2018

A Study of the Judgment Accuracy of Repeated /r/ Stimuli by Graduate Clinicians

Lauren Glover

A Study of the Judgment Accuracy of Repeated /r/ Stimuli by Graduate Clinicians

Lauren Glover

Thesis submitted to the College of Education and Human Services at West Virginia University in partial fulfillment of the requirements for the degree of Master of Science in the Department of Communication Sciences and Disorders

Dennis M. Ruscello, Ph.D., Chair
Jeremy Donai, Ph.D.
Michelle Moore, Ph.D.

Department of Communication Sciences and Disorders

Morgantown, West Virginia
2018

Key Words: Perception, Verbal Transformation Effect, Phonology, Misarticulations

ABSTRACT

A Study of the Judgment Accuracy of Repeated /r/ Stimuli by Graduate Clinicians

Lauren Glover

In the field of speech-language pathology, communication disorders are treated with evidence-based methodologies. Treatments for many of the disorders require clinicians to use their auditory perceptual skills for feedback purposes, so that the client is aware of correct and incorrect treatment responses. It has been reported clinically that repeated listening to client responses over time may result in auditory perceptual confusions. This clinical hypothesis was studied by examining the integrity of judgement accuracy of repeated /r/ stimuli, which varied as a function of correct and incorrect stimuli. Findings showed no statistically significant evidence of auditory perceptual confusions when subjects listened to and evaluated repeated productions of /r/, /w/ for /r/substitution, and /r/ distortion embedded in a CV word.

Acknowledgements

I would like to thank the WVU Speech and Hearing Clinics, located in Allen Hall, for the continued support, equipment that was used for this study, and each of the participants. I would like to thank my thesis advisor, Dr. Dennis Ruscello, for the patience, guidance, and most importantly for being a mentor to me. Thank you to my thesis committee members, Drs. Jeremy Donai and Michelle Moore, for the time you spent investing in my education. Finally, I would like to recognize Dr. Jonah Katz for the preparation of the stimuli used in this study.

Table of Contents

Chapter I: Review of Literature

Treating Communication Disorders

In the field of speech-language pathology, individual clients present with a variety of communication disorders. The communication deficits of the client are assessed and treated through evidence-based methodologies (Orlikoff, Schiavetti, & Metz, 2015). Generally, the speech-language pathologist (SLP) follows a treatment regimen, which involves the collection of baseline data to quantify the disorder, implementation of a treatment, and measurement of the treatment to document a treatment effect. Measurement of a treatment effect typically includes two types of data. The first is generalization data and it is designed to measure transfer to untaught items of the same class being treated (Olswang, & Bain, 1994). For instance, a client who is undergoing treatment for a communication disorder might be assessed periodically to determine if there is a positive change in the behavior of interest via items that have not been introduced in treatment. A positive change would indicate that the client is exhibiting transfer to untreated items.

In addition to generalization testing, there is also measurement of the client's response to individual training trials. During the presentation of the treatment, the client provides different responses, which are subject to some form of accuracy information from the SLP. In an operant paradigm, different positive reinforcement contingencies are provided, so that the client will continue to produce the behavior of interest with greater frequency (Ruscello & Vallino, 2014). In a motor learning paradigm, it is hypothesized that the client receives both internal and external

feedback while practicing a behavior that is being treated. Both types of feedback are important variables of change for a client who is undergoing treatment. They are designed to assist the client in achieving correct use of the treatment target(s) (Wulf, Chiviacowsky, Schiller, & Ávila, 2010).

Measurement of Performance

In the cases of generalization testing and training trial accuracy, the type of information that is provided to the client may be in the form of acoustic, physiologic, or perceptual information (Maryn, 2017; Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995). While there is a lack of isomorphism among the study methods, each provides information that is important in the treatment of clients with communication disorders. For example, acoustic study involves the measurement of frequency, intensity, and time variables. This type of assessment can be performed through spectrographic analysis using programs such as Praat to analyze and synthesize speech (Maryn, 2017). Physiologic examination of speech can be carried out in a number of different ways that include measurement of different speech production parameters associated with the processes of respiration, phonation, resonation, and articulation. An example of this is in an aerodynamic study wherein different parameters of air flow and air pressure are measured to assess a specific physiologic state (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995).

The third and final study method is that of perceptual judgement, and this is the most frequently used in assessing a client's response to individual training trials. SLPs frequently make auditory perceptual judgments concerning the different parameters that comprise speech (Shuster, 1998). When assessing and treating different communication disorders, clinical judgments regarding the accuracy of responses are made throughout assessment and treatment

sessions (Gerrat, Till, Rosenbek, Wertz, & Boysen, 1991). For example, when SLPs provide treatment approaches for clients with different articulatory errors, they generally utilize auditory perceptual judgments to assess the accuracy of the client's individual training trials; however, the use of auditory perceptual judgment can be prone to different perceptual constraints as discussed by different investigators (Kent, 1996).

Issues in Auditory Perceptual Judgments

While this measurement modality, which is often considered the "gold standard" of assessment/treatment, can be economical and convenient since it requires no instrumentation, perceptual measures do have limitations in precision due to their subjective nature. Kent (1996) wrote that SLPs' perceptual judgments are grounded in the following assumptions:

(a) SLPs have a common understanding of perceptual labels such as *hoarse, nasal, rough, monoloud, excess and equal stress or stuttering*;

(b) SLPs use essentially the same verbal descriptors and associated scale values to assess a given sample of speech or voice;

(c) SLPs can isolate for judgment one perceptual dimension from several co-occurring dimensions;

(d) SLPs have a uniform reliability in judging the various dimensions that give a complete clinical portrait of speech or voice disorders; and

(e) SLPs can make perceptual judgments for which the interjudge differences are smaller than the differences needed for clinical classification or to discern changes in clinical status. (p. 7)

The issues identified above are important in developing a perspective for understanding the value and limitations of making auditory perceptual judgements. The limitations are what

Kent (1996) termed potential sources of perceptual inaccuracy. In a review of the literature, the author identified nine variables that may adversely affect auditory-perceptual judgments in the research and clinical domains. The sources include *phonemic restoration, misperceptions of natural fluent speech, phonemic false evaluation, effects of lexical status on phonetic categorization, equivalence classes in phonetic perception, McGurk Effect, prosodic influences on phonetic classification, talker-listener-utterance interactions, and the verbal transformation effect.* In the ensuing discussion, some of the sources will be discussed, because they are more germane to the proposed study.

