5-2016

Effects of Noise, Reverberation and Foreign Accent on Native and Non-Native Listeners' Performance of English Speech Comprehension

Zhao Ellen Peng

University of Nebraska - Lincoln, ellen.z.peng@gmail.com

Lily M. Wang

University of Nebraska - Lincoln, lwang4@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/archengfacpub>

Part of the [Acoustics, Dynamics, and Controls Commons](#), and the [Architectural Engineering Commons](#)

Peng, Zhao Ellen and Wang, Lily M., "Effects of Noise, Reverberation and Foreign Accent on Native and Non-Native Listeners' Performance of English Speech Comprehension" (2016). *Architectural Engineering -- Faculty Publications*. 74.
<http://digitalcommons.unl.edu/archengfacpub/74>

This Article is brought to you for free and open access by the Architectural Engineering at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Architectural Engineering -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Effects of noise, reverberation and foreign accent on native and non-native listeners' performance of English speech comprehension

Z. Ellen Peng^{a)} and Lily M. Wang

Durham School of Architectural Engineering and Construction, University of Nebraska-Lincoln,
1110 South 67th Street, Omaha, Nebraska 68182-0681, USA

(Received 3 July 2015; revised 8 April 2016; accepted 12 April 2016; published online 16 May 2016)

A large number of non-native English speakers may be found in American classrooms, both as listeners and talkers. Little is known about how this population comprehends speech in realistic adverse acoustical conditions. A study was conducted to investigate the effects of background noise level (BNL), reverberation time (RT), and talker foreign accent on native and non-native listeners' speech comprehension, while controlling for English language abilities. A total of 115 adult listeners completed comprehension tasks under 15 acoustic conditions: three BNLs (RC-30, RC-40, and RC-50) and five RTs (from 0.4 to 1.2 s). Fifty-six listeners were tested with speech from native English-speaking talkers and 59 with native Mandarin-Chinese-speaking talkers. Results show that, while higher BNLs were generally more detrimental to listeners with lower English proficiency, all listeners experienced significant comprehension deficits above RC-40 with native English talkers. This limit was lower (i.e., above RC-30), however, with Chinese talkers. For reverberation, non-native listeners as a group performed best with RT up to 0.6 s, while native listeners performed equally well up to 1.2 s. A matched foreign accent benefit has also been identified, where the negative impact of higher reverberation does not exist for non-native listeners who share the talker's native language. © 2016 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4948564>]

[TCB]

Pages: 2772–2783

I. INTRODUCTION

Clear communication is the key to successful learning in classroom settings. ANSI S12.60 (2010) provides design recommendations for background noise level (BNL) and reverberation time (RT) in unoccupied classrooms to ensure good speech intelligibility in core learning spaces, based on optimizing the signal-to-noise ratio (SNR).¹ The majority of the work that ANSI S12.60 references, however, is focused on native English speakers and listeners; this paper seeks to understand how BNL, RT, and talker foreign accent can impact speech comprehension performance by native versus non-native² English-speaking listeners.

Research studies have shown repeatedly that mechanical equipment for a building's heating, ventilating, and air-conditioning (HVAC) system is a major source of background noise that negatively affects students' academic achievement (Nelson and Soli, 2000; Knecht *et al.*, 2002; Nelson *et al.*, 2005). Ronsse and Wang (2010, 2013) studied such impact using *in situ* data and predicted that, with a 1 dBA increase in the unoccupied BNL due to HVAC equipment, the standardized reading comprehension accuracy score may decrease by approximately 1.6% for both second and fourth grade students.

While excessive BNL is unanimously regarded as impairing speech perception, there is less agreement on the

role of reverberation, particularly in the lower range of less than 1.0 s RT. RT is defined as the time it takes for sound energy to decay 60 dB in an enclosed space, and is dependent on both the room volume and amount of sound absorption in the room. Hodgson and Nosal (2002) calculated the optimal RTs to be less than 0.3 s in order to achieve SNRs above +20 dB for classrooms between 300 and 500 m³. In contrast, Bradley and colleagues (1999, 2003, 2008, 2009) argue that early reflections are critical in reinforcing and supporting the directly arriving sound, providing useful sound energy for listeners to resolve auditory information. It was further shown that speech intelligibility performances were at maximum for both adults and children of different ages when RT was at approximately 0.6 s (Yang and Bradley, 2009).

The difficulty in studying the effect of reverberation was recognized by Beaman and Holt (2007) who predicted that, in order to provide statistically significant results (based on an *a priori* statistical power of 0.8), the sample size necessary to study a small difference in RT of less than 0.2 s was as large as 100 participants in a between-subject design. Recent studies on the effect of reverberation have mainly focused on investigating a much larger RT difference of over 0.5 s, though. Ljung and Kjellberg (2009) found that participants performed more poorly and reported investing more effort during recall tasks under 1.2 s than 0.5 s of RT. A study by Klatt *et al.* (2010) using simulated virtual rooms showed that the accuracy of speech perception from word recall tasks was significantly lower under 1.1 s than 0.5 s of RT, for both adults and children in first and third grades. In

^{a)}Current address: Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstrasse 5, 52074 Aachen. Electronic mail: zellenpeng@gmail.com

particular, the main effect of RT had a large calculated effect size in η_p^2 of 0.36. A similar effect size of RT was also realized in research by Valente *et al.* (2012), who tested 8- and 11-yr-old children and adults under two RT conditions of 0.6 and 1.5 s. The main effect of RT was reported in Pearson's r of 0.53, which is equivalent to η_p^2 of 0.31.

There is much support in the existing literature for the RT in classrooms to be less than 1.0 s to facilitate speech intelligibility, but more investigations using smaller intervals of RT in the test design are needed. In the research presented here, five RT scenarios between 0.4 and 1.2 s in approximately equivalent intervals have been simulated and crossed with three BNL conditions of Room Criteria RC-30, RC-40, and RC-50 to create a total of 15 within-subject acoustic combinations. Room Criteria is a method for rating interior noise based on measured octave-band BNLs, as described in ANSI S12.2-2008 (2008).

In addition to investigating a range of RT and BNL, an innovative testing paradigm has been developed and utilized in this investigation to measure speech comprehension performance instead of speech intelligibility or recognition. Prior investigations on classroom acoustics have primarily used speech intelligibility or recognition tasks. These tasks involve short speech stimuli, such as vowels, syllables, words, or single sentences, and measure accuracy performance by calculating the percent of target stimuli correctly identified. In this paper, the testing paradigm measures comprehension performance using longer speech stimuli, requiring more complex cognitive processing in that participants must derive meaning rather than simply repeating the stimulus itself. Comprehension tasks with longer speech stimuli also mimic realistic classroom activities more accurately, as students are asked to do more than simply repeat back the speech stimuli produced by teachers. Klatte *et al.* (2010) and Valente *et al.* (2012) utilized both speech intelligibility and speech comprehension tasks in their investigations. Although no direct comparisons were conducted, results from these two studies implied that background noise and reverberation were more detrimental to speech comprehension performance than speech intelligibility. When studying the effect of foreign accent, Munro and Derwing (1995) noted that listeners might consider accented speech highly comprehensible while in fact performing poorly on intelligibility when transcribing the speech. Based on these previous findings and the authors' intent to simulate realistic classroom activities, speech comprehension tasks were chosen over intelligibility tasks for this investigation.

