INFLUENCE OF DIALECT AND STIMULUS AUDIBILITY ON LISN-S PERFORMANCES

by

Julia Skinner

A Senior Honors Project Presented to the
Honors College
East Carolina University
In Partial Fulfillment of the
Requirements for
Graduation with Honors
by
Julia Skinner
Greenville, NC
May 2018

Approved by:
Andrew J. Vermiglio, AuD
Department of Communication Sciences and Disorders
College of Allied Health Science

Introduction

Speech recognition in noise difficulties is a rising issue in the audiology field. It is the most commonly cited symptom of patients with hearing aids (Kochkin, 2002) and those thought to have an auditory processing disorder (APD; AAA, 2010). It is important to evaluate speech recognition in noise ability in order to quantify the extent of this ability. Understanding the factors that may affect the results is equally as important so that speech recognition in noise problems are not incorrectly diagnosed and treated.

There are a few different tests used to determine if a person has speech recognition in noise difficulties, such as the Hearing in Noise Test (HINT; Nilsson, Soli & Sullivan, 1994; Vermiglio, 2008), the Auditory Figure Ground section of the SCAN Screening Test for Auditory Processing Disorders (Keith, 1995), and the Listening in Spatialized Noise Test – Sentences (LiSN-S; Cameron et al., 2006). All of these test protocols assess the patient’s ability to recognize speech, either words or sentences, in the presence of background noise.

There are a number of factors that may influence speech recognition in noise performance, including working memory (Parbery-Clark et al, 2009), language experience (Weiss and Dempsey, 2008; Nakamura and Gordon-Salant, 2011), and dialect (Dawes & Bishop, 2007). A couple of recent studies have reported that British children and adults scored lower on the American SCAN test than American listeners (Dawes & Bishop, 2007; Dawes, 2011). The poorer performance by British listeners was attributed to the effect of dialect.

Background

Dialect in Speech Recognition Testing

The SCAN-A (Keith, 1994) and SCAN-C (Keith, 1999) tests are used to evaluate adults and children, respectively, for APD. They were developed in the USA. The two versions of the SCAN include four subtests: filtered words, auditory figure-ground, competing words, and competing sentences. Dawes and Bishop (2007) and Dawes (2011) evaluated British children and adults, respectively, with the SCAN tests in order to investigate the possibility of over diagnosis of APD in the UK. Dawes & Bishop (2007) tested 99 children ages 6 to 10 who had normal audiograms, the SCAN-C was administered and the results were compared to US normative data. Across all ages, UK children scored significantly worse than US norms on filtered word subtests and auditory figure-ground subtests as well as on the composite score. Error analysis of the subjects' responses indicated that word familiarity and accent may have affected performances, as UK homophones were typically substituted for US accented words, such as “arm” or “on.” Dawes and Bishop (2007) suggested adjusting the UK raw scores before conversion to standard scores using US norms. The authors found that scores were normally distributed, so they adjusted the scores by adding the difference between UK and US children's raw scores to the UK children's raw scores. When using the adjusted norms, it was found that the majority of participants were no longer considered disordered as compared to the US norms.

Dawes (2011) conducted a similar study but with British adults. Thirty-one UK adults from ages 19 to 64 with normal audiograms and no history of language difficulties participated in this study. The SCAN-A was administered and results were compared to

US normative data. Individual British participant scores on three subtests: filtered words, competing words and competing sentences, and their total composite score were significantly worse than US norms. In this study, Dawes (2011) suggested adjusting UK raw scores by adding a correction factor to UK scores prior to comparison with US norms or create a list of acceptable alternative responses for items that are most likely to be missed by UK participants. Similar to the Dawes and Bishop (2007) study, Dawes (2011) found that accent and word familiarity were most likely the cause of the poorer British scores on the SCAN.

Cokely and Yager (1993) compared word recognition ability for native Spanish speakers, native monolingual English speakers, and English speakers with a few years of college-level Spanish courses. There were 10 adult native Spanish speakers, 15 adult monolingual native English speakers and 15 adult native English speakers with college-level Spanish experience. The authors tested the native Spanish group's word recognition ability by having them write and repeat Spanish words from four 50-word lists of bisyllabic Spanish words. After this, the two groups of English speakers would listen to the recordings of the Spanish speakers and decide whether the Spanish speakers repeated the Spanish words correctly, indicating either "right" or "wrong." The results showed that there was no significant difference between the two native English-speaking groups. This indicates that English-speaking audiologists are competent to judge Spanish listeners' oral responses to word recognition measures.

Woods et al. (2004) conducted a study to see if the SCAN-C showed sociocultural bias in children in second and third grade. Twenty Latino Americans and twenty Anglo-Americans with typical development and proficiency in English participated in the study.

The SCAN-C was administered and results were calculated and compared between the Latino American and Anglo-American groups. It was found that 10% more Latino American kids were categorized as borderline-disordered than Anglo-American kids based on composite scores. Dialect variation was taken into account and alternate responses were determined for target words. Next, the Latino American scores were recalculated. Following the “dialectal scoring,” the distribution of Latino American SCAN-C scores more closely matched that of Anglo American participants.

Shi and Sanchez (2011) conducted a study to compare bilingual Spanish and English speakers’ word familiarity. The authors examined the contribution of word familiarity in both English and Spanish for forty-two bilingual listeners with normal audiograms and no reported language difficulties. There were 22 female and 20 male participants with the average age of 29.6 years. Participants were divided into English dominant and Spanish dominant groups based on self-identification. Word lists from NU-6 and Spanish bisyllabic words were presented to participants binaurally at 45 dB HL through headphones. The listeners repeated the word they heard and rated familiarity on a scale from 1-7. Using this scale, 1 indicates the listener did not recognize the word, 4 indicates the listener recognized the word but not its meaning, and 7 indicates the listener knew the word. Out of the bilingual participants, it was found that the English-dominant listeners were less familiar with Spanish words and Spanish-dominant listeners were less familiar with English words, although familiarity in both languages was consistently high across participants. The authors concluded that clinicians should be aware of the effects of language familiarity on English word tests results.

Shi and Canizales (2013) also conducted a study on bilingual adults. The purpose of this study was to see if there is a dialect effect on a Spanish word recognition test. Forty native adult Spanish speakers with normal hearing participated in the study. There were two groups. The first group of participants (n = 20) spoke with a highland dialect and were from areas in Colombia, Ecuador, Peru, Bolivia, and Chine. The second group (n = 20) spoke with a Caribbean dialect and were from areas of Panama, coastal Venezuela and Colombia. Half of the participants in each group considered themselves to be English-dominant and half in each group were Spanish-dominant. All participants were tested in both quiet as well as three conditions (signal-to-noise ratios [SNRs] of +6, +3, and 0). In all conditions word lists were randomly assigned and presented at 40 dB SL re: pure-tone average of 500-2000 Hz. Participants were instructed to repeat the words they heard and write them down.

