Summer 2021

Lexical Properties of Perceptual Errors Made by Younger and Older Adults Listening to Speech in Multitalker Babble

Hannah Blythe Vickery

Lexical Properties of Perceptual Errors Made by Younger and Older Adults Listening to Speech in Multitalker Babble

by

Hannah Blythe Vickery

Bachelor of Arts
University of South Carolina, 2019

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Speech - Language Pathology

The Norman J. Arnold School of Public Health

University of South Carolina

2021

Accepted by:

Daniel Fogerty, Director of Thesis

Judy R. Dubno, Reader

Beth McCall, Reader

Krystal Werfel, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

DEDICATION

I would like to dedicate this thesis work to my best friend and to my biggest fan. Momma, thank you for always listening and for teaching me all that I know. Daddy, thanks for supporting me through all I do and encouraging me to do more still. I love you both.

ACKNOWLEDGEMENTS

I would like to extend my sincerest thanks to all of those who have helped make this project possible. First and foremost, I would like to thank, Dr. Daniel Fogerty, my graduate academic advisor, thesis director, and research mentor for his continued patience and guidance throughout this journey. His immense knowledge and professional expertise have shaped and supported me throughout my graduate experience. I would also like to thank Dr. Judy R. Dubno, Dr. Beth McCall, and Dr. Krystal Werfel, for their vital feedback and the sacrifice of their time throughout the duration of this study. Finally, I would like to thank the rest of the faculty and staff in the Communication Sciences and Disorders Department at the University of South Carolina for making my time in this program not only wonderfully challenging, but also joyfully fulfilling.

ABSTRACT

Purpose: The purpose of this error analysis was to analyze the lexical properties of misperceptions made when listening to speech in multitalker babble.

Methods: Twenty young adults with normal hearing (YNH), 20 older adults with normal hearing (ONH), and 22 older adults with hearing impairment (OHI) completed a speech-in-babble task. Participants were asked to repeat the final word in 25 high and 25 low context sentences. On each trial, participants either responded with the correct target word, a misperception error, or skipped the trial response. Misperceptions were compiled and analyzed according to their lexical properties.

Results: Results of this study showed that neighborhood density, the proportion of misperceptions within the same phonological neighborhood as the target, and phonologic and semantic similarity measures all demonstrated significant effects of context. Specifically, the older adult groups (ONH and OHI) responded with words in significantly smaller phonological neighborhoods in low context sentences compared to the target, phonological similarity was found to be greater in low context sentences than high, and semantic similarity was found to closer to the target in high context sentences. In addition, group differences were observed in that the YNH group was more likely to guess, resulting in more misperceptions, while the OHI group was more likely to skip response opportunities. Similarly, group differences were noted where older adults produced misperceptions from smaller phonological neighborhoods than the target in low context sentences and the ONH group, specifically, produced more errors that were less

phonologically similar to the target compared to the YNH group in high context sentences.

Conclusions: All groups performed similarly in regard to overall accuracy in recognizing speech in multitalker babble. However, systematic differences were noted between participant groups in the types of misperceptions made across sentence contexts. These patterns suggest that age, hearing loss, and context affect the lexical properties of misperceptions made when listening to speech in multitalker babble.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Large individual differences in the recognition of speech in babble are often observed among older adults, even when audibility is ensured through appropriate amplification. To assess one potential source of variability, this study sought to examine the patterns of word errors made when younger and older adults listen to sentences in babble. Previous literature has suggested that although older adults make more errors, they may use context better than younger adults to fill in final words in sentences (Pichora-Fuller, Schneider, & Daneman, 1995; Speranza, Daneman, & Schneider, 2000). In this way, older adults' purported use of context may be related to their larger and more diverse lexicon. Conversely, Dubno et. al (2000) showed that younger and older adults obtain equal benefit from context when auditory differences are controlled. Therefore, this study seeks to supply additional information about how age and context affects speech perception through the analysis of misperceptions made by younger and older adults. An error analysis will provide specific information about a group's perception by detailing the types of misperceptions made as opposed to a binary coding of accuracy. Thus, when compared to younger adults, perceptual errors by older adults may be less constrained by the target word, due to poorer perceptual encoding related to reduced audibility, but more reflective of the linguistic diversity of their lexicon.

The Speech Perception in Noise (SPIN) Test

Speech understanding arises from a culmination of interacting factors including word recognition through acoustic-phonetic processing, sentence context predictability, vocabulary acquisition, and word meaning. When one or more of these factors does not meet the required thresholds for effective speech understanding, speech perception decreases. Similarly, compounding factors, such as a decrease in the signal-to-noise ratio (SNR) or the introduction of hearing loss, may result in degraded speech perception and understanding. Both of these factors – lexical processing and declines in audibility due to noise and/or hearing loss – results in a greater number of perceptual errors, which may ultimately have consequences as to the lexical characteristics of the misperceptions. One common measure developed to assess speech perception is the Revised Speech Perception in Noise Test (R-SPIN; Bilger et al., 1984). Various speech in noise tests, such as R-SPIN, have become widely utilized in research studies as a means of measuring speech perception and recognition in noisy environments (Billings, 2015; Cherry 1953; Miller, 1947). Measurements of speech recognition are important as scores may indicate how individuals might function with normal and/or reduced hearing acuity in various noisy environments. Research has previously focused on determining the reliability of R-SPIN measurements for clinical practice (Bilger et al., 1984) and the general clinical utility of testing speech in noise (Taylor, 2003), yet the results obtained have yet to specify difficulties subjects may encounter based on the individual characteristics of test items. Consequently, this study seeks to analyze the nature of errors made when individuals listen to speech in babble. By assessing the characteristics of the errors compared to the correct responses, we hope to glean information about how listeners utilize acoustic features, the meaning and context of an utterance, and the word frequency of the target to process information and to perceive speech.

Error Coding

Most commonly, speech recognition tasks define accuracy as a simple binary coding of correct and incorrect responses. However, error analysis offers a richer understanding of what difficulties subjects may experience when completing speech in noise tasks, such as R-SPIN, as it may help to explain partial information the listener used. Error coding sorts data so that patterns of similarity between the target and response emerge and give context to why a subject may struggle with a certain test item, as well as, potentially, the perceptual and cognitive information recruited during the response. Most literature has, to this point, focused on analysis of perceptual errors made while repeating nonsense syllables (Phatak & Allen, 2007; Zaar & Dau, 2015). Examining speech recognition errors in meaningful contexts has been recently utilized to compare the types of speech error patterns between native and non-native English speakers (Zinser et al., 2019) as well as in variable noisy environments (Smith & Fogerty, 2017). According to Smith and Fogerty (2017), errors categorized based on their severity and distance from the target can create a heightened sensitivity to the individual differences within the speech perception and production mechanisms of the listener. Fine-tuned analysis of recognition errors may explain why individuals produce such errors.

