

Speech Recognition in Noise in Adults and Children Who Speak English or Chinese as Their First Language

DOI: 10.3766/jaaa.17066

Erin C. Schafer*
 Katsura Aoyama*
 Tiffany Ho*
 Priscilla Castillo*
 Jennifer Conlin*
 Jessalyn Jones*
 Skyler Thompson*

Abstract

Background: Speech recognition of individuals who are listening to a nonnative language is significantly degraded in the presence of background noise and may be influenced by proficiency, age of acquisition, language experience, and daily use of the nonnative language.

Purpose: The purpose of this study is to examine and compare speech recognition in noise performance across test conditions with varying signal-to-noise ratios (SNRs) as well as the presence of vocal and spatial cues in listeners who speak American English as a native language or Mandarin Chinese as a native language. Self-rated English proficiency and experience were collected for native Mandarin Chinese speakers to determine its relationship to performance on the test measures.

Research Design: A cross-sectional repeated measures design was used for the study.

Study Sample: Four groups of participants were included in the study. The adult groups consisted of 25 adults who speak native English and 25 adults who speak native Mandarin Chinese with English as an additional language. The pediatric groups consisted of 16 children who speak native English and 16 children who speak native Mandarin Chinese with English as an additional language.

Data Collection and Analyses: Percent correct speech recognition in noise was assessed at three SNRs (−3, 0, +3 dB) using the adult or pediatric versions of the AzBio sentence test. The Listening in Spatialized Noise-Sentence (LiSN-S) test was used to determine the effect of providing spatial and vocal cues on the speech recognition in noise performance of the groups of participants. The data for each age group and test measure were analyzed with a repeated measures analysis of variance. Correlation analyses were performed to examine relationships between English proficiency and experience on performance across the speech recognition test conditions.

Results: Analysis of the data from the adult or pediatric AzBio sentence test identified a significant effect of native language for adults but no significant effect for children. The higher SNRs yielded better performance for all listeners. On the LiSN-S test, results for the adult and pediatric groups were similar and showed significantly better performance for the native English speakers in every test condition. The demographic and language characteristics that most affected speech recognition performance across the test measures included the length of time the person lived in the United States, the age of English acquisition, the number of minutes per day English was spoken by the participant, and the self-rated English proficiency.

Conclusions: The findings in this study highlight the importance and benefit of higher SNRs as well as the provision of vocal and spatial cues for improving speech recognition performance in noise of adult and pediatric listeners who speak Mandarin Chinese as a native language.

*Department of Audiology & Speech-Language Pathology, University of North Texas, Denton, TX

Corresponding author: Erin C. Schafer, Department of Audiology & Speech-Language Pathology, University of North Texas, Denton, TX 76203-5017; Email: Erin.Schafer@unt.edu

This project was supported by an American Speech-Language-Hearing Association Multicultural grant.

Key Words: speech recognition, noise, Mandarin, English

Abbreviations: LiSN-S = listening in spatialized noise-sentence; NC = native Chinese; NE = native English; RM ANOVA = repeated measures analysis of variance; SNR = signal-to-noise ratio; SRT = speech recognition threshold

INTRODUCTION

Hearing in noisy environments is often difficult, even for individuals with normal-hearing sensitivity in the presence of high-intensity noise, at a distance from the talker and in reverberant situations (Finitzo-Hieber and Tillman, 1978; Festen and Plomp, 1990; Neuman et al, 2010; Wróblewski et al, 2012). Speech recognition in noise is an even more challenging task when listening to a nonnative language (Crandall and Smaldino, 1996; Nelson et al, 2005; Lecumberri et al, 2010; Nakamura and Gordon-Salant, 2011; Tamati and Pisoni, 2014), likely because of the increased processing demands (Zhang et al, 2016). Given the cultural diversity in the United States, an in-depth understanding of the speech recognition difficulties in noise of nonnative listeners as well as the demographic factors that may influence performance is necessary to consider the listening needs and appropriate speech recognition test materials for various populations of listeners (U.S. Census Bureau, 2011; American Speech-Language-Hearing Association, 2017).

Speech Recognition in a Native Language

When listening in a native language, speech recognition performance of children and adults with normal hearing will often result in 100% accuracy when the stimuli are presented at fixed suprathreshold levels in quiet and at favorable signal-to-noise ratios (SNRs), such as +15 to +2 dB (Finitzo-Hieber and Tillman, 1978; Boothroyd, 2008; McCreery et al, 2010; Spahr et al, 2012; Schafer et al, 2016). However, the addition of higher levels of background noise and reverberation, which is typical in most everyday listening environments, results in substantial declines in speech recognition performance. For example, in a hallmark study by Finitzo-Hieber and Tillman (1978), the authors report that 12 children with normal hearing have excellent word recognition in quiet (i.e., near 100% correct). However, at a 0 dB SNR and reverberation time of 1.2 sec, performance decreased substantially to an average of approximately 30% correct. McCreery et al (2010) reported performance-intensity functions of children ranging in age from 7 to 12 yr as well as data from the same test measure for adults (Boothroyd, 2008). For the word recognition task, most of the children and the adults were able to recognize and repeat the stimuli with 100% accuracy at a +15 SNR. Conversely, when as the SNR decreases below 0 dB SNR, the speech recognition of both age groups decreases significantly with average scores ranging from approximately 20% to 30% correct at a -10 SNR.

In addition to the effects of noise and reverberation, there is a well-known influence of age on speech recognition performance (McCreery et al, 2010; Schafer, Beeler, et al, 2012). In the Schafer et al study, the speech-in-noise thresholds of 68 children, grouped into three-, four-, five-, and six-yr-olds and 17 adults were assessed using simple phrases (e.g., brush his teeth) in the presence of classroom noise. The three-yr-old children had significantly higher (i.e., poorer) thresholds than all other age groups, and the four- and five-yr-old groups also had significantly higher (i.e., poorer) thresholds than the adult group. McCreery et al found a similar effect of age whereby the groups of children who were five to eight yr had significantly poorer performance when compared with the adult data in Boothroyd (2008). According to Nilsson et al (1996), effects of age on speech recognition performance in noise are minimal (i.e., within 1 dB on an adaptive task) for children who are 12 yr and older when compared with adults. Effects of age are also evident at the other end of the spectrum between adults who are younger versus older in age (Dubno et al, 2003; Vermeire et al, 2016). As an example, Vermeire et al (2016) reports that a group of 27 young adults, between the ages of 19 to 25 yr, have significantly better speech recognition (by ~3 dB) in noise than a group of older adults, between the ages of 60 to 82 yr.

The aforementioned studies provide clear evidence that noise, reverberation, and older age may negatively impact speech recognition in adults. In children, noise, reverberation, and a younger age (<8 yr) may influence speech recognition.