Similar to the Shi and Sanchez (2011) study, participants self-rated their Spanish word familiarity on a scale of 1-7. Speech reception testing revealed that Spanish-dominant speakers had a significantly better performance ($p<0.001$) than English-dominant listeners, implying that language dominance has a significant impact on word recognition testing in quiet and noise. An analysis of variance showed a significant difference ($p<0.016$) between the speech perception performances of Caribbean and Highland Spanish dialect groups in the quiet and noise conditions, implying that dialect has an impact on word recognition testing. This study shows that dialect and language dominance significantly affect clinical assessment of word recognition. The authors recommend that clinicians be advised on phonetic features of the dialect when scoring a client’s performance.

Weiss and Dempsey (2008) similarly focused on English-Spanish bilingual speakers, but their goal was to determine whether the Spanish and English versions of the HINT would yield equivalent results in bilingual individuals. This study contained two groups of participants: an early bilingual group of adults who acquired English as their second language before the age of seven, and a late bilingual group of adults who learned English after the age of eleven. The Spanish HINT and English HINT Noise Front conditions were administered to all participants through loudspeakers. All participants performed significantly better on the Spanish HINT than the English HINT (early group, $p<0.0001$; late group, $p<0.004$). For the Spanish HINT, the late bilingual group had an average threshold of 11.0 dB SNR whereas the early bilingual group had an average threshold of 15.1 dB SNR. For HINT thresholds, a more negative threshold represents better performance. This difference was statistically significant ($p<0.007$). A larger difference was found between groups for the English HINT. The early bilingual group had an average threshold of -4.9 dB SNR compared to the late group's average of 4.0 dB. Again, this difference was statistically significant ($p<0.0001$). In this study, the authors concluded that bilingual students would benefit from having a classroom with minimal background noise. Instructions should be given in both languages. The authors recommended that clinicians take language familiarity into account when interpreting HINT performances. Even when the participants are proficient in two languages, HINT scores were significantly better in the native language than the second language of the study participants.

Nakamura and Gordon-Salant (2011) studied bilingual speakers, but focused on adult native Japanese speakers who were fluent in English and had been living in an

English-speaking country since age 12-14 years. Ten Japanese adults who had been living in the United States for more than four years with excellent English word recognition ability participated in the study. Ten adult native English speakers also participated in the study. The English HINT test was administered to the Japanese participants using an adaptive procedure to measure a 50% correct response. It was found that all native Japanese speakers performed significantly poorer for all test conditions of the English HINT than native English speakers ($p < 0.01$). Average scores for native Japanese speakers at +2dB SNR was 79.2% whereas the average score for native English speakers was above 97%. Japanese participants had excellent English word recognition ability in quiet but did not reach native-like speech perception with sentences in quiet. Overall, Nakamura and Gordon-Salant (2011) showed that using speech material with the subject's native language for both quiet and noise testing is important.

Pure-tone Thresholds vs. Speech Recognition Testing

Another factor that may impact speech recognition in noise performance is pure-tone sensitivity. Vermiglio et al. (2012) investigated the relationship between various pure-tone average (PTA) measures and HINT performances. The participants were adults with normal pure-tone thresholds and various degrees of high-frequency hearing losses. The PTA measures used were: PTA (0.5, 1, and 2 kHz), PTA (3, 4, and 6 kHz), and PTA (0.5, 1, 2, 4, and 6 kHz). Significant positive relationships were found between all PTAs measures and the HINT in Quiet. Significant positive relationships were found between PTA (0.5, 1, 2, 4, and 6 kHz) vs. HINT Noise Right, Noise Left, and Noise Composite performances. However, no significant relationships were found between PTA (0.5, 1, and 2 kHz) vs. any of the HINT thresholds or the HINT Composite score.

Wilson, McArdle, and Smith (2007) compared the HINT test, the Bamford-Kowal-Bench Speech-in-Noise Test (BKB-SIN; Bench et al., 1979; Niquette et al., 2003), the Quick Speech-in-Noise Test (QuickSIN; Killion et al., 2004) and the Words-in-Noise Test (WIN; Wilson, 2004; Wilson & Burks, 2005) results. The study included 24 listeners with normal hearing and 72 listeners with hearing loss. Two PTAs were used: three-frequency PTA (0.5, 1, and 2 kHz) and four-frequency PTA (1, 2, 3, and 4 kHz). For the group of individuals with hearing loss, a significant correlation was found between PTA measures vs. all the speech recognition in noise tests. Performances showed that the four-frequency PTA had stronger correlations with all speech recognition performances (r=0.689; $p<0.01$) than the three-frequency PTA (r=0.292; $p<0.05$) on all tests.

Tschopp and Zust (1994) studied the German Speech Performance in Noise (SPIN; Kilakow et al., 1977; Bilger et al., 1984) test, comparing performances between a group of normal hearing adults and a group of adults with hearing loss. Two PTA measures were used: 0.5, 1, and 2 kHz and 0.5, 1, 2, and 4 kHz. For the hearing-impaired group, both PTA measures were found to have a significant positive correlation with the SPIN in all conditions (r=0.76; $p<0.05$). Smoorenburg (1992) evaluated 40 participants' hearing thresholds and compared speech reception thresholds in quiet and in noise. They did multiple comparisons and found a few significant correlations. Using three-frequency PTA (0.5, 1, and 2 kHz), the authors found a significant positive correlation between PTA and speech recognition in quiet (r=0.76; $p<0.05$). Performances also showed that a two-frequency PTA (2 and 4 kHz) has a significant positive correlation with speech recognition in noise thresholds (r=0.725; $p<0.05$).

There are very few previous studies that have focused on the relationship between PTA and LiSN-S performances. Besser et al. (2015) evaluated LiSN-S performances of 26 adults with normal audiograms, separated into a group of younger adults and a group of older adults. Two PTA measures were calculated: four-frequency PTA (0.5, 1, 2, and 4 kHz) and high-frequency PTA (6, 8, 9, and 10 kHz). For the younger group, high-frequency PTA was not significantly correlated to LiSN-S performances. However, a surprising significant negative correlation was found between the four-frequency PTA vs. LiSN-S performances in the Different Voices 0° and Same Voice 0° conditions ($r=0.42$; $p<0.05$). This indicates that poorer PTA leads to better LiSN-S performances, which is counterintuitive. However, for the older adult group, results showed positive correlations between four-frequency PTA vs. Same Voices ±90° and Different Voices 0° thresholds ($p<0.05$). Lastly, for the older group, a significant positive relationship was found between high-frequency PTA vs. Different Voice ±90°, Same Voice ±90°, Different Voice 0° thresholds as well as all derived measure results ($p<0.01$).