As mentioned earlier, recommendations for RT and BNL in the current classroom acoustics standard ANSI S12.60 (2010) are primarily based on results from speech intelligibility studies using native listeners perceiving speech materials produced by native English speakers. However, a recent Institute of Education Sciences survey showed that 21% of students in the U.S. aged 5–17 (or 10.9×10^6 students) speak a language other than English at home (Aud *et al.*, 2010). The population of non-native English-speakers in university classrooms is expected to be similar if not larger, as the U.S. continues to be the most popular destination for international students (Institute for International

Education, 2012). In addition, the population of non-native speakers among university instructors in the fields of science, technology, engineering, and mathematics continues to increase, accounting for 19% of instructors in psychology and 54% in engineering from a recent survey (National Science Board, 2012). To date, speech perception and production by this non-native population under realistic adverse acoustic conditions is not widely understood and consequently is not well-considered in the current design of spaces for speech communication.

Non-native listeners have been found to perform worse than native listeners on speech recognition tasks under extremely adverse listening conditions, not commonly found in the built environment. A number of speech intelligibility studies specifically compared native and non-native listeners' performances on recognition tasks by varying SNRs (mostly below 0 dB) and using white noise or speech-shaped noise as the masker. The stimuli used in the recognition tasks varied between different levels of phonological units including vowels and consonants, words (Rogers *et al.*, 2006; Bent *et al.*, 2010), and sentences (Bradlow and Bent, 2002; Bradlow and Alexander, 2007).

Research involving non-native English talkers has shown that native listeners perform more poorly in understanding foreign-accented speech, compared to speech from native talkers, under the presence of babble noise or reduced SNR (Munro, 1998; Rogers *et al.*, 2004). Another study on non-native speakers found that those who are immersed in English-speaking communities later in life tend to have foreign accents when speaking English that persist throughout their lifetime (Flege *et al.*, 1999).

A finding of particular interest to this investigation is that non-native listeners may have an interlanguage speech intelligibility benefit in perceiving foreign-accented speech from non-native talkers, with whom they share the same native language, than speech from native talkers (Bent and Bradlow, 2003; Imai *et al.*, 2003; Wang and van Heuven, 2015). The non-native listeners achieve better speech intelligibility performance with the matched non-native talker than with a native talker. Such phenomenon has not been studied extensively in the presence of noise or reverberation, though.

Based on the aforementioned considerations, the current project aims to answer the following research questions:

- (1) What are the effects of BNL and RT on English speech comprehension, after accounting for an individual's English language abilities?
- (2) How does a foreign accent affect speech comprehension under adverse acoustic conditions (BNL and RT)?
- (3) How do BNL and RT affect speech comprehension for native and non-native listeners?

Fifty-six normal-hearing adults containing both native and non-native listeners underwent testing with speech comprehension from native talkers. Another 59 listeners were tested with the same speech materials produced by non-native talkers, whose native language is Mandarin Chinese. Other aspects of the methodology and procedure were the same for all tested listeners and are presented in Sec. II.

II. METHODOLOGY

A. Experimental design

1. Acoustic stimuli

All speech comprehension experiments were conducted in the listening chamber at the University of Nebraska. The listening chamber uses a room-in-room design with additional absorptive material introduced to the interior to further reduce the ambient RT to a T_{20} of 0.22 s, averaged across 500–2000 Hz, as measured at the listener position. The ambient BNL of the listening chamber is RC-28 hissy.

To expand beyond research conducted by Klatte *et al.* (2010; 2 noise-type $\times$ 2 SNR) and Valente *et al.* (2012; 2 SNR $\times$ 2 RT), a wider range of acoustic conditions was utilized for this study. A total of 15 acoustic conditions were created from combinations of three levels of BNL (RC-30, RC-40, and RC-50; corresponding to a SNR of +21, +11, and +1 dB, respectively) and five RT scenarios (0.4 to 1.2 s).

Background noise was introduced via a subwoofer in the corner of the chamber and a ceiling panel loudspeaker integrated behind an acoustical panel directly above the listener position. To calibrate the test signals, pink noise was first introduced then digitally filtered to create three conditions of BNL that optimally followed the Room Criteria neutral contours of RC-30, RC-40, and RC-50. The steady-state sound pressure levels in octave bands for the three BNL conditions were measured at the listener position, as seen in Fig. 1; the RC-30 condition was slightly hissy due to ambient conditions at the 4000 Hz octave band. During each main experiment testing session, one of the BNL test signals in WAV format was played back continuously.

The RT scenarios used binaural room impulse responses (BRIRs) simulated from a typical classroom of 260 m³ in the room acoustics modeling program ODEON (Lyngby, Denmark) (version 10). The RT scenarios were varied by combining different ceiling materials with 25-mm acoustical wall panels (NRC 0.70), applied full height on the side and back walls with uniformly scaled absorption coefficients. The BRIRs for the simulated RT scenarios were exported from

FIG. 1. BNLs in the test chamber during ambient and test conditions, as measured at the listener position, from 32 to 4000 Hz.

FIG. 2. RT in T_{20} from 125 to 8000 Hz, measured at the listener position in the listening chamber, for the ambient and five RT scenarios. Error bars indicate one standard deviation from 10 *in situ* measurements. The single numbers given in parentheses on the legend are the average T_{20} measured in the room from 500 to 2000 Hz.

ODEON using the “2D Surround Sound” option, which supported auralization playback via loudspeakers, after adjusting for the distances between the two-channel loudspeakers [Yamaha (Buena Park, CA) HS-50] and the receiver position in the listening chamber. Since the listening chamber was not anechoic, the actual RTs measured at the listener position differed slightly from the simulated RTs and are reported in Fig. 2 across octave band frequencies.

The BRIRs were then digitally convolved in MATLAB with aurally dry speech comprehension materials (discussed further in Sec. II A 2) for playback to listeners during the main experiment. All convolved speech comprehension materials were calibrated at the listener position to playback at 59 dBA (± 2 dB, re 20 μ Pa), across all RT scenarios. The resulting SNR, speech intelligibility index (SII), and speech transmission index (STI)³ are shown in Table I. The test conditions span a wide range of acoustic conditions as may be found in real classrooms.

2. Speech materials

A total of 15 equivalent sets of speech comprehension tests in English were created from preparation materials for

TABLE I. SII, STI, and SNR produced in each acoustic condition.

