Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children^{a)}

Daniel L. Valente,^{b)} Hallie M. Plevinsky,^{c)} John M. Franco, Elizabeth C. Heinrichs-Graham, and Dawna E. Lewis

Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 22 November 2010; revised 17 October 2011; accepted 18 October 2011)

The potential effects of acoustical environment on speech understanding are especially important as children enter school where students' ability to hear and understand complex verbal information is critical to learning. However, this ability is compromised because of widely varied and unfavorable classroom acoustics. The extent to which unfavorable classroom acoustics affect children's performance on longer learning tasks is largely unknown as most research has focused on testing children using words, syllables, or sentences as stimuli. In the current study, a simulated classroom environment was used to measure comprehension performance of two classroom learning activities: a discussion and lecture. Comprehension performance was measured for groups of elementary-aged students in one of four environments with varied reverberation times and background noise levels. The reverberation time was either 0.6 or 1.5 s, and the signal-to-noise level was either +10 or +7 dB. Performance is compared to adult subjects as well as to sentence-recognition in the same condition. Significant differences were seen in comprehension scores as a function of age and condition; both increasing background noise and reverberation degraded performance in comprehension tasks compared to minimal differences in measures of sentence-recognition.

© 2012 Acoustical Society of America. [DOI: 10.1121/1.3662059]

PACS number(s): 43.55.Hy, 43.71.Gv [LMW]

Pages: 232–246

I. INTRODUCTION

Previous studies have shown that children perform more poorly than adults on a variety of speech-perception measures in both noise and reverberation (e.g., [Finitzo-Heiber and Tillman, 1978](#); [Elliott, 1979](#); [Neuman and Hochberg, 1983](#); [Nitttrouer and Boothroyd, 1990](#); [Fallon et al., 2000, 2002](#); [Johnson, 2000](#); [Hall et al., 2002](#); [Jamieson et al., 2004](#); [Bradley and Sato, 2008](#)). Because most communication takes place in environments where noise is present, children may experience more difficulty following conversations with multiple talkers, while listening from a distance and/or acquiring information via overhearing, compared to adults performing the same tasks. The current study examines the effects of room acoustics on children's speech perception and comprehension during complex learning tasks in a simulated classroom environment.

The effects of acoustical environment on speech understanding become important for children in classrooms where it is important to hear and understand verbal information. However, this ability may be compromised as a result of unfavorable classroom acoustics (e.g., [Bradley, 1986](#); [Knecht et al., 2002](#); [Nelson et al., 2008](#)). Speech intelligibil-

ity within a room is primarily determined by (1) the level of the speech, (2) the level and characteristics of the background noise and/or competing talkers, and (3) the reverberant decay/characteristics of the room ([Bradley, 1986](#)). Acoustical standards for classrooms ([ANSI, 2010](#)) recommend maximum background noise levels of 35 dB(A) and a maximum reverberation time (RT) of 0.6 s for typical, medium-sized classrooms. [Bradley \(1986\)](#) recommends RT less than 0.4 s; however, in some classroom configurations, improved performance can be seen by extending RTs beyond 0.6 s (e.g., [Yang and Hodgson, 2006](#)).

A model for determining optimal RTs in classrooms has been proposed by [Hodgson and Nosal \(2002\)](#). This model takes into account background noise sources and location, physical volume of the room, and the interaction among the talker, noise sources, and reverberation parameters. The results of [Hodgson and Nosal \(2002\)](#) counter studies by [Nabelek and Pickett \(1974\)](#), and [Finitzo-Hieber and Tillman \(1978\)](#), who found that optimal speech intelligibility is achieved by reducing the RT to 0.0 s. In real classrooms, which are never noiseless even in the ideal case, reverberation must be optimized rather than minimized for the ideal listening and learning environment. This is made apparent by the ability for strong early reflections (arriving within the first 50 ms after the arrival of the direct sound) to increase a room's effective signal-to-noise ratio, providing a higher level of speech intelligibility given a constant RT. Despite the extensive research regarding the detrimental effects of excessive noise and reverberation in the classroom, and the optimization of these environments, [Knecht et al. \(2002\)](#) reported classroom noise levels ranging from 34 to 66 dB(A) and RTs ranging from 0.2 to 1.27 s.

^{a)}Portions of this work were presented at the 2010/2011 Annual Meetings of the American Auditory Society in Scottsdale, AZ, as well as the 159th meeting of the Acoustical Society of America in Baltimore, MD and the 162nd meeting of the Acoustical Society of America in Seattle, WA.

^{b)}Author to whom correspondence should be addressed. Electronic mail: daniel.valente@boystown.org

^{c)}Present address: Department of Speech and Hearing Sciences, University of Maryland, College Park, MD, 20742.

When compared to older school-aged peers, elementary-aged children require the highest sound-quality environment to achieve near-optimum performance (Bradley and Sato, 2008), yet are attempting to learn in classrooms that have the greatest amount of masking noise (Picard and Bradley, 2001). Within these non-optimal environments, students encounter a variety of complex listening/learning conditions (such as teacher-lectures, multiple-talker discussions, and small-group activities). Learning new information during classroom activities can require divided attention to multiple talkers located to the side or behind, who may or may not be visible, for example, listening to a lecture while taking notes. The effort required to decode the speech signal in unfavorable listening environments may leave fewer resources for other cognitive duties such as comprehension and short- and long-term storage of information (Klatte *et al.*, 2010a; Picard and Bradley, 2001) with potential negative consequences for learning. Poor classroom acoustics have been shown to adversely impact children's educational performance (e.g., Boman, 2004; Dockrell and Shield, 2006; Evans, 2006) as well as conversations among peers in typical multiple-talker discussions (McKellin *et al.* 2007, 2011).

For young children, listening to multiple talkers reduces performance in auditory-only speech-intelligibility studies (Ryalls and Pisoni, 1997). Adding visual cues to auditory stimuli improves speech perception (for a review, see Erber, 1975), but in a multiple-talker scenario (as in a discussion situation in a classroom) where auditory cues are degraded (due to noise or reverberation), having to visualize multiple talkers to follow what is being said may require cognitive resources that might be otherwise allocated for comprehension. The subject's looking behavior during the course of a multiple-talker experimental task may therefore be an important correlate to examine with respect to measures of subject-performance. To evaluate the impact of classroom acoustics on learning, it is important to examine performance in realistic listening environments where the acoustics of the space may affect the ability to hear and understand and where behaviors, such as where the subject chooses to look during a task, can be analyzed.

While previous research has examined classroom acoustics (summarized in Picard and Bradley, 2001), studies that examine speech understanding during realistic learning activities in typical classroom conditions have been limited (e.g., Klatte *et al.*, 2007; Klatte *et al.*, 2010b; Neuman *et al.*, 2010). Much of the previous research in this area has been conducted via laboratory studies using speech-recognition tasks with a single talker and presentation via earphones and/or from a loudspeaker located directly in front of the listener (e.g., Elliott, 1979; Johnson, 2000). Even in studies conducted in real or simulated classrooms, typical speech recognition tasks may not reflect those encountered in actual classrooms (e.g., Nelson *et al.*, 2008; Yang and Bradley, 2009). Klatte *et al.* (2010a) found that sentence recognition was unaffected by reverberation, suggesting that children could compensate for degraded speech for a short time period.

Recent studies have examined children's abilities to perform learning tasks in laboratory (McFadden and Pittman, 2008) or realistic acoustical environments (Klatte *et al.*,

2010a; Klatte *et al.*, 2010b). For example, Klatte *et al.* (2010b) examined speech perception and listening comprehension in children and adults in simulated classroom settings with varying acoustical conditions. Children performed more poorly than adults in a room with a "favorable" RT (0.47 s). When speech and noise maskers were compared, multi-talker babble affected performance on comprehension tasks more than noise, while the reverse was true for speech perception. It was hypothesized that these differences occurred because background speech interfered with children's verbal short-term memory that was needed during the comprehension task.

The directionality of talkers, as well as the length of a listening task, has the potential to affect children's access to speech in a classroom, thereby impacting opportunities for learning. To model classroom activities, tasks should include auditory-visual stimuli with both single and multiple talkers and should represent the complexity and duration encountered in the classroom setting. Likewise, acoustical simulations should include noise and reverberation as actually experienced in classrooms (e.g., direct and indirect sounds, early and late reflections).