The LiSN-S Test & Spatial Advantage

The LiSN-S test is a fairly new hearing in noise test that is used to evaluate the ability to recognize full sentences when there are two voices talking in the background. This test was created in Australian English for Australian listeners (Cameron & Dillon, 2007) and the North American Listening in Spatialized Noise–Sentences Test (NA LiSN-S) was created in a North American dialect for listeners in the United States and Canada (Cameron et al, 2009). The LiSN-S also measures spatial separation advantage, which is the improvement in speech recognition in noise threshold when the competing voices are spatially separated from the target speech versus when the competing voices are coming from the same direction as the target speech (Cameron & Dillon 2007). Spatial advantage is calculated by subtracting the speech reception threshold (SRT) for the 0° condition from the SRT for the ±90° condition. In the development of the LiSN-S, Cameron and Dillon (2007) found that spatial separation advantage for the same voice condition was 11.3 dB, which was significantly higher than the spatial advantage for the different voice conditions, which was 9.3dB ($p < 0.001$).

The LiSN-S Test & Binaural Advantage

Another factor that plays a role in discriminating the speech signal is binaural advantage. Arrival time of sound at the ears plays a role in localization as well as discrimination of speech in noise (Bronkhorst & Plomp, 1988). When interaural time delays are introduced, intelligibility of the target speech increases and the SRTs improve. Bronkhorst and Plomp (1988) evaluated spatial advantage by presenting the target speech from 0° to the participants, with competing noise delivered from seven different directions. Their results indicated that when interaural level differences from the head shadow effect are present, the gain relies on the ear presented with the most favorable SNR. This is consistent with results from the LiSN-S that demonstrate a spatial advantage when the masker is on both sides of the head (Cameron et al., 2006).

Purpose of the Study

This first purpose of this study was to evaluate the differences in LiSN-S performances between two dialect groups: self-identified Southern speakers and non-Southern speakers. The second purpose was to investigate the relationship between PTA measures and LiSN-S performances.

Research Questions

1. What are the differences in LiSN-S performances (thresholds and derived measures) between a group of Southern speakers and a group of non-Southern speakers?
2. What is the relationship between LiSN-S performances (thresholds and derived measures) and PTA measures across all subjects?

Hypotheses

1. There will be significant differences in LiSN-S performances between the two dialect groups.
2. There will be significant positive correlations between PTA measures vs. LiSN-S performances.

Methods

Permission to conduct this research study was obtained from the East Carolina University (ECU) Institutional Review Board. Fifty-six young adults, 1 male and 55 females, from East Carolina University participated in this study. The average age was 20.2 years (SD = 0.64) with an age range of 18 to 22 years. This convenience sample was made up of undergraduate students in the Department of Communication Sciences and Disorders at ECU. All participants were native English speakers and underwent pure-tone audiometry using the Hughson-Westlake method. All participants had normal pure-tone thresholds (≤25 dB HL, 0.25-8 kHz). Participants who reported a history of brain surgery or neurological disorders were excluded from the study.

Experiment 1: Dialect

All participants filled out a questionnaire in which they indicated their self-identified dialect as Southern or non-Southern. Some participants expressed confusion when deciding their dialect, and those who marked “unsure” or “in-between” were removed from the study. There was a group of 37 self-identified Southern (average age 20.7 years) speakers and a group of 19 self-identified non-Southern speakers (average age 20.3 years).

The LiSN-S test was administered to all participants as described in the user manual (Phonak, 2007). Participants were given instructions to ignore the background voices and to focus on the target voice. The background voices run continuously throughout each condition and all the voices are female. The participants’ task was to repeat as many of the words from the target sentence as possible. A 1000 Hz tone at 55 dB SPL was presented before each sentence to alert the participant when the target sentence would be presented. The LiSN-S test uses sentences constructed according to the development criteria for the Bramford-Kowal-Bench sentences (Bench et al., 1979).

All stimuli were presented in a simulated soundfield environment under Sennheiser HD 215 headphones. Knowles Electronic Mannequin for Auditory Research (KEMAR) head-related transfer functions (HRTFs) were used to create a virtual sound field environment. There are four separate listening conditions as seen in Figure 1. In all four conditions, the target sentences were presented from the front (0° azimuth). There are two separate voice conditions, Same Voice and Different Voices. Same Voice means that the background speakers and the target speaker are the same voice. Different Voices means that the background speakers are different voices form that of the target speaker.

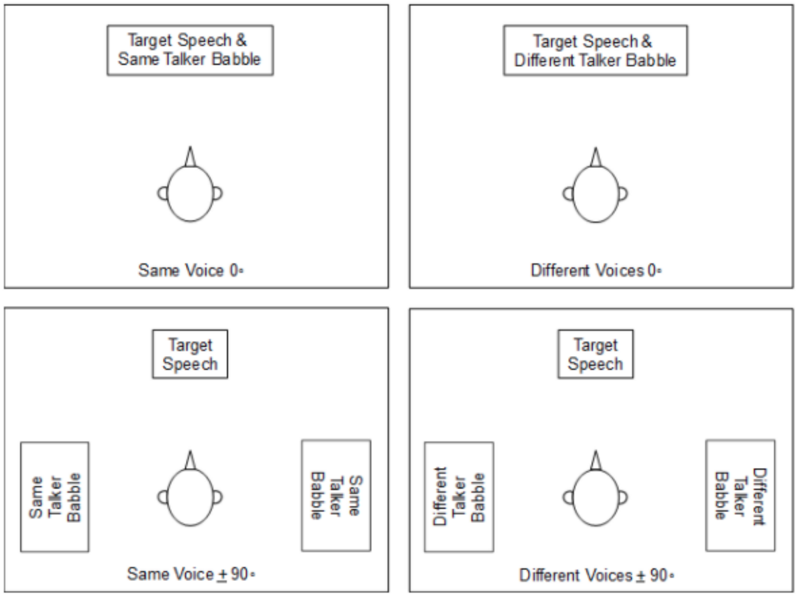

Figure 1. Simulated soundfield environment for the LiSN-S test conditions.