BNLs	SII (top) STI (bottom)					SNR
	RT Scenarios					
	0.4 s	0.6 s	0.8 s	1.0 s	1.2 s	
RC-30 Hissy (38 dBA)	0.89	0.89	0.89	0.89	0.89	+1 dB
RC-40 Neutral (48 dBA)	0.74	0.68	0.65	0.63	0.61	+11 dB
RC-50 Neutral (58 dBA)	0.80	0.84	0.85	0.85	0.85	+21 dB
	0.64	0.60	0.58	0.56	0.54	
	0.52	0.57	0.58	0.58	0.59	
	0.40	0.40	0.39	0.39	0.38	

the listening tests in the Test of English for International Communication (TOEIC). The TOEIC test was used as the template, rather than the TOEFL (Test of English as Foreign Language), because many non-native participants had taken the TOEFL to gain university admission and were thus less familiar with the TOEIC test. Also, the TOEIC requires a slightly smaller vocabulary size of 4000 words compared to 4500 for the TOEFL (Chujo and Oghigian, 2009). The content in the TOEIC test materials covered daily life events using a simple vocabulary that was expected to be understood easily by all non-native listeners who participated in this research. These test items were recorded by native English speakers (one male and four females) in an anechoic chamber and again by two native speakers of Mandarin Chinese using a closely aligned microphone in a sound-attenuated booth. The test items recorded with the native English speakers were later screened by five native listeners, when played back under the ambient condition in the listening chamber, to ensure equivalent difficulty across the 15 test sets. The non-native speakers were one male and one female, who shared a similar degree of accentedness in their spoken English, as screened by the specific skill areas of fluency and pronunciation in the Versant Test (Pearson, 2008). The female Chinese talker received a standardized *t*-score of 55 on fluency and 52 on pronunciation, whereas the male Chinese talker scored 58 and 53, respectively. All talkers were instructed to speak with normal vocal effort and at their normal conversational speed. The speech rate in syllables per second for the native English speakers was calculated to be 5.3 for the male talker and between 3.4 and 5.0 for the four female talkers. The speech rates of the Mandarin Chinese talkers were similar, calculated to be 5.1 for the male talker and 4.0 for the female talker. These speech materials were recorded with a sampling frequency of 44.1 kHz and 16-bit resolution.

Each speech comprehension test set contained four tasks: (1) photograph recognition, in which participants were asked to identify one of the four spoken sentences that matched the photograph displayed on the computer screen; (2) question and response, in which they needed to identify one of the four spoken sentences that best responded to the spoken question; (3) conversation, in which they listened to conversations exchanged between a male talker and a female talker and had to answer spoken questions related to the content with answer options displayed on the computer screen; and (4) paragraph, in which they listened to short paragraphs produced by a single talker and answered questions pertinent to the material. Each test set lasted no more than 15 min and contained 32 multiple-choice items, with performance recorded in percent correct based on accuracy. Each test set was randomly paired with one of the 15 acoustic conditions for each listener; checks were conducted to confirm that the probability of each test set to appear in any serial position during the main experiment was approximately the same across all participants. A custom-made program written in C# was used to present visual prompts and audio for the speech comprehension tests.

3. Composite scale of English proficiency

Conceptually, speech comprehension performance relies heavily on one's target language abilities. An individual's English language abilities can confound speech comprehension performance in acoustic environments, and hence must be controlled in the statistical analysis to better understand the genuine effects of room acoustics. During initial screening, all participants were individually given three tests to establish their English language abilities, both in the cognitive and linguistic domains; these included oral comprehension (Woodcock *et al.*, 2001a), listening span (Woodcock *et al.*, 2001b), and verbal abilities using the English portion of the Bilingual Verbal Ability Tests (Muñoz-Sandoval *et al.*, 1998). Longer speech stimuli were used in this investigation to gauge speech comprehension. A listening span test was included since previous research has shown that listening span affects reading comprehension, which shares similar cognitive processes with speech comprehension (Daneman and Carpenter, 1980).

Both the listening span and oral comprehension tests involved spoken materials. The materials were recorded by a female native English speaker in a sound-attenuated booth with a closely aligned microphone and played back for participants during their initial screen via headphones [Sennheiser (Wedemark, Germany) HD 497]. Participants were encouraged to choose a comfortable listening level of L_{eq} between 65 and 68 dBA (re 20 μ Pa) with L_{max} between 70 and 75 dBA (re 20 μ Pa).

The three tests were used to form a composite scale to measure each individual participant's overall English language abilities, referred to in the remainder of this manuscript as an English proficiency level. The raw scores from each test were first verified to conform to normality before being converted into standardized *z*-scores. The composite score was then calculated by taking the mean of the *z*-scores of the three English language ability tests. The composite scale achieved a high internal consistency with a Cronbach's α of 0.94. More information on the tests used for scoring English proficiency may be found in Peng (2014).

B. General procedure

After providing written consent, each participant was asked to complete an initial screening which lasted a maximum of 2 h; this screening included an orientation, a hearing screening, a demographic survey, and the three English language tests. All participants were screened for hearing thresholds below 25 dB hearing level on both ears from 125 to 8000 Hz. Participants were also asked to complete items adopted from the Language Experience and Proficiency Questionnaire [LEAP-Q (Marian *et al.*, 2007)] to provide data on individual language experience in English. Once they passed the initial screening, participants were invited back for six one-hour long sessions on separate days for the main experiment.

A dual-task paradigm was used in the main experiment with speech comprehension tests as the primary task and an adaptive pursuit rotor (APR) task as the secondary task. The secondary APR task was added to minimize the likelihood

of participants achieving 100% accuracy on the speech comprehension tests. The participants viewed two computer screens, the upper of which showed the graphical user interface for the speech comprehension tests and the lower of which showed the APR task. The APR task required that participants trace a dot as it continuously rotated around a fixed circle, using a stylus on a touchpad controlled by their dominant hand. The dot's speed changed adaptively so that participants would remain approximately 80% on target, and was recorded as an outcome variable (Srinivasan, 2010). A keypad was used by the participants to enter responses for the speech comprehension tests with their non-dominant hand.

Each main experiment session consisted of three speech comprehension test sets, corresponding to testing for three acoustic conditions. From the authors' previous experience, participants tend to be more conscious of the environmental change from varying BNL within a test session. To reduce participants' sensitivity toward the experimental design, the three tests in each hour-long session contained the same BNL condition but with varying RT embedded in the speech materials. A nested Latin square design was utilized to counterbalance the order of presentation for both BNL and RT. A two-factor within-subject design, 3 BNL (RC-30, RC-40, and RC-50) $\times$ 5 RT (five scenarios from 0.4 to 1.2 s), was achieved by exposing each participant to all 15 acoustic conditions. One filler test set and acoustic condition was added to the beginning of each BNL condition as practice trials; these were not entered into analyses.

The study was approved by the Institutional Review Board at the University of Nebraska—Lincoln prior to recruitment. Participants each received a total of \$100 through direct bank deposit or gift cards after completing all test sessions in the study.