The objectives of the current study are to further understand the interaction between the complexity of listening/learning environment and performance in elementary-aged children. Toward that aim, the current experiment provides a realistic model of the challenging situations children face in the classroom by providing subjects with audio-visual stimuli of a teacher and students, longer-duration tasks, and a plausible amount of noise and reverberation based on previous measurements in real classrooms. This approach quantifies the difficulties children have with increasing learning-task complexity, background noise and reverberation, and facilitates the comparison between learning tasks representative of typical classroom activities with those of a standard measure of sentence recognition within the same acoustical environment.

The remainder of this paper will be organized into two parts. In Sec. II, the creation of a simulated classroom will be outlined. The simulated classroom environment was used to create an ecologically valid experimental paradigm that included auditory-visual cues from multiple sources, realistic amounts of background noise, and a RT typical of what would be found in real classrooms. The use of a simulated classroom provides important experimental control that would not be available in a real classroom. Validation measurements are reported here to allow comparison with measurements from actual classrooms. Section III presents a set of two experiments that were conducted in the simulated classroom with an acoustical environment representative of actual classrooms and used as a model for the complex listening conditions and active learning tasks that students encounter on a daily basis. The purpose of these experiments was to examine the performance of children and adults with normal hearing on two listening tasks closely related to tasks performed in the classroom (story comprehension, the length of which was similar to that of a typical classroom lesson, and sentence recognition) presented under different listening conditions in a controlled simulated classroom.

II. CREATION AND VALIDATION OF A SIMULATED CLASSROOM

A simulated classroom environment was created in a room at Boys Town National Research Hospital (Fig. 1). This simulated classroom is composed of a physical room and a virtually modeled room. In the physical room, an array of loudspeakers and LCD monitors reproduced audio and video material. Video recordings were presented so that it appeared as if the people in the recordings were seated at desks around a subject, who was seated in the center of the room. First, acoustical measurements were taken in the real room to quantify background noise and RT. Next, through the array of loudspeakers using a virtual acoustical model, augmentations to the sound field were applied to extend the RT, spatialize individual sound sources in the room, and adjust background noise. Artificial reflections and background noise were radiated in addition to audio recordings of the experimental stimuli to simulate a plausible occupied classroom environment. This environment was created as an alternative to conducting experiments in real classrooms, which introduce issues in regard to repeatability, consistency of noise-sources, and logistics, or conducting experiments in a sound-treated booth, which introduces ecological validity concerns.

A. Measurement of noise characteristics and reverberation time

The physical room used for the simulated classroom had a volume of 86.8 m^3 (dimensions: $2.6 \text{ m} \times 5.3 \text{ m} \times 6.3 \text{ m}$; $h \times w \times l$). The wall surfaces were gypsum, the ceiling was a standard drop construction with acoustical tiles. The flooring was thin, commercial vinyl tile glued to a concrete slab. There was a single door into the room (the door was constructed of solid-core engineered wood). To reduce noise transmission from the adjacent hallway, neoprene gaskets were installed around the door frame and a neoprene sweep/threshold was installed. There were three $1 \text{ m} \times 1 \text{ m}$ interior windows next to the door. For treatment, the cavities in each

FIG. 1. (Color online) The classroom environment used for the experiments. The subject is seated in the center, and five loudspeakers and LCD monitors reproduce speech by one of five talkers. The room is acoustically treated with sound-absorbing foam and a velour curtain to control reverberation.

window box was framed in with 16 mm gypsum board and sealed with silicone caulk, and the cavity between the gypsum board and the glass was filled with a 12.5 mm piece of fiberboard insulation to reduce noise infiltration from the hallway.

The front of the room had a large bookcase filled with irregularly spaced books to provide a diffusing surface and to reduce the possibility of a front-back audible flutter echo. The back, right, and front walls were treated with a regular checkerboard-pattern array of 6.1 cm , $61 \text{ cm} \times 61 \text{ cm}$ sound-absorbing material. This material was applied to each of the exposed walls of the room, as well as the window cavities and door, and covered approximately 50% of the surface area of the room. A 22-oz full-length velour curtain with 50% fullness was positioned on the left side of the room, separating the experimental area from a 2.2-m-wide area where a researcher monitored the experiment over closed-circuit television. This gave the effective simulated classroom a volume of 50.8 m^3 (dimensions: $2.6 \text{ m} \times 3.1 \text{ m} \times 6.3 \text{ m}$; $h \times w \times l$).

Figure 1 provides a graphic representation of the classroom environment as well as the wall-surface treatments and location of the subject in the room. Six tables were positioned in the room to simulate subject-seating locations and for video presentations of 4 students and a teacher. The subject's desk was located in the middle of the room. Video recordings of each of the visual stimuli (students or teacher) were played back on LCD monitors positioned on the tables surrounding the subject (Samsung, 2493HM). The sound stimuli were reproduced over five loudspeakers (M-Audio, AV40), positioned under each LCD monitor.

Figure 2 shows the position of each speaker relative to the center of the room in schematic form. Loudspeakers 1 and 3-5 represented students positioned around the subject. Loudspeaker 2 represented a simulated teacher position. In Fig. 2, distance measurements are given in meters and correspond to a student seated to the left of the subject (position 1; -90° azimuth), students positioned behind and two-desks back from the subject (positions 4, and 5; 135° and -135° azimuth), and a student positioned one row in front and to the right of the subject (position 3; 30° azimuth). Finally the teacher was positioned directly in front of the subject (position 2, 0° azimuth).

Monaural impulse responses (IRs) were measured in the room as well as background-noise measurements taken from a Type I sound-level meter (Larson Davis, model 824; 1/2-in. random incidence capsule). The background noise LA_{eq} (15 min) was 32.4 dB(A). The primary source of background noise was from the HVAC system of the facility. This background noise is considered acceptable according to the ANSI (2010) guidelines for classrooms and was sufficiently low to allow measurement of RT within the space. IRs were measured using a 50-second logarithmic sinusoid-sweep excitation signal with a 24-bit resolution analog-digital (A/D) conversion and 48-kHz sampling rate (Apogee; Ensemble). To generate the impulse, the recorded sweep was convolved in the time domain using an appropriate time-inverse filter with a -6 dB/octave equalization to compensate for the logarithmic frequency-time spacing of the excitation signal. This method is described by Farina (2000) and allows for the

FIG. 2. Plan schematic diagram of the five loudspeakers and desk locations relative to a subject seating in the center for the room. Speakers 1 and 3-5 correspond to simulated student positions in a classroom, while speaker 2 corresponds to a simulated teacher position.

recording of IRs with a very high signal-to-noise ratio (SNR). A single speaker (M-Audio AV40) located in the front of the room at 0° azimuth in the “teacher” position was used to play back the excitation signal. The M-Audio AV40 was chosen because its small, 12.7 cm transducer approximates the directional characteristics of the human voice.

The SNRs of the measured IRs were greater than 70 dB. Two energy decays were calculated from the IRs in accordance with ISO 3382 (1997): the early decay time, or EDT, and the T_{30} . The results of the measurements are presented in Fig. 3. A complete description of the octave-band background noise, RT, early decay time and clarity (C_{50}) is in the Appendix.

For analysis, the broadband IR was filtered using octave-bandwidth, ANSI Type I, sixth order IIR Butterworth filters centered at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. The T_{30} (mid) for the treated room was 0.30 s with longer RT found at the lowest octaves (125, 250 Hz) and a shorter RT at the highest octave-band (8 kHz). The EDT (mid) was 0.24 s. The increased RT at 125 Hz was expected due to the fact that the absorptive materials chosen for treatment were less effective at the lowest-analyzed frequencies. The RT was less than the maximum recommended in the ANSI (2010) guidelines for classrooms at frequencies from 125 to 8000 Hz in all cases.