The Role of Context in Speech Understanding

According to Pichora-Fuller (2008), word understanding is a complex process requiring both high- and low-level cognitive processing. First, the listener must map the lexical features of a word to the appropriate semantic meaning; once complete, the listener then expands their understanding to include other relevant words within an utterance. This process of utilizing surrounding words within an utterance to shape and clarify the meaning of a specific word is described, in this study, as context. In this way, studies found that utilization of context maintained the integrity of some communications, even in the presence of adverse listening conditions, when only glimpses or partial speech information was accessible to the listener (Hutchinson, 1989; Pichora-Fuller, 2008). In an effort to investigate the influence of contextual cues, previous literature explored the benefit subjects derive from utilizing context-dependent cues in determining the final words of high-context sentences (Dubno, Ahlstrom, & Horwitz, 2000; Hutchinson, 1989; Nittrouer & Boothroyd, 1990; Pichora-Fuller, Schneider, & Daneman, 1995; Speranza, Daneman, & Schneider, 2000). Yet a debate remains concerning whether or not older adults are better, worse, or comparable to younger adults when utilizing context during speech in noise tasks. Some research purports that older adults outperform younger adults when context is preserved (Pichora-Fuller, Schneider, & Daneman, 1995; Speranza, Daneman, & Schneider, 2000). On the other hand, similar studies show little or no difference between younger and older listeners (Dubno, Ahlstrom, & Horwitz, 2000; Hutchinson, 1989; Nittrouer & Boothroyd, 1990). The reason for these conflicting conclusions stem from utilization of context and the effects of increased auditory thresholds in the presence of varying noise levels between age groups. The current study seeks to further extend these investigations by analyzing errors made by younger and older adults listening to speech in low and high context sentences. The acoustic and lexical characteristics of these errors may define potential variability in how individuals use context during speech recognition.

The Current Study

In this study, three groups of participants: younger normal-hearing adults (YNH), older normal-hearing adults (ONH), and older adults with hearing impairment (OHI) listened to low-context and high-context sentences in multitalker babble and identified the final word of each sentence. Each participant was encouraged to guess what they heard, even if it did not make sense to them. All responses were scored and then transcribed by trained raters. Errors were extracted and compiled from each listener. Lastly, a detailed analysis of the word frequency, neighborhood density, phonological similarity, and semantic similarity between the target and the error was computed.

Analyses of response errors, referred to here as "misperceptions," in order to distinguish from skipped trial errors, aid in detailing the degree to which an individual's existing knowledge is recruited during the misperception based on partial lexical, semantic, or phonological information processing. Therefore, to assess misperceptions made by younger and older adults, the word frequency of the misperceptions - as well as the phonological and semantic factors of each misperception - are analyzed. This study will determine if there is a varying reliance between older and younger listeners on perceptual and lexical knowledge when listening to speech in babble, as the total amount of vocabulary acquired by either group may be vastly different due to age differences. Furthermore, the misperception analysis of speech in babble will predict how these three factors contribute to speech recognition variability. The three research questions that this study seeks to answer are as follows.

1) Investigators Hartshorne and Germine (2015) found that adults continue to build vocabulary well beyond the typical age of retirement, thus implying that older

adults have larger lexicons compared to those of their younger counterparts. As adults age, they continue to acquire lexical knowledge, and as a result, a larger quantity of words with low word frequencies. In addition, studies have shown that adults' verbal knowledge continues to increase as adults age, despite potential decreases in other cognitive functions (Drag & Bieliauskas, 2010; Park et. al, 2002). To test the premise that the ONH and the OHI groups did, in fact, have larger lexicons, the investigators utilized the Peabody Picture Vocabulary Test, Fourth Edition (PPVT - 4; Dunn & Dunn, 2007), as a measure of lexicon size. Older adults' errors, therefore, may be more likely to contain words that occur with less frequency in the English language compared to younger adults, indicative of their larger and more diverse lexicon. Similarly, given older adults' larger lexicons, the investigators expected their errors to have a smaller neighborhood density compared to younger adults, as older adults' lexicons contain more words with fewer neighbors, despite the context of the sentence.

2) Because response errors may occur due to perceptual confusions with similar phonological characteristics of the target word, we hypothesize that the YNH group and the ONH group will produce errors closer to the phonological neighborhood of the target word in high context sentences, where meaning may be inferred through context. In contrast, the OHI group is expected to produce less phonologically similar errors to the target word in both low and high contexts because they may not have complete auditory access to the target.

3) As semantic context from earlier words in a sentence can increase the predictability of final words (Taylor, 1953), we hypothesize that response errors for all groups will be semantically similar to the target word in high-probability sentences, but

not in low-probability sentences that remove context dependency. Furthermore, we hypothesize that the OHI group will produce more semantically similar errors than the YNH and the ONH groups. Because the phonological properties of the target are degraded due to hearing loss, the contextual cues may remain and therefore, maintain the semantic gist of the sentence leading to semantically similar errors , which are also unconstrained by the phonological characteristics of the target.

CHAPTER 2

METHODS

Subjects

This research study analyzed a dataset previously collected from a larger investigation. Thus, the following information outlines subjects that recruited as part of the larger study. This study consisted of a total of 62 participants. Participants were divided into three groups: 20 younger adults with normal hearing (YNH) (aged 19 through 28 years; 23 years mean) (16 females; 4 males), 20 older adults with normal hearing (ONH) (aged 60 through 74 years; 67 years mean) (17 females; 3 males), and 22 older adults with hearing impairment (OHI) (aged 60 through 85 years; 71 years mean) (14 females; 8 males). Inclusion criteria consisted of audiometric thresholds for YNH adults at or below 20 dB HL for octave frequencies from 0.25 to 8 kHz, and for ONH adults at or below 25 dB HL at 4 kHz and below. OHI adults met inclusionary criteria with thresholds at or below 55 dB HL for octave frequencies from 0.25 to 4 kHz. All participants were self-reported native speakers of American English.

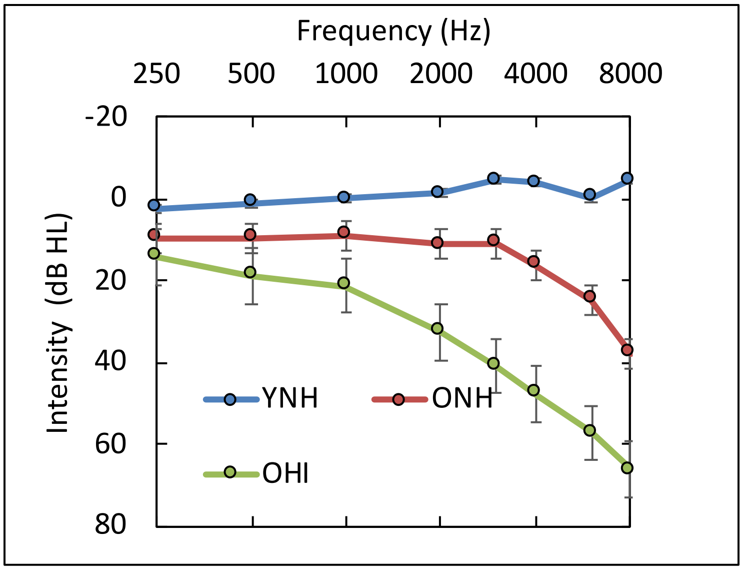

Figure 2.1. Average Audiologic Thresholds by Group

Each participant's vision was screened using the Snellen Visual Acuity Test, with corrective lenses as needed, to ensure that visual acuity was sufficient for perception of visual directions and computer functions. Participants completed a standardized test of receptive vocabulary knowledge, the PPVT - 4 (Dunn & Dunn, 2007): YNH (209.5 mean score), ONH (214.45 mean score), and OHI (218.05 mean score). An independent samples t-test compared PPVT - 4 scores for the YNH group to the combined group of older adults (ONH and OHI). Results showed that the older adults scored significantly higher on the PPVT - 4 compared to their younger adult counterparts, $t(34) = -2.887$, $p = .007$. Note that both the ONH and OHI group did not perform significantly different than the other on the PPVT - 4, $t(31) = -1.389$, $p = .09$.

In addition, both the ONH and OHI groups completed the Mini Mental State Exam, with inclusionary criteria of a score of at least 25, to rule out any cognitive decline (Folstein & McHugh, 1975). Five of the 20 OHI participants reported using hearing aids. Younger adults were recruited at the University of South Carolina and older participants were recruited from the Medical University of South Carolina. All participants provided written informed consent prior to initiation of the experimental procedure. All test procedures received institutional review board approval from the University of South Carolina and the Medical University of South Carolina prior to data collection.