There are also two directional conditions, 0° and ±90°. In the 0° conditions, both the background speakers' and the target speaker's voices are presented from the front (0°), as seen in the top two images of Figure 1. In ±90° conditions, the target speaker's voice is presented from the front but the background speakers are presented from the left (-90°) and right (+90°).

In the first condition, Different Voices ±90°, the background speech consists of voices which are different than the target speech and the background speech is presented from the left (-90°) and right (+90°). In the second condition, Same Voice ±90°, the background and target speakers have the same voice and the background voices are presented from the right and left. In the third condition, Different Voices 0°, the background speakers are different voices than the target speaker and all voices are presented from the front. In the fourth and final condition, Same Voice 0°, all the speakers have the same voice and all the stimuli are presented from the front.

During a test run, the SNR was adaptively adjusted based upon the responses of the participants. The background voices throughout all conditions were presented at 55 dB SPL. The level of the target speech was initially presented at 62 dB SPL and was decreased in 4 dB steps until the first time a participant responded with less than 50% intelligibility of the words. After the first reversal, the level of the target speech was adjusted in 2 dB steps until the completion of the test run. A SRT is defined as the SNR that yields 50% intelligibility of the words. A SRT was found for each of the four listening conditions. Using the SRTs, three derived measures were calculated as follows:

- Spatial Advantage = Same Voice 0° threshold minus Same Voice ±90° threshold
- Talker Advantage = Same Voice 0° threshold minus Different Voices 0° threshold
- Total Advantage = Same Voice 0° threshold minus Different Voice ±90° threshold

Each derived measure "serves to minimize the influence of higher-order language, learning and communication skills on test performance" (Cameron et al., 2009).

After the LiSN-S was administered, a MANOVA was conducted to compare the four listening conditions between the two dialect groups. For all the data combined, a linear mixed model was conducted to investigate the effects of masker talker (same vs. different) and spatial separation of the stimuli (0° vs. ±90°) on LiSN-S thresholds.

Experiment 2: Pure-tone Threshold Averages

After all participants underwent pure-tone testing using the Hughson-Westlake protocol, five different pure-tone averages (PTAs) were calculated: three-frequency PTA (0.5, 1, 2 kHz), four-frequency PTA (0.5, 1, 2, 4 kHz), high-frequency PTA (3, 4, 6 kHz), low-frequency PTA (0.25, 0.5, 1, 2 kHz), and six-frequency PTA (0.5, 1, 2, 3, 4, 6 kHz). PTAs were compared with LiSN-S performances, which were conducted as described above. Pearson correlation coefficients were used to compare participants' LiSN-S performances with the five different PTA measures.

Results

Experiment 1: Dialect

Table 1 shows the LiSN-S descriptive statistics for both dialect groups as well as the differences between the groups. For the Different Voices ±90° condition, the average threshold for Southern speakers was -15.3 dB SNR and the average threshold for non-Southern speakers was -15.7 dB SNR. This means that on average, the subjects in both dialect groups recognized half of the words when the target sentences were presented around 15 dB below the level of background speech. For the Same Voice ±90° condition, the average threshold for Southern speakers was -14.2 dB SNR and the average threshold for non-Southern speakers was -14.3 dB SNR. This means that on average, the subjects in both groups recognized half of the words when the target sentence was presented approximately 14 dB below the level of the background speech. For the Different Voices 0° condition, the average threshold for Southern speakers was -10.1 dB SNR and the average threshold for non-Southern speakers was -10.8 dB SNR. In the last condition, Same Voice 0°, the average threshold for Southern speakers was -1.3 dB SNR and the average threshold for non-Southern speakers was -1.7 dB SNR.

Table 1. LiSN-S Descriptive Statistics and Group Differences for Dialect Groups

Group	Variable	Mean	SD	Min	Max	Range	Difference	*p*
Southern	Different ±90°	-15.3	1.52	-18.6	-11.9	6.7	0.4	0.389
Non-Southern	(dB SNR)	-15.7	1.44	-17.8	-11.8	6.0		
Southern	Same ±90°	-14.2	1.59	-17.4	-10.8	6.6	0.1	0.750
Non-Southern	(dB SNR)	-14.3	1.87	-17.5	-11.4	6.1		
Southern	Different 0°	-10.1	2.44	-13.0	-2.5	10.5	0.7	0.385
Non-Southern	(dB SNR)	-10.8	3.01	-15.8	-4.4	11.4		
Southern	Same 0°	-1.3	0.89	-2.8	0.4	3.2	0.4	0.157
Non-Southern	(dB SNR)	-1.7	0.90	-3.5	0.4	3.9		
Southern	Spatial	12.6	2.20	8.1	17.6	9.5	0.3	0.659
Non-Southern	Advantage	12.3	2.38	5.1	15.8	10.7		
Southern	Talker	9.0	2.45	1.6	12.1	10.5	-0.5	0.463
Non-Southern	Advantage	9.5	2.47	4.8	13.5	8.7		
Southern	Total	14.0	1.63	10.2	17.8	7.6	0.0	0.976
Non-Southern	Advantage	14.0	1.57	10.4	16.0	5.6		

As for LiSN-S derived measures (Table 1), the average Spatial Advantage for Southern speakers was 12.6 dB. This means that on average for the Southern group, there was a 12.6 dB improvement in thresholds from the Same Voice ±90° conditions to the Same Voice 0° conditions. The average Spatial Advantage for non-Southern speakers was 12.3 dB. This means that on average for the non-Southern group there was a 12.3 dB improvement in thresholds from the Same Voice ±90° condition to the Same Voice 0° condition. The average Talker Advantage for Southern speakers was 9.0 dB, meaning that for these subjects there was a 9.0 dB improvement from the Same Voice 0° condition to the Different Voice 0° condition. The average Talker Advantage for non-Southern speakers was 9.5 dB. This means that on average for non-Southern speakers, there was a 9.5 dB improvement from the Same Voice 0° condition to the Different Voice 0°

condition. Lastly, the average Total Advantage was 14.0 dB for Southern speakers. This means that for Southern subjects, there was a 14.0 dB improvement from the Same Voice 0° condition to the Different Voices ±90° condition. The average Total Advantage was 14.0 dB for the non-Southern group. This means that for non-Southern subjects, there was a 14.0 dB improvement from the Same Voice 0° condition to the Different Voices ±90° condition.