B. Creation of the virtual classroom

A real-time virtual acoustical modeling technique was used to adjust the background noise and reverberation

FIG. 3. Measured reverberation time analyzed in octave-bands. Open white bars are early decay times (EDTs) and closed black bars are T_{30} 's.

characteristics of the physical room so that they more closely mimicked conditions found in an average occupied classroom. This technique is similar to artificial- enhancement systems used for extending RT digitally in meeting rooms, rehearsal spaces, or small performance halls (for a review, see Svensson, 1994; Kleiner and Svensson, 1995). The virtual microphone control (ViMiC) system created the spatialized acoustic stimuli of the talkers for the virtual classroom (Braasch *et al.*, 2008). This system is based on the simulation of microphone techniques and acoustic spaces. The virtual room was a shoebox model with dimensions identical to the physical classroom. An array of five virtual microphones positioned at the location of each (real) speaker and five virtual sound sources positioned at the physical location of the LCD monitors were arranged within the virtual room. Using this system, sound sources can be placed virtually in space by positioning and orienting virtual microphones in a 3-dimensional computer-simulated room. For sources: 1 and 3 (see Fig. 2), the orientation was rotated such that the normal angle was facing the front wall of the virtual classroom model to simulate students sitting forward in their desks facing the front of the virtual classroom. All 5 sources used the human voice directivity model within ViMiC, and each of the five virtual microphones in the virtual classroom was omnidirectional.

The modeling software calculates direct sound, first-order reflections, and late reverberation in real-time, based on input parameters for the desired virtual acoustic environment as described in Braasch *et al.* (2008). Briefly, the direct sound is sent into ViMiC modules that process specific parts of the to-be-created IR. Based on input geometric data, the appropriate sensitivity and delay functions for each source and reflection for each loudspeaker are determined. The signal is combined with late reverberation and background noise. This creates the final virtual microphone signals, which are fed to individual loudspeakers. Using these virtual acoustics techniques, a classroom was created with the same

TABLE I. Sound absorption coefficient data for the surfaces used in the virtual classroom model.

Material/frequency (Hz)	125	250	500	1000	2000	4000
Walls: 5/8 in. gypsum on 3-5/8 in. studs + fiberglass	0.16	0.07	0.04	0.04	0.03	0.03
Ceiling: acoustical ceiling tile (2 ft $\times$ 4 ft - sized)	0.70	0.66	0.72	0.92	0.88	0.75
Floor: vinyl tile glued to concrete	0.02	0.03	0.03	0.03	0.03	0.02

dimensions as the physical room (2.6 m $\times$ 3.1 m $\times$ 6.3 m) with the purpose of adding additional reflections and background noise to simulate an actual classroom. The virtual microphone array corresponded to the position of the real speakers in the physical room. In addition, multiple recorded audio tracks were spatialized at the appropriate azimuth angle and distance to correspond to their video recordings. An adaptation of previous custom-designed audio-visual software, described by Valente and Braasch (2008), was used for this simulation and data collection. A number of simplifications in the virtual acoustics model were made to keep the system available to real-time processing of sound sources. The current ViMiC software was implemented as a C-External for the real-time audio/visual programming environment: Max/MSP (CYLING 74; version 5). The ViMiC system was limited to rectangular-shaped rooms with the modeling of surface absorption coefficient (α) values for octave-bands between 250 and 4000 Hz. The absorbing properties of the virtual room environment were taken from averaged α values (see Appendix H in Mehta *et al.*, 1999). To match the real classroom, the surfaces of the walls in the virtual classroom were simulated as 16 mm gypsum board, the ceiling as acoustical tile, and the floor as vinyl glued to concrete. The octave-band α values for each surface are shown in Table I

Another simplification in the simulation environment was that early reflections calculated using the image method (Allen and Berkeley, 1979) were limited to first-order reflections. The remaining energy decay was simulated using the late reverberation module of ViMiC, with the same multi-channel algorithm as described by Braasch *et al.* (2008).

Despite limitations in virtual acoustics, ViMiC affords the ability to create a virtual room in real time as well as the characteristic decays of several sources positioned within that room. These sound sources can be spatialized around an array of loudspeakers without the need for off-line calculations for generating simulated IRs. As such, the virtual environment can be adjusted with respect to the energy decay characteristics (early-to-late energy ratio), RT, and background noise.

C. Validation

To validate the virtual-acoustic system in the sound field, a series of measurements were performed at the subjects' listening position (center of the room, as shown in Fig. 1). Because the physical space was not anechoic, the reflection patterns and energy decay would combine with any artificial reflections and reverberation added by the ViMiC virtual environment (as would be found in a typical artificial enhancement system). The sound field was measured *in situ*, at the

listening position, rather than simply setting internal parameters within the virtual room modeling software to verify the actual sound field at the listeners' ears. The first set of measurements is shown in Fig. 4. For these IRs, the configuration of the ViMiC virtual room model was as follows: The room geometry was set such that the virtual room parameters were the same as the actual physical room. The virtual source was located at 0° azimuth positioned where the front (real) loudspeaker was positioned. The excitation stimulus, deconvolution technique, and methods for calculating RT were the same as used for measurements without the use of the virtual modeling software. Figure 4 shows the effect of adjusting the decay time of the internal RT while holding constant the characteristics of the early reflections (which are based on the geometry of the virtual room and characteristic of the absorbing surfaces) fixed. These measurements show that the T_{30} at the listening position increases monotonically with an increase of the internal RT parameter but with a slope less than 1. RT was calculated at three octave-band centers: 500, 1000, and 2000 Hz. The slope of the growth of reverberation measured in the room was 0.55 for 500 Hz, 0.59 for 1000 Hz, and 0.44 for 2000 Hz. Figure 4 also shows that by using virtual-acoustic techniques, the characteristics of the reverberant decay measured within the room can be adjusted through the range of RTs found in real classrooms (Knecht *et al.*, 2002) without changing the geometry of the virtual room (which would also change the pattern of early reflections). As a result, the virtual classroom can have RTs between 0.3 s (no artificial

FIG. 4. The internal T_{60} parameter of ViMiC versus measured T_{30} at the listening position for 500, 1000, and 2000 Hz octave bands. The “No VA” measurement is the classroom with no virtual acoustics (no artificial reflections or reverberation added).

FIG. 5. Clarity (C_{50}) versus reverberation time (T_{30}) comparing results measured in actual classrooms and auditoria to results in the virtual environment.

electro-acoustic enhancement) and 1.5 s. As the internal RT parameter and the reverberant decay measured at the listening position were slightly different, it was necessary to optimize these for a target RT at the listening position.

Measures of early-to-late energy, based on the source-receiver distance, were analyzed to ensure that the acoustical environment maintained realistic ratios in the simulated IRs. Speech intelligibility is of primary concern in classrooms and the ratio of early-to-late energy is known to affect speech intelligibility (Bistafa and Bradley, 2000). C_{50} has been used as a reference metric as it is a good correlate to the effects of room acoustics on the intelligibility of speech (Bradley *et al.*, 1999). Due to the fact that the simulated classroom was created using a combination of real reflections and reverberation coupled with virtual reflections and reverberation, the impulse response of the simulated classroom was measured at the subject's location and the relative contributions of early and late energy as well as the RT were calculated. Yang and Bradley (2009) have published measurements in both simulated environments (speaker array positioned in an anechoic chamber) and real classrooms and auditoria with respect to comparing C_{50} to a given T_{30} in order to maintain a realistic simulation for a plausible source distance.

Figure 5 shows results of measurements at three octave-band centers taken in the simulated classroom environment. C_{50} values are plotted versus the corresponding T_{30} values plotted on a log scale. The open squares denote values from the best-fit line for real classrooms and auditoria in Yang and Bradley (2009). The filled circles denote average measurements taken in the classroom environment used for this experiment at two RTs that could be measured in an actual classroom: 0.6 s (maximum allowable with respect to the ANSI classroom standard), and 1.5 s (a long RT that might reasonably be expected in a real classroom). This indicates that at the subject's listening position, the virtual environment maintains a ratio of early-to-late energy similar to that found in an actual classroom given a plausible source-receiver distance at a given RT.