Speech Recognition in a Nonnative Language

Similar to performance in a native language, English speech perception in quiet in listeners of a nonnative language is often good to excellent (i.e., 80–100%) at fixed suprathreshold presentation levels (Crandall and Smaldino, 1996; Nelson et al, 2005; Garcia Lecumberri et al, 2010; Lecumberri et al, 2010; Nakamura and Gordon-Salant, 2011; Tamati and Pisoni, 2014). However, as shown in the studies cited previously, the addition of noise results in significantly greater declines in English speech recognition for nonnative speakers of English as compared with NE speakers. For instance, in Nelson et al (2005), investigators assessed word identification in quiet and noise in seven children who spoke English and in 12 children, ages seven to eight yr, who spoke Spanish as a first language and English as a second language (L2). Both groups had better performance

in quiet relative to noise, but the group of Spanish-speaking children had significantly poorer performance in noise by an average of 11% relative to the English-only group. In another study, English sentence recognition in quiet and in noise was evaluated in ten adults, ages 31 to 39 yr, who spoke English as a first language as well as ten adults who spoke Japanese as a first language (Nakamura and Gordon-Salant, 2011). Results showed significantly poorer performance for the native Japanese-speaking group as compared with NE-speaking group participants in quiet as well as noise at every SNR tested (−6, −4, −2, 0, and 2 dB).

For speech recognition in one's nonnative language, numerous other factors that influence their performance, such as context cues, age of English acquisition, amount of exposure to English, type of stimulus, type of noise, and presence of reverberation (Kalikow et al, 1977; Lecumberri et al, 2010; Nakamura and Gordon-Salant, 2011; Rimikis et al, 2013; Zhang et al, 2016). For example, Rimikis et al (2013) evaluated English sentence recognition in noise of 102 nonnative speakers, ages 18 to 50 yr, with various native languages, the age of immigration, number of years in the United States, and the amount of time English was used daily were significant predictors of performance. In another study, Zhang et al (2016) evaluated speech recognition in Mandarin Chinese in quiet and noise conditions in 60 undergraduate students who were native speakers of Japanese. They found that Mandarin Chinese (the Japanese speakers' L2) proficiency and semantic context significantly impacted performance, particularly in noise and when the fundamental frequency information is degraded. See a thorough review by Lecumberri et al (2010) for additional information regarding other factors that may influence nonnative speech recognition in noise. Studies on speech recognition in noise in school-age children who speak English as an L2 are limited.

Study Rationale

Given the language diversity of the U.S. population, whereby 60 million people speak a language other than English at home (U.S. Census Bureau, 2011), audiologists and other health professionals will need to be prepared to appropriately assess and serve patients with varied ethnicities and languages. An in-depth understanding of the effects of nonnative speech recognition as well as factors that may influence the speech recognition performance of a nonnative listener will enable professionals to make appropriate recommendations for patients who have normal hearing or hearing loss.

The purpose of the present study is to examine and compare speech recognition of adults and children who speak English (NE) or Mandarin Chinese (NC) as a native language and to determine how speech recognition is affected by increasing noise levels as well as

the presence of spatial and vocal cues. NC listeners, in particular, were selected for this study because Chinese is widely used in the United States and is spoken by approximately 2.5 million individuals (U.S. Census Bureau, 2011). A secondary goal of the study was to determine how self-rated English proficiency and experience influenced speech recognition in noise performance across the test measures and conditions.

This study will contribute to existing literature by (a) including both adult and pediatric listeners, (b) utilizing a speech recognition test that evaluates the potential benefit of spatial (i.e., noise presented from different locations) and vocal cues (i.e., same or different talkers for speech and noise), and (c) reporting results using commercially available test materials that have not been used previously in studies with nonnative listeners. There are few published studies examining speech-in-noise performance on audiological tests in speakers who speak Mandarin Chinese as a native language, as well as the effects of L2 proficiency on bilingual speakers' performance on audiological tests. Furthermore, to the authors' knowledge, there are no published studies that examined the potential benefit of providing stimuli with combined spatial and vocal cues on this population. However, there are previous studies documenting the separate benefit of vocal cues (i.e., fundamental frequency) (e.g., Zhang et al, 2016) as well as spatial separation of speech and noise sources (e.g., Nilsson et al, 1994; Schafer, Beeler, et al, 2012). Understanding the effects of noise on NC children's speech recognition performance is of great importance given the acoustic characteristics of typical school classrooms where children are expected to learn, which are often plagued with high levels of noise and reverberation (Knecht et al, 2002; Nelson et al, 2007). Furthermore, this study will provide hearing professionals with expectations regarding clinical performance of individuals who have normal hearing and speak Mandarin Chinese as a native language.

METHODS

Participants

Participants included 25 NE adults and 16 NE children as well as 25 adults and 16 children who speak both NC and English. Participants were recruited from the student body at the University of North Texas as well as churches in north Texas. Table 1 provides additional information about the participant demographics. A wide range of ages were included for the children to obtain an adequate sample size, particularly for the children who spoke NC, as well as to follow the recommendation ages for use the speech recognition tests (i.e., a minimum of five to six yr). All NC participants completed an English-proficiency scale that is provided in the Appendix. According to parents of the children or

Table 1. Demographic Information about the Participants

Groups	N	Gender	Age Range (yrs)	Mean Age in Yrs (SD)
English: Adults	25	Male: 8 Female: 17	19–48	25 (5)
Mandarin: Adults	24	Male: 9 Female: 15	26–62	37 (6)
English: Child	16	Male: 9 Female: 7	6–16	10 (3)
Mandarin: Child	16	Male: 4 Female: 12	6–17	9 (3)

self-reports, no participants had a history of speech or language delays, diagnosed disabilities, otitis media, ear surgeries, or academic difficulties. According to a pure-tone hearing screening at octave frequencies ranging from 250 to 8000 Hz, all participants but one (see results section for more information) NC adult had thresholds less than 25 dB HL in each ear for adults and less than 15 dB HL for children. The more stringent threshold criteria for the children was based on evidence from Bess et al (1998) that indicates that children with hearing thresholds greater than 20 dB HL had poorer academic performance and teacher-rated communication abilities relative to peers with better hearing thresholds. In addition, all participants passed a tympanometry screening with pass criteria of pressure between 50 and –150 daPa, compliance between 0.2 and 1.8 mL, and ear canal volume between 0.2 and 2.0 mL. This study was approved by the University of North Texas Institutional Review Board, and written informed consent was obtained from all participants.

Stimuli

AzBio Sentence Test

Speech recognition performance in noise was assessed with the AzBio sentence test for adults and Pediatric AzBio sentence test for children. The open-set AzBio sentence test, developed by Spahr and Dorman (2004) and Spahr et al (2012), was used initially to examine the speech recognition performance of individuals with cochlear implants as part of the Minimum Speech Test Battery. This test had better ecological validity because the speech stimuli from four talkers were more realistic in comparison to other tests utilizing stimuli from only a single talker. In addition, scores from individuals with implants did not reach ceiling (i.e., 100%), which was common when using other frequently used tests (e.g., Hearing in Noise Test) (Gifford et al, 2008). The commercial version of the AzBio sentence test is available on compact disc and consists of 15 lists of 20 fixed-intensity sentences on one channel with continuous ten-talker babble in the opposite channel (Arizona State University Board of Regents, 2006). Each sentence in a particular

list is spoken by one of four talkers (two male; two female talkers), and scores are determined by the percentage of words repeated correctly. Previous research (Schafer, Pogue, et al, 2012) found that 10 of the 15 AzBio sentence lists presented at a 0 and –3 SNR result in similar performance in monolingual English-speaking adults (i.e., five lists were more difficult or easier than the remaining ten lists). Of these, six equivalently difficult and pseudorandomly selected lists (no repeats) were used in the present study (i.e., lists 2, 4, 8, 9, 10, 11).