Design & Stimuli Processing

Twenty-five high probability and twenty-five low probability R-SPIN sentences were used as the experimental stimuli. Sentences were presented in the presence of 12-talker babble at 0 dB signal-to-noise ratio (SNR). The speech, prior to spectral shaping, was calibrated to be presented at 70 dB SPL. All listeners received spectral shaping of stimuli based on their individual audiograms to ensure speech sensation levels of at least 15 dB up to 4 kHz. In actuality, groups with normal hearing (YNH and ONH) received little change via amplification, while the average speech level for the OHI group, post spectral shaping, averaged 82 dB SPL. The spectral shaping procedures outlined here are further detailed in Fogerty et. al, 2020. Following spectral shaping, the stimuli were passed through a linear phase, low-pass filtered, finite-impulse-response, 128th-order filter with a cutoff of 5.623 kHz. This specific filter was designed to ensure that each group received speech within a bandwidth at which audibility could be ensured for the OHI group.

Test Procedure

Participants were seated in a sound-attenuated booth at an individual computer station. All auditory stimuli were presented at a sampling rate of 48.828 kHz through one of a pair of Sennheiser HDA 200 headphones following a TDT System III digital-to-analog processor (RP2/RX6) and headphone buffer (HB7/HB5). Stimuli were presented monoaurally to the right ear unless target sensation levels (i.e., SLs) were better achieved utilizing the left ear. During completion of speech in noise sentences, the responses were audio-recorded for live or off-line scoring and analysis by trained scorers. A rigid scoring procedure was implemented that required a verbatim response for each of the final target words (i.e., no added or missing prefixes, suffixes, or tense markings). No feedback was given to participants and each participant was encouraged to make their best guess. Once scored, a trained research assistant organized the raw response data by the transcribed sentence, the targeted final word, and the orthographic transcription of the participant's response. Skipped trials where the participant did not make a response attempt were notated. The total error proportion was calculated for each participant and compared to the online accuracy scoring to ensure scoring agreement.

Data Analysis

Word Frequency and Density Analysis

The Irvine Phonetic Online Dictionary (IPhOD) is a database composed of English words and pseudowords developed for the purpose of researching speech perception and production through various measures of phonemes (Vaden et al., 2009). This study utilized IPhOD to analyze the word frequencies (SFreq, SUBTLEXus word frequency) (Brysbaert & New, 2009) of both the final target word in each of the 50 experimental sentences as well as the word frequency of the misperception. Each unique word was run through the IPhOD calculator to produce the word's frequency in American English based on the SUBTLEXus database. All raw values were log transformed. The difference was then taken between the target SFreq and the SFreq of the participant's unique misperception to determine the word frequency difference score.

Similarly, the neighborhood density of each unique misperception and target word was measured utilizing the IPhOD database to find the neighborhood density (unsDENS) score. Each unique word was run through the IPhOD calculator to produce the word's neighborhood density in American English based on the SUBTLEXus database. The difference was then taken between the target's unsDENS score and the unsDENS score of the participant's unique misperception to determine the neighborhood density difference score.

Phonological Analysis

To measure the phonemic similarity between each unique target-error pairing, the edit distance, also referred to as the Levenshtein distance, was calculated. First, each target and error were phonetically transcribed via the CELEX database. Next, the transcriptions were compared to calculate the edit distance by determining the minimum number of phonemic omissions, additions, or substitutions that differed between the misperception in relation to the target word. Because the Levenshtein distance is the most common metric used to calculate edit distance, this metric was computed (Navarro, 2001). Therefore, every participant's misperceptions during the 25 high and the 25 low probability sentences received a unique score. Average scores for this phonemic similarity measurement were calculated for the high and low probability sentences, resulting in a total of two scores for each participant.

Next, the second metric utilized to analyze the phonological properties of the misperceptions was calculating the proportion of misperceptions that were within the neighborhood of the target (i.e., had an edit distance equal to 1). After all of the edit distances were compiled, as outlined in the methodology above, the number of misperceptions with an edit distance of 1 was divided by the total number of misperceptions. These values were then taken as the neighborhood proportion value.

Semantic Analysis

The SEMantic simILARity software toolkit, (i.e., SEMILAR) is a corpus of various semantic similarity methods (Rus et al., 2013) and was utilized to determine how closely the misperception compared to that of the target word in terms of semantic similarity. Each of the unique target words and errors were calculated utilizing two types of semantic comparison: path-based (Leacock, Chodorow, & Miller 1998) and information content-based (Jiang & Conrath 1997). Specifically, a similarity score was computed for every response by each participant utilizing both methods of comparison, resulting in two unique similarity scores for each participant. The average was taken for each high and low probability sentence based on both semantic similarity algorithms for each group, resulting in a total of four scores for each participant (i.e., two scores for the path-based algorithm and two for the information content-based algorithm).

Statistical Analysis Model

Once misperceptions were organized according to the methodological protocols outlined above, a mixed model statistical design was implemented, consisting of a within-subjects variable (effect of context) and a between-subjects variable (effect of group). A 2 (context) x 3 (group) analysis of variance (ANOVA) test was completed to determine the main effect of context, group, and potential interactions between the two for each metric discussed. When needed to further explain the results of the ANOVA testing, t-tests were conducted.

CHAPTER 3

RESULTS

Error Analysis Summary

The present study aimed to determine phonemic and semantic properties of speech that may contribute to an individual's ability to recognize speech in babble by examining response errors. Data were compiled from each participant in the three subject groups and organized according to total errors (both skips and misperceptions), the number of skips, and the number of misperceptions during high and low context sentences. Three mixed model 2x3 ANOVA tests were then utilized to examine total errors, skips only, and misperceptions only. This analysis identifies which group performed better or worse compared to the other groups and provides an initial summary of speech recognition.

Total Errors

First, total errors (i.e., the sum of misperceptions and total skips per participant) were calculated to present a summary measure of performance by each participant group. As demonstrated in Figure 3.1 below, more total errors were made in low context sentences than in high context sentences. The mean of the YNH, ONH, and OHI groups' total errors were 13.2, 12.8, and 11.0, respectively, in low context sentences. In comparison, the mean of the YNH, ONH, and OHI groups' total errors were 5.5, 3.6, and 3.0 for high context sentences. In this way, Figure 3.1 visually represents how most errors were produced during low context sentences, where context dependency is absent.

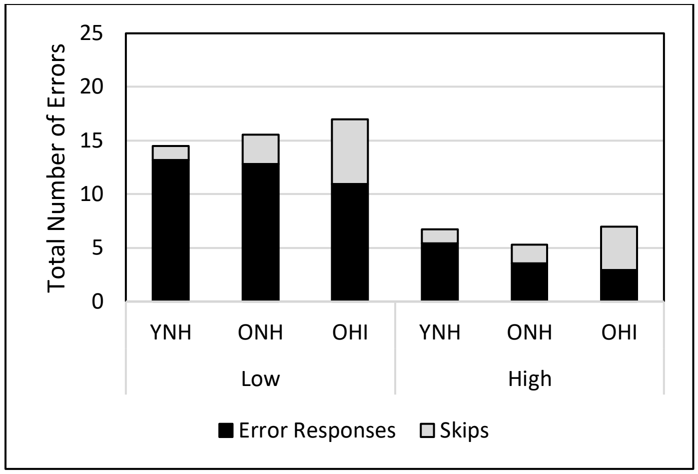

Figure 3.1 Total Number of Errors by Group
This bar graph represents the average total errors made by each participant group for high and low context sentences. The total error calculation is summed by the total average of misperceptions (black portion) and the total average of skips (grey portion).