D. Artificial background noise

Simulated background noise was generated separately and was attenuated such that the sound pressure level [dB(A)] could be adjusted. The noise was band-pass filtered from 31.5 Hz to 16 000 Hz and a low-pass filter of -5 dB per octave was applied, starting with the 63-Hz octave band. The noise was radiated incoherently through the five presen-

tation loudspeakers. This background noise spectrum has been used in previous classroom acoustics research studies where a virtual environment was created (e.g. Yang and Bradley, 2009) and has also been shown to be an effective analog for the indoor noise generated by heating, ventilation, and air-conditioning (HVAC) systems (Blazier, 1981). The noise level was verified by *in situ* measurement with a Type I sound level meter (Larson Davis, model 824; $\frac{1}{2}$ -in. random incidence diaphragm). The dynamic range of the loudspeaker presentation system allows the setting of the background noise between the real room's noise floor [32.4 dB(A)] and the maximum SPL output of the loudspeakers, 93 dB(A) at the listening position. This system, with a dynamic range of 60 dB, is above the range of SNRs found in real classrooms with unamplified speech. The dynamic range for the intelligibility of unamplified speech (from completely unintelligible to completely intelligible) in rooms is typically ± 15 dB related to a long-term speech SPL average of 60 dB(A) (ANSI 1997).

In summary, a simulated classroom environment was created. This classroom consisted of an actual acoustically treated physical space coupled with a virtual model of a classroom. This classroom included plausible locations for simulated students and a teacher positioned around a subject. The real classroom environment was acoustically treated to bring the reverberant decay and background noise to well below what would be found in an average elementary classroom. Augmentations of the sound field by increasing the RT and background noise have been shown to create a simulated classroom sound field that encompasses the range of acoustical environments that would be found in classrooms.

III. SPEECH RECOGNITION AND COMPREHENSION IN A SIMULATED CLASSROOM ENVIRONMENT

Subject performance in the classroom environment is described across two experiments. The first experiment examined how elementary-aged students perform in two plausible classroom learning conditions—a lecture from a teacher and a classroom discussion. All subjects also performed a sentence recognition task. Performance was compared to that of a control group of adult subjects. The classroom environment used for the first experiment was chosen to represent a background noise level and RT that would be found in a typical occupied classroom. In the second experiment, a wider range of classroom environments were evaluated. Performance across

children and adults were examined as a function of increasing RT and decreasing SNR within the environment.

A. General methods

1. Classroom learning task

An elementary-age-appropriate “Reader’s Theater” play (Shepard, 2010) was chosen to provide a novel learning task for all subjects. Initially, 35 questions were created to test comprehension of the play and were piloted on adults ($n=4$) and children ($n=3$). Questions that were always or never answered correctly were eliminated. The final list of 18 questions was designed such that 15 of the questions (three each for the five characters) were based on information given by a single character in the play (either individual talkers in the case of the discussion condition, or played-characters spoken by the teacher alone *in-character* in the case of the lecture condition). Additionally, there were three general questions which pertained to information spoken by multiple characters during the play.

To create plausible classroom lessons for the experiment, individual video recordings were made of: (1) a teacher and four students reading from the script of a play (discussion condition) and (2) only the teacher reading the same play (lecture condition). In the case of the multiple-talker reading, each student acted as a different character, where changes in dialog resulted in different individual active talkers. Only one student/teacher read at a given time; there were no overlapping or competing talkers during the course of the play. As would be found in an active discussion during a classroom lesson, the azimuth of the active talker relative to the subject was quasi-random over the course of the play. There were 163 different changes in dialog which resulted in the active talker being presented at -135 , -90 , 0 , 30 , or 135 degrees azimuth relative to the subject. These positions were chosen to simulate a plausible discussion with the subject sitting in the middle of his/her classroom as indicated in Fig. 2. In the case of the single-talker recording, the teacher read all characters in the play. The level of speech for all talkers was calibrated in the room such that the LA_{eq} (10 min) for each talker equaled 60 dB(A) at the position of the subject. This level represents the typical long-term average SPL of a talker approximately 2 m from the listener within a classroom environment (Picard and Bradley, 2001).

To simulate an actual classroom arrangement, video recordings were made from the perspective of the location of the study participants. Recordings of the teacher and students were made in a sound-treated meeting room at Boys Town National Research Hospital (dimensions: 9.88 m $\times$ 10.24 m $\times$ 3.24 m, 1 $\times$ w $\times$ h). An acoustical analysis of the meeting room can be found in the Appendix. The four students and one teacher were seated in an arrangement similar to that found during the experiment. Five video AVCHD video cameras (JVC Evario hard disk camera, GZ-MG630) were positioned in the relative location where the subject would sit during the experiment. Each camera directly faced an individual talker, such that the recording was made from the subject’s perspective (see Fig. 2). For example, the student seated directly to the left of the subject at 90° azimuth was

filmed in profile. A normal focal length (50 mm; 35-mm film equivalent) was used on each of the video cameras such that when the videos were played back, the image would appear as if the student was sitting in a desk several rows or seats away from the subject. Due to the fact that some unwanted room reflections were present in the recording room, close-microphone techniques using an omni-directional lavalier microphone (Shure, ULX1-J1 pack, WL-93 microphone) was positioned on each talker, and talkers were amplified and recorded with the individual video recording. This technique allowed the talker to be captured with the greatest possible direct-to-reverberant energy ratio. The talker would later be positioned in the virtual room model which would simulate the appropriate source-distance, reflections and reverberation so that the audio and visual stimuli would not be mismatched (i.e., the visual image appears to be coming from several meters away but due to the close microphone techniques, sounds as if the talker is 10 cm away).

2. Sentence recognition task

In an effort to relate comprehension to speech intelligibility in the discussion and lecture conditions, a sentence-recognition task was conducted. A list of 50 meaningful sentences with three key words each (BKB, Bench *et al.*, 1979) was presented. The sentences were produced by a single female talker and digitally recorded in a double-walled sound-treated booth using a condenser microphone (AKG Acoustics, C535 EB) with a flat frequency response (± 2 dB) from 0.2 to 20 kHz.

Similar to the classroom-learning task, the sentence-recognition task was either played (1) quasi-randomly from one of the five loudspeakers positioned around the room (one at a time) or (2) from a single loudspeaker corresponding to the position of the teacher.

3. Head tracker

To collect data on the head-rotation angle of the subject during the course of the experiment, a custom-designed micro-electro-mechanical system (MEMS) gyroscopic head-tracking device was placed on the subject’s head (Analog Devices, EVAL-ADXRS610Z). Sensor data were gathered through a high speed analog-to-digital audio convertor (Electrotap, Teabox). The gyroscope was fitted to a headband. The head tracker was polled every 100 ms for current head rotation angle. Subjects were not informed as to the purpose of the head-worn device and were instructed to look around naturally, as they would in a typical classroom situation.

4. Procedure

The subject was seated in the middle of the simulated classroom and watched the videos of the children and teacher (discussion condition) or of the teacher alone (lecture condition). During the reading of the play, the subject’s head rotation was recorded by means of gyroscopic head tracking. At the end of the play, subjects were asked a series of 18 questions to assess their comprehension of the play. Individual

responses were transcribed and scored by the researcher administering the experiment.

After completing the classroom-learning task, subjects performed the sentence-recognition task in the same condition (i.e., sentences were presented from multiple loudspeakers if the subject completed the classroom learning task in the discussion condition, and sentences were presented from one loudspeaker if the subject completed the classroom learning task in the lecture condition). Spoken sentences were scored on three keywords by a researcher who listened and watched the subject on closed-circuit television. The total number of key words correct was transformed to a percent correct score.

B. Experiment 1: Performance in a favorable acoustic environment

1. Subjects

Fifty children (8-12 yr, 10 children per year of age) and 40 adults (18-58 yr; mean: 25.75 yr, median: 22 yr) with normal hearing (audiometric thresholds ≤ 15 dB HL for octave frequencies from 250 to 8000 Hz) served as subjects for this study. Half of the subjects participated in the discussion condition and half in the lecture condition for the classroom-learning and sentence recognition tasks. Subjects had normal or corrected-to-normal vision. Informed consent procedures were approved by the institutional review board (IRB) at Boys Town National Research Hospital. For participation in the study, subjects were compensated for their time monetarily, and children were also allowed to choose a book to take home.