The Pediatric AzBio sentence test was developed in 2014 (Spahr et al, 2014) and consists of 16 lists of 20 fixed-intensity sentences spoken by one female talker on one channel with ten-talker babble on the opposite channel. For the pediatric test, only sentences that were recognized by normal hearing, typically-developing children were included, and one talker, rather than four, was utilized to make the test less difficult (Spahr et al, 2014). This test consists of 16 equally intelligible and equivalent lists of 20 sentences each. This version of the test is recommended for children as well as adults with poor performance on the original AzBio sentence test and is scored based on the percentage of words repeated correctly. In this study, lists 1 through 6 were pseudorandomly selected (no repeats) and presented to participants.

Listening in Spatialized Noise-Sentence Test

The Listening and Spatialized Noise-Sentence (LiSN-S) test (Cameron et al, 2006; National Acoustic Laboratories, 2011) was used to determine the impact of spatial and vocal cues on the speech recognition performance in NE and NC listeners. The LiSN-S test was presented via computer software and headphones that were set to a comfortable volume as determined by patient preference. The LiSN-S software creates the perception of a three-dimensional acoustic space and evaluates 50% correct speech recognition thresholds (SRTs) in the presence of distracter stories by varying the spatial and pitch (vocal cues) characteristics of incoming stimuli. More specifically, the target sentences are presented at 0° azimuth, and the distracter stories vary in spatial location (0° in front of the listener versus ±90° azimuth on the sides of the listener), vocal characteristics (same versus different

voice than the target sentences), or both. Each participant completed the four listening conditions including: (a) same voice at 0° azimuth (low-cue SRT), (b) same voice at ±90° azimuth, (c) different voices at 0° azimuth, and (d) different voices at ±90° azimuth (high-cue SRT). Upon completion of the four test conditions, the LiSN-S software provides results for two SRT conditions and three advantage conditions. The *low-cue SRT* condition represents a speech-in-noise threshold when no spatial or vocal cues are available (same voice at 0° azimuth) while the *high-cue SRT* represents a speech-in-noise threshold when both vocal (i.e., different talkers) and spatial cues (i.e., stimuli presented at ±90° azimuth) are available. The three advantage conditions represent difference scores that are calculated using two of the conditions described above. The *talker advantage* condition quantifies the listener's ability to use differences in vocal quality to distinguish the signal of interest (talker advantage = same voice 0° – different voices 0°). The *spatial advantage* condition reflects a listener's ability to use differences in the physical location of incoming stimuli to perceive the signal of interest amid competing signals (spatial advantage = same voice 0° – same voice ±90°). Finally, the *total advantage* condition represents the benefit of using vocal and spatial cues over conditions without these cues (total advantage = same voice 0° – different voices ±90°).

English Proficiency Questionnaire

An English proficiency questionnaire (see Appendix) was completed by all NC participants. All adult participants were able to complete the questionnaire independently, and six of the 16 children required some parental assistance to complete the questionnaire. The purpose of this questionnaire was to quantify the relationship between English proficiency and speech recognition performance in noise. The participants were asked about the age of English acquisition, length of English learning, primary language at home, minutes of spoken English per day, and self-rated English proficiency.

Equipment

The hearing screening was conducted using an audiometer (GSI-61), headphones (TDH-50), and a double-walled sound booth. The tympanometry screening was conducted with immittance equipment (Maico MI 34). AzBio test stimuli were presented with the same audiometer, a compact disc player (Sony 5-CD Changer), and one loudspeaker (Grason-Stadler Standard) located at 0° azimuth relative to the listener. The test was presented via sound field loudspeakers to allow for comparison to future studies with participants who have hearing loss and amplification. During the AzBio sentence test, the participant was seated 5' from the head-level loudspeaker, and calibration of the stimuli was conducted with a sound-level

meter (Larson-Davis 824). The LiSN-S stimuli were presented through the LiSN-S software program on a compact disc, a laptop computer (Dell Latitude), and manufacturer-recommended headphones (Sennheiser HD 215).

Procedures

After completing the consent form and answering questions about hearing health, otoscopy as well as the hearing and tympanometry screenings were performed. The English proficiency scale was completed by NC-speaking participants. The order of the AzBio and LiSN-S test measures was counterbalanced across participants. For the AzBio sentence test, participants completed the six lists at three randomly ordered (no repeats) SNRs of -3, 0, and 3 dB. Two of these signal levels, 0 and -3 SNR, were used for the present study because no ceiling or floor effects were obtained with these SNRs in a previous study including normal hearing adult listeners (Schafer, Pogue, et al, 2012). The +3 SNR was added to the present study, given expected performance differences between the groups of listeners, where pediatric or nonnative listeners may need a higher SNR to achieve adequate performance (e.g., 80% correct). The noise level was constant at 60-dBA at the listener's ear while the sentence stimuli ranged from 55 to 65 dBA at the listener's ear depending on the test condition. As stated previously, the LiSN-S test was completed via headphones at a comfortable volume selected by the participant. Before testing, the normal-hearing examiner ensured that stimuli were audible at the volume selected by the patient.

RESULTS

All but one NC participant passed the hearing screening at every frequency. This participant received a hearing test (i.e., threshold search), and the results suggested a mild sensorineural hearing loss only at 2000 (right and left: 30 dB HL) and 4000 Hz (right: 30 dB HL; left: 40 dB HL) with a normal pure-tone average at 500, 1,000, and 2,000 of 18 dB HL. Given the presence of the hearing loss, this participant was not included in the statistical analysis in the results section; however, her speech recognition performance across the test conditions and measures was equal to or better than the average (i.e., within one standard deviation [SD] of the mean) of the adults in the NC group.