ANOVA testing showed a significant main effect of context, $F(1, 59) = 609.799$, $p < .001$, $\eta_p^2 = .912$, where low context sentences had the most total errors. The main effect of group was determined to be non-significant, $F(2,59) = 1.284$, $p = .285$, showing that total errors did not significantly differ between participant groups. A significant interaction between context and group was noted, however, $F(2, 59) = 4.338$, $p = .017$, $\eta_p^2 = .128$. As summarized in the table below, six independent-samples t-tests were performed to compare all three groups to explain the interaction. Results suggest a significant difference between YNH and OHI for total errors in low context sentences where the OHI group had more total errors compared to the YNH group. Overall, the analysis of total errors indicates similar speech recognition of final words for all three groups, with better recognition (i.e., fewer total errors) in high context sentences. The next analyses partition the total errors into misperceptions and skips.

Table 3.1 Total Errors Independent Samples t-tests
This table reports the results of the independent samples t-tests performed on each of the groups in both high and low context sentences.

Groups Compared	Context	*t*	df	p
YNH vs. ONH	HIGH	1.428	38	0.162
YNH vs. ONH	LOW	-1.25	38	0.219
YNH vs. OHI	HIGH	-0.148	40	0.883
YNH vs. OHI	LOW	-2.184	36.609	0.035
ONH vs. OHI	HIGH	-1.31	32.971	0.199
ONH vs. OHI	LOW	-1.309	33.646	0.2

Skips Only

Next, skipped trials (i.e., trials with no response) were analyzed (grey bars in figure 3.1). Results demonstrated a significant main effect of context, where low context sentences had more skips than high context sentences, $F(1,59) = 8.149$, $p = .006$, $\eta_p^2 = .121$. A main effect of group was also noted, $F(2, 59) = 7.502$, $p = .001$, $\eta_p^2 = .086$. No significant interaction was observed, $p = .071$. Six post-hoc independent-samples t-tests were completed to compare groups. The results of each of these tests are displayed in the table below. The results indicate that OHI listeners made significantly more skips than YNH and ONH listeners in both low and high context sentences.

Table 3.2 Skips Only Independent Samples t-tests
This table reports the results of the independent samples t-tests performed on each of the groups in both high and low context sentences.

Groups Compared	Context	*t*	df	p
YNH vs. ONH	HIGH	-0.886	38	0.381
YNH vs. ONH	LOW	-1.802	25.369	0.083
YNH vs. OHI	HIGH	-2.797	25.967	0.01
YNH vs. OHI	LOW	-3.636	23.521	0.001
ONH vs. OHI	HIGH	-2.228	33.716	0.033
ONH vs. OHI	LOW	-2.263	28.567	0.031

Misperceptions Only

Lastly, misperceptions were tallied for each individual participant (black bars in Figure 3.1). ANOVA results indicated a significant main effect of context, with more misperceptions in low context sentences, $F(1,59) = 466.450$, $p = .000$, $\eta_p^2 = .888$, and a significant effect of group $F(2, 59) = 4.082$, $p = .022$, $\eta_p^2 = .022$. No significant interaction was observed ($p > .05$). Six independent samples t-test were utilized to determine which group(s) demonstrated a significant effect.

Table 3.3 Misperceptions Independent Samples t-tests
This table reports the results of the independent samples t-tests performed on each of the groups in both high and low context sentences.

Groups Compared	Context	*t*	df	p
YNH vs. ONH	HIGH	2.343	38	0.024
YNH vs. ONH	LOW	0.376	38	0.709
YNH vs. OHI	HIGH	2.865	40	0.007
YNH vs. OHI	LOW	2.157	40	0.037
ONH vs. OHI	HIGH	0.79	40	0.434
ONH vs. OHI	LOW	1.598	39.562	0.118

Results of independent samples t-tests demonstrated that the YNH group made significantly more misperceptions in high context sentences compared to the ONH and OHI groups and more misperceptions in low context sentences compared to the OHI group. No differences in the number of misperceptions in either context were observed for ONH and OHI groups.

Summary of Errors

In summary, these three analyses describe the general pattern of total errors made by adults listening to speech in multitalker babble. These results demonstrate that more skips and misperceptions occur in low context sentences, which is to be expected as participants have less contextual cues to inform potential responses. More specifically, these results found that groups do not differ significantly on the total number of errors made. Instead, group differences were noted in the pattern of their errors.

First, total errors were dominated by skips for OHI listeners. Thus, these listeners made less guesses, as more trials were skipped, which resulted in fewer misperceptions overall. Consequentially, YNH listeners made more misperceptions since they tended to guess more words, as opposed to skipping, when they were unsure of what they heard. The ONH group, however, varied in their error patterns, indicating that their performance was dependent on the context or specific measure being assessed. All subsequent analyses only consider properties of the misperceptions and do not include skips.

Word Frequency Difference Results

To determine how the lexical property of word frequency may contribute to the number and/or characteristics of errors made by each group, the word frequency of the misperceptions were analyzed. Since the YNH group has a significantly smaller lexicon compared to the ONH and OHI groups, resulting in the YNH group having fewer words with a low word frequency, we predicted that older adults (ONH and OHI) would respond with more misperceptions of a lower word frequency compared to the YNH group.

To test this hypothesis, the investigators calculated a word frequency difference score between each target word and the corresponding misperception for each participant. The word frequency difference score was calculated by subtracting the word frequency of the target from the word frequency of the misperception (word frequency of misperception - word frequency of the target = word frequency difference score). These word frequency difference values are displayed in Figure 3.2 below.

First, to understand the extent to which each group responded with misperceptions of a higher word frequency compared to the target, one-sample t-tests were performed relative to a difference score of zero. Referring to Figure 3.2, both the grey (high context) and black (low context) boxes are situated above the reference line indicating no difference. This indicates that the groups' overall word frequencies of the misperceptions were higher than the targets'. As seen in Table 3.4, the word frequency difference score for all groups was significant in both high and low context sentences.

Table 3.4 Word Frequency Difference One-Sample t-tests
This table reports the results of the one sample t-tests performed on each of the groups in both high and low context sentences.

Group	Context	*t*	df	p
YNH	HIGH	7.525	19	0.001
	LOW	7.379	19	0.001
ONH	HIGH	4.65	19	0.001
	LOW	6.489	19	0.001
OHI	HIGH	4.843	18	0.001
	LOW	6.138	21	0.001

Next, difference scores were examined across context and group. A 2 (context) x 3 (group) ANOVA showed a significant effect of context where participants were more likely to respond with misperceptions of higher word frequency, compared to the target, in high context sentences, $F(1,56) = 9.524$, $p = .003$, $\eta_p^2 = .145$. This effect of context is visualized in Figure 3.2 below. Here, the grey boxes (high context sentences) are situated above the black boxes (low context sentences), indicating that the misperceptions' word frequencies are higher than the targets'. No significant effect of group, $F(2, 56) = .056$, $p = .945$, or interaction , $F(2, 56) = 1.407$, $p = .253$, was observed. Thus, these results indicate similar word frequency effects for all groups, with greater frequency for misperceptions in high context sentences.