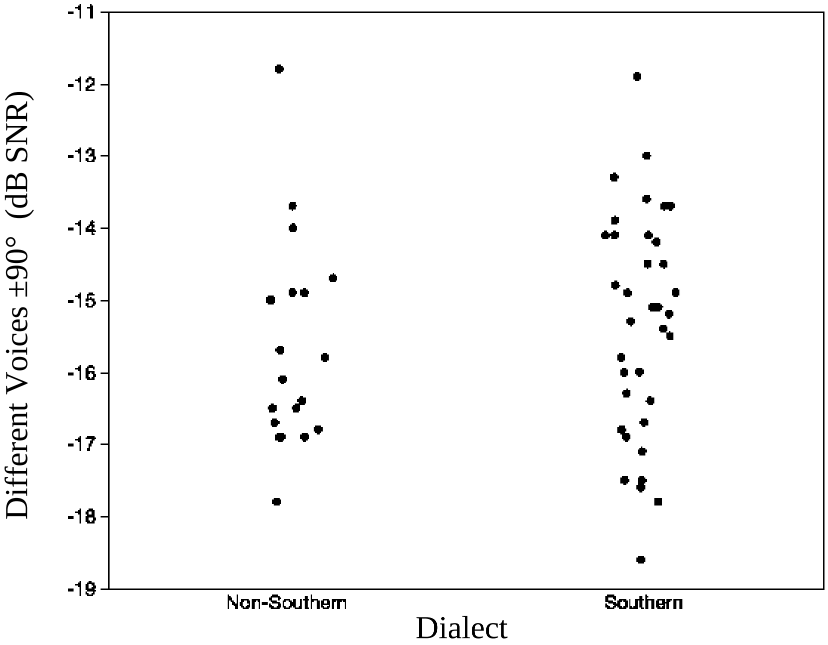

Figure 2. LiSN-S Performance Differences Between Dialect Groups

A MANOVA was conducted to compare LiSN-S results for Southern to non-Southern dialect speakers. No significant group differences were found for the LiSN-S conditions and derived measures. The differences between groups were all small, less than 1 dB. Figure 2 shows an example of the spread in the data found for the non-Southern and Southern dialect groups for the Different Voices ±90° condition. Therefore, the hypothesis that self-identified Southern speakers would perform significantly poorer on the LiSN-S than non-Southern speakers is rejected.

Experiment 2: Pure-tone Threshold Averages

For all data combined, the linear mixed model analysis revealed that both main effects and their interaction were statistically significant (F values are 638.45 for masker talker, 1978.71 for spatial separation of the stimuli, and 368.13 for interaction; all *p*-values < 0.0001). The interaction between voice condition (same vs. different voices) and spatial separation of the target speech and masker stimuli is presented in Figure 3.

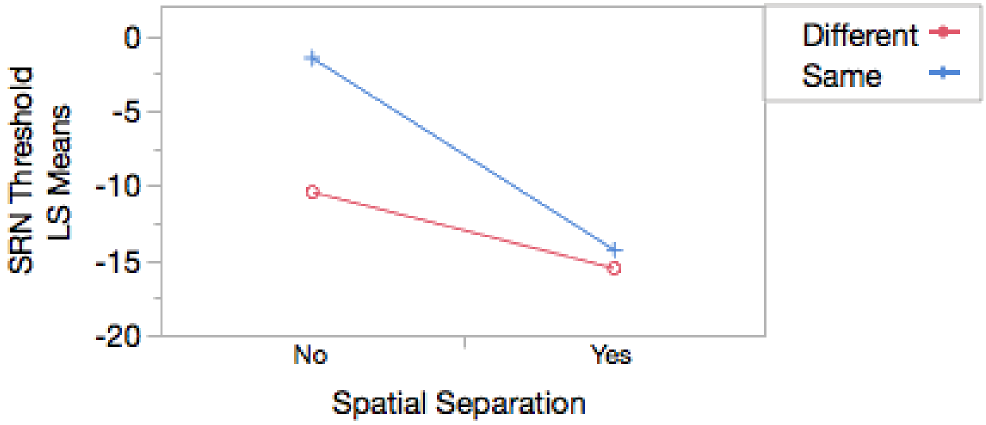

Figure 3. Interaction plot of the same and different voice LiSN-S results and the spatial separation of the target and masker stimuli.

Table 2 shows the descriptive statistics for all LiSN-S performances and derived measures where all participants are combined. The mean SRT for Different Voices ±90° was -15.4 dB SNR (SD=1.50) with a range of 6.8 dB. The mean SRT for Same Voice ±90° was similar, -14.2 dB SNR (SD=1.67) with a range of 6.7 dB. The mean SRT for Different Voices 0° was a bit higher, -10.4 dB SNR (SD=2.65) with a much larger range of 13.3. The mean SRT for the Same Voice 0° condition was much higher, -1.4 dB SNR (SD=0.90) with a very small range of 3.9 dB. For the derived measures, the average

Spatial Advantage was 12.5 dB (SD=2.25) with a rather large rage of 12.5 dB. The average Talker Advantage was lower, 9.2 dB (SD=2.45) with a similar range of 11.9 dB. Finally, the average Total Advantage was the highest, 14.0 dB (SD=1.59) with the smallest range of 7.6 dB.

Table 2. Descriptive Statistics for LiSN-S thresholds for all participants (N=56)

Variable	Mean	SD	Minimum	Maximum	Range
Different Voices ±90° (dB SNR)	-15.4	1.50	-18.6	-11.8	6.8
Same Voice ±90° (dB SNR)	-14.2	1.67	-17.5	-10.8	6.7
Different Voices 0° (dB SNR)	-10.4	2.67	-15.8	-2.5	13.3
Same Voice 0° (dB SNR)	-1.4	0.90	-3.5	0.4	3.9
Spatial Advantage (dB)	12.5	2.25	5.1	17.6	12.5
Talker Advantage (dB)	9.2	2.45	1.6	13.5	11.9
Total Advantage (dB)	14.0	1.59	10.2	17.8	7.6

Pearson Correlation Coefficients were analyzed to determine the relationship between PTA measures vs. LiSN-S thresholds and derived measures. Results are shown in Table 3. Significant positive correlations were found between PTA measures vs. Different Voices ±90° and Same Voice ±90° thresholds. No significant correlations were found between PTA measures vs. the thresholds for the 0° conditions. The strongest correlations were found between four-frequency PTA vs. the Different Voices ±90° (r=0.3969 $p<0.01$) and Same Voice ±90° thresholds (r=0.3905, $p<0.01$). The scatterplots in Figure 4 show the relationships between the LiSN-S SRTs vs. the four-Frequency PTA.