Phonemic Restoration

Grataloup, Hoen, Veuillet, Collet, Pellegrino, and Meunier (2009) discussed the phonemic restoration effect, which was originally reported by Warren (1976). Experiments show that semantic and lexical expectations can sometimes cause illusory auditory perceptions in words that have been acoustically altered. For example, researchers have replaced parts of words with a broadband noise or a cough and listeners identified the words without any perceptual interference. Moreover, they did not report the missing portion of the word, but when the missing portion of the word was substituted by a silent pause, the deleted part of the word was detected by the listeners. The findings suggest that the phonemic restoration effect depends on the type of distortion that is applied to the test stimuli. Kent (1996) indicates that phonemic restoration reflects the top-down processing of linguistic stimuli wherein the listener applies different hypotheses that are a function of the semantic and syntactic features of the utterance.

Misperceptions of Natural Fluent Speech

Some researchers such as Bond and Garnes (1980) reported perceptual experiments in which there were errors in the perception of natural fluent speech. Kent (1996) suggests that

listeners apply neural strategies to comprehend linguistic signals, which can negatively impact their auditory perception of a speech signal. An analysis of errors of misperception identified a number of different sources of error. For example, many of the misperceptions were due to errors of individual sound segments. That is, listeners thought that they heard a sound, but that was not the segment produced in the message. In other cases, speakers simplified utterances by deleting segments, which caused problems for the perception of the listeners. Finally, some additional sources of error were erroneous word boundaries and various lexical uncertainties. Kent (1996) feels that if the perceptual effect is found in normal speech, there is justification to believe that it may also occur in disordered speech.

Phonemic False Evaluation

Phonemic false evaluation is the inaccurate identification of phonemes that were not produced by the speaker. This phenomenon was found by Buckingham and Yule (1987) in the re-analyses of different language corpora of published research studies. According to their analyses, the authors state that phonetic or allophonic variations can lead to phonemic differences in the transcription of elicited utterances. In a summary of this literature, Kent (1996) stated that investigators have implied the following three points: (a) listeners receive messages and apply normalization strategies to those messages, which may limit the perception of subtle errors; (b) speech misarticulations at the phonetic level are problematic to identify reliably, particularly when the listener must extract a message and not just identify segments; and (c) appropriate procedures for transcribing speech are insufficient for subtle variations in atypical sound productions.

Equivalence Classes in Phonetic Perception

An equivalence class is a set of related sounds; however, the constituents of the set vary in their physical characteristics but are able to elicit the class in question. Kent (1996) discusses equivalence class as it applies to the perceptual processing of native (L1) and nonnative languages (L2). When perceiving an auditory stimulus in native and non-native languages, it is common for sounds in the non-native language to become assimilated into sounds in the native language that are perceptually similar. The two sounds from different languages are often treated as equivalent when the speaker produces the sounds. This equivalence is not likely to occur when sounds from different languages are not perceptually/acoustically similar. The basic principles of this concept may be relevant to the SLP who is making clinical judgements when there are speech abnormalities present and there is a contrast between correct sound(s) and nondevelopmental error sound(s), which are often classified as distortions.

McGurk Effect

The McGurk Effect is a very interesting phenomenon that demonstrates how different types of perceptual information can interact to cause perceptual confusions (Strand, Cooperman, Rowe, & Simenstad, 2014). That is, presentation of both competing auditory and visual information can influence auditory perception. For instance, experiments have shown that tokens such as /da/ are perceived as /da/ in an auditory condition alone but may be perceived as /ba/ when supplemented with synchronized visual information demonstrating labial closure. Strand et al. (2014) writes that the McGurk effect is a strong illusion. Research has demonstrated that it can happen when the face and voice are of different gender, when subjects are informed that the auditory and visual stimuli may not correspond, and even when the auditory and visual signals were recorded by the subject experiencing the illusionary effect. This suggests that judgments of speech can be misperceived when the listener utilizes simultaneous

auditory and visual information, which may differ in terms of the perceptual information provided.

Verbal Transformation Effect

The final source of perceptual inaccuracy to be discussed and one that is relevant to the proposed study is the verbal transformation effect (Warren, 1976). This occurs when an identical auditory stimulus is repeatedly heard but is perceived by the listener to be phonetically altered (Sato, Vallee, Schwartz, & Rousett, 2007). Initially, a percept matching of the original prime stimulus is perceived, but at some time, erroneous words or nonwords are identified, which signal a change in the perceived stimulus. The transformation effect continues throughout the presentation of the stimuli, leading to perceptual shifts from one presentation to another or a return to the original form. The study completed in 2007, mentioned above, found that perceptual alterations of the original stimulus could range from small phonetic changes to semantic confusions.

Different explanations have been provided to explain the verbal transformation effect, and they include both linguistic and perceptual processing explanations. Some hypothesize that the verbal transformation effect occurs when lexical activation for a target word extends to phonologically or semantically related words in the person's lexicon. The related words then compete with the presented stimuli as potential perceptual responses. An alternative explanation is that the verbal transformation effect occurs when the peripheral or central processing structures that are triggered in response to the target stimuli are fatigued by repeated stimulation.

The direct investigation of the verbal transformation effect as it applies to the continuous judgment of children's misarticulations has not been studied formally, but there are some

investigations that have examined judgments of repeated similar misarticulated stimuli. For example, Shriberg (1972) assessed the inter-reliability and intra-reliability of articulatory judgments among five experienced SLPs. The SLPs judged audio tapes of four children with nonstandard /r/ and four with deviant /s/. Results suggested that judgements of postvocalic /r/ items had the most varied judgements present. The findings showed the following: listeners shifted their judgments when listening to different but repeated misarticulations, and additionally, there was a greater likelihood for reliability to be better when there was a large percentage of the speech stimuli that was perceived as either correct or incorrect.

Additional work relating to perceptual judgements of the /r/ in repeated tokens was investigated in a cue-trading task reported by Wolfe, Martin, Borton, and Youngblood (2003). The authors used a /r-w/ cue-trading task to examine the perceptual abilities of graduate students with and without clinical experience. Vowel formants were altered by balancing a temporal-spectral cue on F2 against a spectral cue on F3. Subjects listened to the different stimuli and were required to identify the tokens as containing the /r/ or /w/ speech sounds. The participating speech-language pathology graduate students who had previous clinical experience exhibited better perceptual sensitivity in determining when a sound was more perceptually similar to a canonical /r/ or /w/ than graduate students who did not have exposure to the clinical setting. The results of this preliminary study indicated that a task utilizing a cue trading paradigm could be useful in evaluating perceptual sensitivity to the acoustic cues representative of misarticulated speech sounds.

