends similar to those found in previous studies: children's ability to understand speech in noise improved with age and decreasing RT (cf., Bradley 1986; Bradley and Sato 2008; Johnson 2000; Yacullo and Hawkins 1987). Specifically, the

youngest children required an additional 3 dB SNR to reach performance similar to that of the oldest children. The increase in SNR necessary to counteract the negative effects of reverberation (approximately 4 dB) was similar for children in all age groups. Interestingly, the disparity between the anechoic normative data and the RT = 0.3 sec condition was larger than that between RT = 0.3 sec and RT = 0.8 sec. This suggests that anechoic materials may overestimate children's ability to understand speech in classrooms, more so than would be predicted from the difference in performance in "good" versus "poor" acoustical conditions. In turn, this observation would be consistent with the view that it is important to reduce reverberation and maximize SNR in classrooms, especially those designed for early grades.

Klatte et al. (2010b) investigated the effects of type of noise (classroom noise vs. single foreign-language speaker) and reverberation (RT = 0.47 and 1.1 s) on children's (6 – 10 yrs) speech perception and listening comprehension at varying listener distances (3, 6 and 9 m) and SNRs (+3 to −6 dB). The results revealed that participants had poorer speech perception in the more reverberant condition, at a farther distance (poorer SNR), and with the addition of typical classroom noise versus a single speaker as noise. The listening-comprehension task was more challenging with a competing single speaker than with classroom noise, suggesting that, unlike simple word identification, children's higher level processing of acoustic signals may be more adversely affected by informational masking than by energetic masking. In contrast, adults showed no detriment in listening comprehension with a single speaker as a competing distracter, and only a small decrease in performance was observed in the presence of classroom noise. Thus, poor classroom acoustics appear to have a greater negative impact on children's speech perception and listening comprehension relative to adults. As a result, less than favorable classroom acoustics may affect academic achievement. In a related paper (Klatte et al. 2010a), it was reported that poor classroom acoustics also have a negative impact on phonological processing and short-term memory.

The variety of testing materials, RTs, noise types and levels used in the above-mentioned experiments suggest that regardless of classroom conditions, children's speech perception in a reverberant and noisy environment will never be equivalent to that under earphones in quiet, as typically assessed in audiological evaluations. School classrooms are reverberant, and in most cases, two types of competing noise affect children's listening: "stationary noise" from equipment and/or HVAC systems, and "modulated noise" created by the speech of other talkers. The pauses, or "dips", in the amplitude-modulated noise allow for "glimpses" into the signal of interest. Listening in the dips of amplitude-modulated noise provides release from masking, and a resulting improvement in speech recognition, relative to stationary noise (Miller and Licklider 1950). Stuart (2005) examined release from masking in children by comparing their performance in steady-state and interrupted noise, while equalizing the bandwidth of the two types of noise. Using monosyllabic materials presented monaurally under earphones at SNRs ranging from −20 to +10 dB, Stuart showed that children benefit from dips in the noise, as do adults. While this benefit was smaller for children than adults, their results may have been confounded by a floor effect in the continuous-noise condition at −20 dB SNR. The relative improvement in masking release was not evident at −10 dB and higher SNRs. Stuart et al. (2006) found a similar trend for pre-school children, who showed no difference in the amount of masking release relative to adults. However, in both experiments, children's susceptibility to noise may have been overestimated due to the choice of monaural stimulus presentation.

Although, there are no available data from pediatric populations, listening in the dips in a reverberant environment has been investigated in a small group of adult listeners. George et al. (2008) examined sentence intelligibility of young adults in the presence of reverberation (RT = 0.05, 0.25, 0.5, and 1s) and noise with speech-like modulations. Due to temporal smearing of the speech and noise introduced by reverberation, modulated and reverberated

noise provided beneficial masking release only at low RTs (i.e., RT 0.25 s), when both stimulus and noise were reverberated. These results suggest that release from masking would be minimized or absent in the more reverberant conditions typical of classrooms. It is not clear how children would perform on a similar task; nonetheless, one might argue that the threshold for a measurable benefit of release from masking would be even lower for children, since the effects of reverberant and noisy environments on children's speech-recognition performance are more pronounced.

The purpose of the present study was to assess children's and young adults' speech perception under reverberation conditions and noise levels typical of those found in classrooms. Speech recognition was measured in the presence of modulated and steady-state speech-shaped noise in simulated reverberant and pseudo-anechoic conditions to investigate participants' ability to take advantage of the glimpses in the noise caused by the modulation dips. The impact of simulated listener distance on speech perception in a classroom was also examined. Two simulated classroom seating positions were contrasted: 2 meters (the position of a child sitting in the front row of seats), and 6 meters (equivalent to a child occupying a seat in the last row, near the back wall) away from the speaker. Performance-intensity (PI) functions were calculated to estimate the SNR required for a specific level of performance (e.g., 95 % correct). Additionally, the slopes of the PI function were compared to establish the degree of performance shift as a result of the SNR change in different reverberant and noise conditions across age groups. The results of the present study will add to the body of literature which describes how SNR and reverberation affect speech perception as a function of age, how children's performance compares to that of young adults, and how classroom seating location may affect speech perception. The data quantify the extent to which the dips in modulated noise aid performance in a reverberant environment for elementary school-aged children, and how their performance compares to that of young adults.

Materials and Methods

Participants

Forty-eight children (7 to 14 years), and 12 young adults (23 to 30 years) served as listeners. Children were divided into four age groups (6 children per age, 12 per group): 7– 8, 9–10, 11–12, and 13–14 years. All children had hearing within normal limits (20 dB HL; ANSI 2010b) for the octave frequencies from 250 to 8000 Hz. Children were typically developing speakers of American English, and were included only if they had no articulation errors on the Bankson–Bernthal Quick Screen of Phonology (BBQSP; Bankson and Bernthal 1990), and age-appropriate (no more than 1 *SD* below average) or higher scores on the Peabody Picture Vocabulary Test 4 (PPTV-IV; Dunn and Dunn 2006). Young adult monolingual native speakers of American English with normal hearing served as control participants. All participants gave informed consent/assent, and were reimbursed for their participation. Additionally, children received a prize in the form of a book of their choice.

Creation of Reverberant Test Materials

Reverberant materials were created by convolving binaural room impulse responses (BRIRs) with the recordings of the sentence materials and masking noise. This approach (Cameron et al. 2006; Koehnke and Besing 1996; Neuman et al. 2010; Yang and Bradley 2009) allows the creation of "virtual" environments, such as a virtual classroom. CATT-Acoustic software ("CATT-Acoustic™" 2006) was used to create BRIRs to reflect a model of an occupied rectangular classroom identical in size (volume of 198 m^3 (9.3 m [W] $\times$ 7.7 m [D] $\times$ 2.8 m [H])) and with RTs across octave frequencies similar to the average measures from a survey of 30 elementary school classrooms in Ontario, Canada (Sato and Bradley

2008). The average RT for octave frequencies from 125 to 4000 Hz was approximately 0.4 sec. The RTs for the modeled classroom are provided in Table 1.This RT of 0.4 sec is below the recommended maximum allowable (0.6 s) for unoccupied classrooms (ANSI S12.60-2009/Part 1; 2010a), and is representative of an occupied classroom with "good" acoustics.

Three BRIRs were created: 1) RT = 0 s, where the impulse response was truncated before the first reflection in order to remove the effect of reverberation while preserving the acoustic characteristics of the room (pseudo-anechoic (PA), 2) RT = 0.4 s at a distance of 2 m from the speaker (R2m), and 3) RT = 0.4 s at a distance of 6 m from the speaker (R6m). The PA BRIR was designed to mimic the near-anechoic conditions of clinical testing, where reverberation and distance are minimized. The reverberant R2m BRIR was created to represent the position of a child sitting in the first row of a small classroom, outside of this room's critical distance (1.6 m), where direct signal and early reflections are dominant. The early reflections which occur within 30 – 50 ms of the onset of the direct signal, are large in amplitude, and are thought to add "coloration" to and reinforce the direct sound via their integration with the direct signal (Haas 1972; Nabelek 1980; Wallach et al. 1949). R6m BRIR was meant to represent the position of a child sitting in the last row, close to the back wall, in the reverberant field, where the direct signal is attenuated and early and late reflections contribute more equally to the signal at the listeners' ears. The late reflections are smaller in amplitude, more closely spaced than the early ones, and cause temporal smearing of the sound which is detrimental to speech perception.