Table 3. Pearson Correlation Coefficients for PTA measures vs. LiSN-S thresholds (significant correlations in bold; *p*-values in parentheses)

PTA Measures	Different ±90°	Same ±90°	Different 0°	Same 0°
Three-Frequency PTA (0.5, 1, 2 kHz)	**0.3289** **(0.0133)**	**0.3602** **(0.0064)**	0.1670 (0.2185)	-0.1344 (0.3232)
Four-Frequency PTA (0.5, 1, 2, 4 kHz)	**0.3969** **(0.0025)**	**0.3905** **(0.0029)**	0.0902 (0.5087)	-0.1434 (0.2917)
High-Frequency PTA (3, 4, 6 kHz)	**0.3476** **(0.0087)**	**0.2980** **(0.0257)**	-0.0416 (0.7606)	-0.1337 (0.3259)
Low-Frequency PTA (0.25, 0.5, 1, 2 kHz)	**0.3088** **(0.0206)**	**0.3728** **(0.0047)**	0.1825 (0.1782)	-0.2432 (0.2923)
Six-Frequency PTA (0.5, 1, 2, 3, 4, 6 kHz)	**0.3967** **(0.0025)**	**0.3967** **(0.0035)**	0.0686 (0.6154)	-0.1570 (0.2477)

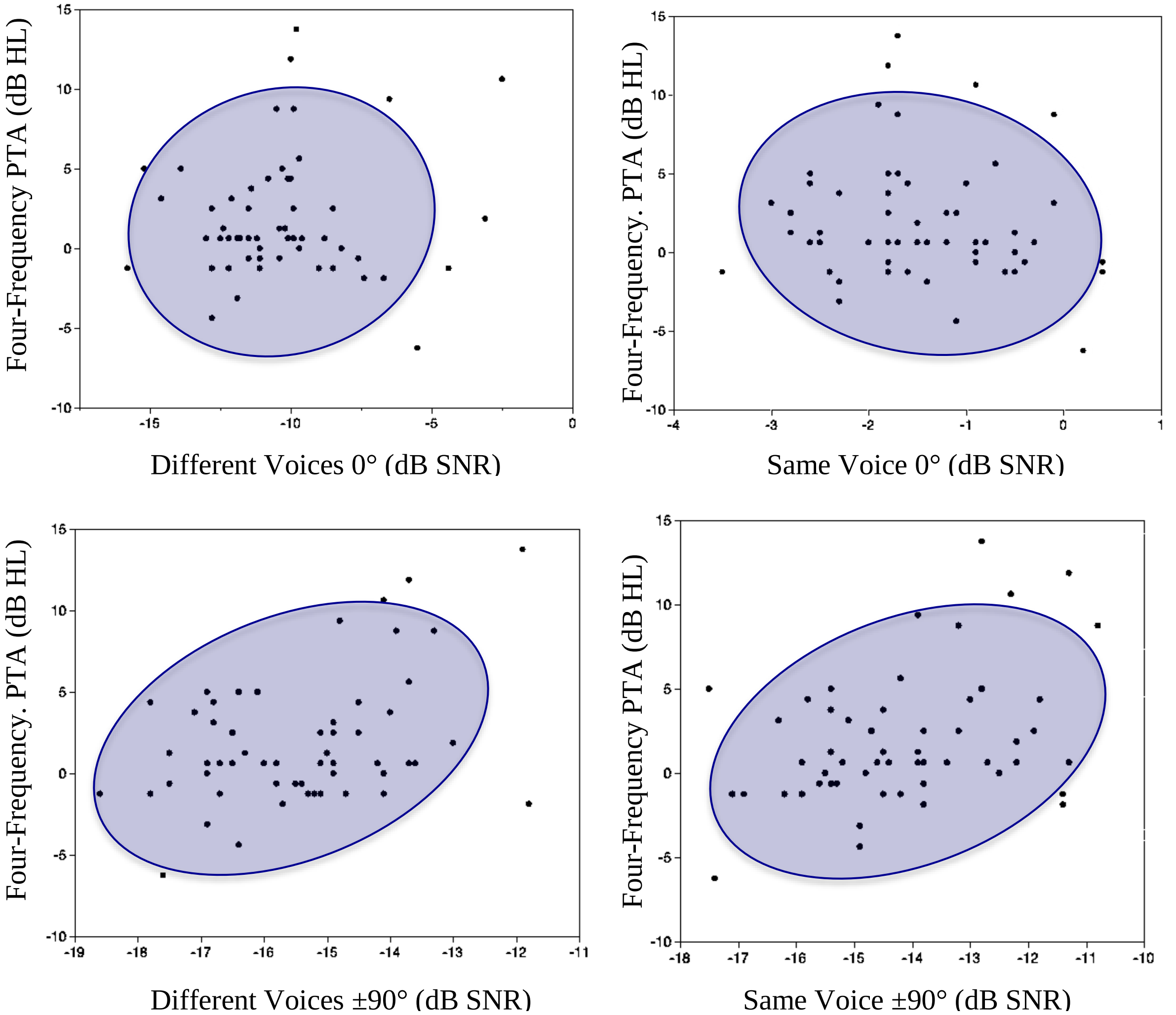

Figure 4. Scatterplots of LiSN-S thresholds vs. four-frequency PTA (0.5, 1, 2, 4 kHz).

Significant positive correlations were also found between the three-frequency PTA vs. Different Voices ±90° thresholds (r=0.3289; *p*<0.05) and between three-frequency PTA vs. Same Voice ±90° thresholds (r=0.3602; *p*<0.01). There were similar significant positive correlations found between the high-frequency PTA vs. Different Voices ±90° thresholds (r=0.3476; *p*<0.01) and between the high-frequency PTA vs. Same Voice ±90° thresholds (r=0.2980; *p*<0.05). Significant positive correlations were also found between low-frequency PTA vs. Different Voices ±90° thresholds (r=0.3088; *p*<0.05) and between low-frequency PTA and Same Voice ±90° thresholds (r=0.3728;

$p<0.01$). Finally, positive correlations were found between six-frequency PTA vs. Different Voices ±90° thresholds (r=0.3967; $p<0.01$) and between six-frequency PTA vs. Same Voice ±90° thresholds (r=0.3967; $p<0.01$).

When comparing PTA measures to LiSN-S derived measures, significant negative correlations were found (Table 4). Significant negative correlations were found between all PTA measures vs. both Spatial Advantage and Total Advantage measures. No significant correlations were found between PTA measures vs. Talker Advantage. The strongest correlations were between four-frequency PTA vs. Total Advantage (r=-0.4540; $p<0.01$) and between six-frequency PTA vs. Total Advantage (r=-0.4635; $p<0.01$).