A final study conducted by Munson and Brinkman (2004) also involved listening to repeated stimuli but the investigators were studying the reliability of perceptual judgments in the administration of an articulation test. That is, they had subjects carry out transcriptions across

multiple sampling periods to determine if such an experimental task posed a threat to the reliability of clinical perceptual judgements. The experimental group transcribed words produced by children with speech sound disorders. The listening conditions included a single presentation of the word and seven presentations of the word. Their data consisted of transcriptions of /s/ across various word tokens. Transcribers participated in two separate transcription sessions, with a week between the two sessions. The experimenters found no systematic effect of presentation condition on either accuracy of judgments or intra-rater reliability. In addition, higher interrater reliability was found in the multiple-presentation condition, especially for /s/ tokens that were inaccurate or acoustically intermediate between correct and incorrect words. The data reported imply that multiple presentations of the stimuli had no measurable effect on correct judgments and intra-rater reliability of children's phonetic accuracy, but they did have a minimal effect on inter-rater reliability.

The studies discussed in this section did not specifically set out to examine the verbal transformation effect but the use of repeated approximate stimuli conditions that could elicit the perceptual illusion are consistent with the study of the effect. The data reported by Shriberg (1972) did show shifts in judgment but the data reported by Munson and Brinkman did not. It is to be noted that the difference may be in the fact that the stimuli were not presented at a high level of repetition in the Munson and Brinkman study; consequently, there was insufficient stimuli to demonstrate the effect.

Statement of the Problem

It is our contention that day to day evaluations of articulatory responses may in fact trigger the verbal transformation effect. SLPs employed in school systems have substantial caseloads of children with various misarticulations and repeatedly listen and evaluate the

responses of children who present with different errors. This potential issue poses a threat to the reliability of clinical judgements, which constitute the data for making decisions regarding continuation and/or dismissal from a treatment program. If the verbal transformation effect is a clinical reality, SLPs need to be aware of such a problem and identify ways to minimize the problem.

In terms of speech sound disorders, 6% of all school-aged children show difficulty in the acquisition of the eight late-acquired sounds and often receive treatment for those errors (Bleile, 2014). The late eight include / θ, ð, s, z, l, r, ʃ, ʒ/, and in particular /r/ is problematic for many children. It can also be problematic for SLPs because its perceptual characteristics differ among correct /r/, distorted or derhoticized /r/, and /w/ in substitution of /r/ (Ohde & Sharf, 1992). It is possible that the different variants of correct and incorrect /r/ may trigger the verbal transformation effect. For instance, an SLP listening repeatedly to tokens of distorted /r/ may shift in her/his perceptual evaluation of stimuli, thus providing incorrect feedback and/or reinforcement to clients. This might also occur for correct /r/ and /w/ in substitution of /r/.

Based on previous studies and clinical reports from SLPs who are employed in the public schools, we predict that repeatedly listening to the same stimuli will trigger the verbal transformation effect and the degree of shifts in perceptual judgment will differ as a function of the token. That is, listeners will experience more perceptual shifts with derhoticized /r/ than /w/ in substitution of /r/, which in turn will result in more perceptual shifts than finally correct /r/. This investigation will examine listeners' ability to make reliable perceptual judgements when presented with repeated speech stimuli that contain the speech sound /r/ or its error variants.

Chapter II: Methods

Subjects

Subjects for this study included 30 graduate students from West Virginia University who were enrolled in the Communication Sciences and Disorders master's degree program. The subjects consisted of a total of twenty-nine females and one male with a mean age of 23 years, 3 months. To qualify for the investigation subjects must have: (1) completed at least one semester of graduate coursework including a course in speech sound disorders, (2) provided or currently provide speech treatment to at least one client with a speech sound disorder, and (3) passed screenings of speech, hearing, and auditory perception. The investigator recruited these subjects by contacting them via written invitation that was posted in the Allen Hall Speech and Hearing Center student clinic preparation room. The nature of the study was explained to the participants, and a written IRB consent form (1803040822) was obtained prior to conduct of data collection. Please refer to Appendix A for a copy of the IRB form.

For each participant in the investigation, the examiner scheduled a time to conduct the preliminary screening procedures (Qualifying Criteria) and administer the experimental stimuli (Listening Task). The following screening procedures were administered by the experimenter:

(1) Pure tone hearing screening of 25dB at 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz;

(2) Articulation screening;

(3) Passing a modified variation of the Speech Production Perception Test for /r/ (Locke, 1980).

Subjects were required to pass all screening measures in order to participate in the investigation. Hearing was screened with a Grason-Stadler GSI 17 portable audiometer and speech was assessed through administrations of the Rainbow Passage (Fairbanks, 1960). The passage provides a representative sample of English speech sounds in a reading context. The examiner requested that the experimental subjects read the passage to screen articulation skills. A modified version of the Speech Production Perception Test (Locke, 1980) was also administered by the experimenter. This test examined a subject's perception of /r/ versus substitution of /w/ and a control sound /l/, which shares phonetic features with /r/. Subjects were required to make perceptual judgments of a target item, which contained /r/. The testing procedure required the experimental subjects to listen and view a picture of the target item (rake). Six random presentations were accompanied by appropriate articulation of /r/ (Is this a rake?), 6 consisted of a /w/ production (Is this a wake?) and 6 consisted of an /l/ production (Is this a lake?). The derhotocized or distorted /r/ was not included in the assessment, because it is not a phoneme of English and would not be consistent with the test format of assessing a target sound versus its English sound substitutions. Subjects were required to pass all screening measures in order to participate in the investigation. Please refer to Appendix B for the screening measures that were employed.

Stimuli

The speech sound stimulus selected for study in the current investigation was the liquid /r/. The rationale for selection was that the speech sound is frequently misarticulated by children. In addition, there are two error variants that are described in the literature, which include a substitution of /w/ in place of /r/ and a distortion of /r/ that is classified as a derhoticized version (Ohde & Sharf, 1992; Rvachew & Brosseau-Lapre, 2018). According to

Ohde and Sharf, /r/ is a speech sound that has rhotic qualities or /r/-coloring, which is the major feature of the auditory percept of /r/. Its lingual production feature is generally described as either bunched or retroflex, although some linguists describe the tongue position as being variable among speakers. Kent and Read (1993) indicated that the most distinctive acoustic feature of /r/ is a lowered F3 that is in close proximity with F2. The /r/ speech sound differs from the other liquid /l/ in terms of the duration of F1(brief steady state and extended transition), the onset frequency of F2 (low onset frequency versus high onset frequency), and the onset frequency of F3 (low F3 versus a high F3).

