2013

Effects of reverberation and noise on speech comprehension by native and non-native English-speaking listeners

Z. Ellen Peng
University of Nebraska-Lincoln, zpeng@huskers.unl.edu

Lily M. Wang
University of Nebraska-Lincoln, lwang4@unl.edu

Siu-Kit Lau
University of Nebraska-Lincoln

Adam M. Steinbach
University of Nebraska-Lincoln

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Noise
Session 4pNSa: Effects of Noise on Human Performance and Comfort II

4pNSa4. Effects of reverberation and noise on speech comprehension by native and non-native English-speaking listeners

Zhao Peng*, Lily M. Wang, Siu-Kit Lau and Adam M. Steinbach

***Corresponding author's address: Durham School of Architectural Engineering and Construction, University of Nebraska-Lincoln, 1110 South 67th Street, Omaha, Nebraska 68182, zpeng@huskers.unl.edu**

Previous studies have demonstrated the negative impact of adverse signal-to-noise-ratios on non-native English-speaking listeners' performance on speech recognition using recall tasks, as well as implied that comprehension skills were more impaired than recognition skills under reverberation and noise. The authors have themselves previously conducted a pilot study on three native and three non-native English-speaking listeners to examine the effects of reverberation and noise using speech comprehension tasks. Those results suggested that speech comprehension performance is worse under longer reverberation times (RT), and that a longer RT is more detrimental to speech comprehension by non-native listeners than native listeners. This paper reports on the refined full study, in which a larger number (up to 30) of each group was tested. Each participant was exposed to 15 acoustic conditions, created from combinations of five RTs (0.4 to 1.2 seconds) and three background noise levels (RC-30, 40 and 50). Speech comprehension performance under each condition was recorded. Confounders related to general speech comprehension abilities were screened for, including listening span, oral comprehension abilities and English verbal skills. Results are presented and compared between native and non-native listeners. [Work supported by a UNL Durham School Seed Grant and the Paul S. Veneklasen Research Foundation.]

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Clear communication is the key to successful classroom learning outcomes, and the room acoustic environment plays a large role in that process. The Institute of Education Science reported that 21% of students in the U.S. age 5-17 (or 10.9 million) speak a language other than English at home (Aud et al., 2010). A number of speech intelligibility studies have specifically compared native and non-native English-speaking listeners' performances on recall tasks by varying signal-to-noise ratios (SNRs), mostly below 0 dB. The stimuli used in the recall tasks varied between different levels of the phonological units, including vowels and consonants (Cutler et al., 2004), single words (Bent et al., 2010; Rogers et al., 2006), and sentences (Bradlow and Bent, 2002). Results from these studies suggest that non-native listeners do perform worse than natives under these extremely adverse listening conditions. However, these studies did not provide further indications on how these non-native listeners perform under more realistic room acoustic environments.

Another set of studies has found that adverse acoustic environments might be more detrimental to speech comprehension tasks, in which listeners are to understand speech information based on context, than to speech intelligibility tasks utilizing simple recall. Klatte et al. (2010) investigated language comprehension in a classroom-like setting under four combinations, comprised of two noise types (activity noise vs. babble noise) crossed with two reverberation times (RTs, 0.5 vs. 1.1 seconds). They conclude that listening comprehension using a paper-pencil instructional task was more impaired than speech recognition using a word-to-picture matching task. This is further supported by Valente et al. (2012), who also tested four combinations comprised of two SNRs (+7 vs. +10 dB) crossed with two RTs (0.6 vs. 1.5 seconds). Although no direct comparison was made for speech comprehension versus recognition, results showed that the detrimental effect of reverberation and noise was more pronounced in speech comprehension tasks than speech recognition tasks. Both the Klatte et al. and Valente et al. studies used digital simulation techniques to create the reverberation conditions, and all participants in these studies were only assigned to one acoustic condition for testing both speech comprehension and speech recognition tasks.

The present study aims at testing both native and non-native listeners across a more comprehensive range of acoustic conditions. A total of 15 realistic acoustic conditions, comprised of RT (five settings with mid-frequency averages between 0.4 and 1.2 seconds) crossed with background noise level (BNL, three settings: Room Criteria (RC) 30, 40 and 50), were developed for testing speech comprehension in the Nebraska Acoustics Chamber. Instead of varying the acoustic conditions between participants, a within-subject design in which each participant is tested for all 15 acoustic conditions was utilized to enhance robustness of the statistical analysis.

