5-2020

Whole Word vs Phoneme Error Scoring for Audiological Word Recognition Testing

Jennifer C. Arnoult

The University of Southern Mississippi

Whole Word vs Phoneme Error Scoring for Audiological Word Recognition Testing

by

Jennifer Arnoult

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

April 2020

Approved by:

____

Jennifer Goshorn, Au.D., Thesis Adviser
Instructor of Audiology, School of Speech and Hearing Sciences

____

Edward Goshorn, Ph.D., Director
School of Speech and Hearing Sciences

____

Ellen Weinauer, Ph.D., Dean
Honors College

Abstract

The purpose of this study was to investigate a possible alternative method of scoring for word recognition testing. The current whole word scoring method is inefficient and often inconsistent with a client's pure tone average scores, and previous research has suggested that phoneme error scoring may be a more effective and accurate scoring method. This project analyzed the audiogram and word recognition test scores of seventeen adult clients. The test used was the standard twenty-five-word list NU-6 test. The researchers analyzed the results from whole word scoring and phoneme error scoring of the same tests and compared the results to the clients' pure tone average scores to look for statistical significance. A significant relationship was found between the phoneme error results of the clients' word recognition tests and the audiograms of their pure tone average tests. Because of this relationship, the research team concluded that the proposed phoneme error scoring model shows promise of being more effective and accurate in showing the extent and type of hearing loss in an individual. In addition, there may be ways this phoneme error scoring could be further developed to show more details of a client's impairment through analysis of distinctive features or other linguistic or acoustic factors.

Keywords: Audiology, Hearing test, Word recognition testing, NU-6, Whole word, Phoneme error

Acknowledgments

The research team would like to acknowledge the staff at the University of Southern Mississippi Audiology Clinic. Participants in this study provided informed consent for their test results to be utilized for research, and this project was conducted with approval from the pertinent Institutional Review Board. Portions of these data were presented in 2019 at the Mississippi Speech-Language-Hearing Association Conference.

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations

CNC	Consonant-Nucleus-Consonant
dB	Decibel
Hz	Hertz
LDL	Loudness Discomfort Level
NU-6	Northwestern University Test Number 6
PB	Phonetically Balanced
PTA	Pure Tone Average
SPSS	Statistical Package for the Social Sciences

Chapter 1: Introduction

There is a plethora of ways and reasons an individual can experience hearing loss. Hearing loss can be conductive (originate from a problem in the outer or middle ear), sensorineural (originate from a problem in the inner ear), or mixed (a combination of the two previously named origins). No matter the origin of an individual's hearing loss, it is likely that this individual may experience difficulties understanding speech. Word recognition testing is a form of audiological testing used to assess an individual's ability to understand speech at sound levels above the individual's threshold (the lowest level at which an individual can understand speech). This type of testing is most useful with clients who state, "I can hear, but I can't understand" (Martin & Greer, 2015). Only testing a client with pure tone tests would limit information an audiologist could glean about the client's abilities to understand speech when trying to make a proper diagnosis. Word recognition testing aids in determining the extent of a client's hearing loss pertaining to their ability to perceive speech. Word recognition testing typically involves the repetition of monosyllabic words in quiet; therefore, it does not show the capabilities of a client in typical conversation (including background noise, etc.), but it does show the extent of each client's capability to understand speech at different levels in controlled conditions. This type of testing can often help an audiologist determine the extent to which amplification could help each individual, if at all (Martin & Greer, 2015).

Word recognition testing methods have evolved over the years, but the central premise of the tests have not changed. Audiologists use these tests to assess an individual's word recognition score by presenting a list of monosyllabic words and asking the individual to repeat the words they hear. These tests are usually performed using an open-response system, which means the clients are not given options of words to choose from, but the clients must produce their responses from their own thoughts. The standard form of word recognition tests contain phonetically balanced word lists of consonant-nucleus-consonant (CNC) words, which were chosen and developed over a period of over 20 years. The CNC words on these lists are composed of two consonants in the initial and final positions and a vowel or diphthong between them. Phonetically balanced word lists “contain all the phonetic elements of connected English discourse in their normal proportion to one another,” (Martin & Greer, 2015), which essentially means that the lists contain the typical distribution of all the phonemes used in the English language.

During word recognition testing, the stimuli are presented at predetermined levels of intensity, usually well above the client’s speech recognition threshold (the lowest intensity level at which the client can discriminate between words). A common level at which the tests would be performed would be thirty to forty decibels above the client’s speech recognition threshold (Martin & Greer, 2015). However, some research has shown intensity levels at that level may be above a client’s loudness discomfort level, or LDL. A client’s loudness discomfort level is the highest intensity level at which sound can be presented to him or her without causing discomfort or pain. Presenting a word recognition test at a level above a client’s LDL would not provide accurate results, because the client would be performing at a level of intensity that he or she is comfortable with (Ricketts, Bentler, & Mueller, 2019). To combat this problem, two researchers named Guthrie and Mackersie established two alternate methods of determining the intensity level at which word recognition tests could be presented (2009).