The derhoticized or distorted /r/ is a substitution that is devoid of an /r/-quality production due to a deviant lingual articulation (Ohde & Sharf, 1992). The descriptions of the production vary by author to include excessive lip rounding, a lack of lingual bunching or retroflexion, and lingual positioning that is either too forward or too far posterior in the oral cavity. Acoustically, comparison with /r/ shows that derhoticized /r/ may present with two different variations that include F3 is too high, or F3 is too high and F2 is lower than expected. In an investigation conducted by Sharf, Ohde, and Lehman (1988), listeners were exposed to synthetic stimuli and required to identify /r/, /w/, and derhoticized /r/. A synthetic token with an F2 onset value of 1725 Hz and F3 onset frequency of 3200 Hz was most often identified as derhoticized /r/. Reliable /r/ identification was found at the onset frequencies of 1325 Hz (F2) and 2600 Hz (F3) and /w/ at the onset frequencies of 925 Hz (F2) and 3800 Hz (F3).

It has been argued that substitution of /w/ in place of /r/ is not the same production as the /w/ used in an appropriate context (i.e. /ret/→/wet/) but most literature does identify the error as one of a sound substitution in place of /r/ (Bleile, 2014). There are a number of acoustic cues that differentiate /r/ from /w/. The most distinctive are the frequencies of F2 and F3 during

placement and onset of the transition into the adjoining vowel. As noted previously in experiments using synthetic stimuli, the second and third formants are very close for /r/ but spaced apart for /w/.

Based on the experiments of Sharf, Ohde, and Lehman (1988), synthetic CV stimuli of /r/, /w/, and derhoticized /r/ with the midfront vowel /e/ were created for this current study using the Pratt 6.0.21 articulatory synthesis version. The development of the stimuli was carried out by a linguist with expertise in articulatory synthesis and with assistance of the thesis Director (DMR). The tokens were then presented to three certified SLPs with a total of 105 years of clinical experience. Each listened individually and were instructed to identify the /r/ stimuli, the /w/ stimuli, and the distorted /r/. A total of 12 stimuli, 4 of each synthetic CV, were played randomly to each of the SLPs. Each identified the stimuli reliably with a total of 3 selection errors among the 36 presentations of the stimuli.

The stimulus tokens were then electronically copied into sets of 30 stimuli at a presentation level of 65db-SPL using cross-platform software for recording and editing auditory stimuli). There was an inter-stimulus pause of 2.5 seconds between each token. Upon completion of the editing, there were 3 sets of 30 stimuli from each of the synthesized stimuli. That is, there was a set of 30 repeated stimuli of the synthesized stimulus ray (/re/) designated R, 30 repeated stimuli of the synthesized stimulus way (/we/) designated W, and 30 repeated stimuli of derhoticized /r/ (/re/) designated WR. In total, there were 90 tokens, 30 from each of the three synthesized tokens. The repeated stimuli were then downloaded to a MacBook Pro.

Procedure

The 3 stimuli sets (R, W, WR) were presented to all subjects, with order of presentation counterbalanced across three subgroups of subjects, each of which contained 10 randomly

assigned subjects from the group of 30 subjects. The rationale for such a design feature was to minimize order effects as a confounding factor, and it allowed the experimenter to control and measure sequencing effects (Orlikoff, Schiavetti, & Metz, 2015). Participants who were presented with presentation order 1 (Group I) were first presented with 30 repeated tokens of the CV syllable beginning with the derhoticized /r/ 'WR', followed by 30 repeated presentations of the CV syllable beginning with the /w/ 'W', and then finally 30 repeated presentations of the CV syllable beginning with the accurately produced /r/ 'R'. The presentation order for Group II was W, R, and WR, and the ordering for Group III was R, WR, and W. Please refer to Table 1-2 for a summary of the counterbalancing.

Table 1-2. Listing of the counterbalanced stimuli presentation by group.

Orders of Presentation	Stimuli	Stimuli	Stimuli
Group I	WR	W	R
Group II	W	R	WR
Group III	R	R	W

Screenings of the subjects were conducted in a clinic room at the Allen Hall Speech and Hearing Center at a time when therapy was not in session in the remaining rooms. The consenting subjects who passed the screening measures completed the listening task order for that randomly assigned subgroup (See Table 1-2). All subjects completed the screening and listening tasks in one session of approximately 30-45 minutes.

The listening task was administered via Beyerdynamic DT 211 headphones through a MacBook Pro. Subjects were instructed that they were going to listen to some words that

contained the /r/ speech sound and that they were going to judge each item as either right or wrong. After listening to each word, they were required to mark a score sheet. The stimuli were presented diotically at 65 dBSPL. Participants were not provided information regarding the number of trials nor their accuracy judgments for individual trials.

The following directions were read to each subject:

> You are going to listen to some words that will be played through the headphones. The words begin with the /r/ sound and you need to judge each word as either right or wrong. After you listen to a word, mark your score sheet as either right or wrong, and then you will hear another word until you are finished. The listening will not take too long so listen carefully. If you are experiencing any problems, stop and let me know.

See Appendix C for a score sheet that was used by the subjects.

Chapter III: Results

Descriptive Statistics

The mean percentage values and the standard deviations for the three subgroups across the three listening conditions are presented in Table 1-3. The scores show that the subjects exhibited a high level of response accuracy for each of the different test tokens. In most cases mean values are at 80% or higher indicating that they were generally correct in their assessment of the different /r/ tokens. The only difference to this trend is Group I when judging the /w/ substitution for /r/ token. Some of the subjects experienced difficulty with the judgment task, which resulted in a mean score of 69%; however, this decrement in performance was not observed in the other groups for that test token or the other test tokens.

Standard deviations show a variable pattern with both limited and substantial dispersion from the mean scores. Those standard deviations associated with substantial dispersion are reflective of some limited subject outliers across the three listening groups. That is, most subjects displayed a high level of judgment accuracy, but some listeners in each group did not and this is reflected in the large standard deviations for some of the test tokens. For instance, Group III listeners demonstrated a mean score value of 82 and a standard deviation of 36 when they judged the /r/ distortion token. Only 50% of the subjects in Group III were accurate in their judgments, which is reflected in the sizable standard deviation. Again, there was no definite trend for a listening group, but /w/ for /r/ substitution showed the most variability across the three listening subgroups.