C. Participants

A total of 117 listeners participated, 58 of whom were tested using speech materials produced by native talkers while the other 59 were tested using speech from native-Chinese-speaking talkers. Participants were recruited via flyers posted on the University of Nebraska at Omaha campus. Based on the native languages reported on the LEAP-Q items in the demographic survey, participants were categorized into three listener groups: (1) native American English speakers (NAE), (2) non-natives who speak native Mandarin Chinese (NNC), and (3) non-natives who speak native languages other than Mandarin Chinese (NNO). The native languages of the NNO listener group spanned a large variety, including Albanian ($n=1$), Arabic ($n=2$), Ewe ($n=1$), Hainanese ($n=1$), Hindi ($n=8$), Kannada ($n=2$), Korean ($n=2$), Nepali ($n=3$), Newari ($n=1$), Persian ($n=1$), Portuguese ($n=6$), Spanish ($n=4$), and Telugu ($n=8$).

It was later found that two listeners (one NAE and one NNC) were unable to complete the dual tasks simultaneously during the main experiment; they were thus removed from data analysis. The final number of participants was then 56 (26 female), who were tested with speech from native talkers. One NNO participant self-identified as a non-native English speaker but scored highly on the English proficiency

tests, within one standard deviation of the proficiency scores among all native listeners. This participant's performance was later found to be an outlier in the speech comprehension performance among other non-native listeners as well. This participant's results were hence only included in statistical models that did not involve comparisons between listener groups. Besides two non-native listeners who reported extensive stays of 20 and 25 yrs in an English dominant community, the average length of immersion in an English-speaking community was 23.6 months (range = 1–90 months).

For all participants tested with speech from the non-native talkers, an additional talker familiarity screen was given during the initial screen to control for possible bias in speech comprehension due to talker voice familiarity since the Mandarin Chinese talkers were recruited from the same community. Among the 59 listener participants (31 female), the male talker was correctly identified only by one listener and the female talker by two listeners. Results from these particular listeners were further analyzed, but no particular bias was found. The average length of immersion in an English-speaking community was 78.1 months (range = 2 to 564 months) for the non-native listeners in this subgroup. No outliers among these non-native listeners were identified as exhibiting exceptional English proficiency level or speech comprehension performance.

Table II reports the average scores of each listener group for each of the administered English language tests and the computed composite English proficiency scale. A simple one-way analysis of variance (ANOVA) indicates that the three listener groups achieve different English proficiency scores, all significantly different from each other at the $p < 0.001$ level as determined from a Tukey HSD *post hoc* test. NNC listeners scored lowest on English proficiency as a group. Listener age was found to significantly predict English proficiency level [$\beta = -0.036$, $t(113) = -2.24$, $p = 0.027$], but its effect on speech comprehension performance was negligible ($p = 0.40$).

III. ANALYSIS AND RESULTS

Prior to analysis, a transformation using rationalized arcsine units (or RAUs) (Studebaker, 1985; Sherbecoe and Studebaker, 2004) was performed on the percent correct data for the speech comprehension measure to adjust for normal distribution. The possible range of RAU scores is between -20 and 120 . The speech comprehension scores in RAU were verified to conform to normality for the majority of the 15 acoustic conditions with non-significant Shapiro-Wilke tests. In the statistical analysis, a mixed-design multivariate analysis of covariance (MANCOVA) was conducted first to examine the room acoustic effects on both the speech comprehension and APR tasks together. Then, the MANCOVA was followed up by analysis of covariance (ANCOVA) using either speech comprehension or APR performance as the univariate dependent variable. In this paper, only ANCOVAs of speech comprehension are reported as these are of the greatest interest; details on the other statistical test results may be found in Peng (2014). Assumptions of sphericity were confirmed for the speech comprehension scores

TABLE II. Means and standard deviations in parentheses for each of the three listener groups with regards to age; scores on the listening span, oral comprehension, and BVAT (English only) tests; and the computed composite standardized z-score representing English proficiency.

Listener group	Age in years	Listening span	Oral comprehension	BVAT (English only)	Composite z-score
NAE $n = 46$ (25 female)	23.4 (5.8)	54.5 (4.6)	29.0 (2.6)	107.6 (9.7)	0.94 (0.39)
NNC $n = 29$ (18 female)	26.4 (3.4)	37.9 (5.4)	15.5 (5.6)	73.6 (11.8)	-0.92 (0.56)
NNO $n = 39$ (14 female)	25.8 (5.9)	43.1 (5.6)	18.1 (6.1)	82.3 (9.2)	-0.48 (0.56)

in RAUs by checking that Mauchly's W was non-significant for BNL and RT in all ANCOVA and ANOVA models reported below.

A. Effects of BNL and RT with English proficiency score as a covariate

ANCOVA models of the results from native talkers ($N = 56$) and non-native talkers ($N = 59$) were first analyzed separately, using BNL and RT as within-subject variables and composite English proficiency score as a covariate.

With speech produced by native talkers, English proficiency was a significant and strong predictor [$F(1,54) = 67.37$, $\eta_p^2 = 0.55$, $p < 0.001$] for speech comprehension score. There were other significant main effects for BNL [$F(2,108) = 36.26$, $\eta_p^2 = 0.39$, $p < 0.001$] and for RT [$F(4,216) = 3.73$, $\eta_p^2 = 0.05$, $p = 0.006$]. It was hypothesized that speech comprehension performance decreases as BNL or RT increases. Therefore, planned comparisons were deemed appropriate using the lowest condition (RC-30 for BNL and 0.4 s for RT) as the reference level to identify a higher level, at which a significant performance deficit was observed. As seen in Fig. 3(a), the results indicate that participants scored significantly higher in the RC-30 BNL condition than in RC-50 (Cohen's $d = 1.18$, $p < 0.001$) but not in RC-40 ($d = 0.23$, $p = 0.093$). For RT, as seen in Fig. 4(a), participants scored significantly higher in the 0.4 s scenario than in the 0.8 s ($d = 0.38$, $p = 0.007$) and in the

1.2 s ($d = 0.42$, $p = 0.003$) scenarios, but not in the 0.6 s ($d = 0.13$, $p = 0.36$) or 1.0 s ($d = 0.13$, $p = 0.32$) scenario. There was a significant interaction between BNL $\times$ English proficiency level [$F(2,108) = 5.72$, $\eta_p^2 = 0.08$, $p = 0.004$]. The performance deficit in speech comprehension with increasing BNL, specifically from RC-30 to RC-50 ($p < 0.004$), was significantly greater for participants with lower English proficiency levels as shown in Fig. 5.