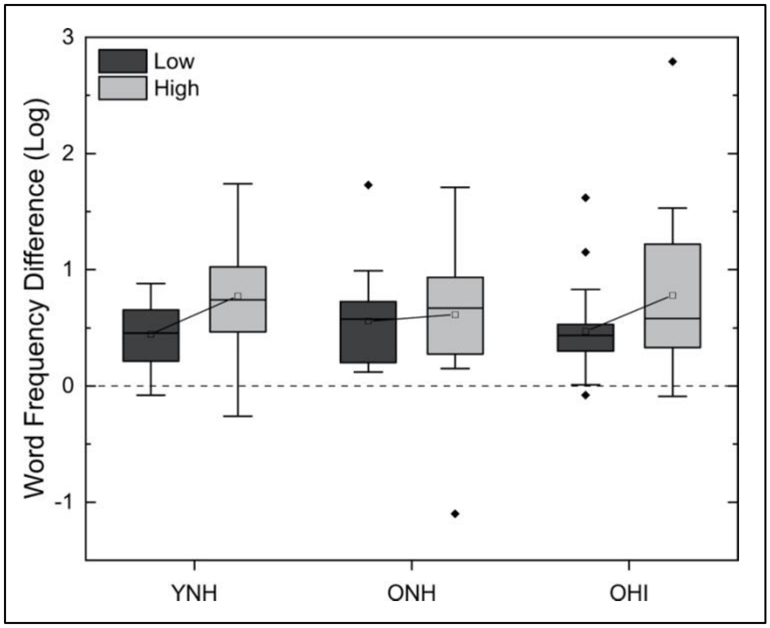

Figure 3.2 Word Frequency Difference Scores by Group
This box and whisker plot visually represents the word frequency difference scores, logarithmically transformed, and divided by high and low context sentences for each participant group. The box demonstrates the interquartile range of the difference scores, while the whiskers demonstrate 1.5x the interquartile range. The box line represents the median value, while the dashed line represents no difference in word frequency between the error and target. Outliers are graphically represented by diamonds, and circles represent the mean of the difference scores.

These results imply that individuals, regardless of age, are more likely to respond with a misperception of a higher word frequency in the English language compared to the target, especially for high context sentences.

Neighborhood Density Difference Results

Like the word frequency difference scoring (reported above), neighborhood density difference scores were calculated by finding the size of the neighborhood for the misperception and the target for all participants across all test items. Difference scores were then calculated by subtracting the neighborhood density score of the target from the neighborhood density scores of the misperception. This metric was chosen to observe whether participants respond more often with errors in smaller or larger phonological neighborhoods. The high confusability of high-density words may suggest that these words have a greater likelihood of being provided as a response when the target is misperceived. Note that this metric does not measure whether or not both the target and the misperception reside in the same phonological neighborhood, but instead, suggests the relative size (i.e. larger or smaller) of the misperception's neighborhood compared to the target's; these values are graphically represented in Figure 3.3.

First, one sample t-tests showed that the YNH group's word density between the target and the misperception were not significant in high context, t(19) = 1.165, *p* = .258, or low context sentences, t(19) -1.410, *p* = .175. The ONH group, t(19) = -2.497, *p* = .022, and the OHI group, t(21) = -3.429, *p* = .003, however, did show a significant difference between the word density of the target and of the misperception in low contexts, but not in high contexts. These results signify that older adults respond with more misperceptions from within smaller phonological neighborhoods (i.e., less confusable) compared to the neighborhoods of the target in low context sentences. This pattern is graphically represented in Figure 3.3. As demonstrated by the figure below, both the ONH and OHI groups show median values (indicated by the box line) and mean values (indicated by the circle) below the reference line in low context sentences.

Table 3.5 Neighborhood Density One-Sample t-tests
This table reports the results of the one sample t-tests performed on each of the groups in both high and low context sentences.

Group	Context	*t*	df	p
YNH	HIGH	1.165	19	0.258
	LOW	-1.41	19	0.175
ONH	HIGH	-1.324	19	0.201
	LOW	-2.497	19	0.022
OHI	HIGH	0.2	18	0.844
	LOW	-3.429	21	0.003

Second, ANOVA testing demonstrated no significant effects of context, F(1,56) = 3.081, p = .085, group, F(2, 56) = 2.415, p = .099, or interaction F(2,56) = 1.366, p = .263. Thus, the relative difference in density between low and high context was not different for any group.

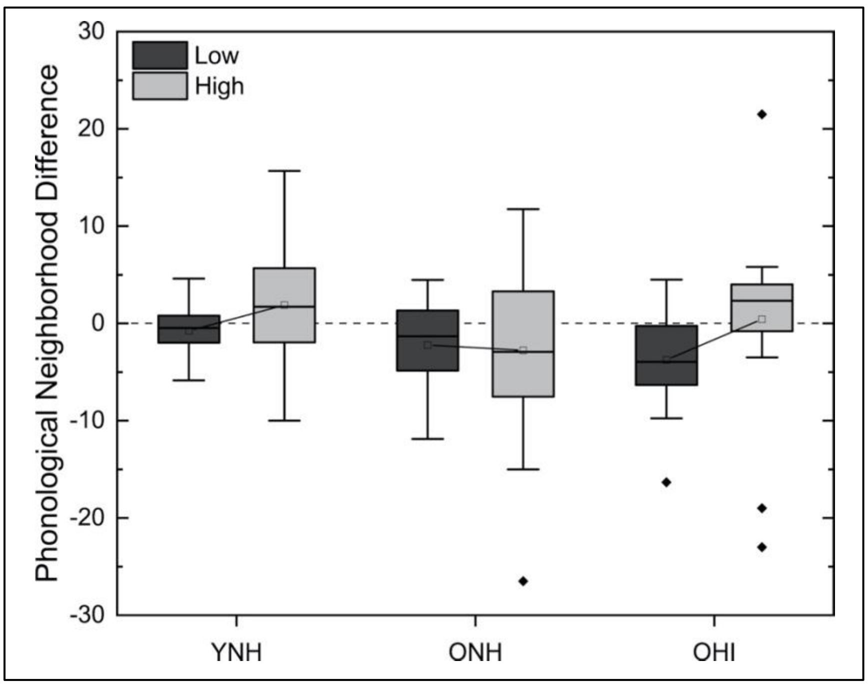

Figure 3.3 Phonological Neighborhood Difference Scores by Group
This box and whisker plot visually represents the phonological neighborhood density difference scores, divided by high and low context sentences for each

participant group. The box demonstrates the interquartile range of the difference scores, while the whiskers demonstrate 1.5x the interquartile range. The box line represents the median value, while the dashed line represents no difference in the phonological neighborhood density difference between the error and target. Outliers are graphically represented by diamonds, and circles represent the mean of the difference scores.

Phonological Analysis Results

This next analysis examined other phonological characteristics of the misperceptions produced. Edit distance was the metric first utilized to explore the hypothesis that the YNH and ONH group could produce phonologically similar misperceptions compared to the target and as opposed to older adults with hearing impairment due to reduced perceptual resolution. The edit distance was calculated for every misperception by determining the minimum number of additions, deletions, and/or substitutions needed to change the target into a given misperception. A 2 x 3 ANOVA showed a significant main effect of context where participants responded with misperceptions with larger edit distances in low context sentences, $F(1, 56) = 40.778$, $p = .000$, $\eta_p^2 = .421$. Similarly, a significant main effect of group was observed, $F(2,56) = 3.636$, $p = .033$, $\eta_p^2 = .115$, and a significant interaction between context and group, $F(2, 56) = 3.985$, $p = .024$, $\eta_p^2 = .125$.

Six independent-samples t-tests were performed on the averaged edit distances to explain the interaction. Results show that the ONH group responded with misperceptions that were less phonologically similar to the target compared to the YNH group in high context sentences, but not in low context sentences. The YNH and OHI groups did not perform significantly different in either high or low context sentences. Similarly, the ONH and OHI groups did not perform significantly different from each other in high or low context sentences.