Stimuli

Bamford-Kowal-Bench (BKB) sentences (Bench et al. 1979) were used in this study. The 320 five- to seven-word sentences were designed to be suitable for first-grade-level children. Each sentence contained 3 to 4 keywords. Three hundred (284 3-keyword and 16 4-keyword) sentences were used in this experiment. In order to maintain an equal number of keywords per condition as well as proper sentence list randomization and condition counterbalancing, only three keywords were scored in each sentence. The personal pronoun at the beginning of a sentence was typically excluded from scoring in the 4-keyword sentences. Ten 4-keyword sentences were used during the practice session, but were not scored.

The BKB sentences, spoken by an adult native female speaker of American English, were recorded in a double-walled sound-treated room. A condenser microphone (Shure Beta 53, Niles, IL) was placed approximately 2 inches from the speaker's mouth and routed to a preamplifier (Shure M267, Niles, IL). Test stimuli were digitally recorded (LynxTWO-B, Costa Mesa, CA) at a sampling rate of 44.1 kHz (32 bits, monaural). The talker was instructed to say each sentence in a clear and natural voice. One of the authors listened to all recorded sentences to ensure that their enunciation was clear and without exaggeration. The process was repeated on a small random selection of the sentences by another author.

Acoustics of the modeled virtual classroom were incorporated into the sentences by convolving the latter with BRIRs described in the preceding paragraph using a standardized linear convolution algorithm in MATLAB® software ("MATLAB" 2011). The convolved stimuli were down-sampled (22.05 kHz, 16 bits). Pauses preceding and following the sentences were removed, and a 500 ms pause was inserted at the beginning of each sentence. This allowed for a 500 ms delay of the sentence onset when it was presented in noise. Although it lead to the removal of level cues normally associated with varying speaker-listener distance, sentence amplitudes were normalized using MATLAB software to equate rms level, in order to control for possible confounding effects of stimulus level. In all conditions, stimuli were presented at 65 dB SPL. Consequently, holding the presentation

level constant allowed an examination of the relative contributions of reverberation and noise to sentence intelligibility.

Noise

Long-term average speech spectrum (LTASS) noise was created in MATLAB software by adding multiple sinusoids (1-Hz spacing) in 1/3-octave bins in the 63 – 16000 Hz range. The shape of the LTASS was based on the combined male and female spectra for speech produced at 70 dB SPL (Byrne et al. 1994). Fifty 3-sec samples of the noise were created, one for each sentence in a given condition. A squared-cosine ramp was applied to the initial and final 50 ms of each noise sample. All LTASS noise samples were convolved with the three BRIRs to incorporate the simulated room acoustics described above.

Modulated LTASS (modLTASS) noise was created by multiplying the LTASS noise segments by the temporal envelope of female speech. The temporal envelope was extracted from a recording of the phonetically-balanced "rainbow passage" (Fairbanks 1960) by applying a Hilbert transform, and a 20-Hz low-pass finite impulse response filter (50000th order, Blackman-Harris). Fifty randomly selected 3-sec segments of the temporal envelope were mixed with the LTASS noise segments. Prior to envelope extraction, pauses longer than 300 ms were removed from the passage, in order to retain only the shorter dips present in the running speech. All signal processing was conducted using MATLAB software. As was done with the sentence materials and LTASS noise, the 50 modLTASS noise samples were convolved with the three BRIRs.

The rainbow passage and BKB sentences were produced by the same female speaker. The analysis of LTASS and modulated LTASS in the frequency domain revealed matched spectra of the two noise types. The frequency spectrum of a few randomly selected BKB sentences showed similarity to that of each noise type (with the exception of the frequency range below the fundamental frequency of the speaker, which was approximately 200 Hz). The use of a speech-shaped noise which approximates the spectrum of the speech signal allowed the effective SNR to be maintained across all frequency bands of the speech. When listening in other noises (e.g., white noise) the SNR typically varies with frequency (Boothroyd 2004).

Test Materials

The BKB sentences were randomly divided into six 50-sentence sets. Each sentence set was assigned to one of the following six conditions: 1) PA + LTASS, 2) R2m + LTASS, 3) R6m + LTASS, 4) PA + modLTASS, 5) R2m + modLTASS, and 6) R6m + modLTASS. For the ease of testing, six sentence lists were created, each containing 300 sentences. The six fifty-sentence sets were counterbalanced across the six conditions. The order of the sentences per condition in each list was randomized prior to the presentation, and remained fixed for all participants. A modified Greco-Latin square design allowed for the counterbalancing of the order of the conditions across presentations, so that no condition occurred in the same place or contained the same set of sentences across the six lists. Because each age group consisted of twelve participants, six of a particular age, each list appeared twice within each age group, but only once per age. None of the sentences appeared more than once per participant. Test stimuli were generated with speech and noise co-located to maintain an equal amount of spatial release from masking in the pseudo-anechoic and reverberant conditions. A reduction in the spatial release from masking has previously been reported in the presence of reverberation (Kidd et al. 2005; Plomp 1976). This approach may not represent typical locations of stimuli and noise sources and theoretically reflects the most difficult spatial listening scenario found in in a classroom. Nevertheless, it allows for better experimental control in determining the contribution of reverberation in the presence of

steady-state and modulated noise relative to the near-anechoic environments of audiology clinics.

The BKB sentences were presented at five SNRs: +10, +5, 0, −5 and −10 dB. This range of SNRs is typical of many classrooms (Blair 1976; Crandell and Smaldino 2000; Knecht et al. 2002; Picard and Bradley 2001; Sato and Bradley 2008). The order of SNRs was fixed across conditions, and was arranged in order from most favorable (+10 dB) to least favorable (−10 dB) to avoid frustration by the youngest participants. Ten sentences (with a total of 30 keywords) served to compute the percentage of correct responses for a given SNR in each condition.

Procedure

Testing took place in a single session for all participants. Hearing screening was performed first (children and adults), followed by the BBQSP and PPVT-IV (children only) in order to determine eligibility. Following hearing, speech and language screenings, data were collected from individual listeners using the BKB sentences in a double-walled IAC sound-treated booth (Bronx, NY) located in a quiet laboratory. Prior to the onset of data collection, calibration was performed to set the stimuli to the desired level. Calibration was checked daily to ensure that these levels did not change throughout the experiment. Stimulus presentation and response acquisition were controlled by a locally developed computer program (Behavioral Auditory Research Tests, BART). Using a personal computer, sentence stimuli were presented at an rms level of 65 dB SPL via Etymotic Research ER-6i isolator earphones (Elk Grove Village, IL). This level was selected to approximate the level of raised voice of a teacher in a noisy classroom (Pearsons et al. 1976), and because it could be held constant over a range of SNRs. Noise level varied from 55 to 75 dB SPL to allow for testing at SNRs ranging from +10 to −10 dB in 5 dB steps. Participants were instructed that they would hear sentences in the presence of noise, and that while many sentences may be easy, some would be difficult to hear. They were asked to repeat the sentences to the best of their ability, and were encouraged to repeat as much as they could hear and/or to guess if they were not sure. Stimulus presentation and response scoring were controlled online.

Ten BKB sentences were used to familiarize the participants with the task prior to data collection. Three of the six conditions (150 sentences, 15 – 20 minutes) were then tested, followed by a 1 – 5 minute break for adults or a 10 – 20 minute break for children. Afterward, the remaining three conditions (150 sentences, 15 – 20 minutes) were presented. Children were allowed to take as many breaks as they chose, but none requested additional breaks. In order to maintain motivation and interest in the task, visual reinforcement (pictures of cute animals, picturesque scenery, etc.) on a computer monitor accompanied the presentation of the stimuli. Each reinforcement image remained on the screen until it was replaced with a different one following every third response. This reinforcement was not contingent on the accuracy of the responses, and was used only to maintain participants' interest in the task. The total testing time ranged from 30 minutes for most adults, to 1 hour for some children.

Data Analysis

Average speech recognition was calculated for each participant, at each of the five SNRs, for each of the six conditions. Data were analyzed using a four-way mixed-model repeated measures analysis of variance (ANOVA), with type of masking noise, reverberation condition (R_C) and SNR as the within-subject factors, and age group as the between-subjects factor. The masking release scores (i.e., the improvement in performance in modulated relative to stationary noise) were analyzed using a three-way mixed-model repeated measures ANOVA, with R_C and SNR as within-subject factors, and age group as a between-

subjects factor. For each participant, the slope of the cubic polynomial function at 70% correct was computed for each of the six conditions, and analyzed using a three-way mixed-model ANOVA, with type of masking noise and R_C as within-subject factors, and age group as a between-subjects factor. Finally, the SNR required for 95% correct performance per age group was estimated from the average of individual cubic polynomial fits for each of the six conditions, and was compared across age groups. This near-optimal level of performance was selected to allow for direct comparisons to previous research.