2. Acoustic conditions

The classroom environment was set such that the levels of the speakers, background noise levels, and RT were that of a typical occupied classroom. These levels were based on comprehensive results from [Bradley and Sato \(2008\)](#), who reported mean SNRs of 11 dB in 41 elementary school classrooms [Mean classroom speech level from teachers to student's desks measured at 59.5 dB(A)]. The level of speech at the subject's listening position in the present study was set to 60 dB(A), with a noise-level of 50 dB(A) (SNR = 10 dB). Based on maximum [ANSI \(2010\)](#) recommendations, the RT (averaged across 500, 1000, and 2000 Hz octave bands), was set to 0.6 s. This condition was not an attempt to create an optimal classroom condition but rather a reasonable real-world level of noise and reverberation as a model of the types of acoustical environments that students are exposed to on a daily basis. Measurements of octave-band background noise, RT, EDT, and C_{50} for this environment can be seen in the Appendix.

3. Results and discussion

For the sentence-recognition task, all subjects in both the discussion and lecture paradigms scored above 95% correct. A paired samples *t*-test revealed no significant difference in scores across age ($P = 0.17$) or listening conditions ($P = 0.14$). These findings are consistent with those of [Bradley \(1986\)](#) for real classrooms with the same RT and SNR conditions used for this experiment.

FIG. 6. Score (% correct) for the classroom comprehension tasks in both discussion and lecture conditions in children and adults.

The results of the classroom-learning tasks are shown in Fig. 6. In general, children performed more poorly than adults on comprehension tasks in both listening conditions (lecture and discussion). Statistical analysis consisted of between-subjects analysis of variance (ANOVA) with effect sizes calculated based on the method by [Rosenthal and DeMatteo \(2001\)](#), reported in *r*. For comprehension scores, a two-way ANOVA with age and condition as independent variables revealed a significant effect of age [$F_{(1,86)} = 23.20$, $P < 0.001$, $r = 0.46$] and condition [$F_{(1,86)} = 10.63$, $P < 0.01$, $r = 0.33$] as well as a significant age $\times$ condition interaction [$F_{(1,86)} = 4.21$, $P = 0.04$, $r = 0.21$]. Children performed more poorly than adults in both listening conditions. Interestingly, children were affected by the discussion condition differently than the adults; the children's comprehension scores were significantly lower for this condition than for the lecture condition. This reduction in comprehension scores across conditions was not seen in the adult group of subjects who had similar mean scores irrespective of condition.

It was hypothesized that subjects who had difficulty determining the active talker during the course of the play and/or who had to turn to look toward the talkers may have demonstrated reduced performance. As such, it was determined if looking behavior during the course of the experiment could account for differences in comprehension performance in the discussion condition. To examine listeners' orientation to the individual talkers during the discussion classroom-learning task, the proportion of events visualized (POEV) was taken by looking at the known visual angle of each of the discrete speakers relative to the listener (recorded as events in time) and comparing them with the gyroscopic data associated with the onset of the event. A 2-s time window (20 samples, 10 before and 10 after the onset of the event) was used, along with an angular window of ± 15 degrees around the position of a talker. The number of events visualized was summed and a proportion was taken comparing the number of events that the subject visualized to the total number of events. In

general, both age groups looked directly at the talkers as they spoke less than 50% of the time with children localizing the active talkers significantly more often than adults [$F_{(1,43)} = 8.98$, $P < 0.01$, $r = 0.41$]. The mean POEV was 35% for children and 28% for adults.

The low POEV was not unexpected given the rapid changes in talkers throughout the task. However, even when listeners do not look directly at all talkers during a group activity, it can be useful to examine the types of overall looking behaviors they exhibit. An analysis of overall looking behavior based on each subject's raw head-angle recordings was performed. It was found that subjects' looking behaviors fell into one of three categories: low (where the subject maintained a low standard deviation of angle from the centerline regardless of the location of the active talker), medium (where the listener first localized each active speaker and then after a period of time reverted to the low-looking behavior), and high (where the listener attempted to follow along with the active talkers throughout the task). For this analysis, the standard deviation of the head track was determined. The standard deviation accounts for the amount of modulation from the midline (a higher standard deviation represents greater head movement). Subjects with $<25^\circ$ SD were grouped into the low category, $25^\circ < \text{SD} < 45^\circ$ were considered medium and $>45^\circ$ were considered high. These angular categories were based on the azimuth angles of the monitors and speakers within the classroom (see Fig. 2).

Looking behavior category as a function of age is shown in Fig. 7. Results indicated that children were more likely to demonstrate high looking behavior than adults (44% of children versus 5% adults) while adults were more likely to exhibit low (28% of children versus 55% of adults) and medium behaviors (28% of children versus 40% of adults). A regression analysis was performed to determine if the raw looking behavior value (SD of the head track) was a significant predictor of comprehension score in the discussion classroom-learning task. The dependent variable was the

comprehension score of subjects with the SD of the head track as the independent variable. Looking behavior was a significant predictor of comprehension score [$F_{(1,42)} = 4.20$, $P = 0.046$, $r = 0.30$]. Subjects who exhibited a higher looking behavior performed more poorly on the task than subjects with lower looking behaviors.

In summary, a simulated classroom was created in Experiment 1 where subjects completed two listening tasks in one of two plausible listening conditions in a classroom environment that had SNR and reverberant-decay properties representative of those elementary-aged students typically experience while learning. Measurements in the simulated classroom were consistent with real classrooms in terms of background noise and the ratio of the C_{50} metric for a given T_{30} (Yang and Bradley, 2009). Results of a test of speech recognition in the classroom were consistent with results seen in real classrooms with similar acoustics (Bradley, 1986).

Despite similarity between children and adults and between the two listening conditions in terms of sentence-recognition scores, significant differences were seen in the results of the classroom-learning task. As might be expected, children performed more poorly than adults in both the discussion and lecture conditions. Unlike adults, children also performed more poorly in the discussion condition than in the lecture condition. Although children were more likely to look toward talkers during the discussion condition, their scores were poorer than adults. This finding could indicate that attempting to visualize the talker utilizes cognitive resources that would otherwise be allocated for short- and long-term storage of information (e.g., Klatte *et al.*, 2010a; Picard and Bradley, 2001). Indeed subjects who exhibited higher looking behaviors had lower comprehension scores than those who exhibited low looking behavior. None of these differences manifested themselves in the reference measure of sentence recognition, where all subjects performed at or above 95% correct. The results of this experiment indicate that examining measures of learning can explain differences in performance for children that measures of speech recognition do not explain.

Experiment 1 revealed differences in comprehension between children and adults in both discussion- and lecture-type classroom lessons. It is known that classroom environments vary with respect to SNR and RT and often do not conform to the ANSI (2010) guidelines of 35 dB(A) for unoccupied rooms with only background noise and a maximum RT of 0.6 s (Knecht *et al.*, 2002). In most classrooms, measured occupied conditions are worse. Knecht *et al.* (2002) reported that in 32 unoccupied classrooms, background noise was as high as 66 dB(A) with background noise above 50 dB(A) in 9 of 32 classrooms. Knecht *et al.* (2002) also reported RTs greater than 1.0 s in 7 of 32 classrooms. These data are in agreement with previous studies (Kodaras, 1960; Sanders, 1965; McCroskey and Devens, 1975; Bradley, 1986).

C. Experiment 2: Performance in adverse classroom acoustics

Experiment 2 examined how children and adults perform in both the sentence-recognition and the classroom-learning

FIG. 7. The percentage of subjects in each looking behavior category for children and adults while performing the discussion classroom-learning task.

tasks as acoustic conditions in the simulated classroom degrade. Further, the relationship of age to performance was examined (e.g. Yang and Bradley, 2009).

1. Subjects

Participants included 60 children (8 and 11 yr, 30 for each year of age) and 30 adults (19-32 yr; mean: 25.75 yr, median: 22 yr) with normal hearing (audiometric thresholds ≤ 15 dB HL for octave frequencies from 250 to 8000 Hz). Eight- and 11-yr-olds were selected as a subset of the ages chosen in Experiment 1 because they represent both younger and older elementary-aged children who often learn in dynamic listening environments. The 60 children and 30 adults were divided into two groups of equal participants in which half performed the classroom-learning and sentence-recognition tasks in a discussion condition and the other half in a lecture condition. Within each condition, equal numbers of listeners from each age group ($n = 5$) performed the two listening tasks in one of three acoustical environments.