With speech produced by native-Chinese-speaking talkers, there were significant main effects for English proficiency level [$F(1,57) = 20.49$, $\eta_p^2 = 0.25$, $p < 0.001$], BNL [$F(2,114) = 122.85$, $\eta_p^2 = 0.67$, $p < 0.001$], and RT [$F(4,228) = 6.12$, $\eta_p^2 = 0.09$, $p < 0.001$]. Again, planned comparisons were used to compare higher levels in BNL and RT with the lowest levels as the reference to identify the level at which significant performance deficit occurs. For BNL, as seen in Fig. 3(b), listeners performed significantly better in the RC-30 condition than in the RC-40 ($d = 31$, $p = 0.022$) and RC-50 ($d = 1.8$, $p < 0.001$) conditions. For RT, as shown in Fig. 4(b), listeners scored significantly higher under the 0.4 s scenario than in the 0.8 s ($d = 0.32$, $p = 0.02$), 1.0 s ($d = 0.42$, $p = 0.002$), and 1.2 s ($d = 0.45$, $p = 0.001$) scenarios, but not in the 0.6 s scenario ($d = 0.04$, $p = 0.74$). The results show that listeners' speech comprehension performance begins to decline significantly at the RC-40 BNL condition and the 0.8 s RT scenario, respectively. No other significant

FIG. 3. Speech comprehension performance (in RAU) as a function of BNL with English speech from (a) native talkers and (b) non-native talkers. The darker horizontal lines in the boxplots indicate medians of the speech comprehension scores, averaged across five RT scenarios for each individual listener, with 99% of the mean scores lying within the whiskers. Note: "n.s." for $p > 0.05$, * for $p < 0.05$, *** for $p < 0.001$.

FIG. 4. Speech comprehension performance (in RAU) as a function of RT with English speech from (a) native talkers and (b) non-native talkers. The darker horizontal lines in the boxplots indicate medians of the speech comprehension scores, averaged across three BNL conditions for each individual listener, with 99% of the mean scores lying within the whiskers. Outliers are shown as solid dots outside the whiskers. Note: "n.s." for $p > 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$.

FIG. 5. Scatter plot of speech comprehension versus English proficiency scores for each BNL condition (RC-30, RC-40, and RC-50) for listeners tested with speech produced by native talkers. Linear regression lines were fitted to each BNL condition and were all found to be significant at $p < 0.001$. The standardized coefficient β was estimated as 0.59, 0.65, and 0.79 for RC-30, RC-40, and RC-50, respectively.

interactions were found. Also noteworthy is that no significant interaction was found between BNL and RT from either native or non-native talkers, suggesting no interdependence between BNL and RT on speech comprehension performance.

B. Effect of talker accent

To examine the effect of talker accent, a mixed-design ANOVA was fitted to the full dataset ($N = 114$), with speech comprehension performance as the dependent variable. For independent variables, the new ANOVA model included two between-subject variables of listener group (NAE vs NNC vs NNO) and talker accent (NAE vs NNC) and two within-subject variables of BNL and RT.

The ANOVA model revealed several significant main effects and interactions. The significant main effects included talker accent [$F(1,108) = 48.62$, $\eta_p^2 = 0.30$, $p < 0.001$], listener group [$F(1,108) = 26.12$, $\eta_p^2 = 0.31$, $p < 0.001$], BNL [$F(2,216) = 146.38$, $\eta_p^2 = 0.57$, $p < 0.001$], and RT [$F(4,432) = 8.42$, $\eta_p^2 = 0.06$, $p < 0.001$]. The following two-way interactions were found to be significant: BNL $\times$ talker accent [$F(2,216) = 7.82$, $\eta_p^2 = 0.06$, $p = 0.001$] and BNL $\times$ listener group [$F(4,216) = 2.55$, $\eta_p^2 = 0.03$, $p = 0.04$]. The only significant interaction involving RT was a three-way interaction of RT $\times$ talker accent $\times$ listener group [$F(8,432) = 2.38$, $\eta_p^2 = 0.02$, $p = 0.016$].

For the talker accent main effect, a simple *post hoc* comparison showed that listeners' comprehension was lower for non-native talkers ($M = 73.3$, $SE = 1.1$) than native talkers ($M = 84.3$, $SE = 1.2$) [$d = 0.65$, $p < 0.001$]. The speech comprehension deficit due to foreign accent was 11 RAU, or approximately 11% accuracy. For the significant interaction of talker accent $\times$ BNL, planned comparisons showed that the comprehension deficit for the Chinese-accented talkers was significantly greater under the RC-50 than the RC-30 condition, $p = 0.001$. This interaction is shown in Fig. 3 when comparing the downward slopes of the speech

comprehension performance between native and non-native talkers. BNL, specifically the RC-50 condition, was more detrimental to the comprehension of non-native talkers for all listeners.

For the listener group main effect, pairwise comparisons using Tukey's HSD test revealed that differences between all possible pairs of NAE ($M = 85.3$, $SE = 1.2$), NNC ($M = 78.5$, $SE = 1.6$), and NNO ($M = 72.5$, $SE = 1.3$) listeners were statistically significant ($d = 0.43$, $p < 0.001$ for NAE vs NNC; $d = 0.23$, $p = 0.045$ for NNC vs NNO; $d = 0.72$, $p < 0.001$ for NAE vs NNO). Despite scoring the lowest on English proficiency as a group, NNC listeners performed significantly better on speech comprehension than NNO listeners when averaged across talker accents.

The significant three-way interaction between RT $\times$ talker accent $\times$ listener group was slightly more difficult to interpret. Planned contrast comparisons revealed statistically significant differences between 0.4 versus 0.8 s ($p = 0.013$) RTs and 0.4 versus 1.2 s ($p = 0.019$) RTs. In Fig. 6, the mean difference of speech comprehension performance between native and non-native talkers is plotted as a performance deficit for the three listener groups in the 0.4, 0.8, and 1.2 s RT scenarios. The deficit is calculated by subtracting the performance with Chinese talkers from the performance with native talkers, so that larger values indicate a larger detrimental effect from the non-native talkers. The significant three-way interaction suggests that the variations in performance deficit due to talker accent differed across listener groups. For instance, NAE listeners experienced a significantly greater performance deficit under 0.8 and 1.2 s than in the 0.4 s RT (pane labeled "NAE" in Fig. 6). For NNC and NNO listeners, the non-native accent did not incur a significantly greater performance deficit with increasing RT. NNO listeners experienced the greatest performance deficit among all three listener groups under all scenarios in RT.

C. Differences between listener groups

To investigate differences between listener groups, the full dataset ($N = 114$) was divided into three subsets (i.e., NAE, NNC, and NNO) as described in Table II. An ANCOVA model was fitted separately to each listener group, with BNL and RT as within-subject variables, talker accent as a between-subject variable, and composite English proficiency score as a covariate. The effect sizes of BNL and RT on speech comprehension scores were compared empirically across the three listener groups. The effect size is commonly used to describe the predictor variable's strength on the dependent variable. In this paper, the effect size is expressed in η_p^2 , calculated as the ratio of variance in the performance outcome explained by a single variable (BNL, RT, English proficiency score, or interactions) while controlling for all other independent variables. Only statistically significant main effects and interactions from the factorial ANCOVA are reported for the three listener groups.