Results

Speech recognition in Noise and Reverberation

Average speech-recognition scores as a function of age group are plotted in Fig. 1. Performance in LTASS noise is shown on top, while performance in modulated LTASS noise is shown below. The four columns show the results for SNRs ranging from −10 to +5 dB, while the parameter in each panel is R_C. Inspection of the figure reveals an overall improvement in speech recognition scores with increasing SNR and age, regardless of the noise type. At the two poorest SNRs, performance in the two reverberant conditions (R2m and R6m) was worse than that in the PA condition, particularly for the modLTASS condition at −10 dB SNR. A ceiling effect was present for all conditions and age groups at +5 dB SNR, and at 0 dB SNR for most conditions and age groups. Exceptions at 0 dB SNR are the three youngest groups in the R6m LTASS condition. Because +10 dB SNR data did not differ from +5 dB SNR data for any condition, they were not plotted. These patterns are supported by the following statistical analyses.

Prior to statistical analyses, mean percent correct scores were transformed to rationalized arcsine unit (RAU) scores to make them more suitable for statistical analysis (Studebaker 1985). Mauchley's Test of Sphericity detected a violation of the homogeneity of variance (R_C, SNR, and their interaction), and hence a Greenhouse-Geisser correction was applied to a four-way mixed-model repeated measures ANOVA using the RAU-transformed scores of speech recognition performance. Due to the ceiling effect and lack of variance in PA conditions at +5 and +10 dB SNR, only −10, −5, and 0 dB SNR were included in this analysis. Statistical analyses are summarized in Table 2.

Main effects and two-way interactions should be interpreted with caution because of the significant three-way interactions. Tukey Honestly Significant Difference (HSD) post hoc testing was conducted for the statistically significant three-way interactions. For the R_C, SNR and age interaction, performance of the 7 – 8 year olds was significantly poorer than that of young adults in all reverberation conditions at −5 and −10 dB SNR, and at 0 dB SNR in the R6m condition. Additionally, the youngest group of listeners had significantly poorer speech recognition compared to the 11 – 12 year olds in the PA and R2m condition at −5 dB SNR, as well as that of 13 – 14 year olds at −10 dB SNR in the PA condition and at −5 dB SNR in the R2m condition. The group of 9 – 10 year olds showed statistically poorer performance than that of adults at −10 dB SNR in the PA and R2m conditions, and at −5 dB SNR in the R6m condition. Within each group of children, performance in the PA condition was statistically better than that in the R2m and R6m conditions at −5 and −10 dB SNR, but did not differ significantly between the R2m and R6m conditions. For the youngest group of participants, a significant improvement was also observed between the R6m and PA conditions at 0 dB SNR. In contrast, speech recognition for the adults was significantly better in the PA condition than in R2m and R6m only at −10 dB SNR. Orthogonal contrasts for the R_C, SNR and age interaction obtained from post hoc testing are presented in Tables 3 and 4. The type of masking noise, R_C, and SNR interaction revealed that speech recognition in modulated LTASS noise was statistically better than that in LTASS noise for the PA and R6m conditions at −5 and −10 dB SNR, and for the R2m condition at −10 dB SNR. Within

each masking noise type, performance in the PA condition was significantly improved in comparison to the R2m and R6m conditions at −5 and −10 dB SNR, but no statistically significant differences were observed between the two reverberant conditions. Additionally, when listening in competing modulated LTASS noise, performance was significantly worse in the R2m condition at 0 dB SNR and the R6m condition at −10 dB SNR, relative to the PA condition at the same SNR. Finally, when LTASS noise was the competing distracter, speech recognition was significantly poorer in the R6m condition than in either the PA or R2m conditions at 0 dB SNR. Tables 5 and 6 show the orthogonal contrasts for the three-way interaction between the type of masking noise, R_C, and SNR, obtained from post hoc testing.

In order to determine whether performance in either reverberant condition (R2m, R6m) could be predicted from the PA condition, Pearson product-moment correlation coefficients were computed using RAU-transformed scores. Results revealed that a decrement in performance due to the combination of noise and reverberation at −10, −5 and 0 dB SNR could not be accounted for by the scores obtained in the PA condition (average $r = 0.37$, $SD = 0.19$).

Release from Masking

Release from masking resulting from the ability to access speech during pauses in the modulated noise was determined by calculating the difference between speech recognition in the LTASS and modLTASS conditions. Figure 2 illustrates the magnitude of masking release as a function of age, for −10, −5, 0 and +5 dB SNR. Positive scores indicate masking release, while negative scores represent an increase in masking. Masking release is highest at the least favorable SNR (−10 dB), where it is reduced with the addition of reverberation (more so for the R2m condition), and varies with both age and SNR. Specifically, masking release increases with age at −10 dB SNR in the two reverberant conditions, while the inverse was observed at −5 dB SNR, where it appears to decrease as a function of age, and where the largest benefit from listening in the dips was observed for the R6m condition. Release from masking is absent due to the ceiling effect in each RT condition at 0 and +5 dB SNR, and at −5 dB SNR for all but the youngest group in the PA condition. At 0 dB SNR, the 9 – 10 and 11 – 12 year olds show small amounts of masking increase in the presence of modulated noise in the R2m condition. A masking release of similar magnitude was present for the R6m condition for those same groups. As in the previous figure, +10 dB SNR did not differ from +5 dB SNR for any condition, and therefore was not plotted.

A three-way mixed-model repeated measures ANOVA was carried out for the masking release scores. Mauchley's Test of Sphericity detected a violation of the homogeneity of variance (R_C, SNR and their interaction), and hence a Greenhouse-Geisser correction was applied to the ANOVA. As was done in the analysis described above, only −10, −5, and 0 dB SNR were included, due to the ceiling effects and lack of variance in all reverberation conditions at +5 and +10 dB SNR. A summary of the results of the analysis is shown in Table 7.

The repeated measures ANOVA revealed statistically significant two-way interactions which help interpret the driving forces behind the main effects. Tukey HSD post hoc testing demonstrated that, for the SNR and age group interaction, 7–8 year olds had significantly less masking release than 11–12 and 13–14 year olds at −10 SNR ($p < 0.05$). At an SNR of −10 dB, masking release differed significantly across the three reverberation conditions (PA > R6m > R2m; $p < 0.05$), while at −5 dB SNR, performance in the R6m condition was significantly better than in the R2m or PA conditions ($p < 0.02$), but did not differ between the PA and R2m conditions ($p = 1$). Within each reverberation condition, masking release increased significantly with decreasing SNR (−10 > −5 > 0 dB; $p < 0.02$). The only

exception was the PA condition where a statistically significant difference was present between −10 and −5 dB SNR ($p < 0.001$), but not between −5 and 0 dB SNR ($p = 0.21$).

Performance-Intensity Slope

To predict speech recognition performance for a greater number of SNRs than tested, the individual mean data were fitted with a cubic polynomial. Equations for polynomial orthogonal fits are useful because the goodness of fit can be described, any point on the function can be calculated precisely, and the first derivative of a polynomial defines the slope of a function at any selected point (Wilson and Strouse 1999). The cubic polynomial was selected in this study because it can closely approximate the S-shape of a PI function derived from speech recognition performance. A goodness-of-fit analysis between the observed scores and the scores predicted by the polynomial regression at −10 and −5 dB SNR revealed statistically significant Pearson product-moment correlation coefficients (average $r = 0.97$, $SD = 0.04$), and an average deviation (AD)[1] of 1.47 ($SD = 0.72$). A goodness-of-fit analysis was not performed for higher SNRs, where the slope of the polynomial function was not stable and performance often reached ceiling. The predicted speech recognition performance (in percent correct) as a function of SNR for each age group is presented in Fig. 3. The three columns represent the reverberation conditions (PA, R2m, and R6m), while the top and bottom rows show predicted scores for the LTASS and modulated LTASS noise, respectively. Predicted speech recognition performance increases as a function of age, regardless of the R_C and masking noise type. In the PA condition, the predicted SNRs required for a score of 90% correct do not differ between the three oldest groups of children and the group of young adults. The separation in predicted performance between the above mentioned groups occurs at less favorable SNRs than that for the youngest group of children. In the R2m and R6m conditions, the predicted speech recognition scores for all children (regardless of group) are lower than those of the adult group, until asymptotic performance is reached. In all but one case where the combination of R_C and SNR resulted in sub-optimal performance, the youngest listeners required higher SNRs for equivalent performance relative to other participants. For the sentence and noise materials used in this study, the estimated SNRs required for approximately 95% correct performance were computed and are given in Table 2. On average, the youngest children would need approximately 2.5 dB greater SNR to attain adult-like speech recognition performance in either PA condition. In contrast, in the R6m + LTASS condition, the SNR increase needed for 95% correct performance by the youngest group relative to the adults was 4.5 dB. The estimated SNRs required for optimal performance increase with decreasing age and with the addition of reverberation. The slope of the PI function, which represents the shift in predicted performance with a given change in SNR, differs across conditions. For example, it appears to be greater for the LTASS noise than for the modulated LTASS noise. Additionally, under LTASS noise, all groups of children show a decrease in slope with the addition of reverberation and distance (PA > R2m > R6m). When modulated LTASS noise was used as the competing masker, this was no longer apparent. In fact, for the two oldest groups of children and the adults, this relationship was reversed (R2m = R6m > PA).