METHODOLOGY

Acoustic Conditions

Each participant was individually tested in an immersive acoustics environment in the Nebraska Acoustics Chamber. The test chamber has a volume of 25 m^3 and rather absorptive materials, including acoustical ceiling tiles, thin carpet on a floating floor system, gypsum walls with acoustically absorptive panels, and corner bass traps. One side wall and the back wall are slightly slanted to reduce flutter echoes and standing waves. The ambient BNL of the test chamber was measured to be RC-26 hissy at the listener position with air-conditioning, lighting and electrical systems in regular operation. To create the three BNL conditions, white noise was digitally filtered by the audio program Cool Edit 2000, and then introduced through a concealed ceiling speaker and a corner subwoofer to comply with the RC-30, 40 and 50 contours, as measured at the listener position.

Five scenarios of RT between 0.4 and 1.2 seconds were produced using the computer auralization program ODEON. A classroom of 260 m^3 was simulated in ODEON with permanent gypsum board finishes on the front wall, thin carpet on the floor, and fixed desks and chairs. The five RT scenarios were achieved by choosing combinations of different types of ceiling tiles and scaled absorption coefficients on the back and the side walls. The ceiling tiles varied among gypsum board, NRC 0.55, or NRC 0.70 acoustical ceiling tiles, while a single scaling factor was applied to vary the baseline 1-inch absorptive wall panel absorption coefficients across octave band frequencies. The five simulated RTs, reported in the form of T_{20}, averaged across 500, 1000 and 2000 Hz octave bands are indicated in Table 1.

TABLE 1. Simulated reverberation time scenarios with T_{20} averaged across 500, 1000 and 2000 Hz octave band frequencies.

RT Scenario	Simulated T_{20} [seconds]
1	0.39
2	0.60
3	0.81
4	1.01
5	1.18

Sets of binaural room impulse responses (BRIR) for the five RT scenarios were exported after adjusting for the loudspeaker-listener configuration in the test chamber. Each BRIR, corresponding to one RT scenario, was then digitally convolved with the anechoically recorded speech materials for playback via a two-channel loudspeaker system (distinct from the BNL playback system), flanking the computer monitor over which the speech comprehension tasks are presented.

Testing Materials

Speech Comprehension Tasks

Fifteen sets of speech comprehension tests with equivalent content difficulty were developed following the format of the Test of English for International Communications (TOEIC). Each set of tests takes approximately 15 minutes to complete and contains four types of speech comprehension tasks: (1) photo recognition (in which participants must identify the sentence that best describes the photo displayed among four sentences spoken); (2) question and response (in which participants must select the most appropriate response which follows a spoken question); (3) conversations between two talkers (in which participants must answer a series of aural questions about the conversation after it is finished); and (4) talks by a single talker (in which participants must answer a series of aural questions about the talk after it is finished). Each test contains a total of 32 items from these four tasks. Speech comprehension scores were calculated as percent correct based on the number of items answered accurately.

To avoid the learning effect, each test was only presented once to each participant while testing across the assorted acoustic conditions. The order of presentation for the tests was counter-balanced across all participants. All speech materials were recorded by native English speakers in an anechoic chamber and were convolved digitally with the RT scenarios described earlier. A computer program was created to prompt audio playback and record participant responses on a 24-inch monitor screen.

Adaptive Pursuit Rotor Task

During pilot testing, the authors found that speech comprehension scores reported in percent correct did not appropriately account for the additional cognitive load that might take place under adverse acoustic conditions. That is, a listener may actually still obtain high speech comprehension scores in an adverse acoustic condition, but at the cost of increased cognitive load or stress. Consequently, a competing adaptive pursuit rotor (APR) task (Srinivasan, 2010) was introduced to measure the cognitive load while simultaneously testing for speech comprehension. The APR task adaptively changes the speed of a dot moving along a fixed circle path, so that the participants traced the dot with 70% accuracy using a stylus and pad while performing the speech comprehension tasks under the assorted acoustic conditions.

NASA Task Load Questionnaire

After the speech comprehension tasks for each acoustic condition were completed, a NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988) was administered to solicit participants' subjective perceptions on the task load for the simultaneous tasks completed. These subjective ratings are correlated with the objective performance under the acoustic conditions tested.

Procedure

Screening

Participants were recruited on the University of Nebraska at Omaha campus. Those eligible to participate in the full study must have normal hearing with 25 dB HL or lower on both ears from 125 to 8000 Hz as indicated in the audiometric hearing screen. A demographic survey and a slightly modified Language Experience and Proficiency Questionnaire (Marian et al., 2007) were then administered to eligible participants. Sets of English proficiency test were also administered on both native and non-native listeners to screen for confounders related to general speech comprehension performances, including listening span, oral comprehension abilities and English verbal skills. The listening span and oral comprehension abilities were tested using subtests from the Woodcock-Johnson III Tests of Achievement and Tests of Cognitive Abilities. The English portion of the Bilingual Verbal Ability Tests was administered for testing English verbal skills.