Speech Recognition in Noise on the AzBio Sentence Test

The average AzBio sentence recognition performance of the adults and children is shown in Figures 1 and 2, respectively. Given the excellent performance (i.e., near ceiling) of many of the adults and children, particularly in the +3 dB SNR condition, the data were transformed

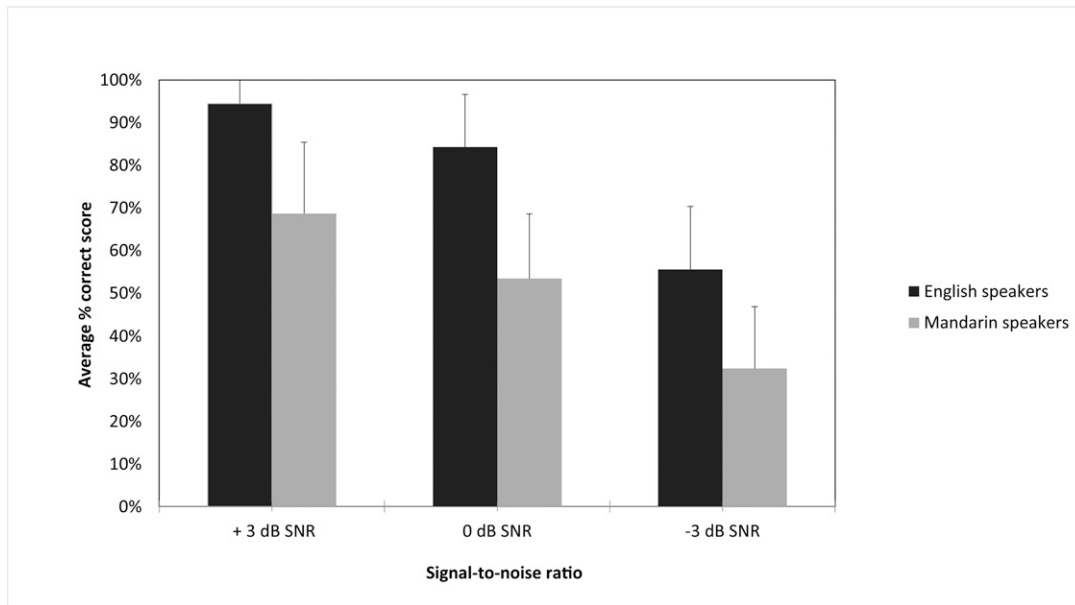


Figure 1. Average speech recognition on the AzBio sentence test in adults. Vertical lines represent SD.

into rationalized arcsine transformed units before the statistical analysis to ensure homogeneous variance across the scores (Studebaker, 1985). The data for each separate age group (i.e., adults; children) were analyzed using a two-factor repeated measures analysis of variance (RM ANOVA) with the independent variables of group (NE; NC) and listening condition (+3, 0, -3 dB SNR). Data from adults and children were not compared given the unequal sample sizes as well as the different versions of the AzBio sentence test that were used.

The analysis, with a Greenhouse–Geisser Correction, for the adults yielded a significant main effect of group, $F_{(1,147)} = 77.6, p < 0.0001$, a significant main effect of

listening condition, $F_{(2,47)} = 367.9, p < 0.0001$, and a significant interaction effect between group and listening condition, $F_{(2,147)} = 20.7, p < 0.0001$. To further examine the significant main effects and interaction effect, a Tukey–Kramer Multiple Comparisons Test was conducted, which accounts for the multiple post hoc comparisons (adjusts the p value). For the main effect of group, the NE group showed significantly higher average performance at each SNR ($p < 0.05$). For the main effect of listening condition, each 3-dB increase in SNR resulted in significantly higher performance across both groups of listeners ($p < 0.05$). When examining the significant interaction effect, group differences were detected at

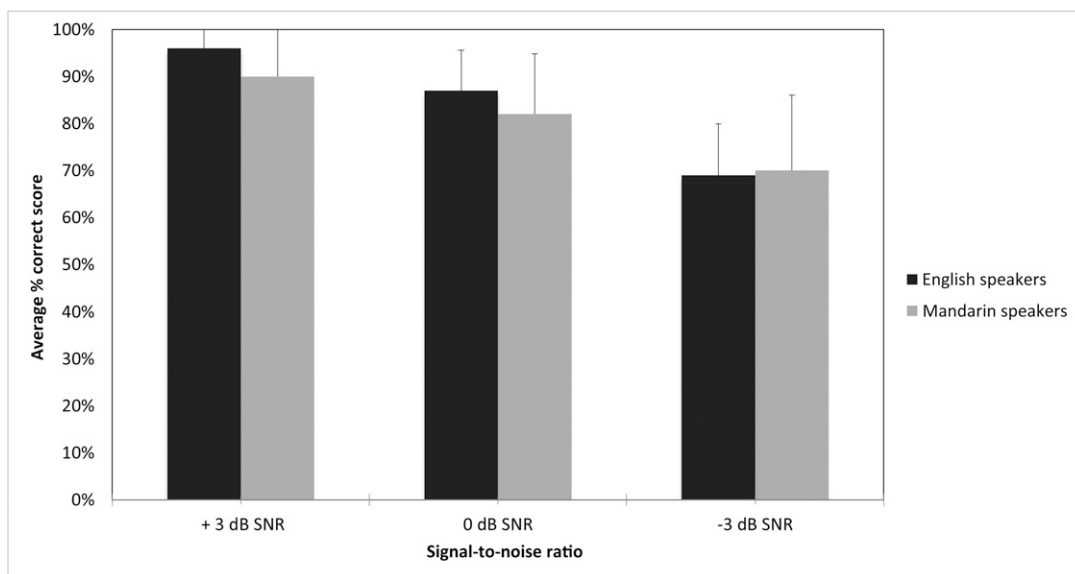


Figure 2. Average speech recognition on the pediatric AzBio sentence test in children. Vertical lines represent SD.

each separate SNR ($p < 0.05$). In addition, all conditions and comparisons across both groups were significantly different with the exception of the performance of the NE group at a -3 SNR compared to the NC group at a 0 SNR ($p > 0.05$).

Results for the pediatric group were much different than those of the adult group. The RM ANOVA, with a Greenhouse–Geisser Correction, on the pediatric data yielded no significant main effect of group, $F_{(1,96)} = 0.72$, $p = 0.40$, a significant main effect of listening condition, $F_{(2,96)} = 115.9$, $p < 0.00001$, and a significant interaction effect, $F_{(2,96)} = 3.3$, $p = 0.04$. The post hoc analysis for the main effect of SNR showed that each 3-dB increase in SNR resulted in significantly better speech recognition performance across both groups. The post hoc analysis on the interaction effect suggested that, at each separate SNR, the talker groups had similar performance. However, comparisons within and between group for pairs of the different SNRs yielded significant differences ($p < 0.05$), with the exception of no significant difference between the NE group at a 0 dB SNR and the NC group at a $+3$ dB SNR ($p > 0.05$).

Performance on the LiSN-S Test

The average performance of both age groups on the LiSN-S test is shown in Figure 3 for the SRT conditions and in Figure 4 for the advantage conditions. Data from both age groups were combined and analyzed with two separate three-factor RM ANOVA, one for the SRT conditions and one for the advantage conditions. The independent variables for each RM ANOVA included age group (adult; child), talker group (NE; NC), and test condition.