---

[1]

$$AD = \sum_{i=1}^{N} \frac{|Ei - Oi|}{N - 1}$$

N – number of observations

Ei – predicted value of case i

Oi – observed value of case i

The slopes of the polynomial fit at 70% correct predicted performance are plotted for each reverberant condition as a function of age group in Fig. 4. The left panel shows the slopes for predicted performance under LTASS noise, while the right panel shows the slopes for the modulated LTASS noise. The slopes in the LTASS condition are steeper than those in the modLTASS condition for all groups of children. In the LTASS condition, there is a decrease in the steepness of the slopes of the predicted PI function with increasing reverberation and distance (PA > R2m > R6m). Additionally, there appears to be little change in the slope as a function of age in the PA + LTASS condition, while in the R2m + LTASS and, even more so in the R6m + LTASS condition, the data show an overall increase in slope with increasing age. A similar pattern of slope change in the two reverberant conditions is also apparent in modulated LTASS noise, even though the magnitude of the slope is similar for the two reverberant conditions (R2m = R6m). Contrary to trends in other conditions, in the PA + modLTASS condition, there is an overall decrease in slope with increasing age. Namely, the slope in the PA condition increases only from 7–8 to 9–10 years, and then decreases with increasing age.

A three-way mixed-model ANOVA with repeated measures was conducted to quantify the magnitude of slopes of the polynomial fit at 70% correct predicted speech recognition performance for all age groups across the reverberation and masking noise conditions. Specifically, in the PA + modLTASS condition, the slopes at 70% correct could not be calculated for seven out of 60 participants (with no more than two per age group) due to near-optimal performance at −10 dB SNR. Because it had no measurable influence on the overall means or variance for all age groups in all conditions, a mean imputation approach was chosen to substitute for the missing data. Consequently, the arithmetic averages per age group were computed and used to replace the missing data points in the PA condition. The results of the analysis are summarized in Table 8.

The repeated measures ANOVA revealed statistically significant two-way interactions which help to interpret the main effects. Tukey HSD post hoc testing revealed that 7–8 year olds had significantly shallower slopes than adult listeners in the R2m ($p < 0.05$) and R6m ($p < 0.001$) conditions, and 11–12 year olds in the R2m condition ($p < 0.01$). There were no statistically significant differences between the slopes of the predicted PI functions across the three reverberation conditions within different age groups, with the exception of 9 – 10 year olds, where the slope was significantly steeper in the PA than the R6m condition ($p < 0.02$). For the masking noise and R_C interaction, slopes of the predicted PI functions were significantly steeper under LTASS when compared to modulated LTASS noise in all reverberation conditions. In the LTASS condition, the slope decreased significantly with increasing reverberation and distance (PA > R2m > R6m; $p < 0.05$), while in the modLTASS condition a different, and almost opposite, trend emerged (R6m = R2m > PA; $p < 0.05$).

Discussion

The present study was conducted to document the developmental effects of masker noise type, reverberation and SNR on speech recognition. It is the first study where the magnitude of masking release in stationary noise was quantified across age in non-reverberant and reverberant listening conditions. Likewise, it is the first effort to investigate the change in predicted performance at 70% correct in a reverberant environment in the presence of modulated and stationary noise for elementary school aged children and young adults.

The results of this study show that type of masker, reverberation, SNR and age all affect speech recognition. In all conditions, performance improved as a function of age at the two poorest SNRs. In all, these results support previous findings that children need more

favorable SNRs than adults for equivalent speech recognition in reverberant and noisy listening conditions (Bradley and Sato 2008; Johnson 2000; Klatte et al. 2010b; Neuman and Hochberg 1983; Neuman et al. 2010; Yang and Bradley 2009). Somewhat unexpectedly, at −10 and −5 dB SNR, modulated LTASS noise affected performance in the R2m condition more negatively than in the R6m condition. This reversal in performance relative to performance in LTASS noise might be explained by the presence of direct signal and high-amplitude early reflections in the R2m condition, and the reduced amplitude of the direct signal in the R6m condition. Specifically, while the integration of early reflections with the speech signal has been reported to increase the robustness of speech (Haas 1972; Yang and Bradley 2009), a similar integration of direct signal and early reflections with modulated LTASS noise in the R2m condition could cause the peaks in the noise waveform to become higher in amplitude and longer in duration, causing the dips in the noise to become shorter relative to the modulated noise in the R6m condition. In the R6m condition, the absence of the direct signal in the presence of early and late reflections would result in shallower noise peaks relative to the R2m condition. In the present study, the reduction in the amplitude of the modulated masker peaks resulted in a less pronounced decrease in performance. A measure of C7 was computed to approximate the direct-to-reverberant energy ratios for the two reverberant conditions. C7 is expressed in dB and is the ratio of energy in the first 7ms from the direct sound to the end of the impulse response. The results revealed that C7 in the R2m condition was, on average, 3.12 dB greater than that in the R6m condition, indicating a higher level of the direct signal in the R2m condition.

The interaction between masking release and noise and reverberation is apparent from the present results. On average, participants performed more poorly under LTASS than modulated LTASS noise, but only at negative SNRs. This performance gap widened as SNR decreased. The comparison of results for the two maskers revealed that release from masking decreased with increasing SNR and the addition of reverberation. In similar investigations, Stuart (2005) and Stuart et al. (2006) found that young children benefited from masking release on par with adults in anechoic conditions. However, the addition of mild reverberation in the present study revealed that the listening conditions typical of acoustically "good" classrooms (ANSI S12.60-2010/Part 1 2010a) reduce the amount of masking release. The performance in the PA condition improved 48–56% as a result of masking release, whereas only 19–30% (R2m) and 26–36% (R6m) improvements could be seen when reverberation was introduced at a −10 dB SNR. Furthermore, with the addition of reverberation at −10 dB SNR, masking release in the modulated noise showed incremental increases as a function of age until 11–12 (R2m) or 13–14 (R6m) years, a finding that was not apparent in the anechoic condition in this or previously-reported studies.

In this study, speech recognition decreased only at negative SNRs, where it was adversely affected by the addition of reverberation representative of "good" classroom acoustics (ANSI S12.60-2010/Part 1 2010a). Negative SNRs are typical of many classrooms (Blair 1976; Crandell and Smaldino 2000; Knecht et al. 2002; Picard and Bradley 2001; Sato and Bradley 2008). In the current study, speech recognition decreased in reverberation relative to the non-reverberant environment, more so in the modLTASS than the LTASS conditions. The discrepancy between the reduction in speech recognition in the two masker conditions could be partially attributed to a possible floor effect at the lowest SNR in the LTASS conditions. However, the fact that the floor effect was not present in the group of young adults, and possibly 13–14 year olds, suggests that reverberation affects speech perception in modulated noise to a greater extent than when a stationary noise masker is used. This is to be expected, as the dips in amplitude-modulated noise become temporally smeared by reverberation, thereby reducing the effective SNR (steady-state noise should not be affected in a similar manner, as it lacks temporal fluctuations.) It has been suggested that the increased susceptibility to temporal smearing, such as that introduced by reverberation, may

be the result of variable and often immature development of temporal resolution in children (Hartley et al. 2000; Jensen and Neff 1993). This developmental immaturity could result in the reduction in masking release in the reverberant conditions as seen in the current study. The data from the children and adults who participated in this study are similar to those obtained by George et al. (2008) for a group of adult listeners, who found that the benefit of listening in the dips of a modulated noise was minimal when both speech and noise were reverberated and spatially co-located. Similarly, in the present study the masking release benefit was reduced in comparison to that observed when stimuli and noise were presented in an anechoic environment. However, in this study masking release was still observed at an RT of 0.4 sec. In contrast, George et al. (2008) reported no release from masking once RT reached 0.25 sec. Differences in the methods used by George et al. (speech recognition threshold for short sentences measured in the presence of maximum-length sequence noise) may have contributed to the discrepancies in findings. In the present study, the release from masking across conditions was lower for adults than for the two older groups of children. Given superior overall speech recognition by adults, it appears that the reduction in release of masking was not caused by the inability to use the glimpses in the modulated noise, but rather by lower susceptibility to masking by stationary noise at the lowest SNR.