Speech Comprehension Tests

There were 15 acoustic conditions tested, from combinations of three BNLs and five RTs. An additional three tests were developed for training purposes and thus not graded. These training tests were presented under optimum RT (lower than the lowest RT scenario tested) for each BNL condition; and they always appeared first during testing of a specific BNL condition. All 18 tests were separated into six one-hour sessions scheduled on separate days, with three tests completed in each session. During each hour-long session, the BNL did not change; however, the RT of the auralized signals changed between the three tests within the session. The specific order of room acoustic conditions was randomly assigned to each participant.

RESULTS

A pilot study was previously conducted on three native (all male) and three non-native listeners (2 males). The non-native pilot participants were chosen based on their comparable performances on listening span, oral comprehension and English verbal abilities tests. All pilot participants age between 24 and 26 years old and had normal hearing. All non-native pilot participants speak Mandarin Chinese as their native language and had an average length of English immersion of 2.6 years (range: 24-46 months) at the time of testing. The competing APR task and the NASA-TLX were not yet incorporated in the pilot study. All pilot participants were only tested for speech comprehension under 15 acoustic conditions. During pilot study testing, temporary absorptive materials were placed at the corners to reduce the room reverberation to a mid-frequency RT of 0.35 seconds (prior to installation of permanent materials). The resulting perceived RTs measured at the listener position were measured at 0.41, 0.57, 0.80, 1.08 and 1.20 seconds for the five RT scenarios, respectively. Even with a small sample of only six participants, significant main effects were found for nativeness ($F(1, 4) = 20.24$, $p = .011$, effect size in partial $\eta^2 = .84$) and RT ($F(4, 16) = 3.88$, $p = .022$, partial $\eta^2 = .49$). Non-native pilot participants' speech comprehension scores were significantly lower than those of native pilot participants. Specifically, results showed that participants scored significantly lower in the longest RT scenario (1.2 seconds) than in two RT scenarios of 0.57 and 1.08 seconds ($F(1, 4) = 34.00$, $p = .004$ and $F(1,4) = 9.72$, $p = .036$, respectively). No significant main effect was found for BNL ($p = .28$, partial $\eta^2 = .28$).

In the full study, up to 30 participants in each listener group (native and non-native) were tested for speech comprehension simultaneously with the APR task under the assorted acoustic conditions. The results of the objective performance and subjective ratings will be presented and compared between native and non-native listeners during the presentation at ICA Montreal.

DISCUSSIONS

Analysis of the pilot study results indicated that the sets of speech comprehension tests might not be as equivalent as initially intended. Additional material validation screening was thus conducted with five native English-speaking listeners. None of these validation testees participated in the pilot study or the full study. There were more items recorded in the anechoic speech comprehension material database than the actual items assigned to create the 18 sets for the pilot study. During the validation screening, all items were played back under the ambient environment in the Nebraska Acoustics Chamber, with a BNL of RC-26 hissy and a mid-frequency RT of 0.35 seconds (using temporary absorptive materials, prior to final installation of permanent materials).

Speech comprehension test items receiving 80% or lower scores across all validation testees and pilot participants were identified and reviewed for their content. For Task 1 (photo recognition) and Task 2 (question and response), most items were answered with high accuracy regardless of the acoustic conditions. For Task 3 (conversation) and Task 4 (talk), items requiring complex memorization and calculation were removed because these do not test participants' ability to interpret meaning based on context. In addition, items which two or more validation testees answered incorrectly were not utilized in the full study. All validated items in the four tasks utilized for testing acoustic conditions were at least 73% correct by validation testees who were tested under the test chamber ambient acoustic environment. Although the removal of low-scoring items might increase the possibility of subjects reaching the ceiling of 100% correct on the speech comprehension task, the competing APR task complements the results by accounting for the partial task load during the test. More discussions on the simultaneous tasks will be presented at ICA Montreal.

ACKNOWLEDGEMENTS

This project has been supported by a UNL Durham School of Architectural Engineering and Construction Faculty Seed Grant and by the Paul S. Veneklasen Research Foundation. The authors would like to thank these sponsors for their support.

The authors are also grateful to our collaborators at Armstrong World Industries (particularly Dr. Kenneth P. Roy and Sean Browne) for their assistance in recording the anechoic speech comprehension material and lending monitor speakers for the testing.

REFERENCES

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)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

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. **5**, 3142-3151.

Bent, T. and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **3**, 1600-1610.

Bradlow, A. R. and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **1**, 272-284.

Cutler, A., Weber, A., Smits, R. and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **6**, 3668-3678.

Hart, S. G. and Staveland, L. E. (**1988**). "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," In *Human Mental Workload* (Eds. P. A. Hancock and N. Meshkati), Amsterdam: North Holland Press, 139-183.

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 and Health **49**, 270-282.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (**2007**). "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," Journal of Speech, Language and Hearing Research, **50**(4), 940-967.

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," Applied Psycholinguistics **3**, 465-485.

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.

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., 232-246.