The first of these methods is simply to determine the client's LDL and present the stimuli at a level five decibels (dB) below the LDL. The second method does not require a test to determine the client's LDL. Guthrie and Mackersie analyzed the phoneme recognition scores of numerous clients and produced a table of the best intensity levels at which to test individuals based on their threshold at a frequency of 2,000 Hz. These scores are represented in Table 1.

2,000 Hz thresholds < 50 dB HL	25 dB SL
2,000 Hz thresholds 50-55 dB HL	20 dB SL
2,000 Hz thresholds 60-65 dB HL	15 dB SL
2,000 Hz thresholds 70-75 dB HL	10 dB SL

Table 1. *Source:* (Guthrie & Mackersie, 2009)

In this table, the values on the right side represent the number of decibels the researchers recommend adding to the client's threshold to determine the correct intensity for the word recognition test. However, because of the variability of clients' auditory systems, none of these three methods of determining the proper intensity for a word recognition test is foolproof. Audiologists must be adaptable in each testing situation, and they should communicate effectively with clients about their comfort level during testing (Ricketts, Bentler, & Mueller, 2019).

After the word recognition test is performed, it is scored using a percent correct scoring measure based on a whole-word scoring system. For each word incorrectly repeated, the score is decreased by a certain percentage, typically 4% for a twenty-five-word list and 2% for a fifty-word list. This scoring method used for word recognition tests yields a whole word score rather than a score that considers specific phoneme errors made by the client. The whole word scoring method does not allow audiologists to obtain as much information from the test as they could using more specific methods (Martin & Greer, 2015). This means that the audiologist may not gain enough information about a client's specific word recognition difficulties to make an informed decision about the client's speech perception and the types of features a client would need in his or her amplification system. Additionally, most word recognition tests are performed in quiet. This method of testing does not approximate real-world conversation, and therefore may not accurately predict the success of the client in more difficult speech situations or environments. As stated previously, word recognition testing simply provides audiologists with an estimate of a client's best performance in quiet, or the client's "PB Max." A goal of word recognition testing is to accurately determine each client's PB Max and to aid him or her in achieving the closest performance to that PB Max in daily life (Ricketts, Bentler, & Mueller, 2019).

In addition to lacking in specific phoneme scoring, word recognition testing has been shown to have inconsistent in test-retest reliability, especially for individuals with sensorineural hearing loss. In a research study conducted by Thorton and Raffin (1978), numerous inconsistencies were discovered in the results of the word recognition tests of individuals with and without hearing loss. Thorton and Raffin noticed that these inconsistencies had a measurable pattern. A pure tone hearing test is a test in which an individual's hearing abilities are evaluated through the presentation of and response to standard tones. When an individual's pure tone hearing test results were at the extremes of hearing severity, word recognition scores were more reliable and consistent, and they better represented an individual’s word recognition abilities. In other words, clients with normal hearing and with severe to profound hearing loss are more likely to produce an accurate and consistent result on word recognition tests than those with mild to moderately severe hearing loss. Because of this, word recognition test results for those with levels of hearing loss on the two extremes of the spectrum hold more weight with audiologists, and they may more confidently use word recognition tests to make decisions about diagnosis and treatment of a client’s hearing loss. Results for those with hearing loss in the middle range can vary immensely, and results may need to differ by 20% or more to be considered significantly different (Thorton & Raffin, 1978).

When conducting word recognition testing, one of the decisions the audiologist must make is which word list is appropriate for the client. Researchers have compiled numerous word lists over the years, each with different goals. The standard (phonetically/phonemically balanced) word recognition test lists, such as the Northwestern University Test Number 6 (NU-6) test and the W-22 test are used to gain a general idea of a client’s ability to understand speech. However, generalized measures of word recognition tend to grossly underestimate or overestimate the severity of an individual’s hearing loss and word recognition ability (Ricketts, Bentler, & Mueller, 2019). This is especially true of individuals with high frequency hearing loss, so researchers have developed word recognition test lists called “high frequency emphasis lists.” These tests better represent the true word recognition abilities of clients with high frequency hearing loss. An individual with high frequency hearing loss may sometimes have word recognition scores that do not correlate with his or her pure tone thresholds nor their word recognition scores on standard lists. High frequency emphasis lists better represent their true difficulty recognizing speech. However, it is not recommended that these tests be used in the place of traditional word recognition tests. Instead, they are often used as follow-up assessments with those known to have high frequency hearing loss or sensorineural hearing loss, as these two factors can be and often are indicative of one another. The results of a high frequency emphasis test are often used to help audiologists determine the best approach to treatment for clients diagnosed with hearing loss.

An example of such a test would be the California Consonant Test. The California Consonant Test is a hearing test that is composed of one hundred monosyllabic items randomly split into two fifty-word lists. This test uses a closed-response system, which means that the client must choose between multiple given responses. On this test in particular, clients must choose between four items for each stimulus (Martin & Greer, 2015). The closed-set response makes this test favorable for use with both children and adults. As one might infer from the name, the California Consonant Test is primarily used to identify consonant confusions in clients with high frequency hearing loss. Additionally, this test could be used to help an audiologist decide if a client with high frequency hearing loss should be treated unilaterally (using a hearing aid in only one ear) or bilaterally (using hearing aids in both ears) (Auditec, Inc, 2015).