In the present study, speech recognition decreased with decreasing SNR, but not until the signal and noise were at equivalent levels. At 0 dB SNR, the small but measurable effects of age became apparent in the R6m condition, where speech recognition improved as a function of age group until 13–14 years, when it became adult-like. In the PA condition at −5 dB SNR, only the youngest group of participants had poorer performance than adults. Once reverberation was introduced, however, all groups of children performed more poorly. In contrast, the scores of adults were similar across the three reverberation conditions, especially under modulated LTASS noise. At −10 dB SNR, the effect of age was more pronounced, and it extended to all reverberation conditions in all age groups. Consequently, the present findings show that children's ability to understand short sentences is unaffected until speech is presented at negative SNRs, in both LTASS and modulated LTASS noise. This is in contrast to previous work which has shown that higher SNRs are necessary for optimal or near optimal speech recognition. The estimates of SNR needed for 95% correct performance found by Bradley and Sato (2008), given the 0.3 – 0.7 s RT range, suggested that eleven-year olds would require +8.5 dB, eight-year olds would require +12.5 dB, and six-year olds +15.5 dB SNR. Moreover, the scatter in speech recognition scores suggests that a large number of students would have more difficulty understanding speech in noisy and reverberant conditions than the mean data would imply. In fact, SNRs of +15 (11 yrs), +18 (8 yrs), and +20 dB (6 yrs) would be required for 80% of the children to achieve near-optimal performance (95% correct). The room acoustics used in the present study represented an acoustic approximation of the average of the 30 rooms where Bradley and Sato (2008) conducted their experiment. Hence, large differences in predicted SNRs needed for 95% correct performance reported in the present study relative to the Bradley and Sato investigation, are more likely the result of stimulus and masker selection, rather than the differences in acoustic conditions. In the Bradley and Sato study, a closed set of monosyllabic rhyming words were presented in HVAC noise. In contrast, sentence materials presented in LTASS and modulated LTASS noise were used in the present study. In close agreement with the findings of Bradley and Sato, Neuman et al. (2010) estimated that when reverberation time was 0.8 sec, nine- and twelve-year-olds would require a +15 dB SNR in order to achieve 95% correct performance. Six-year-olds, however, would require that same 15 dB SNR to achieve the same level of performance when the reverberation time was at the ANSI-recommended upper limit of 0.6 sec (ANSI, 2009a). As in the present study, Neuman and colleagues used BKB sentence test materials. However, unlike the present investigation, the sentences were spoken by a male speaker, and the masker competition was multi-talker babble, which could potentially explain the differences in both the results and SNR

predictions. The estimated SNRs for 95% correct accuracy for test materials used in this study ranged from −5 to +1.75 dB SNR (see Table 9).

Estimates of the slope of the performance-intensity function at 70% correct speech recognition revealed that the degree of change in performance with a 1 dB shift in SNR was less when the masking noise was modulated rather than steady-state. This suggests that children and adults can obtain a sufficient amount of acoustic information from the brief pauses in the noise to ameliorate the decline in speech recognition with decreasing SNR relative to a non-modulated masker. In the LTASS condition, the addition of reverberation reduced the slope of the predicted PI function, suggesting that a larger increase in SNR was needed than in the PA condition, for equivalent improvement in both reverberant conditions (more so in the R6m than the R2m condition.) In both masker conditions, the performance slope increased as a function of age in the two reverberant conditions. This finding is in agreement with the analysis of the PI function slopes by Neuman et al. (2010), which revealed that a larger increase in SNR was needed for equivalent improvement in performance in younger versus older children under similar reverberation conditions. In the present study, the PA + modLTASS condition for 11–12 and 13–14 year olds and adults showed an inverse trend for the slope of the PI function. Given the higher overall performance at similar SNRs in this condition, this finding suggests a more gradual decrease in performance with an equivalent negative SNR shift than that of the two youngest groups of children. This implies that the effective use of the dips in an anechoic competing masker improves with age.

Conclusions

The present study describes speech recognition of normal-hearing school-aged children and adults under reverberation and noise. The results document the negative effects of reverberation and noise, and reveal an overall improvement in performance with increasing age. The youngest children, whose classrooms tend to be noisiest (Picard and Bradley 2001), were most affected by the combination of noise and reverberation. The present study was also the first attempt to document children's release from masking in a reverberant environment. The findings indicate that reverberation significantly reduces the benefit of glimpses in the noise caused by modulations, and that this decrement is greatest for the youngest children. The use of LTASS and modulated LTASS noise allowed for experimental control of effective SNR across frequency bands, and eliminated the possible effects of informational masking. The decrease in speech recognition due to reverberation could not be predicted from testing in non-reverberant environments. Hence, this study supports previous recommendations that reverberation and noise should be incorporated into clinical speech recognition testing in order to provide an accurate predictor of real-world performance (Nabelek and Robinette 1978; Neuman et al. 2010; Yacullo and Hawkins 1987), especially for young children who are most susceptible to noise, reverberation and their combination.

Acknowledgments

References

American National Standards Institute. ANSI S12.60-2009/Part 2. New York: ANSI; 2009. Acoustical Performance Criteria, Design Requirements and Guidelines for Schools, Part 2: Relocatable Classroom Factors.

American National Standards Institute. ANSI S12.60-2009/Part 1. New York: ANSI; 2010a. Acoustical Performance Criteria, Design Requirements and Guidelines for Schools, Part 1: Permanent Schools.

American National Standards Institute. ANSI S3.6-2010. New York: ANSI; 2010b. Specifications for audiometers.

Bankson, N.; Bernthal, J. Bankson–Bernthal Test of Phonology. San Antonio, TX: Special Press; 1990. Bankson-Bernthal Quick Screen of Phonology.

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]

Bitan T, Cheon J, Lu D, et al. Developmental increase in top–down and bottom–up processing in a phonological task: an effective connectivity, fMRI study. J Cogn Neurosci. 2009; 21:1135–1145. [PubMed: 18702576]

Blair, J. Unpublished doctoral dissertation. Evanston, IL: Northwestern University; 1976. The contributing influences of amplification, speechreading, and classroom environments on the ability of hard-of-hearing children to discriminate sentences.

Boothroyd A. Room acoustics and speech perception. Semin Hear. 2004; 25:155–166.

Bradley JS. Speech intelligibility studies in classrooms. J Acoust Soc Am. 1986; 80:846–854. [PubMed: 3760338]

Bradley JS, Sato H. The intelligibility of speech in elementary school classrooms. J Acoust Soc Am. 2008; 123:2078–2086. [PubMed: 18397015]

Byrne D, Dillon H, Tran K, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994; 96:2108–2120.

Cameron S, Dillon H, Newall P. Development and evaluation of the listening in spatialized noise test. Ear Hear. 2006; 27:30–42. [PubMed: 16446563]

CATT-Acoustic™ [computer program] Version 8.0e. Gothenburg, Sweden: CATT; 2006.

Crandell CC, Smaldino JJ. Classroom acoustics for children with normal hearing and with hearing impairment. Lang Speech Hear Serv Sch. 2000; 31:362–370.

Dunn, L.; Dunn, D. Peabody Picture Vocabulary Test IV. Circle Pines, MN: American Guidance Service; 2006.

Etymotic Research. Bamford-Kowal-Bench Speech in Noise Test (Version 1.03). Elk Grove Village, IL: Etymotic Research; 2005.

Fairbanks, G. Voice and articulation drillbook. New York: Harper; 1960.

Finitzo-Hieber T, Tillman TW. Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. J Speech Hear Res. 1978; 21:440–458. [PubMed: 713515]

Gelfand SA. Optimizing the reliability of speech recognition scores. J Speech Lang Hear Res. 1998; 41:1088–1102. [PubMed: 9771631]

George EL, Festen JM, Houtgast T. The combined effects of reverberation and nonstationary noise on sentence intelligibility. J Acoust Soc Am. 2008; 124:1269–1277. [PubMed: 18681613]

Haas H. The influence of a single echo on the audibility of speech. J. Audio Eng. Soc. 1972; 20:146–159.

Hartley DE, Wright BA, Hogan SC, et al. Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children. J Speech Lang Hear Res. 2000; 43:1402–1415. [PubMed: 11193961]

Jamieson DG, Kranjc G, Yu K, et al. Speech intelligibility of young school-aged children in the presence of real-life classroom noise. J Am Acad Audiol. 2004; 15:508–517. [PubMed: 15484600]

Jensen JK, Neff DL. Development of basic auditory discrimination in preschool children. Psychol Sci. 1993; 4:104–107.

Johnson CE. Children's phoneme identification in reverberation and noise. J Speech Lang Hear Res. 2000; 43:144–157. [PubMed: 10668658]

Kidd G, Mason CR, Brughera A, et al. The role of reverberation in release from masking due to spatial separation of sources for speech identification. Acta Acust United Ac. 2005; 91:526–536.