2. Acoustical conditions

Three environments with varying SNRs and RTs were simulated: SNR = +10 dB, RT = 1.5 s; SNR = +7 dB, RT = 1.5 s; SNR = +7 dB, RT = 0.6 s. These represent adverse classroom environments with increasing RT, background noise or both compared to the baseline classroom used in Experiment 1. The +7 dB SNR was chosen based on pilot testing with adults that resulted in floor effects at poorer SNRs. For comparison, 10 subjects per age group ($n = 30$) from Experiment 1 were included in the data analysis for the baseline acoustical environment (SNR = +10 dB, RT = 0.6 s). Measurements of octave-band background noise, RT, EDT, and C_{50} for these three additional environments can be seen in the Appendix.

3. Results and discussion

Mean sentence-recognition scores are shown in Table II. All subjects scored above 82% on the task, irrespective of condition or acoustical environment, and all but one 8-yr-old

and one 11-yr-old scored above 92%. The scores were converted into rationalized arcsine units (RAU) to equalize the variance across the range of scores (Studebaker, 1985). A between-subjects ANOVA with age, condition, RT, and SNR as the independent variables revealed significant main effects of age [$F_{(2,96)} = 15.30$, $P < 0.001$, $r = 0.49$], RT [$F_{(1,96)} = 38.03$, $P < 0.001$, $r = 0.53$], and SNR [$F_{(1,96)} = 34.24$, $P < 0.001$, $r = 0.51$]. Even though scores for all listeners decreased in the more adverse acoustical environments in both conditions, the 8-yr-old listeners performed more poorly than the 11-yr-olds and adults regardless of condition and acoustical environment. No other comparisons reached significance.

The mean scores on the classroom-learning task by age, condition, and acoustical environment are shown in Fig. 8. A between-subjects ANOVA with age, condition, RT, and SNR as independent variables revealed a main effect of age [$F_{(2,96)} = 19.54$, $P < 0.001$, $r = 0.54$] and interactions between condition and RT [$F_{(1,96)} = 8.01$, $P < 0.006$, $r = 0.28$] as well as condition and SNR [$F_{(1,96)} = 8.57$, $P < 0.004$, $r = 0.29$], both shown in Fig. 9. The 8-yr-old children performed more poorly than the 11-yr-olds and adults in both conditions and all acoustical environments. Comprehension performance was poorer in the discussion condition than the lecture condition both at the longer RT (1.5 s) and lower SNR (+7 dB). In the baseline condition (RT 0.6 s and SNR + 10 dB), listeners performed similarly across conditions. Additional significant main effects of condition [$F_{(1,96)} = 25.23$, $P < 0.001$, $r = 0.46$], RT [$F_{(1,96)} = 17.94$, $P < 0.001$, $r = 0.40$], and SNR [$F_{(1,96)} = 11.33$, $P < 0.001$, $r = 0.32$] were found but must be interpreted with respect to the significant interactions of condition $\times$ RT and condition $\times$ SNR. The decrease in comprehension scores across condition was dependent on RT. In the longer RT condition, the decrease in scores as a function lecture versus discussion task was greater than in the more favorable RT. The same trend was seen as a function of decreasing the SNR.

The subjects' looking behavior was analyzed to examine any potential differences between the age groups, as well as whether these differences in head movement were related to

TABLE II. Mean sentence recognition scores (percentage correct with SDs).

Acoustical environment (SNR; RT)	Age, ($n = 10$ per age)	Discussion condition	Lecture condition
+7 dB; 1.5 s	8	91.87 (5.34)	95.47 (1.59)
	11	97.87 (1.28)	95.47 (5.32)
	Adult	98.40 (0.76)	99.20 (0.73)
+10 dB; 1.5 s	8	98.00 (1.63)	98.27 (2.14)
	11	98.93 (0.89)	98.27 (1.30)
	Adult	99.33 (0.82)	99.20 (0.56)
+7 dB; 0.6 s	8	98.27 (0.90)	97.47 (1.96)
	11	98.93 (1.21)	98.67 (1.05)
	Adult	99.20 (0.56)	99.33 (1.16)
+10 dB; 0.6 s	8	98.67 (2.62)	99.20 (0.39)
	11	100.00 (0)	100.00 (0)
	Adult	99.58 (0.76)	100.00 (0)

FIG. 8. Mean comprehension scores on the classroom-learning task in the discussion (left) and lecture (right) conditions as a function of acoustical environment for the three age groups tested.

the variance in comprehension scores in the discussion condition. As in Experiment 1, the standard deviation (SD) of the listeners' head movements from the 0° centerline was used to analyze looking behavior. A higher SD would indicate a greater amount of head movement during the classroom task.

Results of looking behavior were analyzed using a three-way ANOVA, between-subjects design with age, RT, and SNR as independent variables. The sample means are displayed in Fig. 10. This analysis determined if looking behavior was significantly different across age or acoustical environment. While there was a general trend of decreasing looking behavior as a function of age in all but the most adverse environment, none of the main effects or interactions reached significance ($P > 0.05$).

The impact of the amount of head movement during the course of the experiment on a subject's comprehension score was determined. A scatter plot with looking behavior on the

abscissa and comprehension score on the ordinate is shown in Fig. 11. While the amount of looking is negatively correlated with comprehension score (1 additional degree of SD in looking behavior yields a 0.2% drop in comprehension score), this regression analysis was not significant, ($P > 0.05$). For this set of subjects, comprehension score was not significantly related to looking behavior. While increased looking may suggest that listeners are expending more cognitive resources processing the stimuli, the wide variability across age group and tested environment suggest that more research is necessary to determine the role of looking behavior during a complex learning task.

Finally, the looking behavior tracks were analyzed to determine the percentage of time during the course of the experiment that the subject was looking at the active talker. The sample means of this analysis are shown in Fig. 12. This

FIG. 9. Comprehension scores as a function of RT (left) and SNR (right) for the discussion and lecture conditions.

FIG. 10. Average looking behavior during the classroom-learning task as a function of acoustical environment for the three age groups (discussion condition only).

FIG. 11. Comprehension score as a function of looking behavior for all subjects.

analysis revealed that, irrespective of looking behavior, subjects visualized the active talker less than 50% of the time across all ages and conditions. It appears that even in the age group that exhibited the highest looking behavior (8-yr-old children), the amount of time that looking behavior actually resulted in visualizing the active talker was low. The dynamic nature of the task, with fast moving dialogue changes among talkers, similar to a realistic classroom discussion, revealed that subjects did not visualize a high percentage of the active talkers during the task. It is possible that the poor relationship between looking behavior and comprehension is related to the fact that even when listeners tried to look at talkers in an effort to improve understanding, they were unable to actually visualize the talker each time he/she spoke.

Further research is necessary to determine if looking behavior can be a significant predictor of comprehension score across a larger set of acoustical environments if subjects are able to correctly visualize the active talker at a higher rate. In addition, tasks that allow for within-subject

FIG. 12. Proportion of events visualized during the classroom-learning task as a function of acoustical environment for the three age groups (discussion condition only).

comparisons across multiple acoustical environments may provide additional information regarding the relationship between looking behavior and comprehension in active listening tasks.

Results suggest that reducing the SNR from +10 to +7 has a more adverse effect than increasing the RT from 0.6 to 1.5 s. The addition of excessive noise on reduced task performance compared with excessive reverberation is consistent with results presented by others (e.g., Bradley, 1986; Bradley *et al.*, 2003; Yang and Hodgson, 2006).

In summary in Experiment 2, subjects were tested in acoustical environments that elementary-aged children might encounter in schools. Children and adults performed listening tasks similar to those found in classrooms to better understand how children listen and learn. Despite differences in performance across ages and acoustical environments, sentence recognition was not affected by listening condition (lecture versus discussion), and scores for all subjects were high in all acoustical environments.

For the comprehension task, the younger elementary-aged children performed more poorly than the older elementary-aged children and adults in all conditions and acoustical environments. This finding supports previous research and suggests that younger students require a more favorable acoustical environment for optimal academic achievement (e.g., Crandell and Smaldino, 2000). Unlike the sentence-recognition task, comprehension scores were more adversely affected by listening in an active discussion compared to a lecture in acoustical combinations of unfavorable SNRs, RTs, or both. This finding suggests that the acoustical conditions needed for understanding in a dynamic classroom environment may vary depending on the task. It further suggests that a comprehension task, rather than a sentence-recognition task, is a more appropriate measure of children's performance in an active discussion. Including multiple talkers and locations, as well as using a task-length similar to classroom lessons, also may provide a more accurate representation of performance in active discussions than tasks that only require repetition of words and/or sentences. It was of interest to examine the listeners' looking behavior in the discussion condition to account for the differences in comprehension scores. Although a relationship was observed, looking behavior was a significant predictor of comprehension scores only in the baseline condition.