The RM ANOVA on the SRT conditions showed a significant main effect of talker group, $F_{(1,162)} = 23.0$, $p < 0.0001$, significant main effect of age group, $F_{(1,162)} = 7.7$, $p = 0.007$, and a significant main effect of test condition, $F_{(1,162)} = 172.1$, $p < 0.0001$. Significant interaction effects, with a Greenhouse–Geisser Correction, were obtained between talker group and condition, $F_{(1,162)} = 18.0$, $p < 0.001$, as well as between age and condition, $F_{(1,162)} = 0.0$, $p = 0.99$. No significant interaction effect was detected between the talker group and the age group, $F_{(1,162)} = 4.0$, $p = 0.05$. Post hoc analyses on the main and interaction effects were conducted with the Tukey–Kramer Multiple Comparisons. For the main effects of age group and talker group, adults showed significantly lower (better) thresholds than the children ($p < 0.05$), and the NE participants had significantly lower ($p < 0.05$) thresholds than the NC participants. Regarding the main effect of condition, the high-cue SRT resulted in lower thresholds than the low-cue SRT ($p < 0.05$). For the interaction effect between talker group and condition, the NE group was significantly better than the NC group only in the high-cue condition ($p < 0.05$). Similarly, for the interaction effect between age and condition, the adults only showed better performance than the children in the high-cue condition ($p < 0.05$).

The average results for the advantage conditions are shown in Figure 4. The RM ANOVA on the advantage conditions yielded a significant main effect of talker group, $F_{(1,243)} = 39.3$, $p < 0.0001$, significant main effect of age group, $F_{(1,243)} = 4.8$, $p = 0.03$, and a significant main effect of test condition, $F_{(2,243)} = 95.4$, $p < 0.0001$. A significant interaction effect, with a Greenhouse–Geisser Correction, was obtained between talker group and condition, $F_{(1,243)} = 3.4$, $p = 0.04$. No significant

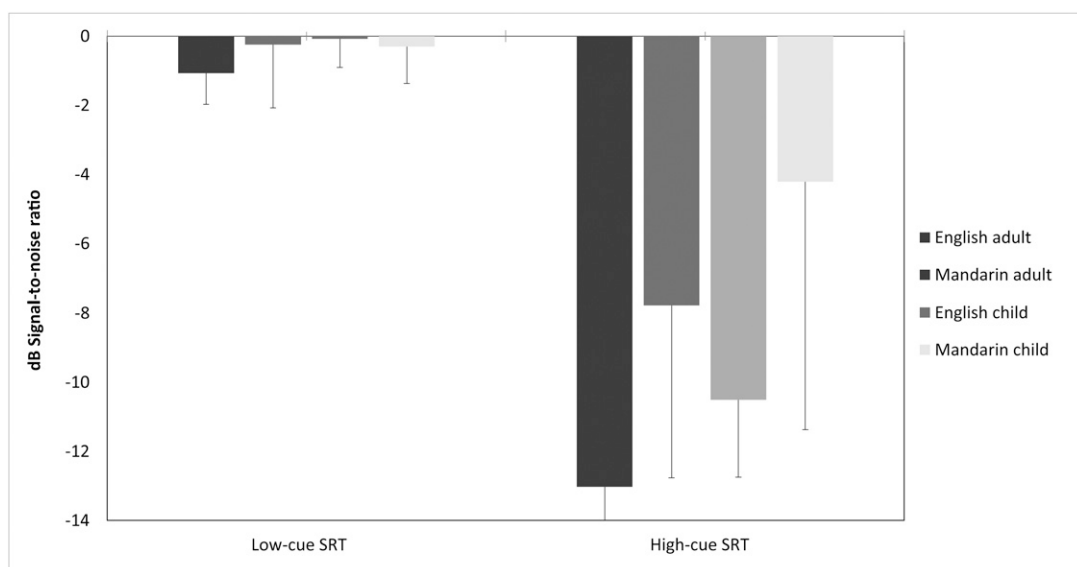


Figure 3. Average speech-in-noise thresholds on the listening and spatialized noise-sentence test in for the children and adults. Vertical lines represent SD.

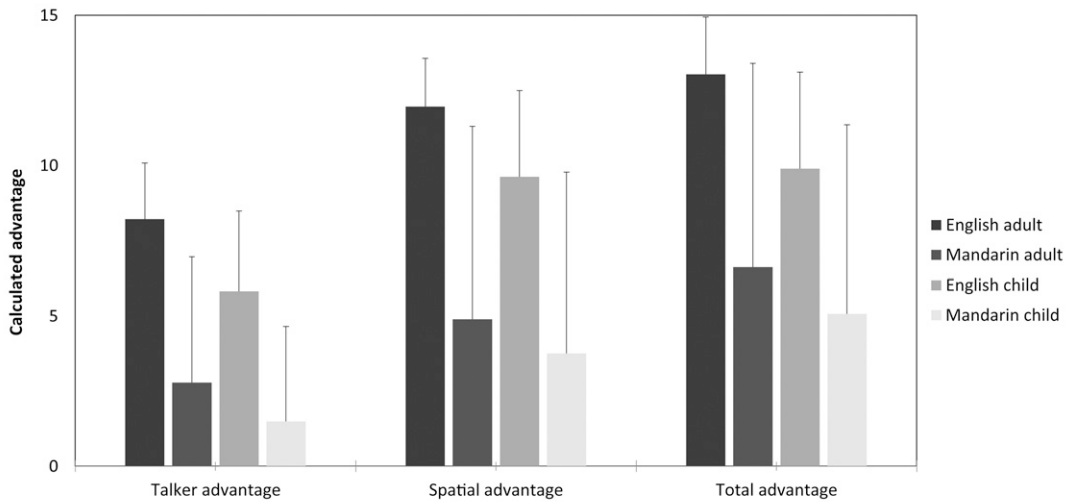


Figure 4. Average advantage scores on the listening and spatialized noise-sentence test for the children and adults. Vertical lines represent SD.

interaction effect was calculated between the age group and condition, $F_{(1,243)} = 0.56, p = 0.56$ or between the talker group and the age group, $F_{(1,243)} = 0.52, p = 0.47$. The post hoc analyses on the main effects of age group and talker group showed significantly larger advantage scores for the adults over the children ($p < 0.05$) and the NE participants over the NC participants ($p < 0.05$). Regarding the main effect of condition, each comparison between conditions yielded significant differences ($p < 0.05$) with the total advantage resulting the highest advantage followed by the spatial advantage and talker advantage.

When examining the interaction effect between the talker group and the condition ($p < 0.05$), there were several noteworthy findings. First, in each condition, the NE speakers had larger advantages than the NC speakers ($p < 0.05$). Second, the talker advantage and spatial advantage conditions for the NC speakers resulted in significantly smaller advantages than all remaining conditions ($p < 0.05$). Finally, the total advantage condition for the NC speakers did not differ ($p > 0.05$) from the spatial advantage condition for the NE speakers.

Correlations: Performance and Demographic Factors

The Pearson’s product-moment correlation coefficient was used to examine how the data collected from the English Proficiency Questionnaire (Table 2) are related to speech recognition performance on the AzBio (arcsine transformed data) and LiSN-S tests. Correlation coefficients for adult and children are provided in Tables 3 and 4, respectively. The significance of the moderate (≥ 0.3) and strong (≥ 0.05) correlations (Cohen, 1988) were tested with *t*-tests. Results of the correlation analyses for the adult data showed some moderate and strong significant correlations between the demographic factors and performance on both test measures. In particular, for the adults, the demographic factors that most influenced performance on both test measures were the length of time the person lived in the United States and the age of English acquisition. For the children, four factors were related to performance in the test measures: the child’s age at testing, length of time the child lived in the United States, the number of minutes English is

Table 2. Results of English Proficiency Questionnaire for Mandarin/Chinese Participants

Group	Age: English Acquisition (yrs)	Months in United States	Minutes of English Per Day	Self-Rated English Proficiency (1–10)	
NC adults	Range	5–40	1–468	10–600	2–10
	Mean	13	77	174	7
	SD	6	99	188	2
NC children	Range	1–9	1–144	60–720	2–9
	Mean	5	53	401	7
	SD	2	51	187	3

Note: English proficiency ratings were on a scale with 1 as the lowest rating and 10 as the highest rating relative to native English speakers.