Klatte M, Hellbrück J, Seidel J, et al. Effects of classroom acoustics on performance and well-being in elementary school children: A field study. Environ Behav. 2010a; 42:659–691.

Klatte M, Lachmann T, Meis M. Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. Noise Health. 2010b; 12:270–282. [PubMed: 20871182]

Knecht HA, Nelson PB, Whitelaw GM, et al. Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. Am J Audiol. 2002; 11:65–71. [PubMed: 12691216]

Kodaras M. Reverberation times of typical elementary school settings. Noise Control. 1960; 6:17–19.

Koehnke J, Besing JM. A procedure for testing speech intelligibility in a virtual listening environment. Ear Hear. 1996; 17:211–217. [PubMed: 8807263]

Martin FN, Armstrong TW, Champlin CA. A survey of audiological practices in the United States. Am J Audiol. 1994; 3:20–26.

MATLAB [computer program] Version 7.12.0.635. Natick, Massachusetts: The MathWorks Inc.; 2011.

Miller GA, Licklider J. The intelligibility of interrupted speech. J Acoust Soc Am. 1950; 22:167–173.

Nabelek, AK. Effects of room acoustics on speech perception through hearing aids. In: Libby, ER., editor. Binaural Hearing And Amplification. Chicago: Zenetron; 1980. p. 123-145.

Nabelek AK, Robinette L. Reverberation as a parameter in clinical testing. Audiology. 1978; 17:239–259. [PubMed: 666659]

Nabelek AK, Robinson PK. Monaural and binaural speech perception in reverberation for listeners of various ages. J Acoust Soc Am. 1982; 71:1242–1248. [PubMed: 7085996]

Neuman AC, Hochberg I. Children's perception of speech in reverberation. J Acoust Soc Am. 1983; 73:2145–2149. [PubMed: 6875100]

Neuman AC, Wroblewski M, Hajicek J, et al. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. Ear Hear. 2010; 31:336–344. [PubMed: 20215967]

Pearsons, K.; Bennett, R.; Fidell, S. Speech levels in various noise environments (Report No. EPA-600/1-77-025). Washington, DC: Environmental Protection Agency; 1976.

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

Plomp R. Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). Acustica. 1976; 34:200–211.

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]

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]

Stuart A, Givens GD, Walker LJ, et al. Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise. J Acoust Soc Am. 2006; 119:1946–1949. [PubMed: 16642807]

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

U.S.Access.Board. [Retrieved 7/29/2011] Implementing Classroom Acoustics Standards: a Progress Report. 2011. from http://www.access-board.gov/acoustic/.

Wallach H, Newman EB, Rosenzweig MR. The precedence effect in sound localization. Am J Psychol. 1949; 62:315–336. [PubMed: 18134356]

Wilson, R.; Strouse, A. Auditory measures with speech signals. In: FE, M.; WF, R., editors. Contemporary perspectives in hearing assessment. Boston: Allyn & Bacon; 1999. p. 21-99.

Yacullo WS, Hawkins DB. Speech recognition in noise and reverberation by school-age children. Audiology. 1987; 26:235–246. [PubMed: 3632478]