It is important to note that research on the topic of speech perception has evolved to be what it is today. In their comprehensive evaluation of research on speech perception, Jusczyk and Luce (2002) discovered that there are still many topics lacking sufficient research. They highlighted three major topics related to speech perception:

invariance, constancy, and perceptual units of speech. For each of these topics, they provided a summary of the research in speech perception. Jusczyk and Luce (2002) related that one goal in speech perception was the identify the invariant acoustical characteristics which an individual uses to perceive speech sounds. Through their literature review, it was found that speech is not invariant. However, it has many changing characteristics, such as co-articulation. Co-articulation is the influence of surrounding speech sound on the target speech sound, which influences and changes the acoustic signal of speech. Although speech does not seem to have an invariant set of characteristics by which we identify phonemes, individuals nevertheless are able to correctly perceive speech. This ability supports the Motor Theory of Speech Perception (Liberman, Cooper, Harris, & MacNeilage, 1963 as cited in Jusczyk and Luce, 2002), which states that a speech sound is perceived by relating speech sounds to the articulatory gestures that made them. When conducting word recognition testing, it is crucial to remember that most consonant and vowel segments are produced simultaneously rather than sequentially and independently from one another. This alters the way clients perceive different words and may account for many errors in speech perception. Additionally, Jusczyk and Luce stated that the concepts and research behind the Motor Theory of Speech Perception and the Direct Perception Theory are important to speech perception. Both of these theories highlight the importance of articulatory gestures used to produce speech pertaining to speech perception.

Another significant portion of past research in the field of speech perception states that constancy is present in the perception of speech sound, but not in the production of speech sound (Jusczyk and Luce, 2002). There are differences between different speakers, including the pitch of the voice and speaking rate that affect the constancy of the signal by altering the acoustic characteristics. Despite the variability in the production of the speech signal, individuals with normal hearing typically have no difficulty correctly perceiving the speech signal. Therefore, it must be a goal of audiologists to understand what acoustic and other characteristics impact speech perception. Understanding speech perception allows audiologists to better help clients find ways to improve errant speech perception with hearing aids and other means.

Lastly, Jusczyk and Luce (2002) summarized research on the topic of perceptual units of speech. The principal unit of speech perception is a phonetic segment. Each of the distinctive features of individual phonemes combines into phonetic segments. A phonetic segment is a unit of perception, and that is what an individual perceives during a word recognition assessment (Jusczyk & Luce, 2002). Research in speech perception is an ongoing process that continues to build on the foundation set by core research discoveries of the past.

Several previous studies in this field have examined the idea that phoneme scoring would be advantageous for audiologists. One such study was performed in 2016 by Billings, Penman, Ellis, Baltzell, & McMillan. This study included speech-in-noise testing variables, but their findings strongly supported utilizing phoneme error scoring rather than whole word scoring techniques. The phoneme based scoring system was noted to be 30% better than the whole word scoring system, and it was said to “provide several advantages over word scoring” (Billings, Penman, Ellis, Baltzell, & McMillan, 2016).

The current study aims to examine a possible change in the way traditional CNC word recognition tests are scored. The focus of this experiment was on the weakness in the whole-word scoring system commonly utilized during these tests. Determining an alternative method of scoring that provides a more comprehensive representation of a client's hearing ability in controlled conditions was the goal of this study. The method chosen was a phoneme-based scoring system in which each errant phoneme response recorded and scored. It was hypothesized that this scoring method will give audiologists more information about the root of clients' incorrect responses than the methods that are presently used.

Chapter 2: Methods

Subjects for the study were obtained from the clinic archive records of the University of Southern Mississippi Speech and Hearing Clinic. From this archival data, the research team retrieved the results of seventeen adults' pure tone audiograms and word recognition test scores from prior testing. The clients studied were all between the ages of forty-nine and ninety-one, with an average age of seventy-five years old. Scores were retrieved from the clients who signed consent forms at the time of their audiological evaluation, allowing scores to be utilized for research studies. Records of both ears of each client were accessed, therefore the research team analyzed a total of thirty-four ears. Each client was tested using the standard pure tone testing procedures and two twenty-five-word lists, with one for each ear, from the NU-6 word recognition tests. The CNC words are arranged on the list in order of difficulty, from easiest to hardest. The word recognition tests had been scored using the traditional whole-word scoring method. The research team documented the results of the clients' pure tone hearing tests, as well as the whole word scores of the NU-6 tests.