Table 3. Correlation Coefficients between Demographic Variables and Test Results for the Adult Group

	AzBio: +3 dB SNR	AzBio: 0 dB SNR	AzBio: -3 dB SNR	LiSN-S: Low-Cue SRT	LiSN-S: High-Cue SRT	LiSN-S: Talker Adv.	LiSN-S: Spatial Adv.	LiSN-S: Total Adv.
Age	0.09	0.14	-0.40*	-0.22	-0.10	0.14	-0.23	-0.16
Age: English acquisition	-0.33*	-0.26	-0.24	-0.52*	0.12	-0.30*	-0.43*	-0.31*
Length of time in U.S.	0.49*	0.22	-0.20	0.06	-0.35*	0.35*	0.30*	-0.31*
Minutes English/Day	0.40*	0.26	-0.23	-0.02	-0.23	0.23	0.24	0.21
Proficiency	0.59*	0.35*	-0.08	-0.08	-0.11	0.11	0.15	0.17

Note: * = significant at the < 0.01 probability level; Adv. = advantage.

spoken per day, and the child's self-rated English proficiency.

DISCUSSION

Performance on the AzBio Sentence Test

The goal of this study was to compare speech recognition in noise in adults and children who speak American English and Mandarin Chinese with and without the presence of spatial and vocal cues. The second goal of this study was to examine the effects of English proficiency and experience on speech recognition for the native Mandarin Chinese speakers. On the AzBio sentence test, a fixed-intensity measure that yielded percent correct scores, the NE adults' scores were significantly higher than the NC adults at all three SNRs. This finding is similar to previous studies that showed better speech recognition performance of NE adults relative to nonnative adults (Nakamura and Gordon-Salant, 2011). However, unlike previous studies in children (Crandall and Smaldino, 1996; Nelson et al, 2005), there were no significant differences between the groups across at the three SNRs. The authors of this study hypothesize that the differences between findings for the adults and children stem from the difficulty of the adult version of the AzBio (higher level vocabulary) relative to the simplicity of the pediatric version of the AzBio (lower vocabulary level). There is a possibility that NE and NC group differences would have been detected in the study if the adult AzBio sentence test was used. There is also a possibility that no group differences between the adult groups would have been detected if the Pediatric version

of the AzBio was used for the pediatric as well as adult groups.

When examining the data from the English Proficiency Questionnaire (Table 2), there are two factors that likely contributed to the different findings between the NC adult and child groups. First, the average age at English acquisition was eight yr for the children as compared with 13 yr for adults. Second, given that all the children are currently living and attending school in the United States, the child group is exposed to a greater number of spoken-English minutes per day. On average, children spoke English an average of 410 min per day (SD = 187), whereas adults only heard an average of 186 min per day (SD = 194).

It has been reported that the age of L2 acquisition and amount of L2 use affect the L2 speakers' overall proficiency as well as perception of individual segments (Flege et al, 1999; Piske et al, 2001). For example, Italian-English bilingual speakers who started learning English (L2) later in life (<14 yr) had lower scores in English vowel discrimination task than Italian-English bilingual speakers who started learning English earlier in life (<7 yr) (Flege et al, 1999). In addition, there were differences among the bilinguals who started learning English (L2) earlier in life (<7 yr) based on the amount of daily L2 use. Those who used English frequently, scored similar to the monolingual English speakers, whereas those who used English less (and Italian more), scored lower than the monolingual English scores on vowel perception. In the present study, the NC children were using English on a daily basis and started learning English at a younger age than the NC adults. These two factors provide a reasonable explanation for the NC

Table 4. Correlation Coefficients between demographic Variables and Test Results for the Pediatric Group

	AzBio: +3 dB SNR	AzBio: 0 dB SNR	AzBio: -3 dB SNR	LiSN-S: Low-Cue SRT	LiSN-S: High-Cue SRT	LiSN-S: Talker Adv.	LiSN-S: Spatial Adv.	LiSN-S: Total Adv.
Age	0.43*	0.30*	0.40*	-0.26	0.53*	0.08	0.07	0.13
Age English	-0.29	-0.14	0.03	0.10	0.19	-0.19	-0.25	-0.20
Length in U.S.	0.81*	-0.57*	0.45*	-0.35*	0.00	0.58*	0.52*	0.60*
Minutes English/Day	0.71*	0.57*	0.62*	-0.64*	0.22	0.17	0.15	0.23
Proficiency	0.54*	0.39*	0.47*	-0.58*	0.01	-0.08	-0.18	-0.03

Note: * = significant at the <0.01 probability level; Adv. = advantage.

children's performance that was similar to that of the NE children for the sentence recognition task at three fixed SNRs.

For the correlations between demographic variables and AzBio test results at the +3 dB SNR (see Table 3), the adult NC speakers' scores were significantly correlated with every demographic variable except age at testing. However, at 0 dB SNR and -3 dB SNR, only their self-related proficiency or age at testing, respectively, was significantly correlated with their performance. These results suggest that, for the NC adult speakers, the factors related to their English (L2) proficiency and experience with English affected their performance only at the more favorable conditions (+3 dB and 0 dB SNR). When the condition is not favorable (-3 dB SNR), English proficiency did not seem to affect their performance. These findings are consistent with Shi and Sánchez (2010), who found that Spanish-English bilingual speakers' language dominance affected their performance on audiological tests. Specifically, they found that Spanish-English bilingual speakers performed better on audiological tests in their dominant language (Spanish or English). The current study also showed that the NC adults' scores were lower than the NE adults, and their performance was correlated with their English (L2) proficiency at +3 dB and 0 dB SNR. These results indicate that the NC adult speakers' results on the AzBio test were negatively affected because it was in their L2, and the results at more difficult SNRs need to be interpreted with caution.

For the NC children, all of the demographic variables, with the exception of age at English acquisition, were significantly correlated with performance on the Pediatric AzBio sentence test (Table 4). These results suggest that, for the NC children, their L2 use and proficiency influenced the test results more than the age at which they started learning English. These results coincide with previous research in L2 speech perception and production (Flege et al, 1999; Piske et al, 2001) in that the amount of L2 use (and L2 proficiency as a result) as well as the length of time in the United States are important factors. Nishi et al (2017) also reported that Spanish-English bilingual children can recognize English consonants in quiet and in noise as accurately as English monolingual children, if their dominant language is English. In the current study, the NC-speaking children were using English >6 h per day and had been in the United States for >4 yr on an average (Table 4). The results of Nishi et al and the current study both indicate that bilingual children's performance on the audiological tests in English may not differ from those of monolingual children, as long as they are dominant or proficient in English.