Yang W, Bradley JS. Effects of room acoustics on the intelligibility of speech in classrooms for young children. J Acoust Soc Am. 2009; 125:922–933. [PubMed: 19206869]
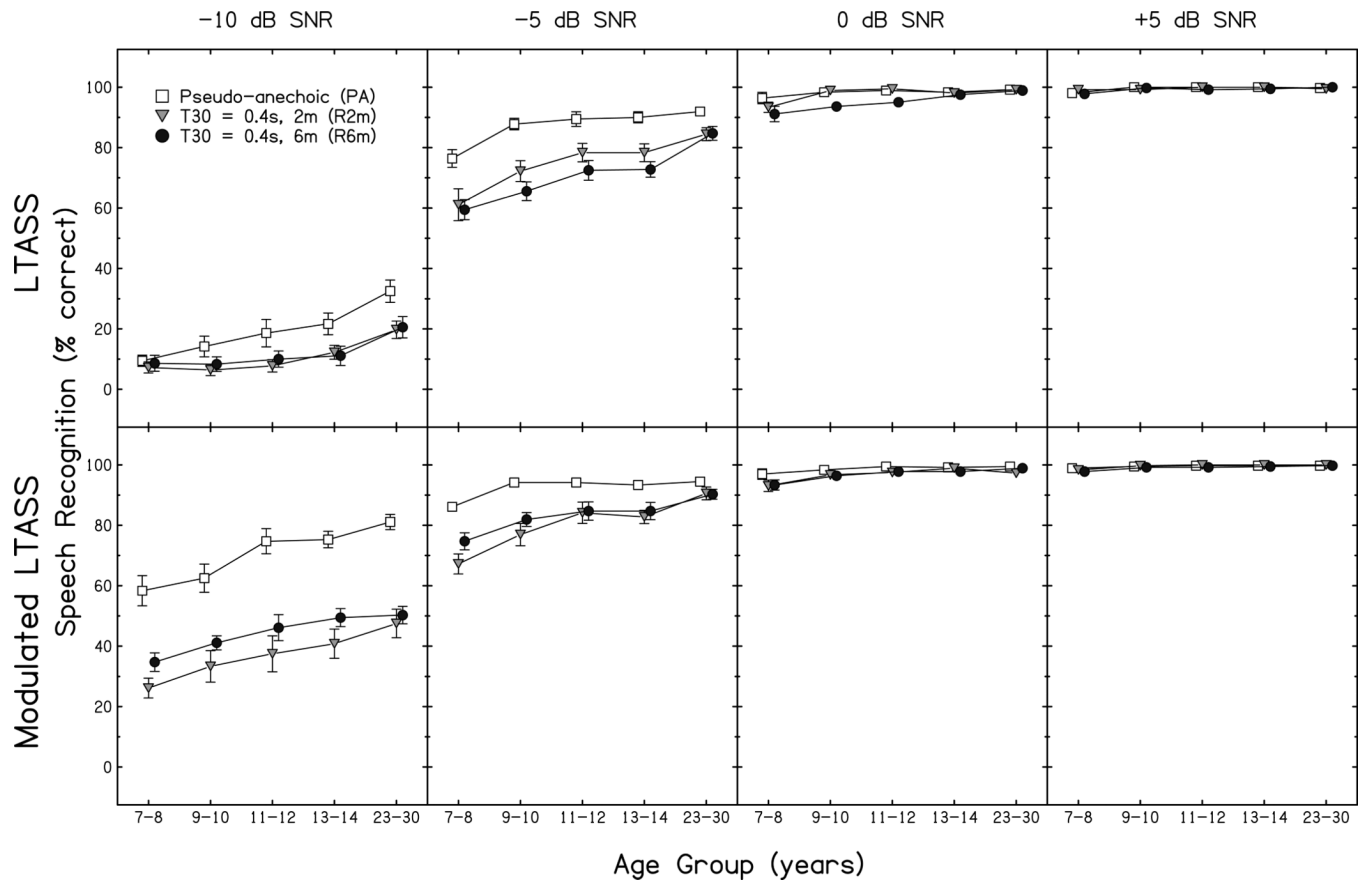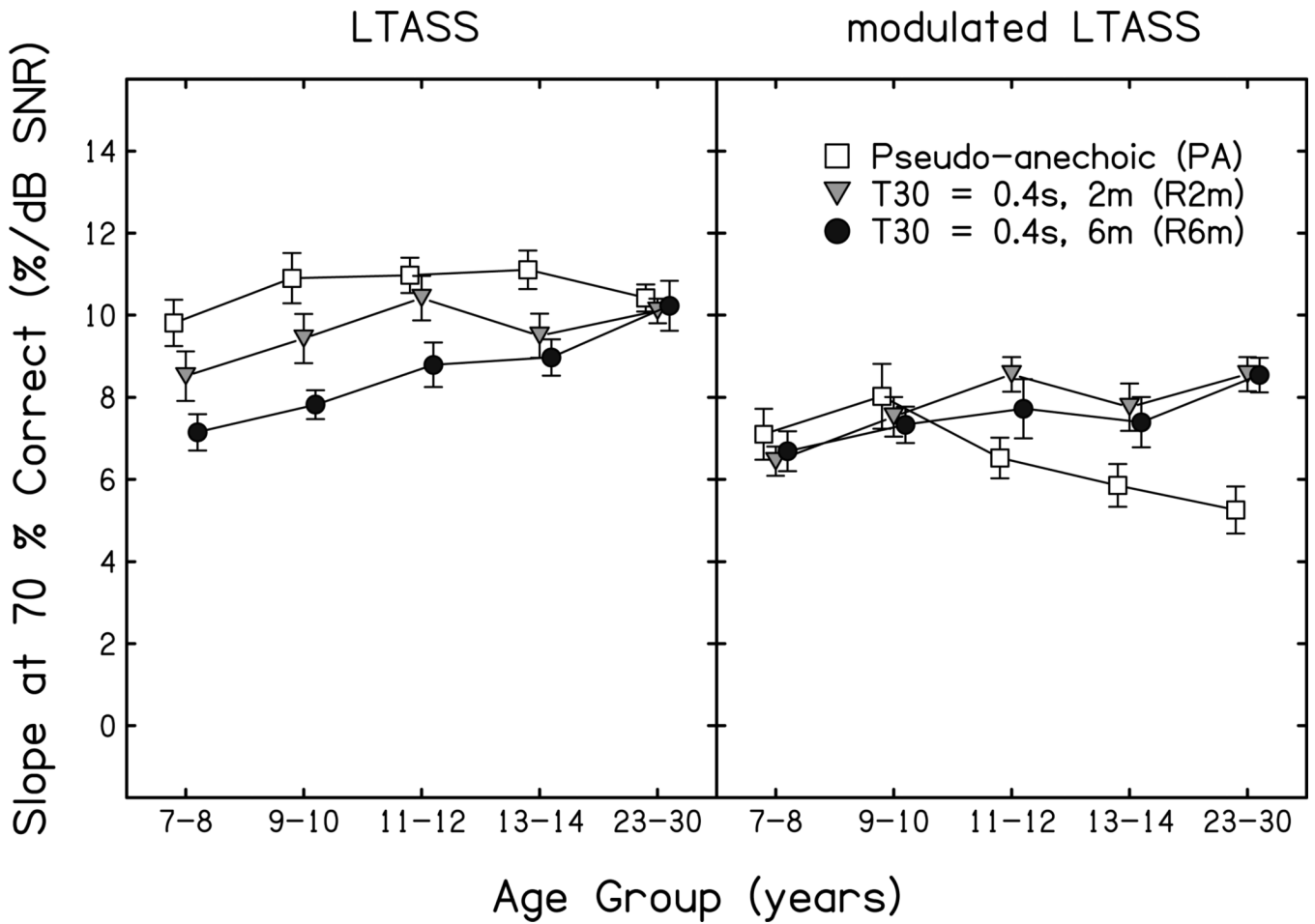
Figure 1.
Percent correct speech recognition (and standard error) as a function of age, with reverberation condition as the parameter. Different columns show the scores for the SNRs tested. Top and bottom rows show the scores obtained under LTASS and modulated LTASS noise, respectively.
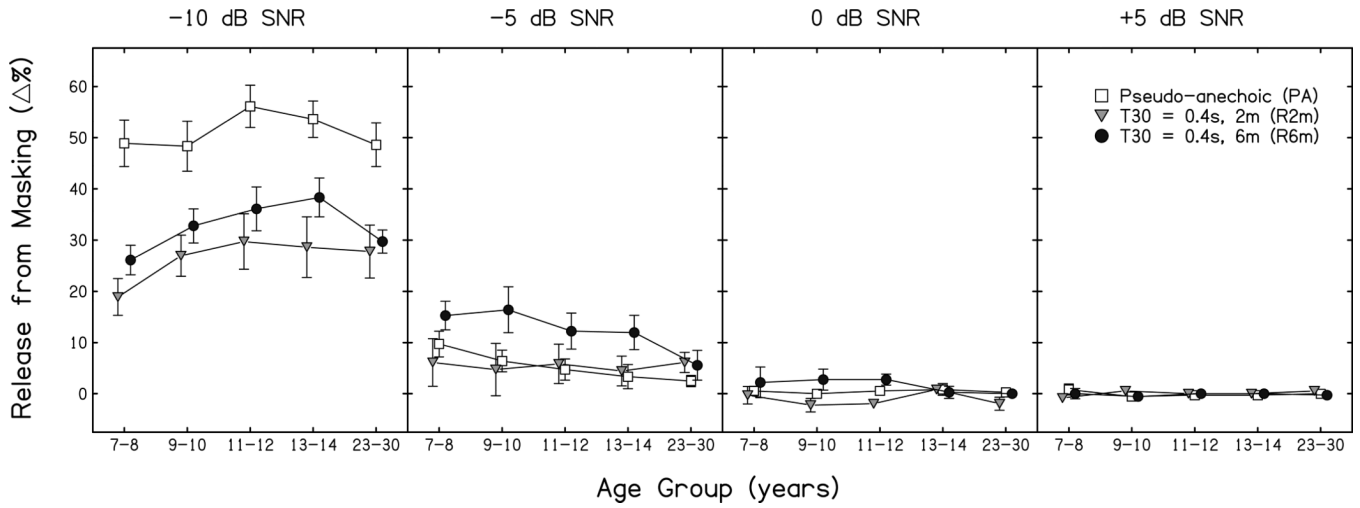
Figure 2.
Mean release from masking (and standard error) as a function of age, with reverberation condition as the parameter. Each panel show the masking release scores for a different SNR. The scores were obtained by subtracting mean individual scores obtained under LTASS noise from those obtained under modulated LTASS noise, to quantify the benefit from listening in the "dips" of the modulated noise.
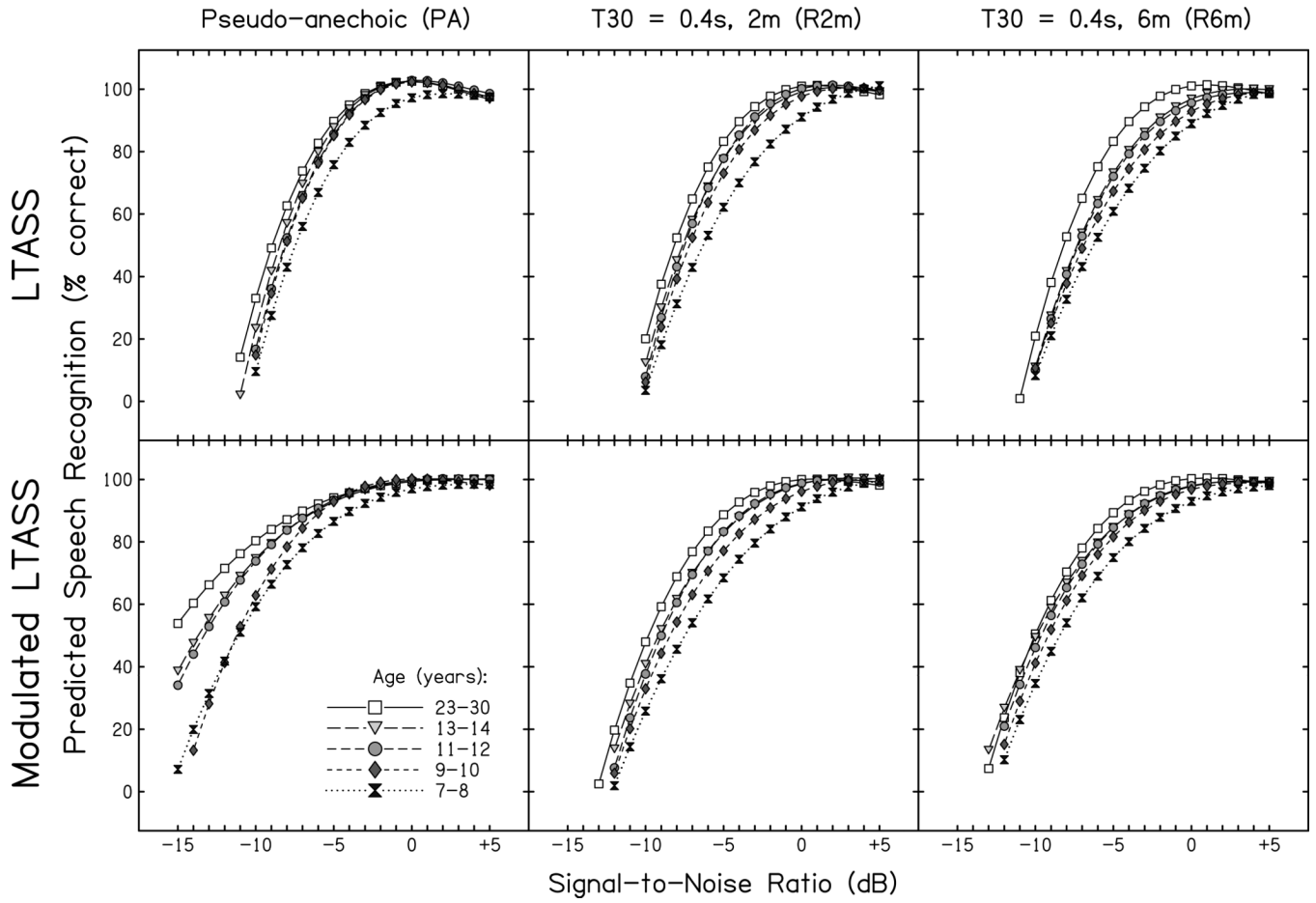
Figure 3.
Cubic polynomial fit of the individual data representing predicted speech recognition scores as a function of SNR, with age group as the parameter. Each column shows the predicted scores for one reverberation condition. Top and bottom rows show the predicted performance for LTASS and modulated LTASS noise, respectively.