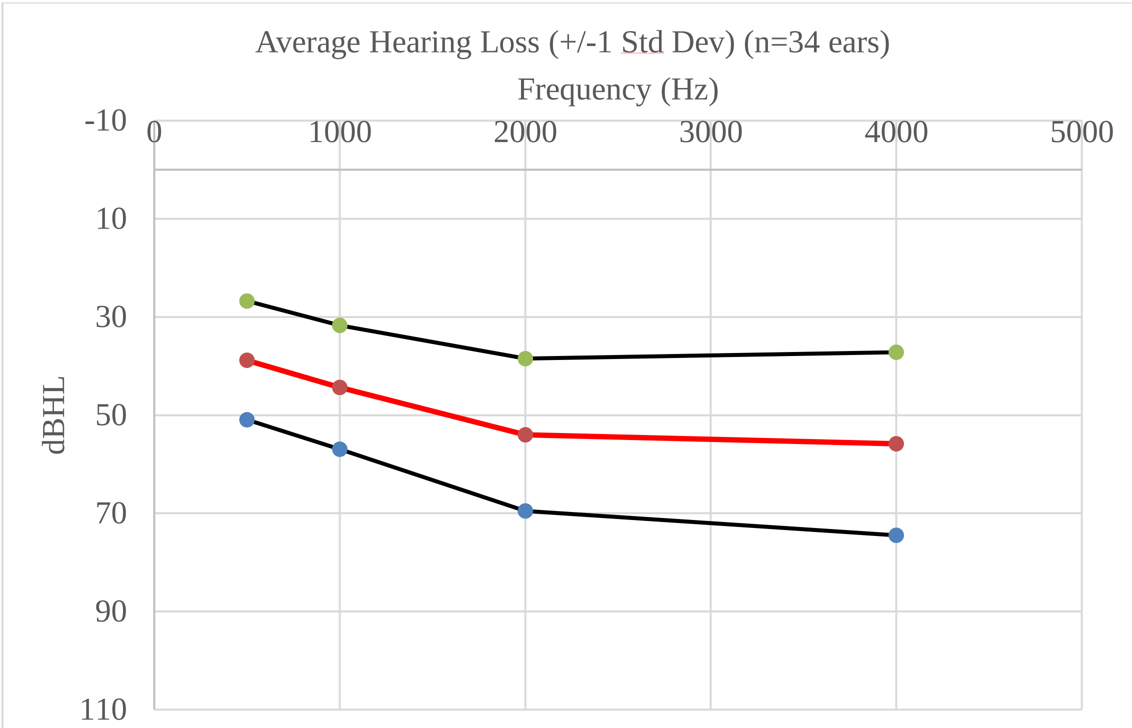

Figure 1. Average hearing loss with +/- 1 Standard Deviation

Above is a graph representing the calculation of the mean hearing loss of all of the participants of the study, with the extreme plot points representing +/- one standard deviation from the mean. Two key values were calculated for each participant, the three frequency pure tone average – PTA3 (average of pure tone threshold at 500, 1000, and 2000 Hz) and the four frequency pure tone average – PTA4 (average of pure tone threshold at 500, 1000, 2000, and 4000 Hz). The mean PTA3 is 46 dBHL, and the mean PTA4 is 49 dBHL. Throughout the study, these dependent variables were compared with other variables, and those comparisons were analyzed for significant relationships.

Each NU-6 word on the list contains three phonemes presented in the CNC format discussed previously. In the whole word scoring method, any errant response in a stimulus word caused the audiologist to mark the entire word incorrect, and each word was an equal percentage of the total score. In the new scoring method, each errant response to the presented stimuli is categorized by the number of phonemes and the type of phonemes that were not perceived accurately. They can also be more simply categorized into categories of mild, moderate, and severe, as shown below.

Mild	One consonant error
Moderate	Two consonant errors Vowel error
Severe	Three phoneme errors

Table 2. *Classification of severity of phoneme errors*

This table shows the simplest way to classify the new scoring methods, but the actual scoring process requires a more specific system. The results of these tests are best represented by a numerical scoring system. These four experimental scoring models are named the total phoneme error model, the severity model, the square model, and the exponential model. In each of the four models, a zero represents no errant responses, and this represents the score a client would receive if the client responded correctly to the stimulus. Each model exhibits a different way to represent the severity of a client's errant responses. In the total phoneme error model, each errant phoneme is assigned one point to the client's score. This model is the one most commonly studied in experiments similar to the present study (Billings, Penman, Ellis, Baltzell, & McMillan, 2016; Schlauch, Anderson, & Micheyl, 2014; Gelfand & Gelfand, 2012). In the severity model, each error is weighted according to its severity. Each misperceived consonant adds one point to a client's score, while each misperceived vowel adds two points to an individual's score. In the square model, the total phoneme error score is raised to the second power. The exponential involves the variable “e,” which has a value of 2.718. To calculate a client’s score using the exponential model, one must raise both the number of phoneme errors and the number of vowels misperceived to the power of “e,” and those two values are added together to find the final score for that word. The two tables below provide more information on the different models of scoring. The first is a table that displays the calculation formula of the different scoring methods, and the second table shows how different types of errors affect a client’s score using each of the different scoring methods.