Performance on the LiSN-S Test

Unlike the AzBio test, the LiSN-S test uses an adaptive paradigm to determine two speech-in-noise thresh-

olds with and without vocal and spatial cues as well as three advantage conditions that provide an estimate of relative contributions from vocal cues, spatial cues, or both cues. For the analysis of the LiSN-S test data, results from the adults and children were combined because the same version of the test was administered to both age groups. Overall, adults and NE speakers had better speech-in-noise thresholds. However, a closer look at the post hoc analysis suggested that the age and language group differences only occurred in the high-cue condition.

This is a noteworthy finding because it suggests that, the two NC groups were unable to gain as much benefit from the presence of the vocal (i.e., different talkers) and spatial cues (i.e., stimuli presented at $\pm 90^\circ$ azimuth) as the two NE age groups. The performance similarities of NE and NC age groups in the low-cue condition also highlight the importance of providing spatial and vocal cues to NE and NC listeners, regardless of age.

When examining the data across the three advantage conditions, similar to the SRT conditions, adults and NE speakers had better speech-in-noise thresholds relative to the child and NC groups, respectively. In addition, the total advantage condition resulted in the largest advantage relative to the spatial and talker advantage conditions. The most noteworthy finding for the post hoc analysis was that the NE speakers had larger advantages than the NC speakers within each of the advantage conditions. To summarize, the NE adults and children made greater use of both vocal and spatial cues when compared with the NC adults and children. Similar findings were reported by Cooke et al (2008) where groups of native and nonnative listeners benefitted equally to the presence of vocal cues (i.e., speech and masker that differed in fundamental frequency), but the native listeners had significantly better overall performance.

Comparisons between Test Measures

When examining the results across groups and the two test measures, it is clear that the fixed-intensity procedure on the AzBio sentence test yields different outcomes, in terms of the group comparisons, than the adaptive procedure used for the LiSN-S test. As discussed above, two different versions of the AzBio sentence test were used for the adults and children, which likely resulted in different outcomes for the two age groups. It is also important to note that the noise stimulus for the AzBio sentence test is 10-talker babble whereas the noise stimulus/masker for the LiSN-S test is the same or a different competing talker. The differences in the noise stimuli between the tests is notable given that multitalker babble is considered more of an energetic masker, where the noise degrades with the perception of the speech stimuli, but contains few, if any, contextual cues (Lecumberri et al, 2010). On the other hand, the competing talker for the LiSN-S test is considered an informational

masker (i.e., a second talker), where the competing signal interferes with the target signal, and contextual information is available. Previous research suggests that nonnative listeners' speech recognition is negatively affected by both energetic and informational maskers, but the energetic masker results in even greater decrement than the informational masker (Garcia Lecumberri and Cooke, 2006). It is also possible that the use of the simpler pediatric stimuli as well as the smaller sample size contributed to the lack of significant differences between pediatric groups when using the energetic masker.

Clinical Implications and Study Limitations

Evaluation of speech recognition performance in noise is an ecologically valid clinical measure and should be used for routine clinical evaluations. Although clinical measures of speech recognition are often conducted in well-controlled environments, they provide an estimate of a patient's ability to recognize speech stimuli in challenging environments. Not only can speech-in-noise measures be used for counseling purposes, but they can be used for comparison with other populations, as was shown in the present study. However, it is pertinent for audiologists to be aware of demographic factors that could influence results. Therefore, the most effective use of English speech-in-noise measures for nonnative listeners may be for comparison purposes, such as performance in quiet versus performance in noise or with and without a hearing aid with a directional microphone.

The findings in this study contribute to the published literature on speech recognition in noise of nonnative listeners, and in particular highlight the importance and benefit of higher SNRs, vocal cues, and spatial cues for improving performance of adult and pediatric NC listeners in the presence of background noise. The results in the present study provide evidence regarding the difficulties that NC listeners may encounter in noisy environments, such as school classrooms, social situations, and workplaces. When testing the speech recognition of patients who speak Mandarin Chinese or another native language, audiologists and health professionals will need to have realistic expectations regarding clinical findings. Shi and Sánchez (2010) suggested that bilingual dominance should be considered especially for adults who started learning English later in life. Bilingual speakers' performance may be affected by their proficiency in English if English is their L2 and less dominant language. Professionals also need to keep in mind that any degree of hearing loss would likely result in even greater difficulty (i.e., double jeopardy) than a patient who speaks English as a native language (Nelson et al, 2005).

Future research will need to be conducted to compare the speech recognition performance of children who speak American English, Mandarin Chinese, and other languages as their native language in order to examine

the effects of energetic and information maskers more closely. In the opinion of the authors, both types of noise stimuli are beneficial to use during clinical evaluations given that both are encountered in everyday environments. In addition, use of test measures that use fixed-intensity as well as adaptive stimuli provide opportunities to examine the effects of different SNRs as well as a sensitive measure (i.e., 50% correct speech-in-noise thresholds) to compare conditions with stimuli variations.

Limitations in the present study relate to the test measures, sample sizes, and varied demographic characteristics. First, as discussed previously, different versions of the AzBio sentence test were used to ensure the use of stimuli with appropriate vocabulary levels for each age group. Therefore, comparisons between the two age groups were difficult; however, results showed some noteworthy differences in performance across age groups. Second, the use of unequal sample sizes precluded any statistical comparison between the age groups on either measure to avoid violation of the homogeneity of variance assumption for the ANOVA. Future research will use identical test measures and sample sizes to further examine the effect of age. Use of larger sample sizes would also be beneficial to further examine and control for demographic characteristics as well as English proficiency and experience. In the present study, the samples included listeners with highly variable ages of English acquisition, lengths of residence in the United States, and minutes of English spoken per day.

CONCLUSIONS

Results of this study showed better speech recognition in the NE adult group relative to the NC adult group on both test measures. However, when comparing the child groups, the NE group had better speech recognition on only the LiSN-S test, which examined the benefit of providing vocal and spatial cues. The findings in this study contribute to the published literature on speech recognition in noise of nonnative listeners, and in particular highlight the importance and benefit of higher SNRs, vocal cues, and spatial cues for improving performance of adult and pediatric NC listeners in the presence of background noise.

REFERENCES

- American Speech-Language-Hearing Association. (2017) Issues in ethics: cultural and linguistic competence. Position Statement. <http://www.asha.org/Practice/ethics/Cultural-and-Linguistic-Competence/>. Accessed September 1, 2017.
- Arizona State University Board of Regents. (2006) *AzBio Sentence Test*. Tempe, AZ: Auditory Potential, LLC.
- Bess FH, Dodd-Murphy J, Parker RA. (1998) Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. *Ear Hear* 19(5):339–354.
- Boothroyd A. (2008) The performance/intensity function: an underused resource. *Ear Hear* 29(4):479–491.