Table 3.6 Edit Distance Independent Samples t-tests
This table reports the results of the independent samples t-tests performed on each of the groups in both high and low context sentences.

Groups Compared	Context	*t*	df	p
YNH vs. ONH	HIGH	-2.98	38	0.005
YNH vs. ONH	LOW	-0.278	38	0.783
YNH vs. OHI	HIGH	-1.691	37	0.099
YNH vs. OHI	LOW	-0.031	40	0.976
ONH vs. OHI	HIGH	1.637	37	0.11
ONH vs. OHI	LOW	0.16	40	0.874

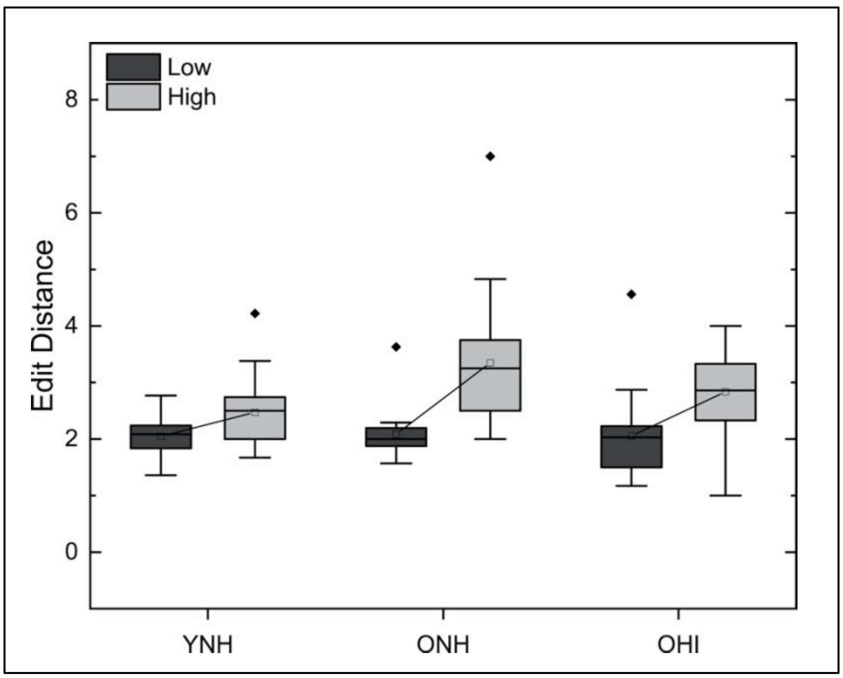

Figure 3.4 Edit Distance Values by Group
This box and whisker plot visually represents the edit distance between targets and errors, divided by high and low context sentences for each participant group. The box demonstrates the interquartile range of the difference scores, while the whiskers demonstrate 1.5x the interquartile range. The box line represents the median value, outliers are graphically represented by diamonds, and circles represent the mean of the difference scores.

Next, the second metric calculated the proportion of misperceptions that were within the neighborhood of the target (i.e., edit distance equal to 1). A 2 x 3 ANOVA showed a significant main effect of context $F(1, 59) = 55.093$, $p = .000$, $\eta_p^2 = .483$, where a greater proportion of misperceptions in low context sentences were within the neighborhood of the target. No significant main effect of group, $F(2, 59) = .774$, $p = .466$, or interaction, $F(2, 59) = 1.001$, $p = .374$, was observed. Overall, both analyses demonstrate greater phonological similarity of misperceptions compared to the target in low context sentences. These results suggest that the ONH group may be less constrained by the phonological properties of the target in high contexts.

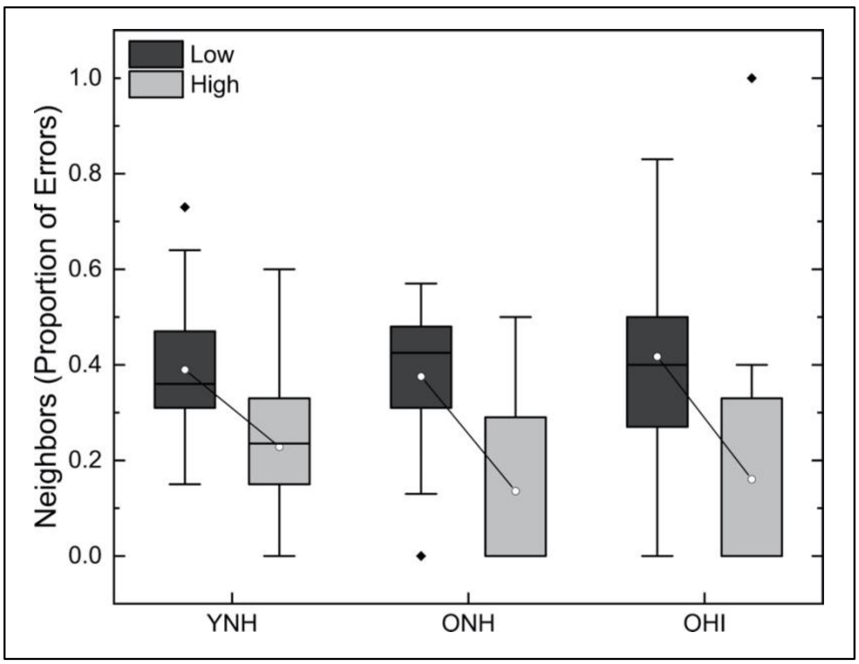

Figure 3.5 Proportion of Errors by Group
This box and whisker plot visually represents the proportion of errors, divided by high and low context sentences for each participant group. The box demonstrates the interquartile range of the difference scores, while the whiskers demonstrate 1.5x the interquartile range. The box line represents the median value, outliers are graphically represented by diamonds, and circles represent the mean of the difference scores.

Semantic Similarity Results

Since earlier words in a sentence may impact the predictability of final words (Taylor, 1953), the investigators hypothesized that response errors for all groups would be semantically similar to the target word in high-probability sentences, but not in low-probability sentences that remove context dependency. Furthermore, we hypothesized that older adults with hearing loss would produce more semantically similar errors than the younger and older normal-hearing groups due to the OHI group's decreased hearing acuity of phonological cues. To test these hypotheses, data were analyzed utilizing two types of semantic comparison, path-based (Leacock, Chodrow, & Miller, 1998) and information content-based (Jiang & Conrath, 1997), and analysis was completed on both data sets. As results were consistent across the path-based and information content-based metrics, only the information content-based results (Jiang & Conrath, 1997) are reported here, which demonstrated a greater range across the conditions.

Results of the 2 x 3 ANOVA showed a main effect of context, $F(1,56) = 13.301$, $p = .001$, $\eta_p^2 = .192$, with greater semantic similarity in high context sentences. No significant main effect of group $F(2,56) = .018$, $p = .982$, or interaction was observed, $F(2,56) = .036$, $p = .965$. These results suggest that all participants were more likely to respond with an error semantically similar to the target in high context sentences, when context dependency was intact.