Method	**Formula**
Severity Model	cons1+cons2+(Vx2)
Square Model	total phoneme error^2
Exponential Model	phoneme error^e+V^e (e=2.718)

Table 3. *Calculation formula for each severity model.*

Stimulus	Response	Whole Word Error	Total Phoneme Error	Severity Model	Square Model	Exponential Model
KNOCK	KNOCK	0	0	0	0	0
THOUGHT	FOUGHT	1	1	1	1	1
GIN	JIM	1	1	1	1	1
PIKE	KITE	1	2	2	4	6.58
SHACK	SHUCK	1	1	2	1	2
KNOCK	MAP	1	3	4	9	20.81

Table 4. *Examples of phoneme errors in each scoring method.*

Once all the word recognition tests were scored utilizing the new scoring methods, the data compiled and analyzed using regression analysis techniques with the Statistical Package for the Social Sciences, or the SPSS. Regression analysis is a form of statistical analysis that allows a researcher to see relationships in data. Regression analysis can also enable a researcher to describe statistical relationships between two variables, predict outcomes based on these relationships, and even predict outcomes while accounting for extraneous variables (in the more complex forms) (Schneider, Hommel, & Blettner, 2010). There are numerous different types of regression analysis, but in this experiment, the research team utilized one of the more simple methods of regression analysis. This type of regression analysis is called linear regression, and it simply displays the relationships and trends when comparing two variables. The strength of the relationship can also be shown by the calculation of the line of best fit in a linear regression analysis.

Chapter 3: Results

Table 5 is a comprehensive summary of the data gathered in the research project. The minimum, maximum, mean, and standard deviation of the study variables were reported. These are referred to as the descriptive statistics of the experiment.

Variable	n	Minimum	Maximum	Mean	Standard Deviation
PTA3	34	27	65	46	13
PTA4	34	25	70	49	14
PB%	34	20	100	76	23
Whole Word	34	0	20	6.1	5.7
Total Number Phoneme Error	34	0	42	11.3	12.6
Severity Model	34	0	58	15.4	17.3
Square Model	34	0	110	24.9	32
Exponential Model	34	0	244	51.7	69.9
Consonant 1	34	0	15	3.4	4.2
Nucleus	34	0	16	4.1	4.8
Consonant 2	34	0	15	3.8	4.1

Table 5. *Descriptive Statistics*

Once the data was gathered and documented numerically, the research team was able to perform a regression analysis on the data in order to see the relationships between all of the relevant variables tested in this experiment.