Table 1-3. Mean values and standard deviations for Groups across the different test tokens.

Groups	Group I	Group II	Group III
Listening Token			
/r/	$\overline{x} = 100$ $\sigma = 1$	$\overline{x} = 81$ $\sigma = 33$	$\overline{x} = 100$ $\sigma = 2$
/r/ → /w/	$\overline{x} = 69$ $\sigma = 47$	$\overline{x} = 89$ $\sigma = 27$	$\overline{x} = 87$ $\sigma = 18$
/r/ Distortion	$\overline{x} = 98$ $\sigma = 4$	$\overline{x} = 99$ $\sigma = 2$	$\overline{x} = 82$ $\sigma = 36$

See Figure 1-3 for a graphic representation of the results described herein. Note upon inspection that the tokens beginning with the accurately produced /r/ and /r/ distortion appear to have the overall highest judged correctness, while the tokens with the /w/ for /r/ substitution appear to have the lowest percent correctness. However, despite the lower values

listeners also exhibited a high level of judgment accuracy for that token.

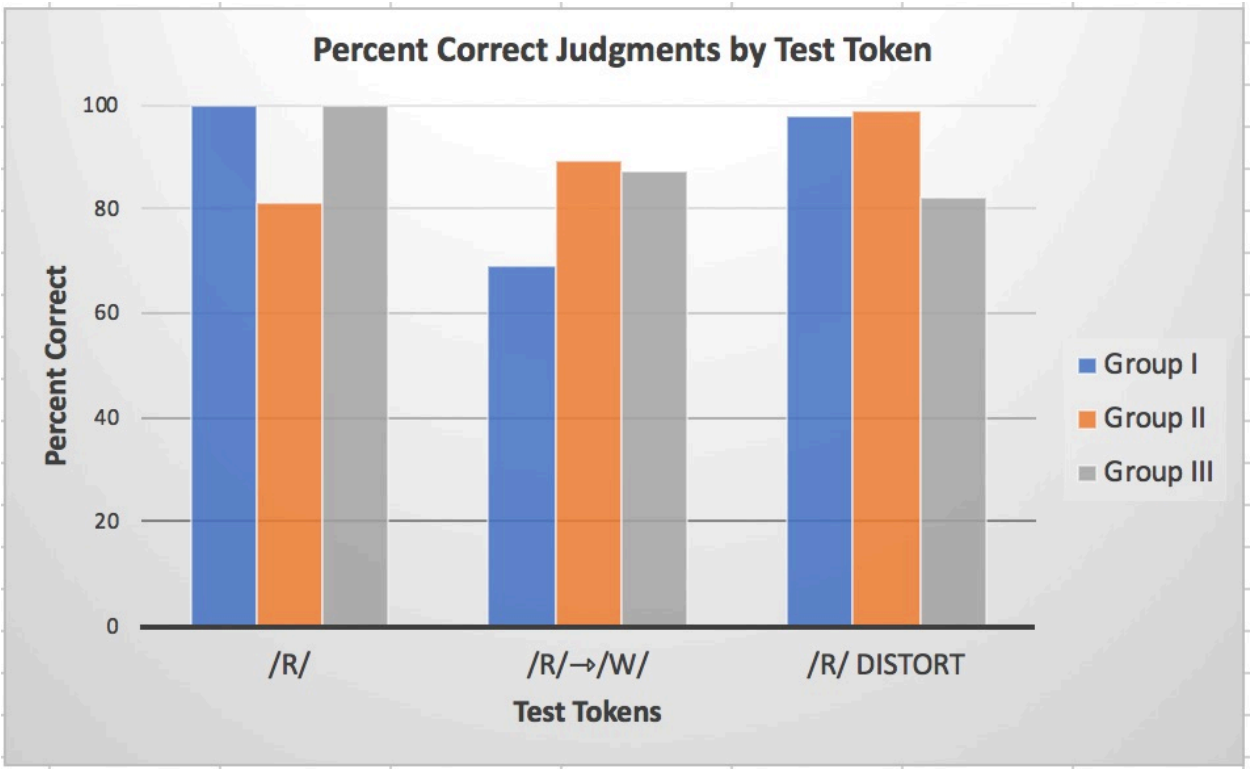

Figure 1-3. Mean percentage scores of the different groups for the three test tokens.

The final graphic shown in Figure 2-3 is that of mean scores by order of presentation. Recall that listeners judged the different test tokens in blocks of 30 trial presentations that were administered in counterbalanced order. It would appear that, across the different orders of presentation, the second condition resulted in slightly lower accuracy judgment scores regardless of the stimuli; however, the reduction in performance is minor when compared with the first and third conditions. Overall the descriptive data show some slight drops in performance that appear to be subject specific. That is, 6 subjects experienced difficulty with the listening tasks but most did not, as shown by aggregate performance typically at the level of 80% or higher.