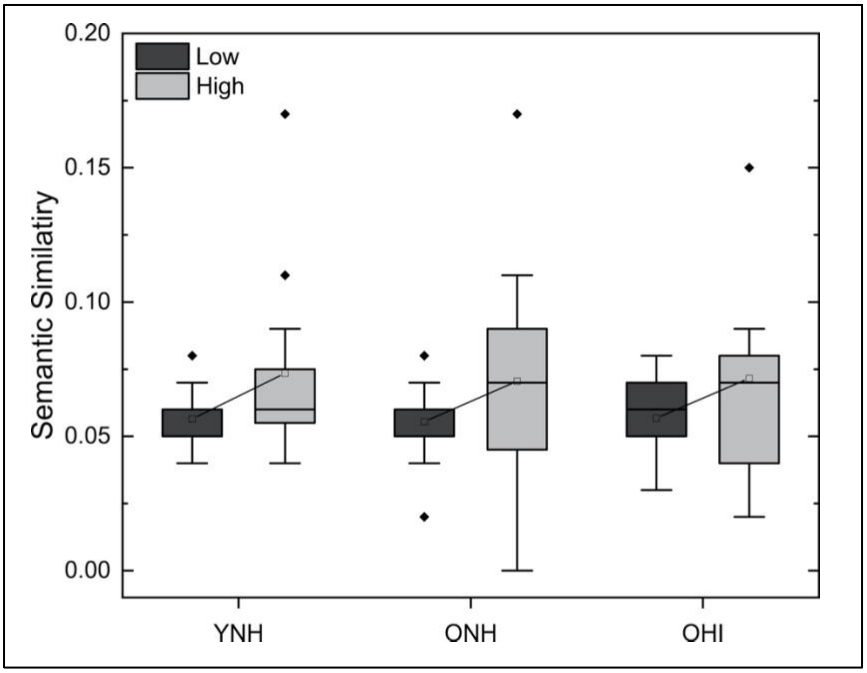

Figure 3.6 Semantic Similarity Values by Group
This box and whisker plot visually represents the semantic similarity between targets and errors, divided by high and low context sentences for each participant group. The box demonstrates the interquartile range of the difference scores, while the whiskers demonstrate 1.5x the interquartile range. The box line represents the median value, outliers are graphically represented by diamonds, and circles represent the mean of the difference scores.

CHAPTER 4

DISCUSSION

The purpose of this study was to describe the various lexical properties of errors made by participants during a speech perception in babble task. Upon initial analysis of participants' overall accuracy, all groups performed relatively the same, while the follow-up analysis revealed differences between groups for skips and misperceptions. Specific analyses then focused on the phonemic, semantic, word frequency, and neighborhood density components of the misperceptions to determine potential patterns of these misperceptions. Results of this study showed that age and context may contribute to the types of misperceptions a participant makes when listening to speech in background noise.

Error Properties

Analysis of errors show that more misperceptions occurred in low context sentences. This suggests that context facilitated accuracy as the reduction of skips increased perception in high context sentences. In addition, older adults with hearing impairment demonstrated a greater propensity for skipping responses in low context sentences. Conversely, the YNH group guessed more often than the OHI group, leading to more misperceptions overall. Therefore, the OHI group's responses were more influenced by the context of the sentence compared to the younger listening group, since they responded more often in high context sentences and skipped more opportunities in low. There are two potential explanations for why the OHI group performed in this way.

First, the poorer resolution of the stimulus brought about by reduced hearing acuity and combined with less top-down contextual information may lead to insufficient information needed to formulate a response. Second, while many factors may contribute to the response criterion an older adult needs to make a decision or response, a review of research proposes that an individual's motivation, emotional state, cognition, and more may contribute to the outcome of an older adult's response (Löckenhoff, 2018). However, other studies propose that older adults are more likely to avoid or defer decision-making more than their younger adult counterparts (Finucane et. al, 2002). In the context of this study, one must remember that only the OHI group performed significantly in this way – not both the OHI and the ONH groups (or all the "older adult" participants). For this reason, this specific response criterion does not fully explain the OHI group's performance. The following sections further explain how individual access to acoustic and lexical properties of speech contributes to recognition.

Word Frequency Difference

The findings of this study are consistent with previous studies where participants responded with misperceptions of a higher frequency in the language than the target (Cooke et. al, 2019; Felty et. al, 2013; Pollack et.al, 1960; Savin, 1963; Vitevitch, 2002). Note that these studies focused on younger participants only who had normal hearing thresholds and that completed the task with contextually independent stimuli, such as single words or isolated non-words. The findings of the current study, therefore, suggest that the results of earlier studies (Cooke et. al, 2019; Felty et. al, 2013; Pollack et.al, 1960; Savin, 1963; Vitevitch, 2002) extend to the groups assessed in this analysis, who showed the same effect.

Results of the word frequency difference analysis, however, suggested the opposite outcome. Results indicated similar word frequency effects for all groups. Therefore, lexicon size appeared to have little bearing on misperceptions as measured by this word frequency metric.

Another effect noted in this study was that high context sentences resulted in higher word frequencies compared to low context sentences for all groups. Because sentences with high context retain informational cues, which may lead participants to make an informed guess under degraded conditions such as noise, participants may be more likely to respond with words of a higher word frequency as these words occur most often and may more easily fill in the blank left by such top-down information. Alternatively, since low context removes any information with which to narrow down potential guesses, participants are more likely to respond with any word – either high or low word frequency.

Neighborhood Density Difference

As neighborhood density is another common metric used to analyze patterns of misperceptions (Felty et. al, 2013; Vitevitch, 2012), the investigators determined if any of the groups, in either high or low context sentences, produced misperceptions with smaller or larger phonological neighborhoods compared to the target. The investigators expected the ONH and OHI groups to respond with words from denser phonological neighborhoods, since these neighborhoods possess more words with which they may be perceptually confused.

Results of the neighborhood density difference analyses show that the YNH group typically responds with a word in a phonological neighborhood of the same size. Both older adult groups (ONH and OHI), however, respond with words in significantly smaller phonological neighborhoods in low context sentences compared to the target. This suggests that older adults respond with words which have fewer possible confusions since they belong to neighborhoods of a smaller size, relative to the target, when context dependency is removed.

In previous studies, investigators have found that misperceptions tend to come from denser phonological neighborhoods as they possess more words which have the potential to be perceptually confusing (Vitevitch, 2012). In this study, the investigators extended the scope of that research to include older adults with normal hearing and with hearing impairment. These older adult groups, however, demonstrate a different pattern than the younger adults as outlined by Vitevitch (2012).

What then caused the ONH and OHI groups to behave differently than their younger counterparts? One plausible explanation is that presenting words in sentences, compared to isolated words in Vitevitch (2012), introduced a new compounding factor, which now acts on the groups in a novel way. Another explanation takes into consideration the implications of the investigators' earlier error analysis, where it was found that older adults (both ONH and OHI) tend to produce more skips than the younger listening group. As older adults do not make as many guesses during speech in noise tasks, it may be that these individuals only respond when they are relatively certain as to the misperception they believe they heard. Their level of certainty may increase when they believe they "hear" a word belonging to a smaller phonological neighborhood, as less perceptually confusable words are found within these smaller neighborhoods.

Phonological Similarity

This study's phonological similarity analysis specifically assessed phonemic changes between targets and corresponding misperceptions as measured by edit distance scores. Results demonstrated that phonological similarity was greater in low context sentences, as indicated by smaller edit distance scores and the proportion of responses that were neighbors. In other words, misperceptions are guided more by phonological confusions in low context conditions.

A difference is noted between groups, however, when analyzing the edit distance scores in high context sentences compared to low context sentences for each group. Specifically, the ONH group responded with misperceptions of a greater edit distance to the target compared to the YNH group in high context sentences only. These results suggest that older adults with normal hearing are less constrained by the phonological aspects of the target compared to the younger group. This may be due, in part, to perceptual encoding brought about by aging, although unlikely, as the other "older adult" group (i.e., OHI) did not perform in the same way. This conclusion then suggests that the ONH group simply places more weight on the context of the sentence when formulating a response.