IV. SUMMARY

The present study investigated how well school-age children perform on speech- perception/intelligibility tasks similar to the types of tasks they experience in classrooms. These tasks included both a discussion (where a teacher and multiple students presented information from multiple locations around the subject) and a lecture (where a teacher alone presented material from a location in front of the subject). By using a simulated classroom that included virtual room modeling techniques as well as auditory-visual stimuli, the experimental paradigm simulated a realistic environment to a greater degree than previous studies that have presented

test materials in sound booths, over headphones, or from a single loudspeaker.

Not hearing the words a teacher is saying during the course of class will make learning difficult. The implications of excessive reverberation or a poor SNR are apparent; the audibility of speech will be reduced; this has an additive effect over the course of a typical-length classroom lesson or discussion. Indeed, in the results presented in this experiment, variance in performance in both children and adults was seen across condition, RT, and SNR scenarios. This effect was less apparent when subjects were merely required to complete a sentence-recognition task, where subjects were not required to listen, comprehend, and recall concepts after a period time but rather repeat material quickly after hearing it without long-term comprehension.

The lack of decrements in sentence recognition in unfavorable environments may result in an underestimation of the deleterious effects of poor classroom acoustics in daily learning activities for students. The sensitivity to small changes in SNR is apparent in the range of the comprehension scores in Experiment 2 for children and adults, which were not seen to the same degree in the baseline classroom tested in Experiment 1. Finally, the low POEV in all conditions across both children and adults in both experiments could indicate that subjects had a hard time keeping up with the rapidly changing dialog of the learning task, but the extent to which this would be expected to influence their comprehension scores is still an open question. These results highlight the need for diligent work during classroom design to minimize the amount of masking background noise and employ reverberation optimization techniques such those described by [Hodgson and Nosal \(2002\)](#), [Yang and Hodgson \(2006\)](#), and [Bradley et al. \(2003\)](#) for example. Results of the decrement in performance reveal that depending on the complexity of the learning task, even classrooms that meet ANSI standards still may not be an optimal environment for efficient listening and learning.

Continued research is needed to extend the age range to include younger children and children with hearing loss as well as to include tasks that involve multitasking (such as listening to a lecture and taking notes) and divided attention. By conducting studies that maintain a greater degree of ecological validity, the impact of acoustical environment on a child's ability to learn and understand in the classroom can be more fully understood. The impact of this ecological validity can be investigated by another

potential topic for future study: determining the extent including visual cues of the talkers has the ability to influence comprehension. This could be investigated by having conditions that include auditory-only versus audio-visual stimuli. Further analysis of looking behavior, such as the ability to correctly identify talkers as they speak, may provide additional information about the effects of looking behavior on comprehension.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support for this research through NIH grants R03 DC009675 (D. L., E. H., J. F.), T32 DC000013 (D. V.), T35 DC08757 (H. P.), and P30 DC004662. The authors thank two anonymous reviewers for helpful comments and suggestions on an early version of this paper. Finally, the authors thank Roger Harpster at the Boys Town National Research Hospital for assistance with the setup of the laboratory space and videos.

APPENDIX

The following section describes the background noise, clarity, and reverberation measurement for the current paper. The background noise measurements and IRs were measured and analyzed using the same methods described in [Sec. II A](#). [Table III](#) describes the octave-band background noise measurements for the simulated classroom with no virtual acoustics added (labeled: "No VA"), and with the +10 and +7 dB SNRs used for the experiment. In addition, the background noise levels for the sound-treated meeting room used for recording the discussion and lecture-type stimuli are provided.

[Table IV](#) shows the octave-band RT (both EDT and T_{30}) and C_{50} for the simulated classroom with no virtual acoustics added and in the two RT-conditions used for the experiment (0.6 s and 1.5 s). The internal T_{60} for the ViMiC software's reverberator module is also provided when applicable. For each of the measurements in the simulated classroom, the values are the mean and SD of the five sound-source locations (see [Fig. 2](#)). Finally, data are provided for the meeting room used to record the discussion and lecture-type stimuli, both using a far-field source-receiver distance (3.5 m source-receiver distance) and a close-microphone source-receiver distance as was used in recording the individual stimuli (10 cm).

TABLE III. Octave-band filtered background noise measurements for the acoustical environments used for the study.

Measurement/location	Octave-band center (Hz)						
	125	250	500	1000	2000	4000	8000
Meeting Room	45.0	38.2	33.9	25.4	19.9	19.5	20.2
Simulated classroom (No VA)	52.9	40.3	29.0	22.8	20.5	19.6	20.4
Simulated classroom + 10 SNR	58.0	55.1	45.7	43.9	38.4	34.4	25.3
Simulated classroom + 7 SNR	59.9	58.1	49.4	47.2	41.9	37.6	28.0

TABLE IV. Octave-band filtered C_{50} , EDT, and T_{30} measurements for the environments used for the study. For the simulated classroom measurements, the mean and SD for each of the five sources is given. The meeting room close-microphone measurement was made using a 10 cm source-receiver distance similar to how stimuli recordings were captured within the space, as well as a far-field 3.5 m source-receiver distance. The VA T_{60} is the internal T_{60} value used with the ViMiC software, if applicable.

Location/condition	Measure	VA T_{60}	Octave band (Hz)						
			125	250	500	1000	2000	4000	8000
Simulated classroom, no VA	C_{50} (dB)	N/A	7.32 (1.44)	9.14 (1.40)	12.30 (2.55)	14.40 (1.95)	14.00 (1.80)	13.00 (0.73)	14.80 (0.88)
	EDT (s)		0.40 (0.10)	0.35 (0.04)	0.25 (0.08)	0.25 (0.03)	0.23 (0.04)	0.24 (0.02)	0.22 (0.01)
	T_{30} (s)		0.53 (0.01)	0.51 (0.02)	0.40 (0.01)	0.35 (0.01)	0.4 (0.03)	0.4 (0.03)	0.34 (0.02)
Simulated classroom, 0.6 RT	C_{50} (dB)	0.64	1.96 (3.03)	2.36 (1.93)	2.6 (1.75)	3.62 (2.21)	4.84 (1.20)	6.36 (1.80)	11.0 (1.86)
	EDT (s)		0.58 (0.14)	0.59 (0.08)	0.62 (0.05)	0.6 (0.02)	0.6 (0.04)	0.5 (0.05)	0.37 (0.08)
	T_{30} (s)		0.57 (0.03)	0.58 (0.03)	0.62 (0.03)	0.62 (0.03)	0.56 (0.04)	0.48 (0.02)	0.41 (0.01)
Simulated classroom, 1.5 RT	C_{50} (dB)	1.40	-1.80 (1.71)	-2.76 (1.19)	-2.52 (0.75)	-1.80 (0.98)	-0.66 (1.23)	1.82 (1.44)	7.62 (1.77)
	EDT (s)		0.96 (0.12)	1.10 (0.12)	1.44 (0.04)	1.43 (0.11)	1.17 (0.01)	0.84 (0.04)	0.47 (0.02)
	T_{30} (s)		1.24 (0.04)	1.32 (0.10)	1.62 (0.02)	1.63 (0.03)	1.42 (0.03)	1.03 (0.05)	0.60 (0.03)
Meeting room (3.5 m)	C_{50} (dB)	N/A	1.70	0.87	2.13	3.37	4.07	4.90	7.20
	EDT (s)		0.60	0.73	0.74	0.68	0.81	0.75	0.56
	T_{30} (s)		1.12	1.12	0.93	0.99	1.13	1.21	0.98
Meeting room (10 cm)	C_{50} (dB)	N/A	17.60	13.40	10.70	13.00	13.40	16.90	19.90
	EDT (s)		0.05	0.05	0.03	0.17	0.05	0.01	0.01
	T_{30} (s)		1.11	1.06	0.83	0.86	0.99	0.98	0.71