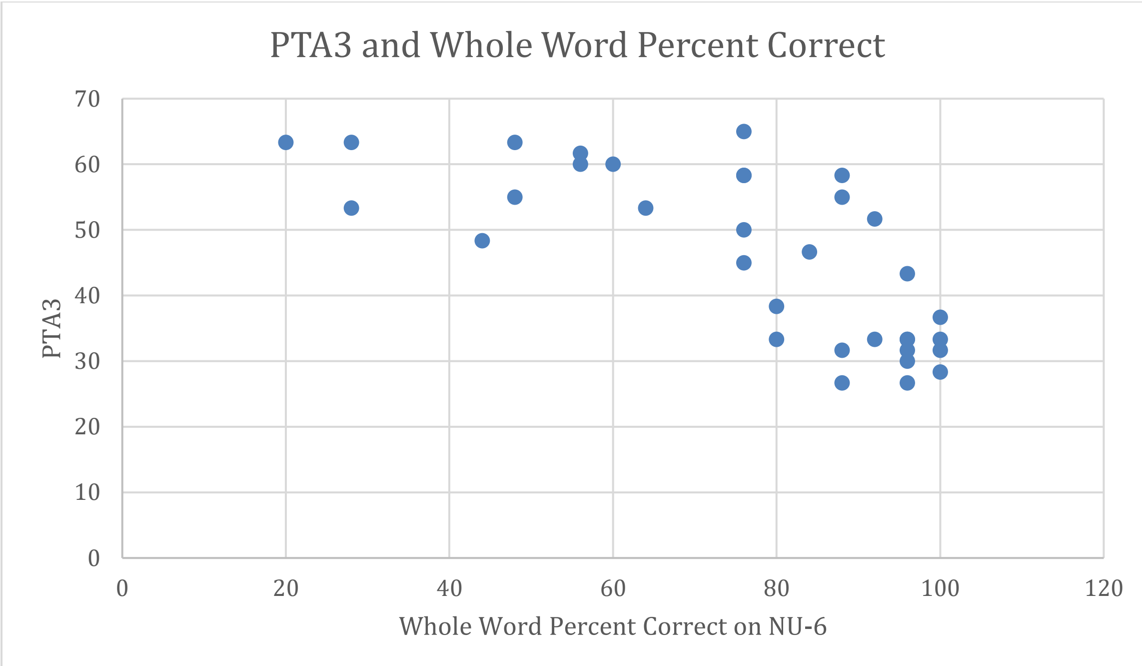

Figure 2. *Relationship between PTA3 and Whole Word Percent correct*

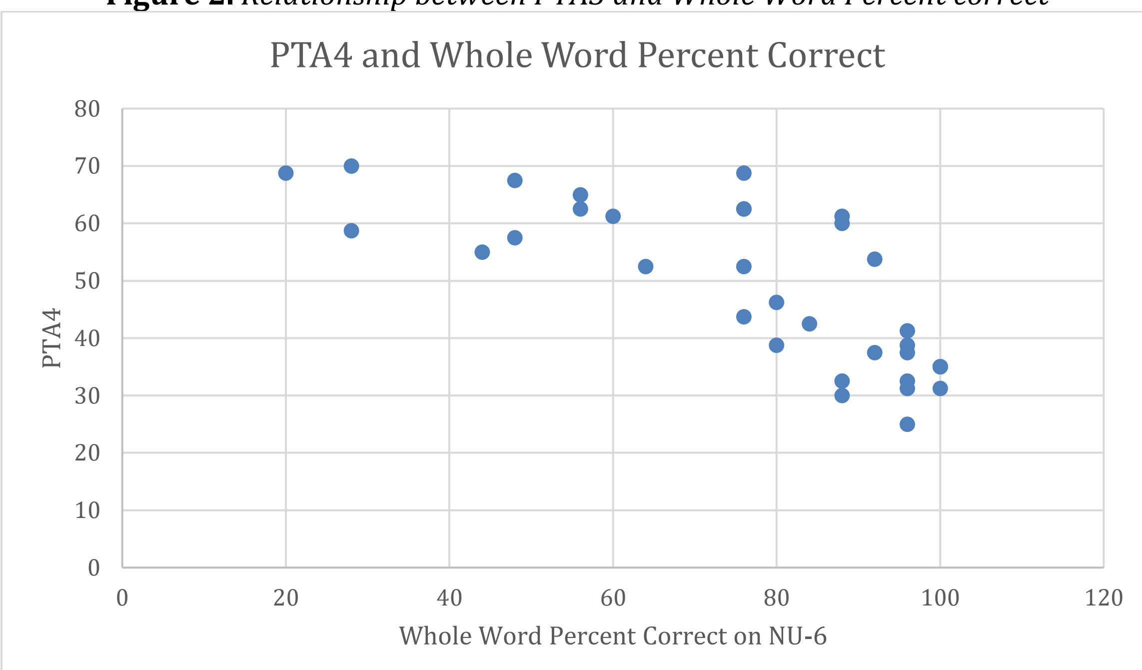

Figure 3. *Relationship between PTA4 and Whole Word Percent correct*

First, the research team analyzed the relationship between the pure tone averages of the clients (both the PTA3 and the PTA4 values) and the original whole word score results. There was not a strong statistical relationship between PTA3 or PTA4 and the

Whole Word scoring method. This relationship is shown in the scatter plots below.