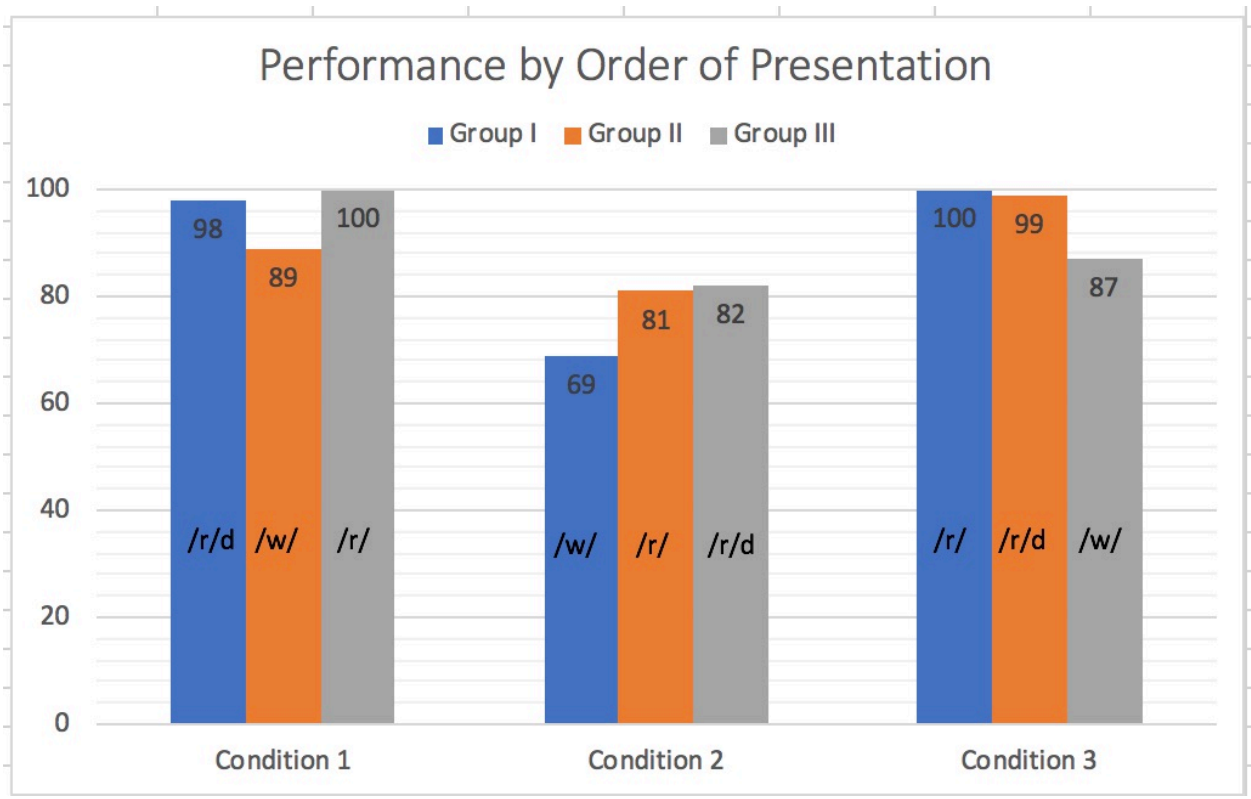

Figure 2-3. Mean percentage scores of the different groups by order of presentation.

Inferential Statistics

Inferential statistical analysis was conducted to determine if the descriptive differences observed among groups with respect to the different listening conditions were statistically significant. The first analysis consisted of the computation of a one-way ANOVA using SPSS statistical software to determine whether there were significant differences across the different tokes. The three different tokens or levels of the independent variable were correct /r/, /w/ for /r/ substitution, and distorted /r/, and the dependent measure was the judgment accuracy. There were no statistically significant differences among the different tokens, F (2, 58) = 1.941, p = .153 ηp^2 = .063.

To examine possible accuracy judgments shifts within a listening condition, a 2 x 3 repeated measures analysis of variance (ANOVA) was computed. The analysis was done to test for differences between the first half of the trials within a specific token listening condition compared to the second half of the trials across the three-listening conditions. It was hypothesized that accuracy judgments for a token may change as subjects judged individual presentations of that particular token. The main effect of listening condition approached statistical significance, but did not reach a level of significance to reject the null hypothesis, $F(2, 58) = 2.47$, $p = 0.093$, $\eta p^2 = 0.079$. There was no significant main effect for the first fifteen presentations of a token versus the last fifteen presentations of the same token (referred to in this paper as split-half), $F(1, 29) = .020$, $p = .889$, $\eta p^2 = 0.001$. In addition, there was no significant interaction effect between test token and split-half, since the listeners did not perform differently on the first fifteen presentations compared to the last fifteen presentations of a token as a function of condition, $F(2, 58) = .512$, $p = .602$, $\eta p^2 = 0.017$.

Because the main effect of listening condition trended toward significance, pairwise comparisons and Cohen's *d* were computed for the data. The difference between /r/ and /w/ for /r/ substitution trended toward significance ($p = 0.06$) but did not reach a level for rejection of the null hypothesis. The other comparisons were also not significant ($p \geq 0.12$). Cohen's *d* is a statistic that is used to estimate an effect size of the difference between two means, and thus was computed to evaluate the clinical relevancy of the results. The Cohen's *d* values comparing /w/ for /r/ substitution and derhoticized /r/ and /w/ for /r/ substitution and /r/ demonstrated what would be characterized as small to moderate effect sizes as shown in Table 2-3.

Table 2-3. Cohen's *d* values.

	Cohen's *d* value
/w/ - derhoticized /r/	.39
/w/ - /r/	.40

Summary of Results

Most of the subjects exhibited a high percentage of correctness when judging the accuracy of the different test tokens. There were some subjects in the three different subgroups who did not, resulting in a limited number of outliers. Upon inspection of the results respective to the different test tokens, there were no statistically significant differences. In addition, the findings did not show a significant difference between the first half of the presentations of a test token compared to the second half. That is, there did not appear to be a perceptual shift in the judgment accuracy of the different tokens with repeated presentations.

Chapter IV: Discussion and Conclusions

Study Findings

Based upon the findings, the investigator could not reject the null hypothesis. That is, statistical testing indicated no significant differences among groups when subjects evaluated the different test tokens. Moreover, the majority of subjects performed at a very high level of response accuracy. It was hypothesized that judgment shifts would occur across the stimuli with repeated listening conditions that are similar to an actual clinical situation. This was based on clinical reports of judgment error when engaging in actual clinic treatments that utilized

continuous feedback given by an SLP. That is, SLPs have reported that they are unable to discriminate auditorily among correct target sounds and misarticulations in some situations when required to listen and evaluate repeated target stimuli. In order to make the task as difficult as possible, we used /r/ stimuli because the error variants include both phonemic (sound substitution) and nonphonemic (/r/ distortion) misarticulations. Additionally, /r/ is frequently in error and proven to be a difficult remediation target (Shuster, 1998), so we hypothesized that it would be challenging for the study participants.

We had proposed that the shifts in judgment that would be identified would be theoretically consistent with the verbal transformation effect (Warren, 1976) which has consistently demonstrated shifts in judgment when individuals repeatedly listen to various linguistic stimuli. This would indicate that judgments of correct and incorrect stimuli would show the same results as Warren's subjects who listened to stimuli that did not include speech sound errors. However, our findings are not consistent with previous clinical reports and certainly do not support performance explained by the verbal transformation effect. However, the current findings do not negate the clinical hypothesis, but rather suggest the need for further study.

Extension of the Current Research

It is to be noted that this study is the first that we are aware of that empirically examined the experimental question in the theoretical context of the verbal transformation effect. The findings are preliminary in nature and the trends toward statistical significant indicate the need for further study. The early investigation that was reported by Shriberg (1972) did report judgment shifts in perceptual accuracy, but that was a clinical study which included 5 experienced judges. His findings indicated that judgements of postvocalic items had the most

varied judgements of accuracy. He also found that listeners shifted their judgments when listening to different but repeated misarticulations. These variables should be considered in an extension of the current study. For instance, an alteration in sound position such as presenting stimuli in postvocalic position might be implemented along with an increase in the number of stimuli presented. Each of the current stimuli involved 30 presentations and it may be that an increase of 75 to 90 stimuli of each would trigger accuracy judgment changes. In actual clinical situations the SLP is evaluating a number of different clients with speech sound disorders and judging numerous target productions. The employ of added stimuli would be consistent with typical clinical practice and may allow for a finer grained assessment of a practitioner's judgment skills to determine if perceptual shifts are a clinical reality.