- Cameron S, Dillon H, Newall P. (2006) The listening in spatialized noise test: an auditory processing disorder study. *J Am Acad Audiol* 17(5):306–320.
- Cohen J. (1988) The analysis of variance. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 273–406.
- Cooke M, Garcia Lecumberri ML, Barker J. (2008) The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am* 123(1):414–427.
- Crandell CC, Smaldino JJ. (1996) Speech perception in noise by children for whom English is a second language. *Am J Audiol* 5(3):47–51.
- Dubno JR, Horwitz AR, Ahlstrom JB. (2003) Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *J Acoust Soc Am* 113(4 Pt 1):2084–2094.
- Festen JM, Plomp R. (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am* 88(4):1725–1736.
- Finitzo-Hieber T, Tillman TW. (1978) Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *J Speech Hear Res* 21(3):440–458.
- Flege JE, MacKay IRA, Meador D. (1999) Native Italian speakers' perception and production of English vowels. *J Acoust Soc Am* 106(5):2973–2987.
- Garcia Lecumberri ML, Cooke M. (2006) Effect of masker type on native and non-native consonant perception in noise. *J Acoust Soc Am* 119(4):2445–2454.
- Garcia Lecumberri ML, Cooke M, Cutler A. (2010) Non-native speech perception in adverse conditions: a review. *Speech Commun* 52:864–886.
- Gifford RH, Shallop JK, Peterson AM. (2008) Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurootol* 13(3):193–205.
- Kalikow DN, Stevens KN, Elliott LL. (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61(5):1337–1351.
- Knecht HA, Nelson PB, Whitelaw GM, Feth LL. (2002) Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *Am J Audiol* 11(2):65–71.
- Lecumberri M, Cooke M, Cutler A. (2010) Non-native speech perception in adverse conditions: a review. *Speech Commun* 52(11–12):864–886.
- McCreery R, Ito R, Spratford M, Lewis D, Hoover B, Stelmachowicz PG. (2010) Performance-intensity functions for normal-hearing adults and children using computer-aided speech perception assessment. *Ear Hear* 31:95–101.
- Nakamura K, Gordon-Salant S. (2011) Speech perception in quiet and noise using the hearing in noise test and the Japanese hearing in noise test by Japanese listeners. *Ear Hear* 32(1):121–131.
- National Acoustic Laboratories. (2011) *Listening in Spatialized Noise-Sentences Test, LISN-S Software*. Stafa, Switzerland: Phonak.
- Nelson EL, Smaldino J, Erler S, Garstecki D. (2007) Background noise levels and reverberation times in old and new elementary school classrooms. *J Educ Audiol* 14:16–22.
- Nelson P, Kohnert K, Sabur S, Shaw D. (2005) Classroom noise and children learning through a second language: double jeopardy? *Lang Speech Hear Serv Sch* 36(3):219–229.
- Neuman AC, Wroblewski M, Hajicek J, 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.
- Nilsson M, Soli SD, Sullivan JA. (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95(2):1085–1099.
- Nilsson MJ, Soli SD, Gelnett DJ. (1996) *Development of the Hearing in Noise Test for Children (HINT-C)*. House Ear Institute.
- Nishi K, Trevino AC, Rosado Rogers L, García P, Neely ST. (2017) Effects of simulated hearing loss on bilingual children's consonant recognition in noise. *Ear Hear* 38(5):e292–e304.
- Piske T, MacKay IRA, Flege JE. (2001) Factors affecting degree of foreign accent in an L2: a review. *J Phonetics* 29(2):191–215.
- Rimikis S, Smiljanic R, Calandruccio L. (2013) Nonnative English speaker performance on the Basic English Lexicon (BEL) sentences. *J Speech Lang Hear Res* 56(3):792–804.
- Schafer EC, Anderson C, Sullivan J, Wolfe J, Duke M, Osman H, Wright S, Dyson J, Bryant D, Pitts K. (2016) Children's auditory recognition of digital stimuli. *J Educ Ped Rehab Audiol* 22:1–11.
- Schafer EC, Beeler S, Ramos H, Morais M, Monzingo J, Algier K. (2012) Developmental effects and spatial hearing in young children with normal-hearing sensitivity. *Ear Hear* 33(6):e32–e43.
- Schafer EC, Pogue J, Milrany T. (2012) List equivalency of the AzBio sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants. *J Am Acad Audiol* 23(7):501–509.
- Shi LF, Sánchez D. (2010) Spanish/English bilingual listeners on clinical word recognition tests: what to expect and how to predict. *J Speech Lang Hear Res* 53(5):1096–1110.
- Spahr AJ, Dorman MF. (2004) Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Arch Otolaryngol Head Neck Surg* 130(5):624–628.
- Spahr AJ, Dorman MF, Litvak LM, Cook SJ, Loiselle LM, DeJong MD, Hedley-Williams A, Sunderhaus LS, Hayes CA, Gifford RH. (2014) Development and validation of the pediatric AzBio sentence lists. *Ear Hear* 35(4):418–422.
- Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, Loiselle LM, Oakes T, Cook S. (2012) Development and validation of the AzBio sentence lists. *Ear Hear* 33(1):112–117.
- Studebaker GA. (1985) A “rationalized” arcsine transform. *J Speech Hear Res* 28(3):455–462.
- Tamati TN, Pisoni DB. (2014) Non-native listeners' recognition of high-variability speech using PRESTO. *J Am Acad Audiol* 25(9):869–892.
- U.S. Census Bureau. (2011) Language use in the United States: 2011 [American Community Survey Reports]. <http://www.census.gov/prod/2010pubs/acs-12.pdf>. Accessed September 1, 2017.
- Vermeire K, Knoop A, Boel C, Auwers S, Schenus L, Talaveron-Rodriguez M, De Boom C, De Sloovere M. (2016) Speech recognition in noise by younger and older adults: effects of age, hearing loss, and temporal resolution. *Ann Otol Rhinol Laryngol* 125(4):297–302.
- Wróblewski M, Lewis DE, Valente DL, Stelmachowicz PG. (2012) Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear Hear* 33(6):731–744.
- Zhang L, Li Y, Wu H, Shu H, Zhang Y, Li P. (2016) Effects of semantic context and fundamental frequency contours on Mandarin speech recognition by second language learners. *Front Psychol* 7:908.

University of North Texas Speech & Hearing Center

Questionnaire for English Proficiency

Please complete this questionnaire if you are an English-as-a-second-language learner.

Name	Age
Occupation	Date
1 At what age did you learn English?	
2 How long have you lived in the United States?	
3 On average, how many minutes or hours each day do you speak English?	
4 What language is spoken in your home most of the time?	English Chinese/ Mandarin
5 Are you proficient at reading English? <small>If no, the researcher will summarize the consent form for you to make sure you understand this study.</small>	Yes No
6 If you are an IELI student, what is your current level?	Communication : Level _____ Writing : Level _____
7 Which dialect of Chinese do you mainly speak?	Chinese Cantonese Other _____ Mandarin Taiwanese
8 Please list all dialects you can speak. _____	

Self-rated English listening proficiency

Please rate your English Proficiency

11-point scale with 0 as the least proficient; 10 as the most proficient.



The least
proficient

The most
proficient

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.