Table 4. Pearson Correlation Coefficients for PTA vs. LiSN-S advantage measures (significant correlations in bold; *p*-values in parentheses).

PTA Measures	Spatial Advantage	Talker Advantage	Total Advantage
Three-Frequency PTA (0.5, 1, 2 kHz)	**-0.3673** **(0.0054)**	-0.1595 (0.2403)	**-0.3822** **(0.0036)**
Four-Frequency PTA (0.5, 1, 2, 4 kHz)	**-0.4073** **(0.0018)**	-0.0780 (0.5678)	**-0.4540** **(0.0004)**
High-Frequency PTA (3, 4, 6 kHz)	**-03195** **(0.0164)**	0.0303 (0.8244)	**-0.4080** **(0.0018)**
Low-Frequency PTA (0.25, 0.5, 1, 2 kHz)	**(-0.3771)** **(0.0042)**	(-0.1827) (0.1777)	**(-0.3685)** **(0.0052)**
Six-Frequency PTA (0.5,1,2,3,4,6)	**-0.4012** **(0.0042)**	-0.0712 (0.1777)	**-0.4635** **(0.0003)**

As seen in Table 4, significant negative correlations were also found between Spatial Advantage vs. three-frequency PTA (r=-0.3673; $p<0.01$), four-frequency PTA (r=-0.4073; $p<0.01$), high-frequency PTA (r=-0.3195; $p\leq0.05$), low-frequency PTA (r=-0.3771; $p<0.05$), and six-frequency PTA (r=-0.4012; $p<0.01$). There were also significant negative correlations between Total Advantage vs. three-frequency PTA ($r=-0.3822$; $p<0.01$), four-frequency PTA ($r=-0.4540$; $p<0.01$), high-frequency PTA ($r=-0.4080$; $p<0.01$), low-frequency PTA ($r=-0.3685$; $p<0.01$) and six-frequency PTA ($r=-0.4635$; $p<0.01$).

Discussion

In this study, there were two experiments. For the first experiment involving dialect, no significant differences were found between dialect groups for LiSN-S thresholds or derived measures. Previous research by Dawes (2011), Dawes & Bishop, (2011), Shi & Canizales, (2013), and Woods et al, (2004) demonstrated that participants have poorer performances in speech recognition testing when the target speech is presented in a dialect different from the participant's own dialect.

However, none of these studies used the LiSN-S test. It is possible that the LiSN-S is unique and is not as affected by dialect as the previously studied tests. Cameron et al. (2006) developed the LiSN-S test, and stated that the derived measures were specifically designed to cancel out any dialect or language effect. This was not demonstrated since no significant dialect effects were found in the present study.

The present study also compared two versions of North American dialects, which were self-identified as "Southern" and "non-Southern." Since these categories were self-reported by the participant, there is room for error because participants could have misperceptions about themselves and their memories (Stone et al., 1999). The limitations of self-report might have led to the lack of statistically significant results. It is also possible that the Southern and non-Southern dialects are too similar to cause a significant difference between the two groups' performances.

In experiment 2,the present study found significant positive correlations between PTA measures and ±90° conditions. Besser et al. (2015) found a significant positive relationship between PTAs and LiSN-S performances in older adults but a significant negative relationship in younger adults. The present study found a significant positive relationship between PTAs and LiSN-S performances in the ±90° conditions in younger adults, which is direct contrast with this the study by Besser and colleagues.

However, there are some similarities between the results of the present study and those of Besser et al. (2015). Besser et al. (2015) found significant positive correlations between PTA measures vs. Same Voice ±90° and Different Voice ±90° for older adults. The present study found similar correlations in younger adults. The present study revealed a significant negative correlation between PTAs vs. Spatial Advantage and Total Advantage. Besser et al. (2013) found this same relationship between PTA measures and all three derived measures.

There are a few reasons why the relationship between PTA and LiSN-S performances varied between Besser et al. (2015) and the present study. One possibility is that their sample size (n=26) was less than half the number of participants as in the present study (n=56). A smaller sample size may produce tenuous correlations. Also, for the participants in the Besser et al. (2015) study, the average four-frequency PTA was -0.4 dB HL, whereas the average PTA in the present study was 2.12 dB HL. Also, the younger adults in the Besser et al. (2015) study had a much narrower range of hearing thresholds (-6.3 to 6.3 dB HL) compared to the present study (-6.3 to 13.8 dB HL). In addition, the LiSN-S performances in the Besser et al. (2015) study were 1-2 dB better in all conditions than in the present study. The contrasting results may have been due to differences in research participants. The difference between the ages in groups may also have an influence on the results. Besser et al. (2015) had a slightly larger range of young adult ages (18-27) for their participants than the young adults in the present study (18-22). It is also possible that the participants' dialect had an influence as well. Participants in the present study were all from the United States, whereas participants in the Besser et al. (2015) study were Canadian. More research is needed on the potential effect of dialect on LiSN-S performances for Canadian vs. United States dialects.

PTA measures are also used to identify normal vs. elevated hearing thresholds. Recall that overall, significant positive correlations were found between PTA vs. ±90° conditions in the present study. This implies that better hearing ability leads to better LiSN-S performances, meaning that stimulus audibility may significantly impact these performances. Previous research has shown this to be true in regards to individuals with elevated PTAs. Smoorenburg (1992) used PTAs and found that once hearing loss exceeds 10-15 dB HL, the presumed lack of stimulus audibility results in elevated speech reception thresholds in noisy conditions. Vermiglio et al. (2012) evaluated HINT performances between a group of normal hearing adults and a group of adults with various degrees of hearing loss. The results showed significant positive relationships between PTAs and HINT performances in quiet and in noise, indicating that hearing ability impacts these performances.

Tschopp and Zust (1994) investigated the relationship between PTA and speech recognition on the German SPIN test. In a group of individuals with hearing impairment, PTA measures were found to have a significant positive correlation with the SPIN results for all listening conditions ($p<0.05$). Wilson et al. (2007) evaluated the relationship between pure-tone thresholds and speech recognition in noise using the HINT test, the BKB-SIN test, the QuickSIN test, and the WIN Test. Wilson and colleagues compared performances between a group with normal hearing and a group with hearing loss. They reported that the group of listeners with hearing loss had poorer speech recognition performances on all tests than the group with normal hearing, however these differences were not quantified. All of these previous studies indicate that pure-tone sensitivity may affect speech recognition in noise ability. The results of the present study indicate this significance as well for the LiSN-S ±90° conditions.