Figure 4.
Mean slopes (and standard error) of the cubic polynomial at 70% correct predicted speech recognition performance. Left and right columns show the slopes of predicted performance for LTASS and modulated LTASS noise, respectively. The parameter within each panel is reverberation condition, as noted by the inset. The value of the slope represents the change in percentage correct predicted performance with a 1 dB change in the SNR.

Table 1

Reverberation time (in sec) at octave band frequencies for three conditions: pseudo-anechoic (PA), reverberant at a 2 m distance (R2m) and reverberant at a 6 m distance (R6m).

Condition	Frequency						
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	Average
PA	0.04	0.02	0.02	0.02	0.01	0.01	0.02
R2m	0.41	0.45	0.43	0.37	0.33	0.35	0.39
R6m	0.41	0.45	0.43	0.35	0.31	0.34	0.38

Table 2

Summary table of a four-factor mixed-model repeated measures analysis of variance. The differences in speech recognition as a function of Age Group, noise masker type (Noise), reverberation condition (R_C), signal-to-noise ratio (SNR) and their interactions were investigated. Degrees of freedom (df) and p-values are adjusted with the Greenhouse-Geisser correction where necessary (asterisks).

Factor	df	F	p	η_p^2
Age Group	4.00	19.34	< 0.001	0.585
Noise	1.00	719.49	< 0.001	0.929
Noise × Age Group	4.00	2.19	0.082	0.137
*R_C	1.72	125.99	< 0.001	0.696
*R_C × Age Group	6.88	0.54	0.803	0.037
*SNR	1.51	2566.68	< 0.001	0.979
*SNR × Age Group	6.05	2.35	0.038	0.146
*Noise × R_C	1.84	19.69	< 0.001	0.264
*Noise × R_C × Age Group	7.37	0.64	0.731	0.044
*Noise × SNR	1.98	379.47	< 0.001	0.873
*Noise × SNR × Age Group	7.93	0.79	0.613	0.054
*R_C × SNR	3.09	24.46	< 0.001	0.308
*R_C × SNR × Age Group	12.37	2.06	0.020	0.131
*Noise × R_C × SNR	3.59	10.76	< 0.001	0.164
*Noise × R_C × SNR × Age Group	14.35	0.57	0.892	0.040

Table 3

Orthogonal contrasts of age groups (post hoc Tukey HSD) indicate significant differences in speech recognition performance as a function of reverberation condition and signal-to-noise ratio (SNR). Performance in the pseudo-anechoic (PA) condition, reverberant condition at a 2 m distance (R2m) and reverberant condition at a 6 m distance (R6m) were compared at −10, −5, and 0 dB SNR.

Age Groups (years)		Reverberation condition								
		PA SNR (dB)			R2m SNR (dB)			R6m SNR (dB)		
		−10	−5	0	−10	−5	0	−10	−5	0
7–8	9–10									
7–8	11–12		*			†				
7–8	13–14	*				*				
7–8	23–30	‡	*		‡	‡		*	‡	†
9–10	11–12									
9–10	13–14									
9–10	23–30	‡			*				*	
11–12	13–14									
11–12	23–30									
13–14	23–30									

*
Significant at $p < 0.05$

†
Significant at $p < 0.01$

‡
Significant at $p < 0.001$

Table 4

Orthogonal contrasts of reverberation conditions (post hoc Tukey HSD) indicate significant differences in speech recognition performance as a function of age group and signal-to-noise ratio (SNR). Performance in the pseudo-anechoic (PA) condition, reverberant condition at a 2 m distance (R2m) and reverberant condition at a 6 m distance (R6m) were compared within each age group at −10, −5, and 0 dB SNR.

Reverberation Conditions		Age Groups (years)														
		7–8			9–10			11–12			13–14			23–30		
		SNR (dB)			SNR (dB)			SNR (dB)			SNR (dB)			SNR (dB)		
		−10	−5	0	−10	−5	0	−10	−5	0	−10	−5	0	−10	−5	0
PA	R2m	‡	‡		‡	‡		‡	‡		‡	‡		†		
PA	R6m	†	†	*	‡	‡		‡	‡		‡	‡		†		
R2m	R6m															

*
Significant at *p* < 0.05

†
Significant at *p* < 0.01

‡
Significant at *p* < 0.001

Table 5

Post hoc comparison (Tukey HSD) of noise masker types (Noise) indicates significant differences in speech recognition performance in LTASS versus modulated LTASS (modLTASS) noise as a function of reverberation condition and signal-to-noise ratio (SNR). Performance in the pseudo-anechoic (PA) condition, reverberant condition at a 2 m distance (R2m) and reverberant condition at a 6 m distance (R6m) were compared at −10, −5, and 0 dB SNR.

		Reverberation Condition								
		PA			R2m			R6m		
Noise		SNR (dB)			SNR (dB)			SNR (dB)		
		−10	−5	0	−10	−5	0	−10	−5	0
LTASS	modLTASS	†	†		‡			‡	‡	

†Significant at $p < 0.01$

‡Significant at $p < 0.001$

Table 6

Orthogonal contrasts of reverberation conditions (post hoc Tukey HSD) indicate significant differences in speech recognition performance as a function of noise masker type (Noise) and signal-to-noise ratio (SNR). Performance in the pseudo-anechoic (PA) condition, reverberant condition at a 2 m distance (R2m) and reverberant condition at a 6 m distance (R6m) were compared within each age group at −10, −5, and 0 dB SNR.

Reverberation Conditions		Noise					
		LTASS			modLTASS		
		SNR (dB)			SNR (dB)		
		−10	−5	−0	10	−5	0
PA	R2m	‡	‡		‡	‡	*
PA	R6m	‡	‡	‡	‡	‡	
R2m	R6m			*	†		

*
Significant at *p* < 0.05

†
Significant at *p* < 0.01

‡
Significant at *p* < 0.001

Table 7

Summary table of a three-factor mixed-model repeated measures analysis of variance. The differences in release from masking in modulated LTASS relative to LTASS noise as a function of Age Group, reverberation condition (R_C), signal-to-noise ratio (SNR) and their interactions were investigated. Degrees of freedom (df) and p-values are adjusted with the Greenhouse-Geisser correction for all factors and interactions except Age Group.

Factor	df	F	p	η_p^2
Age Group	4.00	1.38	0.252	0.091
R_C	1.92	24.30	< 0.001	0.306
$R_C \times$ Age Group	7.68	0.54	0.817	0.038
SNR	1.70	560.23	< 0.001	0.911
SNR $\times$ Age Group	6.79	2.32	0.033	0.144
$R_C \times$ SNR	3.19	33.71	< 0.001	0.380
$R_C \times$ SNR $\times$ Age Group	12.78	0.34	0.983	0.024

Table 8

Summary table of a three-factor mixed-model repeated measures analysis of variance. The differences in slopes of the cubic polynomial at 70% correct predicted speech recognition performance as a function of Age Group, noise masker type (Noise), reverberation condition (R_C) and their interactions were investigated.

Factor	df	F	p	η_p^2
Age Group	4	4.57	0.003	0.250
Noise	1	133.54	< 0.001	0.708
Noise × Age Group	4	1.46	0.226	0.096
R_C	2	4.64	0.012	0.078
R_C × Age Group	8	4.44	< 0.001	0.244
Noise × R_C	2	23.89	< 0.001	0.303
Noise × R_C × Age Group	8	0.95	0.477	0.065

Table 9

Estimated SNRs (in dB) required for approximately 95% correct speech recognition performance for age group (columns) and condition (rows).

Noise Masker	RT Condition	Age					
		7–8 yrs	9–10 yrs	11–12 yrs	13–14 yrs	23–30 yrs	
LTASS	PA	−1.33	−3.17	−3.33	−3.67	−3.83	
	R2m	0.33	−1.17	−2.17	−1.92	−2.75	
	R6m	1.75	0.67	−0.33	−0.92	−2.75	
modLTASS	PA	−1.58	−4.17	−5.00	−3.92	−4.25	
	R2m	1.33	−0.92	−2.42	−1.92	−3.25	
	R6m	1.00	−0.25	−2.08	−2.00	−3.42	