As shown in Table III, English proficiency had a statistically significant main effect on speech comprehension performance with comparable effect sizes according to η_p^2

FIG. 6. Three-way interaction between talker accent, listener group (NAE vs NNC vs NNO), and RT (0.4 vs 0.8 vs 1.2 s), shown in terms of performance difference due to the Mandarin Chinese accent. Error bars indicate one standard error.

across all listener groups. Talker accent was a significant and strong predictor of performance for both NAE and NNO listener groups, with worse performance for the non-native talker. NNC listeners, however, were much less affected by the Mandarin Chinese accent; talker accent did not have a significant main effect on this listener group, suggesting a benefit due to having a matched accent with the talker.

The significant main effect of BNL on speech comprehension performance was observed in all listener groups. Based on η_p^2 interpretations recommended by Cohen (1977), the effect size is moderate for NAE listeners and strong for both NNC and NNO listeners. A significant and moderate main effect of RT on speech comprehension performance was only found for NNO listeners. In general, the comparison of effect sizes implies that the speech comprehension performance of native listeners is less impacted by BNL and RT than for the non-native listener groups.

NAE listeners are the only group to show a significant interaction between BNL and talker accent. The speech comprehension performance deficit from RC-30 to RC-50 was significantly greater for NAE listeners when listening to the

non-native talker than for the native talker, as shown in Fig. 7. This interaction is not significant for the NNC and NNO listener groups though, because those groups experienced a similar performance across BNL conditions with both the NAE and NNO talkers. In general, NAE listeners were also less affected by BNL and RT, with smaller effect sizes than the two non-native listener groups.

Comparing the effect sizes between the two non-native listener groups provides an opportunity to examine the benefit of having a matched accent with the talker for speech comprehension in background noise and reverberation. The main effect of BNL was both statistically significant and strong for both NNC and NNO listeners, as indicated by the similar η_p^2 values of 0.37 and 0.44, respectively. However, the main effect of RT was statistically non-significant and weaker for NNC listeners with a lower η_p^2 of 0.06 (within the small effect range), than for NNO listeners who show a statistically significant effect with η_p^2 of 0.12. This suggests that NNC listeners were able to overcome the negative impact of RT when sharing the same native language with the talker, distinguishing them from their non-native peers in the NNO listener group.

TABLE III. Effect sizes of significant main effects and interactions in the ANCOVA model to predict speech comprehension performance, separately for the three listener groups (total $N=114$). Bold values indicate statistically significant results. Note: N_1 = Number of listeners tested with speech from native talkers; N_2 = Number of listeners tested with speech from non-native talkers.

	NAE Listeners		NNC Listeners		NNO Listeners	
	$\left(\begin{smallmatrix} N_1 = 26 \\ N_2 = 20 \end{smallmatrix} \right)$		$\left(\begin{smallmatrix} N_1 = 10 \\ N_2 = 19 \end{smallmatrix} \right)$		$\left(\begin{smallmatrix} N_1 = 19 \\ N_2 = 20 \end{smallmatrix} \right)$	
	<i>p</i> -value	η_p^2	<i>p</i> -value	η_p^2	<i>p</i> -value	η_p^2
English proficiency (Covariate, Between-subject)	<0.001	0.27	0.002	0.33	<0.001	0.46
Talker Accent—NAE vs. NNC (Between-subject)	<0.001	0.36	0.056	0.13	<0.001	0.68
BNL (Within-subject)	0.004	0.12	<0.001	0.37	<0.001	0.44
RT (Within-subject)	0.38	0.02	0.18	0.06	0.001	0.12
BNL $\times$ Talker accent (Two-way interaction)	<0.001	0.20	0.51	0.03	0.068	0.07

FIG. 7. Two-way interaction between talker accent (native vs non-native) and BNL on speech comprehension performance plotted individually for listener groups NAE, NNC, and NNO. Error bars indicate one standard error.

IV. DISCUSSION

The investigations presented in this paper aim to address three research questions on the impacts of realistic adverse room acoustic conditions, talker accent, and listener native language on speech comprehension. The analyses and results provide evidence that first, after accounting for English proficiency, both BNL and RT negatively affect speech comprehension performance, even at levels found in realistic rooms which are at a higher SNR than tested in previous research. Second, as found by other researchers, foreign-accented speech negatively impacted speech comprehension relative to native-accented speech. However, foreign accent had different effects on speech comprehension performance with increasing BNL versus increasing RT. For RT, a significant drop in performance was observed at the 0.8 s RT condition compared to the lowest RT of 0.4 s when perceiving either native American English or native Mandarin Chinese talkers. For BNL, however, a significant performance drop occurred at a BNL level of RC-40 when perceiving non-native talkers, compared to a higher RC-50 for native talkers. Third, the speech comprehension performance of all listener groups was negatively affected by BNL, but the negative effect of RT was only found to be significant for non-native listeners who speak native languages other than Mandarin Chinese (NNO).

Two factors may potentially contribute to the improved ability among NNC listeners to suppress the negative effects of longer RTs, in comparison to the NNO listeners: either better English proficiency or an intelligibility benefit due to sharing the same accent with the non-native talkers. First of all, the two groups of NNC listeners tested with native and non-native talkers did not significantly differ in English proficiency, as indicated in independent *t*-tests for the three English language tests (all *p*'s > 0.05). Also, as shown previously, NNC listeners as a group scored lowest on the

composite scale of English proficiency, eliminating the possibility of this group benefiting from a higher level of English proficiency. The results instead point to an intelligibility benefit of matched accent, so that the performance deficit due to a Chinese-accented talker was smaller for NNC listeners than for NNO listeners, as shown in Fig. 6. Indeed, the performance of NNC listeners under adverse reverberation was similar to native listeners, while that of their other non-native peers (NNO listeners) suffered. The benefit gained by NNC listeners having a matched accent with the non-native talker in the tested acoustic conditions does not appear to be large enough for them to exceed the performance in perceiving native talkers. However, in this investigation, the NNC listeners still benefited in perceiving foreign-accented speech in reverberation by sharing the same accent with the non-native talkers, although such benefit was not available to other non-native listeners who had a mismatched accent.

As seen in the effect size comparison on the full dataset, BNL had a much stronger detrimental effect on speech comprehension performance than RT, particularly for those who were less proficient in English. Furthermore, adverse acoustic conditions affected the speech comprehension performance of native and non-native listener groups differently. Higher BNLs were equivalently detrimental to both non-native listener groups, as indicated by the similar effect sizes for the main effects. Non-native listeners with no matched accent benefit were more adversely affected by increasing reverberation than native listeners. The interaction between BNL and RT was not found to be significant in any of the factorial models tested, suggesting that the impact of these acoustic conditions on speech comprehension performance are independent from each other.

The levels of BNL and RT that produced significant deficits in speech comprehension performance can furthermore

be identified from the results of this study to provide supplementary guidelines for the design of spaces where speech communication is important, depending on the English-speaking skills of the occupants. For BNL, when compared to the lowest condition of RC-30, significant deficits in English speech comprehension were identified at RC-50 for native-accented speech and at RC-40 for non-native-accented speech. This observation held for both native and non-native listeners. For RT, significant performance deficits occurred at 0.8 s for both talkers with and without a Mandarin Chinese accent, when compared against the lowest 0.4 s scenario. However, this observation was only true for non-native listeners and did not hold for native listeners who were able to comprehend speech equivalently well across all tested RT scenarios. Conservatively then, RT and BNL design criteria are recommended at the level below which significant comprehension deficit was first observed, as summarized in Table IV.