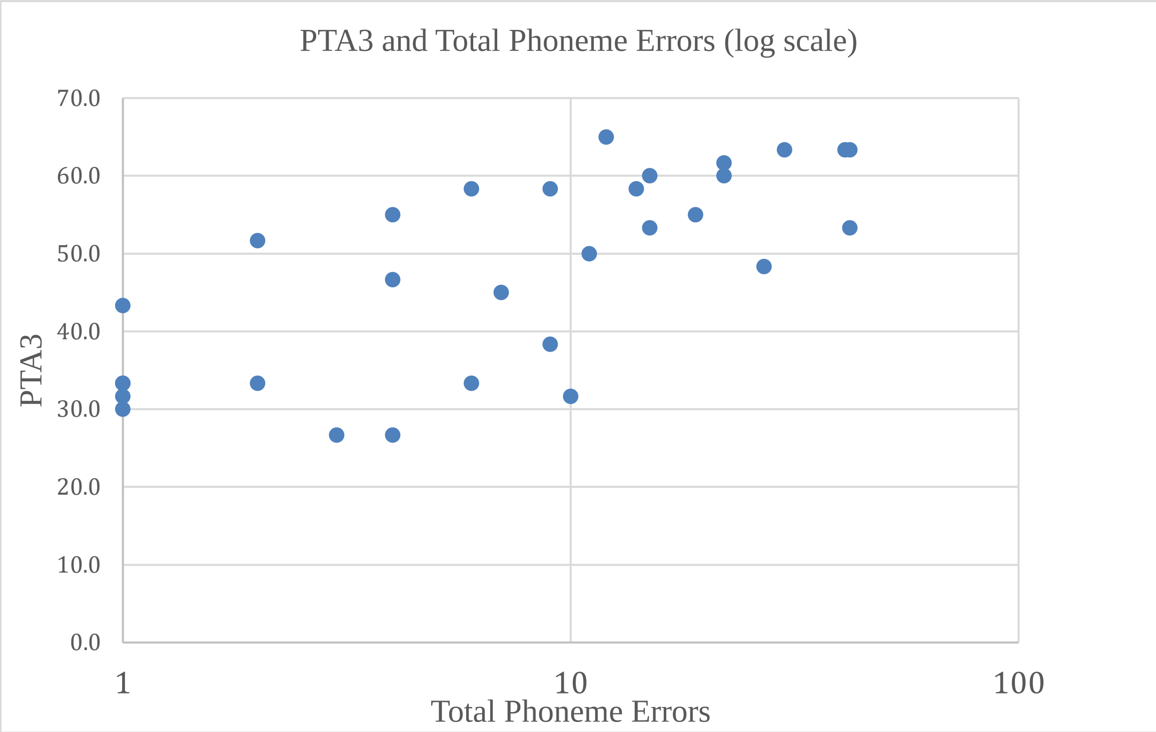

Figure 4. *Relationship between PTA3 and total phoneme errors*

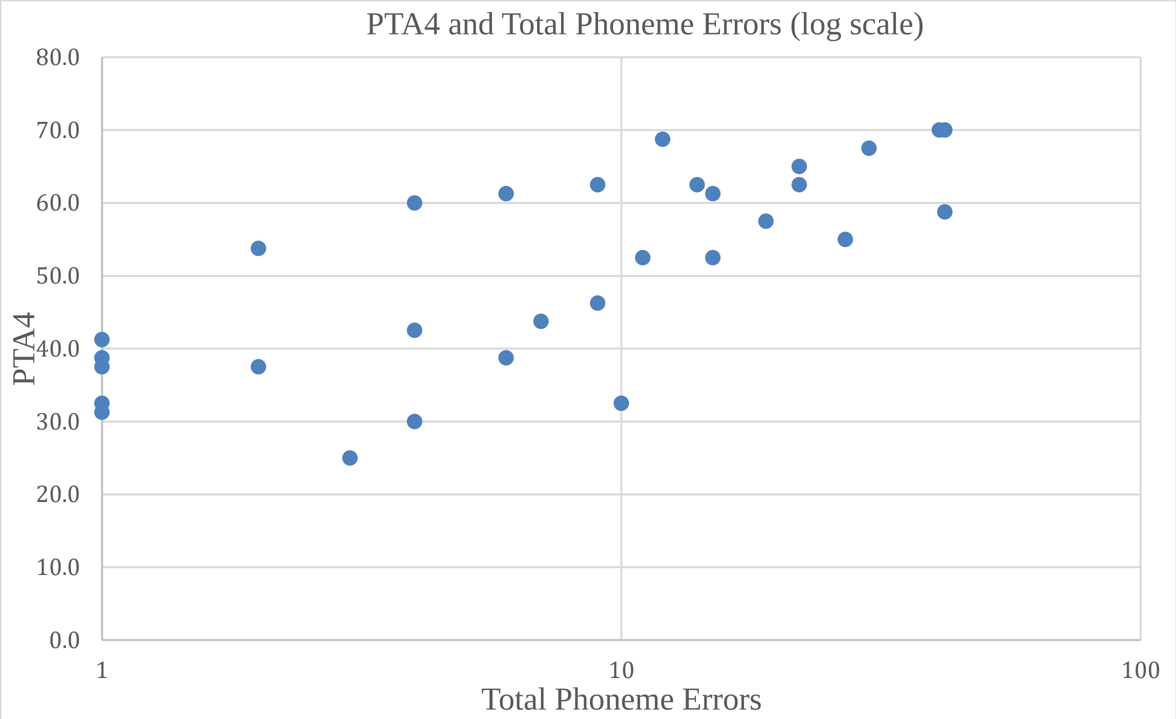

Figure 5. *Relationship between PTA4 and total phoneme errors*

Next, the relationship between PTA3 and the PTA4 and the new scoring methods were examined. The relationship between PTA3 and phoneme scoring did show a statistically significant relationship. When compared with the PTA4 results, phoneme scoring did not have a significant relationship. Still, the p score of this relationship was 0.054 which is near a statistically significant result, and this supports the trend seen with PTA3 and phoneme scoring. These result of the regression analysis shows that the total phoneme error method of scoring was, in fact, a better representation of the clients' hearing abilities than whole word scoring, and could be used to gain more information about a client's type and level of hearing loss. This would be more useful to an audiologist than the whole word results of a client's word recognition test.

For a relationship between to variables to be statistically significant, the p value between two variables needs to be below the level of 0.05. A p value below 0.05 indicates that there is less than a 5% chance that the correlation of the two variables occurred by chance. There is also a possibility that the correlation could be contributed to another variable. Therefore, this value shows that there is a high probability that the relationship between the two variables is strong and not caused by chance. The tables below show the numerical values of regression analysis performed on the data represented in the previous scatter plots.

Variable	Standard Coefficient Beta	t	Significance (p)
Constant		2.201	.036
PB%	-1.4	-1.5	.143
Whole Word Score	-2.4	-1.5	.155
Total Phoneme Error Score	**9.2**	**2.1**	**.045**
Severity Model	-4.4	-1.8	.076
Exponential Model	-3.1	-1.4	.161

Table 6. *PTA3 and Phoneme Error Scoring Regression Analysis*

Variable	Standard Coefficient Beta	T	Significance (p)
Constant		2.321	.028
PB%	-1.4	-1.6	.121
Whole Word Score	-2.4	-1.6	.130
Total Phoneme Error Score	8.4	2.016	.054
Severity Model	-3.9	-1.7	.099
Exponential Model	-2.8	-1.4	.187

Table 7. *PTA4 and Phoneme Error Scoring Regression Analysis*

All other variables tested in this experiment were also not statistically significant. In conclusion, the results of the study show a statistically significant relationship between PTA3 and the Phoneme Error Scoring Method. These results support further research into phonemic scoring methods.