Clinical Implications

There are several clinical implications of the study that are relevant to both students and practitioners. First, the vast majority of our student subjects were accurate in their judgments of the different stimuli. They all had clinical experience but not a vast amount of clinical experience when compared to a practitioner. Their performance certainly indicates that they were capable of completing the listening tasks successfully; however, there was a small subgroup who did experience problems and could probably benefit from additional perceptual training in graduate school. Educational programs in the major might want to include additional listening experiences for student clinicians who demonstrate difficulties in the perceptual judgment of correct and incorrect articulatory responses (Dudding & Notingham, 2018). This could also be extended to practitioners in a variety of ways that might include continuing education credit for participating in a refresher activity to re-calibrate with colleagues. Such experiences can assist those SLPs who must work in such venues as schools and preschool clinics. They are exposed to

a variety of speech sound disorders and may vary in their assessment of different clients on certain days or in relation to a specific speech sound error. These are important concerns since their perceptual judgments are the basis for who will continue treatment or be dismissed.

Conclusions

1. The data generated from this investigation do not support the rejection of the null hypothesis.

2. Generally, subjects performed at a high level of accuracy in the judgment of /r/, /w/ for /r/ substitution, and distorted /r/. There were only six subjects (20%) of the 30 participants who experienced some difficulty with the listening tasks, but there was no trend of poorer performance associated with one of the three stimuli.

3. Although the investigation was inconclusive, there are empirical and clinical implications.

    a. This was a preliminary investigation and further study is needed to verify the clinical hypothesis that listening to repeated stimuli may cause changes in perceptual judgment.

    b. Future SLPs and current practitioners need to be cognizant of this potential problem and take action based on student mentoring and/or inservice activities.

References

Bernthal, J. E., Bankson, N. W., & Flipsen, Jr. (2017). *Articulation and phonological disorders (8th ed.).* Boston, MA:Pearson.

Bleile, K. M. (2014). *The late eight (2nd ed.).* San Diego: Plural Publishing.

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.), Perception and production of fluent speech (pp. 115–132). Hillsdale, NJ: Erlbaum.

Buckingham, H. W., & Yule, G. (1987). Phonemic false evaluation: Theoretical and clinical aspects. Clinical Linguistics and Phonetics, 1, 113–125.

Donai, J. J. & Halbritter, R. M. (2017). Gender identification using high-frequency speech energy: effects of increasing the low-frequency limit. *Ear and Hearing*, 38, 65-73.

Dudding, C. C. & Nottingham, E. E. (2018). Simulation use in university programs in communication sciences and disorders. *American Journal of Speech-Language Pathology*, 27, 71-81.

Fairbanks, G. (1960). Voice and articulation drillbook, 2nd edn. New York: Harper & Row. pp124-139.

Gerratt, B. R., Till, J. A., Rosenbek, J. C., Wertz, R. T., & Boysen, A. E. (1991). Use and perceived value of perceptual and instrumental measures in dysarthria management. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), Dysarthria and Apraxia of Speech: Perspectives on Management (pp. 77–93).

Grataloup, C., Hoen, M., Veuillet, E., Collet, L., Pellegrino, F., & Meunier, F. (2009). Speech restoration: an interactive process. *J Speech Lang Hear Res*, 52(4), 827-838.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *J Speech Hear Res*, 38(6), 1212-1223.

Kent, R. D. (1996). hearing and believing some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech Lang Pathol*, 5(3), 7-23.

Kent, R. D. & Read, C. (1992). The acoustic analysis of speech. San Diego: Singular Publishing.

Maryn, Y. (2017). Practical acoustics in clinical voice assessment: a praat primer. *Perspect ASHA SIGs, 2* (SIG 3), 14–32.

Munson, B. & Brinkman, K. N. (2004). The influence of multiple presentations on judgments of children's phonetic accuracy. *Am J Speech Lang Pathol*, 13(4), 341-354.

Ohde, R. N. & Sharf, D. J. (1992). *Phonetic analysis of normal and abnormal speech.* New York: Merrill.

Oller, D. K. & Ramsdell, H. L. (2006). A weighted reliability measure for phonetic transcription. *J Speech Lang Hear Res*, 49(6), 1391-1411.

Olswang, L. B. & Bain, B. (1994). Data collection monitoring children's treatment progress. *Am J Speech Lang Pathol*, 3(3), 55-66.

Orlikoff, R.F., Schiavetti, N., & Metz, D.E. (2015). Evaluating research in communication disorders (7th ed.) (pp. 10-11). Upper Saddle River, New Jersey: Pearson Education, Inc.

Ruscello, D. & Vallino, L. (2014). The application of motor learning concepts to the treatment of children with compensatory speech sound errors. *Perspect Speech Sci Orofac Disord*, 24(2), 39-47. doi: 10.1044/ssod24.2.39.

Sato, M., Vallée, N., Schwartz, J., & Rousset, I. (2007). A perceptual correlate of the labial-coronal effect. *J Speech Lang Hear Res*, 50(6), 1466-1480. doi: 10.1044/1092-4388(2007/101).

Sharf, D. J. Ohde, R. N., & Lehman, M. E. (1988). Relationship between the discrimination of /w-r/ and /t-d/ continua and the identification of distorted /r/. *Journal of Speech and Hearing Research*, 31, 193-206.

Shriberg, L. D. (1972). Articulation judgments: some perceptual considerations. *J Speech Hear Res*, 15(4), 876-882. doi: 10.1044/jshr.1504.876.

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research*, 41, 941-950.

Strand, J., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual differences in susceptibility to the mcgurk effect: links with lipreading and detecting audiovisual incongruity. *J Speech Lang Hear Res*, 57(6), 2322-2331. doi: 10.1044/2014_JSLHR-H-14-0059.

Tuller, B. & Kelso, J. A. S. (1991). The production and perception of syllable structure. *J Speech Hear Res*, 34(3), 501-508. doi: 10.1044/jshr.3403.501.