- Allen, J., and Berkley, D. (1979). "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65**(4), 943–950.
- ANSI (1997), S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).
- ANSI (2010), 12.60-2010/Part 1, *Acoustical Performance Criteria, Design Requirements and Guidelines for Schools, Part 1: Permanent Schools* (Acoustical Society of America, New York).
- Bench, J., Koval, A., and Bamford, J. (1979). "The BKB (Bamford-Koval-Bench) sentence lists for partially-hearing children," *Br. J. Audiol.* **13**, 108–112.
- Bistafa, S. R., and Bradley, J. S. (2000). "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," *J. Acoust. Soc. Am.* **107**(2), 861–875.
- Boman, E. (2004). "The effects of noise and gender on children's episodic and semantic memory," *Scand. J. Psychol.* **45**, 407–416.
- Blazier, W. (1981). "Revised noise criteria for application in the acoustical design and rating of HVAC systems," *Noise Control Eng.* **16**, 64–73.
- Braasch, J., Peters, N., and Valente, D. L. (2008). "A loudspeaker-based projection technique for spatial music applications using virtual microphone control," *Comput. Music J.* **32**(3), 55–71.
- Bradley, J. (1986). "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.* **80**, 837–845.
- Bradley, J., and Sato, H. (2008). "The intelligibility of speech in elementary school classrooms," *J. Acoust. Soc. Am.* **123**(4), 2078–2086.
- Bradley, J., Sato, H., and Picard, M. (2003). "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.* **113**(6), 3233–3244.
- Bradley, J. S., Reich, R., and Norcross, G. N. (1999). "A just noticeable difference in C_{50} for speech," *Appl. Acoust.* **58**, 99–108.
- Crandall, C. C., and Smaldino, J. J. (2000). "Classroom acoustics for children with normal hearing and with hearing impairment," *Lang. Speech Hear. Serv. Schools* **31**, 362–370.
- Dockrell, J., and Shield, B. (2006). "Acoustical barriers in classrooms: The impact of noise on performance in the classroom," *Br. Educ. Res. J.* **32**, 509–525.
- Elliott, L. L. (1979). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," *J. Acoust. Soc. Am.* **66**, 651–653.
- Erber, N. P. (1975). "Auditory-visual perception of speech," *J. Speech Hear. Disord.* **40**, 481–492.
- Evans, G. (2006). "Child development and the physical environment," *Annu. Rev. Psychol.* **57**, 423–451.
- Fallon, M., Trehub, S. E., and Schneider, B. A. (2000). "Children's perception of speech in multitalker babble," *J. Acoust. Soc. Am.* **108**, 3023–3029.
- Fallon, M., Trehub, S. E., and Schneider, B. A. (2002). "Children's use of semantic cues in degraded listening environments," *J. Acoust. Soc. Am.* **111**(1), 2242–2249.
- Farina, A. (2000). "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *108th AES Convention*, Paris, France, Preprint No. 5093 (D-4), pp. 1–24.
- Finitzo-Hieber, T., and Tillman, T. (1978). "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," *J. Speech Hear. Res.* **21**, 440–458.
- Hall, J. W. III, Grose, J. H., Buss, E., and Dev, M. B. (2002). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," *Ear Hear.* **23**, 159–165.
- Hodgson, M., and Nosal, E.-M. (2002). "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," *J. Acoust. Soc. Am.* **111**(2), 931–939.
- International Organization for Standardization (1997), *Measurement of the reverberation time of rooms with reference to other acoustical parameters*. ISO 3382-Acoustics (ISO, Geneva, Switzerland).
- Jamieson, G., Kranjc, G., Yu, K., and Hodgetts, W. (2004). "Speech intelligibility of young school-aged children in the presence of real-life classroom noise," *J. Am. Acad. Audiol.* **15**, 508–517.
- Johnson, C. E. (2000). "Children's phoneme identification in reverberation and noise," *J. Speech Lang. Hear. Res.* **43**, 144–157.
- Klatte, M., Hellbrück, J., Seidel, J., and Leistner, J. (2010a). "Effects of classroom acoustics on performance and well-being in elementary school children: A field study," *Environ. Behav.* **42**, 659–692.
- Klatte, M., Lachmann, T., and Meis, M. (2010b). "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting," *Noise Health Noise, Mem. Learn.* **12**, 270–282.
- Klatte, M., Meis, M., Sukowski, H., and Schick, A. (2007). "Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children," *Noise Health* **9**, 64–74.
- Kleiner, M., and Svensson, P. (1995). "A review of active systems in room acoustics and electroacoustics," *Proceedings of Active95, International Symposium on Active Control of Sound and Vibration*, July 6-8, Newport Beach, CA, pp. 39–54.
- Knecht, H., Nelson, P., Whitelaw, G., and Feth, L. (2002). "Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements," *Am. J. Audiol.* **11**, 65–71.

- Kodaras, M. (1960). "Reverberation times of typical elementary school settings," *Noise Control* **6**, 17–19.
- Mehta, M., Johnson, J., and Rocafor, J. (1999). *Architectural Acoustics: Principles and Design*, edited by E. Francis (Prentice-Hall, Upper Saddle River, NJ), Appendix H, pp. 407–411.
- McFadden, B., and Pittman, A. (2008). "Effect of minimal hearing loss on children's ability to multitask in quiet and in noise," *Lang. Speech Hear. Serv. Schools* **39**, 342–351.
- McCroskey, F., and Devens, J. (1975). "Acoustic characteristics of public school classrooms constructed between 1890 and 1960," in *Proceedings of Noise Expo, National Noise and Vibration Control Conference* (Acoustical Publications, Bay Village, OH), pp. 101–103.
- McKellin, W. H., Shahin, K., Hodgson, M., Jamieson, J., and Pichora-Fuller, K. (2007). "Pragmatics of conversation and communication in noisy settings," *J. Pragmat.* **39**, 2159–2184.
- McKellin, W. H., Shahin, K., Hodgson, M., Jamieson, J., and Pichora-Fuller, K. (2011). "Noisy zones of proximal development: Conversation in noisy classrooms," *J. Sociolinguist.* **15**, 65–93.
- Nabelek, A., and Pickett, J. (1974). "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids," *J. Acoust. Soc. Am.* **56**, 628–639.
- Nelson, E., Smaldino, J., Erler, S., and Garstecki, D. (2008). "Background noise levels and reverberation times in old and new elementary school classrooms," *J. Educ. Audiol.* **14**, 12–18.
- Neuman, A., and Hochberg, I. (1983). "Children's perception of speech in reverberation," *J. Acoust. Soc. Am.* **74**, 215–219.
- Neuman, A., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2010). "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," *Ear Hear.* **31**(3), 336–344.
- Nittrouer, S., and Boothroyd, A. (1990). "Context effects in phoneme and word recognition by young children and older adults," *J. Acoust. Soc. Am.* **87**, 2705–2715.
- Picard, M., and Bradley, J. S. (2001). "Revisiting speech interference in classrooms," *Audiology* **40**, 221–244.
- Rosenthal, R., and DiMatteo, M. R. (2001). "Recent developments in quantitative methods for literature review," *Annu. Rev. Psychol.* **52**, 59–82.
- Ryalls, B. O., and Pisoni, D. B. (1997). "The effect of talker variability on word recognition in preschool children," *Dev. Psychol.* **33**, 441–451.
- Sanders, D. (1965). "Noise conditions in normal school classrooms," *Except. Child.* **31**, 344–353.
- Shepard, A. (2010). "Aaron Shepard's RT Page: Scripts and tips for reader's theater," <http://www.aaronsherp.com/rt/index.html#RTE> (Last viewed February 22, 2010).
- Studebaker, G. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Svensson, P. (1994). "On reverberation enhancement in auditoria," Ph.D. thesis, Chalmers University of Technology, Gothenburg, Sweden.
- Valente, D. L., and Braasch, J. (2008). "Subjective expectations adjustments of early-to-late reverberant energy ratio and reverberation time to match visual environmental cues of a musical performance," *J. Acoust.* **94**, 840–855.
- Yang, W., and Bradley, J. (2009). "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *J. Acoust. Soc. Am.* **125**(2), 922–933.
- Yang, W., and Hodgson, M. (2006). "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," *J. Acoust. Soc. Am.* **120**(2), 801–807.