Limitations

A possible limitation in this study was the selection of dialect groups. Dialect was self-identified by the participants through a questionnaire, and Southern and non-Southern dialects were not defined for the participants. Therefore, there is room for error, as participants may consider themselves Southern or non-Southern speakers for different reasons. For example, where the participants were born or raised, where their parents are from, etc. may impact their decision. Just because a participant was born in the southern United States does not necessarily mean they have a Southern dialect. Some participants expressed confusion and were not sure which dialect to choose. There are a wide variety of Southern dialects in the United States. Future researchers should clearly define the parameters of the specific dialects being studied.

It is also possible that most Southern speakers have access to Standard American English (Edwards, 2003). This dialect is prevalent in news, films, and other media, so most southern Americans probably have daily access to this dialect. Many southern Americans also have access to non-Southern dialects through their community as well, because there are many non-Southern speakers living in the south. This could possibly impact their performance on the LiSN-S, which is presented in a General American English dialect, since they are accustomed to hearing this dialect.

Another limitation was the group of participants who took part in the study. All of the participants except one were female. The age range for these participants was rather narrow (age 18-22) and over 40 of the 56 participants were age 20. Therefore, this sample may not be generalized to older and younger populations.

Conclusions and Future Directions

In conclusion, our first hypothesis was that Southern speakers would have significantly poorer LiSN-S performances than non-Southern speakers. This hypothesis was rejected since there was no statistically significant difference found between dialect and LiSN-S performances. The second hypothesis was that there would be significant positive correlations between PTAs vs. LiSN-S SRTs and between PTAs vs. derived measures. This hypothesis was partially supported with statistically significant positive correlations found between PTAs vs. the SRTs for the LiSN-S ±90° conditions and PTAs vs. Spatial Advantage and Total Advantage. These significant correlations imply that audibility of the stimulus may have an impact on performances for the LiSN-S ±90° conditions even for individuals with normal pure-tone thresholds. Therefore, clinicians should be aware of the influence of hearing sensitivity for LiSN-S ±90° conditions when inferring diagnoses from LiSN-S results. Future research should use well-defined parameters for dialects as well as investigate the effects of other United States and Canadian dialects. Future research should also investigate the effect of stimulus audibility on LiSN-S performances for children and older adults.

References

American Academy of Audiology (AAA). (2010). Practice Guidelines for the diagnosis, treatment, and management of children and adults with central auditory processing disorder. Retrieved from: https://www.audiology.org/publications-resources/document-library/central-auditory-processing-disorder

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108-112.

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., & Pichora-Fuller, M. K. (2015). Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies. *Ear & Heraing, 36*(1), 24-41.

Besser, J. Pichora-Fuller, M. K., & FEsten, J. M. (2013). Advantage of talker differences and spatial separation for speech-on-speech listening in younger and older adults with good audiograms. *Proceedings of Meeting on Acoustics, 19*, 1-7.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research*, *27*(1), 32-48.

Bronkhorst, A. W. & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America, 63*(4),1508-1516.

Cameron, S., Brown, D., Keith, R., Martin, J. Watson, C. & Dillon, H. (2009). Development of the North American Listening in Spatialized Noise-Sentences Test (NA LiSN-S): Sentence equivalence, normative data, and test-retest reliability studies. *Journal of the American Academy of Audiology, 20*(2), 128-146.

Cameron, S. & Dillon, H. (2007). Development of the listening in spatialized noise-sentences test (LISN-S). *Ear and hearing*, *28*(2), 196-211.

Carhart, R. Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America, 45*(3), 694-703.

Cokely, J. A. & Yager, C. R. (1993). Scoring Spanish word-recognition measures. *Ear & Hearing, 14*(6), 395-400.

Dawes, P. (2011). The SCAN-A in testing for auditory processing disorder in a sample of British adults. *International Journal of Audiology, 50*(2), 107-111.

Dawes, P. & Bishop, D.V.M. (2007). The SCAN-C in testing for auditory processing disorder in a sample of British children. *International Journal of Audiology, 46*(12), 780-786.

Edwards, H. T. (2003). *Applied phonetics: the sounds of American English*. United Nations Publications.

Keith, R. W. (1994). *SCAN-A: A test for auditory processing disorders in adolescents and adults*. Psychological Corporation.

Keith, R. W. (1995). Development and standardization of SCAN-A: test of auditory processing disorders in adolescents and adults. *Journal-American Academy of Audiology, 6*, 286-286.

Keith, R. W. (2000). Development and standardization of SCAN-C test for auditory processing disorders in children. *Journal of the American Academy of Audiology*, *11*(8), 438-445.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337-1351.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395-2405.

Kochkin, S. (2002). Marketrak VI: Consumers rate improvements sought in hearing instruments. *Hearing Review*, *9*(11), 18-22.

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 & Hearing, 32*(1), 121-131.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085-1099.

Niquette, P., Arcaroli, J., Revit, L., Parkinson, A., Staller, S., Skinner, M., & Killion, M. (2003). Development of the BKB-SIN Test. In *annual meeting of the American Auditory Society, Scottsdale, AZ.*

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and hearing*, *30*(6), 653-661.

Phonak (2007) *Listening in Spatialized Noise-Sentences Test: Instruction Manual.* Phonak LLC, Warrenville, IL: Author.

Shi, L. & Canizales, L. A. (2013). Dialectal effects on a clinical Spanish word recognition test. *American Journal of Audiology, 22*(1), 74-83.

Shi, L. & Sanchez, D. (2011). The role of word familiarity in Spanish/English bilingual word recognition. *International Journal of Audiology, 50*, 66-76.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (1999). The science of self-report: Implications for research and practice. Psychology Press.

Vermiglio, A. J. (2008). The American English hearing in noise test. *International Journal of Audiology*, *47*(6), 386-387.

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, Hearing in Noise Test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology, 23*(10), 779-788.

Weiss, D. & Dempsey, J. J. (2008). Performance of bilingual speakers on the English and Spanish versions of the Hearing in Noise Test (HINT). *Journal of the American Academy of Audiology, 19*(5), 5-17.

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14*(9), 453-470.

Wilson, R. H., & Burks, C. A. (2005). Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *Journal of rehabilitation research and development, 42*(6), 839.

Woods, A. G., Peña, E. D., & Martin, F. N. (2004). Exploring possible sociocultural bias on the SCAN-C. *American Journal of Audiology*, *13*(2), 173-184.