Semantic Similarity

Previous literature suggests that semantic context from earlier words in a sentence can increase the predictability of final words (Taylor, 1953). With this knowledge, the investigators expected all groups, regardless of age and hearing acuity, to produce misperceptions semantically similar to the target in high context sentences. This hypothesis was supported by the results of the semantic similarity analysis. A significant effect of context was noted as participants produced more misperceptions that were semantically related to the target in high context sentences compared to low context sentences. These results suggest that context dependency directly impacts misperceptions made by all participant groups.

Future Research & Implications

This study utilized established methodology to extend analyses of misperceptions to older adults and to the characteristics of errors made by participants, where real words are assessed within the context of a sentence. Previous studies have specifically focused on misperception analyses of non-words, isolated words, and isolated words in noise (Cooke et. al, 2019; Felty et. al, 2013; Marxer et. al, 2016; Vitevitch, 2002). Therefore, research should continue to expand upon this work by broadening the scope of the analysis to assess the acoustic-phonetic characteristics of the misperceptions. Specifically, misperceptions should be analyzed to assess the sound features of the phonetic errors within the words, such as the place of articulation, voicing patterns, and manner of production. In turn, these acoustic-phonetic characteristics may predict what types of misperceptions are typically produced by those with normal hearing (and have full access to the speech spectrum) and those with hearing loss, where reduced temporal processing of phonetic features may be present (Gordon-Salant et. al, 2006). In this way, a clinical analysis of a client's misperceptions could provide better detail regarding the specific difficulties and, therefore, encouraging precision intervention to address those specific errors. This would enable clinicians to target specific phonetic characteristics rather than simply trying to improve intelligibility in a general way due to poorer accuracy. This study found that accuracy was the same across groups – but misperceptions lead to group differences – and identification of these differences has the potential to differentiate intervention to improve intelligibility.

Furthermore, most previous studies have focused on the analysis of misperceptions produced by younger adults with normal hearing. While these studies have laid the foundation for analyses of misperceptions in this sample of the population, further study should focus on expanding this work to older adults and how the aging process influences their misperceptions. Similarly, cognition, as a mental function affected by aging in the course of a typical lifespan, would be a valuable component to investigate. In fact, its findings may hold great significance for the population as individuals continue to live longer and cognition continues to change.

Lastly, the general study of misperceptions holds significance as it helps us understand more about how communication breaks down in a variety of contexts and environments. By exploring the patterns of these errors, we can begin to formulate predictors for why certain errors occur and then explore ways in which we can best resolve the ensuing miscommunication caused by the misperception.

Conclusions

Results of this study revealed that overall accuracy was similar for all groups. However, analysis of the misperceptions demonstrated systematic differences across sentence contexts and between groups:

1) More skips and misperceptions occurred in low context sentences.
2) OHI listeners produced the greatest proportion of skips in the total error count.
3) YNH listeners made more misperceptions as they tended to guess more words instead of skipping them.

4) All groups produced misperceptions of a higher word frequency than the target.
5) Older adults produced misperceptions from a smaller phonological neighborhood than the target in low context sentences.
6) A greater proportion of misperceptions were found to be within the same neighborhood as the target in low context sentences for all groups.
7) The ONH group, specifically, produced more errors that were less phonologically similar to the target compared to the YNH group in high context sentences.
8) For all groups, the presence of a high predictability sentence context resulted in misperceptions that had greater semantic similarity to the target.

These findings delineate patterns of misperceptions that may further inform how individuals, based on their age and hearing ability, might perceive speech in babble as a function of the sentence context. In summary, these results highlight the importance of analyzing misperceptions as the error patterns detail differences made during routine SPIN tasks, which are not typically noted when looking at overall accuracy.

REFERENCES

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.

Billings, C. J., Penman, T. M., McMillan, G. P., & Ellis, E. M. (2015). Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment and age. *Ear and Hearing*, *36*(6), 710–722.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41*(4), 977-990.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*(5),975–979.

Cooke, M., Lecumberri, M.L.G., Barker, J., & Marxer, R. (2019). Lexical frequency effects in english and spanish word misperceptions. *The Journal of the Acoustical Society of America*, *145*(2), EL136-EL141.

Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: cognitive aging. *Journal of Geriatric Psychiatry and Neurology, 23*(2), 75–93.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*(1), 538–546.

Dunn, L. M. & Dunn, D.M. (2007). Peabody picture vocabulary test (PPVT - 4). Forms L and M. Circle Pines, Minn.: American Guidance Service.

Felty, R.A., Buchwald, A., Gruenenfelder, T.M., & Pisoni, D.B. (2013). Misperceptions of spoken words: data from a random sample of American English words. *The Journal of the Acoustical Society of America*, *134*(1), 572-585.

Finucane, M.L., Slovic, P., Hibbard, J.H., Peters, E., Mertz, C.K., & MacGregor, D.G. (2002). Aging and decision-making competence: an analysis of comprehension and consistency skills in older versus younger adults considering health plan options. *Journal of Behavioral Decision Making*, *15*(2), 141–164.

Fogerty, D., Madorskiy, R., Ahlstrom, J.B., & Dubno, J.R. (2020). Comparing speech recognition for listeners with normal and impaired hearing: simulations for controlling differences in speech levels and spectral shape. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4289-4299.

Folstein, M., & McHugh, P. (1975). Mini mental state a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *The Journal of the Acoustical Society of America*, *119*(4), 2455-2466.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433–443.

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journal of Gerontology, 44*(2), P36–P44.

Jiang, J. J., & Conrath, D. (1997) "Semantic similarity based on corpus statistics and lexical taxonomy", Proceeding of International Conference on Research in Computational Linguistics, 19-33.

Leacock, C., Chodorow, M., and Miller, G. (1998). Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics, 24*(1), 147-165.

Löckenhoff, C.E. (2018). Aging and decision making: a conceptual framework for future research – a mini-review. *Gerontology, 64,* 140-148.

Marxer, R., Barker, J., Cooke, M., & Garcia Lecumberri, M. L. (2016). A corpus of noise-induced word misperceptions for English. *The Journal of the Acoustical Society of America*, *140*(5), EL458-EL463.

Miller, G.A. (1947). The masking of speech. *Psychological Bulletin,44*(2), 105-29.

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys, 33*(1), 31 – 88.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, *87*(6), 2705–2715.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*(2), S72–S82.

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *The Journal of the Acoustical Society of America*, *121*(4), 2312–2326.

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *The Journal of the Acoustical Society of America*, *32*(4), 454-457.

Rus, V., Lintean, M., Banjade, R., Niraula, N., & Stefanescu, D. (2013). “SEMILAR: The semantic similarity toolkit.” Proceeding of the 51st Annual Meeting of the Association for Computational Linguistics, August 4-9, 2013, Sofia, Bulgaria. 163-168.

Savin, H.B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America*, *35*(2), 200 – 206.

Smith, K. G., & Fogerty, D. (2017). Speech recognition error patterns for steady-state noise and interrupted speech. *The Journal of the Acoustical Society of America,* *142*(3), EL306-EL312.

Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging*, *15*, 253–258.

Taylor, B. (2003). Speech-in-noise tests: how and why to include them in your basic test battery. *The Hearing Journal, 56*(1), 40-43.

Taylor, W. L. (1953). "Cloze Procedure": a new tool for measuring readability. *Journalism Quarterly, 30*(4), 415–433.

Vaden, K.I., Halpin, H.R., & Hickok, G.S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com.

Vitevitch, M.S. (2002). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and Speech, 45*(4), 407-434.

Zaar, J., & Dau, T. (2015). Sources of variability in consonant perception of normal-hearing listeners. *The Journal of the Acoustical Society of America*, *138*(3), 1253-67.