Chapter 4: Discussion

The main purpose of this research project was to test a possible solution to the ineffectiveness of current scoring procedures on audiological word recognition testing. Many research studies have proven that there are problems with the way word recognition testing is scored. This is a problem of long standing. Textbooks such as *Introduction to Audiology* by Martin and Greer include sections explaining the shortcomings of word recognition testing (2009). Research has proven that there is potential for audiologists to obtain a great deal more information about the hearing capabilities of a client through word recognition testing. Additionally, utilizing scoring methods based on phonemic errors rather than whole-word errors could aid audiologists in accomplishing this goal. Some studies have suggested that phoneme scoring could increase the efficiency and decrease the variance present in word recognition testing (Billings, Penman, Ellis, Baltzell, & McMillan, 2016). There is a significant amount of research on this topic and countless potential solutions, but none of the research is conclusive enough for the audiological community to implement new scoring methods. This project sought to supplement the research already existing on this topic in hopes that it will encourage a much-needed change in the field of audiology.

The results of the current study state that the only scoring method that was significantly related to the extent of a client's hearing loss (or the PTA3) is the total phoneme error score. The whole word scoring method did not produce a significant relationship with either the PTA3 or the PTA4 score. To put it simply, this means that the phoneme error scoring method provided a more accurate measure of each client's hearing ability, and therefore is a better method of scoring for word recognition testing. Based on this result, it is recommended that word recognition test scoring methods should include the phoneme scoring method.

A regression analysis between the dependent variables (PTA3 and PTA4) and the other variables was performed in the SPSS system to investigate the possibility of significant relationships between PTA and other variables such as age, PB max, and vowel/consonant relationships. Most of these variables either had no relationship or a very weak relationship with PTA. Additionally, there was no significant relationship between age and PB max, nor was there a significant relationship between age and PTA. However, there was a slight relationship between vowel and consonant errors. Clients overall tended to make more vowel errors than consonant errors, but again this relationship did not reach statistical significance. The sole significant relationship found in this study was the relationship between total phoneme errors and PTA3.

Chapter 5: Conclusion

Based on the findings in this study and numerous others, it is likely that phoneme error scoring can and will be useful to audiologists as a supplemental measure of hearing capability. However, this research project raises the idea that further research on different scoring methods for word recognition testing would be necessary.

In the future word recognition scoring models should include the severity of errors in word recognition scoring. The severity of a client's hearing loss is often revealed by the severity of the client's word recognition errors rather than the degree of hearing loss. For example, an error in the perception of a vowel is more severe than an error in the perception of a consonant, and a lack of response is more severe than a response with all three phonemes misperceived. It is important to account for the severity of errors like those so audiologists can obtain all of the information possible from the results of an individual's word recognition test.

Similarly, a distinctive feature analysis of errors may be useful to audiologists, as it could provide information about trends in the misperception of phonemes. For example, if a client misperceived all or most fricative sounds present in the words of the word recognition test, that may indicate that the client suffers from high frequency hearing loss. Phoneme based scoring methods could be developed that account for these variables and help audiologists identify the type and severity of a client's hearing loss with more accuracy and specificity. Lastly, more information from the results of an individual's word recognition test may be gained through the analysis of other linguistic components. Some examples of linguistic components that may be useful to analyze may include reversals, opposites, and rhyming. The results of this study have sparked interest in pursuing ways that phoneme error scoring and other scoring methods can be incorporated into current clinical practice.

References

Auditec, Inc. (2015, September 22). *California Consonant Test.* Retrieved from auditec.com: https://auditec.com/2015/09/22/cct/

Billings, C. J., Penman, T. M., Ellis, E. M., Baltzell, L. S., & McMillan, G. P. (2016, March). Phoneme and Word Scoring in Speech-in-Noise Audiometry. *American Journal of Audiology, Vol. 25*.

Gelfand, S. & Gelfand, J.. Psychometric Functions for Shortened Administrations of a Speech Recognition Approach Using Tri-Word Presentations and Phonemic Scoring. Journal of Speech, Language, and Hearing Research, Vol. 55, June 2012.

Guthrie, L. A., & Mackersie, C. L. (2009). A comparison of presentation levels to maximize word recognition scores. *Journal of the Academy of Audiology, 20,* 381-390.

Jusczyk, P. & Luce, P. Speech Perception and Spoken Word Recognition: Past & Present. Ear & Hearing, Vol. 23 #1, 2002.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1963). A motor theory of speech perception. Paper presented at the Proceedings of the Speech Communication Seminar, Stockholm: Royal Institute of Technology, Speech Transmission Laboratory. as cited by Jusczyk and Luce (2002)

Martin, F. N., & Greer, J. C. (2015). Speech Audiometry. In F. N. Martin, & J. C. Greer, *Introduction to Audiology, 12th edition* (pp. 98-125). Upper Saddle River, NJ: Pearson Education.

Ricketts, T., Bentler, R. A., & Mueller, H. G. (2019). *Essentials of modern hearing aids: selection, fitting, and verification.* San Diego, CA: Plural Publishing.

Schlaucb, R. S., Anderson, E. S., & Micheyl, C. (2014). A demonstration of improved precision of word recognition scores. *Journal of Speech, Language, and Hearing Research,* 543-555.

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, *107*, 776-782.

Thorton, A., & Raffin, M. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research, 21,* 507-518.