The different impacts of BNL and RT on comprehending native versus foreign-accented speech by native and non-native listeners may lie in the top-down and bottom-up processes in speech processing (Zekveld *et al.*, 2006, 2009). In top-down processing, listeners actively search for their internal representation of the target words or sentences to match the incoming speech stimuli. In bottom-up processing, listeners parse out meaning from the amount of information contained in the degraded speech.

Reverberation, noise, and foreign accent are all sources of speech degradation in this investigation. When the incoming native speech stimuli are masked in reverberation, the temporal fine structure provides “glimpses” into usable information from the modulation for listeners to achieve release from masking which lead to improved speech comprehension (Gnansia *et al.*, 2008). Naturally, native listeners outperform the non-native listeners in both top-down and bottom-up processing due to better language proficiency. It is speculated that the native listeners benefited from recovering information effectively from these glimpses in the reverberant conditions. As a result, the negative effect of RT was not apparent for them up to the highest RT tested of 1.2 s, unlike non-native listeners who performed worse at conditions above RT of 0.6 s. However, louder background noise conditions do not provide such glimpses, so no additional information can be retrieved by the native listeners to capitalize on such processing. Both native and non-native listeners performed significantly worse above the same BNL.

In understanding speech with foreign accent, lowering BNL improves comprehension by reducing the overall intensity of spectral masking due to the broadband noise used in this study. However, it seems that increasing modulation glimpses by lowering RT did not provide the same amount

of additional resources as it did for native speech, even for native listeners to gain enough release from masking to counter the negative effect of foreign accent. As a result, the recommended design guideline only varies for BNL but not RT between including native and non-native talkers in the classroom.

V. CONCLUSION

In this investigation, the effects of BNL, RT, and talker accent on speech comprehension by native and non-native listeners have been examined. Using laboratory-controlled experiments, a total of 15 acoustic conditions comprised of three BNL conditions (RC-30, RC-40, and RC-50) and five RT scenarios (from 0.4 to 1.2 s) were created to simulate realistic classroom acoustic environments. To measure listeners’ performance when exposed to these acoustic conditions, a dual-task paradigm was adopted for testing speech comprehension and an APR (dot-tracing) task simultaneously. Speech comprehension was tested using the same experimental design, but different talkers (native American English vs native Mandarin Chinese) for recording the same speech comprehension materials in English, to study the effect of talker foreign accent.

Based on the results, recommended design criteria for BNL and RT have been proposed for spaces in which speech communication is important, as summarized in Table IV, beyond which adult listeners began to experience significant performance deficits on the speech comprehension tasks. A matched accent benefit was identified for non-native listeners who share the same native language as a non-native talker. Sharing the same native language helped non-native listeners overcome the negative effects of reverberation on speech comprehension performance, to be on par with that of native English listeners.

The recommended values for BNL and RT provided in Table IV are not stricter than those listed in the existing ANSI S12.60-2010 classroom acoustics standard (i.e., 35 dBA BNL and 0.6 s RT), but this research was conducted on normal-hearing adults, for whom the speech materials were expected to be relatively easy to comprehend. More investigation is needed to determine what values are suitable for non-native English-speaking children; children have generally been shown to require better acoustic conditions than adults to achieve similar speech perception performance (Yang and Bradley, 2009). Further subjective testing could use additional BNLs with finer resolution, as well as lower values of BNL (below RC-30) and RT (below 0.4 s), to further improve the recommended guidelines. Future work is also recommended to verify the results from this research using simulations from other locations in the classroom (e.g., side and

TABLE IV. Recommended guidelines for BNL and RT in spaces where speech communication is important, based on the English-speaking skills of adult occupants.

	Native English <i>Talkers</i> Only	Both Native and Non-native English <i>Talkers</i>
Native English <i>Listeners</i> Only	BNL $\leq$ RC-40 (48 dBA), RT $\leq$ 1.2 s	BNL $\leq$ RC-30 (38 dBA), RT $\leq$ 1.2 s
Both Native and Non-native English <i>Listeners</i>	BNL $\leq$ RC-40 (48 dBA), RT $\leq$ 0.6 s	BNL $\leq$ RC-30 (38 dBA), RT $\leq$ 0.6 s

back of the room), which exhibit lower interaural cross-correlation due to the proximity of reflecting surfaces.

In conclusion, designers should be aware of the linguistic diversity among occupants in when designing for classroom acoustics. Depending on whether non-native English speakers exist among listeners and talkers, more stringent acoustic requirements may be necessary to attain optimal speech comprehension performance.

ACKNOWLEDGMENTS

This research was funded by the Paul S. Veneklasen Research Foundation and by a research seed grant from the Durham School of Architectural Engineering at the University of Nebraska—Lincoln. The authors are grateful to Dr. Thomas Carrell and Dr. Nirmal Srinivasan for the implementation of the dual-task paradigm and to Won Seok Jang for the computer test program. We would also like to thank Dr. Kanae Nishi and Dr. Siu-Kit Lau for providing constructive feedback on the project; Dr. Kenneth P. Roy, Sean Browne, and anonymous volunteers from Armstrong World Industries for recording the speech comprehension materials and providing equipment; and all the volunteers who participated in the study. Brenna Boyd, Laura Brill, Jennifer Francis, Kristin Hanna, and Adam Steinbach are also acknowledged for their assistance with the experiments and data analyses.

¹For typical core learning spaces of less than 283 m³, the acoustic conditions shall not exceed 35 dBA BNL and 0.6 s RT in the unoccupied mode (2010 revision).

²Unless otherwise noted, the terms “native” and “non-native” refer to native and non-native English-speaking hereafter.

³SII and STI are both frequency-weighted physical measures of acoustic signals that strongly correlate with intelligibility of speech. ANSI S3.5 (1997) defines that SII accounts for speech and noise levels and talker vocal effort; while STI accounts for speech and noise levels and temporal modulation due to room reverberation (Steeneken and Houtgast, 1980). A value greater than 0.5 corresponds to over 80% sentence intelligibility in SII (Hornsby, 2004) and 90% in STI (Houtgast and Steeneken, 1984).

ANSI (1997). ANSI/ASA S3.5. *Methods for calculation of the speech intelligibility index* (American National Standards Institute, New York).

ANSI (2008). ANSI/ASA S12.2-2008, *Criteria for Evaluating Room Noise* (American National Standards Institute, New York).

ANSI (2010). ANSI/ASA S12.60, *Performance Criteria, Design Requirements, and Guidelines for Schools, Part 1: Permanent Schools* (American National Standards Institute, New York).

Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., Frohlich, L., Kemp, J., and Drake, L. (2010). “*The condition of education 2010 (NCES 2010-028)*” (U.S. Department of Education, Institute of Education Sciences, Washington, DC), pp. 32–33.

Beaman, C. P., and Holt, N. J. (2007). “Reverberant auditory environments: The effects of multiple echoes on distraction by ‘irrelevant’ speech,” *Appl. Cognit. Psychol.* **21**, 1077–1090.

Bent, T., and Bradlow, A. R. (2003). “The interlanguage speech intelligibility benefit,” *J. Acoust. Soc. Am.* **114**, 1600–1610.

Bent, T., Kewley-Port, D., and Ferguson, S. H. (2010). “Across-talker effects on non-native listeners’ vowel perception in noise,” *J. Acoust. Soc. Am.* **128**, 3142–3151.

Bradley, J. S., Reich, R. D., and Norcross, S. G. (1999). “On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility,” *J. Acoust. Soc. Am.* **106**, 1820–1828.

Bradley, J. S., and Sato, H. (2008). “The intelligibility of speech in elementary school classrooms,” *J. Acoust. Soc. Am.* **123**, 2078–2086.

Bradley, J. S., Sato, H., and Picard, M. (2003). “On the importance of early reflections for speech in rooms,” *J. Acoust. Soc. Am.* **113**, 3233–3244.

Bradlow, A. R., and Alexander, J. A. (2007). “Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners,” *J. Acoust. Soc. Am.* **121**, 2339–2349.

Bradlow, A. R., and Bent, T. (2002). “The clear speech effect for non-native listeners,” *J. Acoust. Soc. Am.* **112**, 272–284.

Chujo, K., and Oghigian, K. (2009). “How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary,” *J. Asia TEFL* **6**, 121–148.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*, revised ed. (Lawrence Erlbaum Associates, Inc., Mahwah, NJ).

Daneman, M., and Carpenter, P. A. (1980). “Individual differences in working memory and reading,” *J. Verbal Learn. Verbal Behav.* **19**, 450–466.

Flège, J. E., Yeni-Komshian, G. H., and Liu, S. (1999). “Age constraints on second-language acquisition,” *J. Memory Lang.* **41**, 78–104.

Gnansia, D., Jourdes, V., and Lorenzi, C. (2008). “Effect of masker modulation depth on speech masking release,” *Hear. Res.* **239**, 60–68.

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**, 931–939.

Hornsby, B. W. Y. (2004). “The Speech Intelligibility Index: What is it and what’s it good for?” *Hear. J.* **57**, 10–17.

Houtgast, T., and Steeneken, H. (1984). “A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria,” *Acta Acust. Acust.* **54**, 185–199.

Imai, S., Flège, J. E., and Walley, A. (2003). “Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners,” in *Proceedings of the International Congress on Phonetic Science*, Barcelona, Spain.

Institute for International Education (2012). “Open doors 2012 fast facts,” p. 2.

Klatte, M., Lachmann, T., and Meis, M. (2010). “Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting,” *Noise Health* **12**, 270–282.

Knecht, H. A., Nelson, P. B., Whitelaw, G. M., and Feth, L. L. (2002). “Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements,” *Am. J. Audiol.* **11**, 65–71.

Ljung, R., and Kjellberg, A. (2009). “Recall of spoken words presented with a prolonged reverberation time,” *Build. Acoust.* **16**, 301–312.

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). “The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals,” *J. Speech, Lang. Hear. Res.* **50**, 940–967.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., and Ruef, M. L. (1998). *Bilingual Verbal Ability Tests: Comprehensive Manual* (Riverside Publishing, Itasca, IL).

Munro, M. J. (1998). “The effects of noise on the intelligibility of foreign-accented speech,” *Studies Sec. Lang. Acquis.* **20**, 139–154.

Munro, M. J., and Derwing, T. M. (1995). “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners,” *Lang. Learn.* **45**, 73–97.

National Science Board (2012). *Science and Engineering Indicators 2012* (National Science Foundation, Arlington, VA), pp. 3.1–3.65.

Nelson, P., Kohnert, K., Sabur, S., and Shaw, D. (2005). “Classroom noise and children learning through a second language: Double Jeopardy?” *Lang., Speech, Hear. Serv. Schools* **36**, 219–229.

Nelson, P. B., and Soli, S. (2000). “Acoustical barriers to learning: Children at risk in every classroom,” *Lang., Speech, Hear. Serv. Schools* **31**, 356–361.

Pearson (2008). *Versant English Test* (Pearson Education, Inc., New York).

Peng, Z. E. (2014). “Designing acoustics for linguistically diverse classrooms: Effects of background noise, reverberation and talker foreign accent on speech comprehension by native and non-native English-speaking listeners,” Ph.D. dissertation, University of Nebraska, Lincoln, NE.

Rogers, C. L., Dalby, J., and Nishi, K. (2004). “Effects of noise and proficiency on intelligibility of Chinese-accented English,” *Lang. Speech* **47**, 139–154.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). “Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing,” *Appl. Psycholing.* **27**, 465–485.

Ronsse, L. M., and Wang, L. M. (2010). “Effects of noise from building mechanical systems on elementary school student achievement,” *ASHRAE Trans.* **116**, 347–354.

Ronsse, L. M., and Wang, L. M. (2013). “Relationships between unoccupied classroom acoustical conditions and elementary student achievement measured in eastern Nebraska,” *J. Acoust. Soc. Am.* **133**, 1480–1495.

- Sherbecoe, R. L., and Studebaker, G. A. (2004). "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," *Int. J. Audiol.* **43**, 442–448.
- Srinivasan, N. K. (2010). "The perception of natural, cell phone, and computer-synthesized speech during the performance of simultaneous visual-motor tasks," Ph.D. dissertation, University of Nebraska, Lincoln, NE.
- Steeneken, H. J. M., and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **67**, 318–326.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech, Lang. Hear. Res.* **28**, 455–462.
- Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., and Lewis, D. E. (2012). "Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children," *J. Acoust. Soc. Am.* **131**, 232–246.
- Wang, H., and van Heuven, V. J. (2015). "The interlanguage speech intelligibility benefit as bias toward native-language phonology," *i-Perception* **6**, 2041669515613661.
- Woodcock, R. W., McGrew, K., and Mather, N. (2001a). *Woodcock-Johnson Tests of Achievement* (Riverside Publishing, Itasca, IL).
- Woodcock, R. W., McGrew, K. S., and Mather, N. (2001b). *Woodcock-Johnson III Tests of Cognitive Abilities* (Riverside Publishing, Itasca, IL).
- Yang, W., and Bradley, J. S. (2009). "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *J. Acoust. Soc. Am.* **125**, 922–933.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., and Schoonhoven, R. (2006). "Top-down and bottom-up processes in speech comprehension," *NeuroImage* **32**, 1826–1836.
- Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S., and Houtgast, T. (2009). "User evaluation of a communication system that automatically generates captions to improve telephone communication," *Trends Amplif.* **13**, 44–68.