Warren, R. (1976). Auditory illusions and perceptual processes. In N. J. Lass (Ed.), Contemporary issues in experimental phonetics (pp. 389–418). New York: Academic Press.

Wolfe, V., Martin, D., Borton, T., & Youngblood, H. C. (2003). The effect of clinical experience on cue trading for the /r-w/ contrast. *Am J Speech Lang Pathol*, 12(2), 221-228. doi: 10.1044/1058-0360(2003/068).

Wulf, G., Chiviacowsky, S., Schiller, E., & Ávila, L. T. G. (2010). Frequent external-focus feedback enhances motor learning. *Frontiers in Psychology*, *1*, 190. http://doi.org/10.3389/fpsyg.2010.00190

Appendix A

Only Minimal Risk
Consent Information and HIPAA Form

Principal Investigator	Dennis Ruscello-Dennis.Ruscello@mail.wvu.edu
Department	Department of Communication Sciences and Disorders
Protocol Number	1803040822
Study Title	**A Study of the Judgment Accuracy of Repeated /r/ Stimuli by Graduate Clinicians**
Co-Investigator(s)	Mrs. Lauren Glover- lmckee1@mix.wvu.edu

Contact Persons

Mrs. Lauren Glover, B. S.
In the event you experience any side effects or injury related to this research, you should contact Mrs. Lauren Glover at (304) 293-2894. (After hours contact: Mrs. Glover at (304) 276-2721. If you have any questions, concerns, or complaints about this research, you can contact Dr. Dennis Ruscello at (304) 293-2894.

For information regarding your rights as a research subject, to discuss problems, concerns, or suggestions related to the research, to obtain information or offer input about the research, contact the Office of Research Integrity and Compliance at (304) 293-7073.
In addition if you would like to discuss problems, concerns, have suggestions related to research, or would like to offer input about the research, contact the Office of Research Integrity and Compliance at 304-293-7073.

Introduction

You, _____, have been asked to participate in this research study, which has been explained to you by Mrs. Lauren Glover or Dr. Ruscello. This study is being conducted by Mrs. Glover and Dr. Ruscello in the Department of Communication Sciences and Disorders at West Virginia University.

Purpose(s) of the Study

The purpose of this investigation is to study your listening skills for words that contain the /r/ speech sound. This speech sound is frequently a problem for children to produce, and we want to know how a listener will judge a list of words that have the /r/ sound. This task is similar to the judgments that you make when providing treatment services to the children who attend the Allen Hall Speech and Hearing Center. This study will help us understand how listeners judge pronunciations of the /r/ speech sound, since it can differ in its pronunciation.

Description of Procedures

This study involves listening to a list of 90 words. You will listen to the words through headphones in a quiet room. After each word is presented, you will judge the word as either right or wrong and mark your response on score sheet. Before you listen to the words, your hearing, speech, and auditory perception will be screened. Pure tone hearing screening will be carried out with an audiometer. When you hear a tone in your ear, you will raise your hand. Your speech will be screened by reading a paragraph that contains all the sounds of the English language. Finally, we will check your auditory discrimination by having you point to a picture of an item that is shown to you. The listening task and the screenings will take approximately 60 minutes for you to complete.

Discomforts

There are no known or expected risks from participating in this study, except for the mild frustration associated with the listening task.

Alternatives

You do not have to participate in this study. You are under no obligation to participate and there is no penalty for not participating.

Benefits

You may not receive any direct benefit from this study. The knowledge gained from this study may eventually benefit others.

Financial Considerations

There is no financial payment for participation in this study.

Confidentiality

Any information about you that is obtained as a result of your participation in this research will be kept as confidential as legally possible. Your research records and test results, just like hospital records, may be subpoenaed by court order or may be inspected by the study sponsor or federal regulatory authorities without your additional consent.

Audiotapes or videotapes will be kept locked up and will be destroyed as soon as possible after the research is finished.
In any publications that result from this research, neither your name nor any information from which you might be identified will be published without your consent.

HIPAA

We know that information about you and your health is private. We are dedicated to protecting the privacy of that information. Because of this promise, we must get your written authorization (permission) before we may use or disclose your protected health information or share it with others for research purposes.
You can decide to sign or not to sign this authorization section. However, if you choose not to sign this authorization, you will not be able to take part in the research study. Whatever choice you make about this research study will not have an effect on your access to medical care.

Persons/Organizations Providing the Information

You as the participant in the study.

Persons/Organizations Receiving the Information

The research site(s) carrying out this study, which includes WVU and the research investigators.

The Following Information Will Be Used

The response judgments to the stimuli (90 /r/ words) will be tabulated and group aggregate data will be used in the analysis. The judgments of the different /r/ productions will be identified. Individual subject response data will not be used.

The Information is Being Disclosed for the Following Reasons

Publication of study results (without identifying you)

You May Cancel this Authorization at Any Time by Writing to the Principal Investigator

Dennis M. Ruscello, Ph.D.
805 Allen Hall, P.O. 6122
West Virginia University
Morgantown, WV 26506-6122
Denis.Ruscello@mail.wvu.edu

If you cancel this authorization, any information that was collected already for this study cannot be withdrawn. Once information is disclosed, according to this authorization, the recipient may redisclose it and then the information may no longer be protected by federal regulations.

You have a right to see and make copies of your medical records. You will not be able to see or copy your records related to the study until the sponsor has completed all work related to the study. At that time you may ask to see the study doctor's files related to your participation in the study and have the study doctor correct any information about you that is wrong.

This authorization will expire at the end of the study unless you cancel it before that time (or has a specific expiration date).

Voluntary Participation

Participation in this study is voluntary. You are free to withdraw your consent to participate in this study at any time.

Refusal to participate or withdrawal will not affect your class standing or grades, as appropriate and will involve no penalty to you. Refusal to participate or withdrawal will not affect your future care at West Virginia University. In the event new information becomes available that may affect your willingness to participate in this study, this information will be given to you so that you can make an informed decision about whether or not to continue your participation.
You have been given the opportunity to ask questions about the research, and you have received answers concerning areas you did not understand.

Upon signing this form, you will receive a copy.

I willingly consent to participate in this research.

Signatures

Signature of Subject

_____

Printed Name Date Time

_____

The participant has had the opportunity to have questions addressed. The participant willingly agrees to be in the study.

Signature of Investigator or Co-Investigator

_____

Printed Name Date Time

_____

Appendix B

Subject ID Number _____ Date _____

Group Assignment 1 ___ 2 ___ 3 ___

1. Hearing Screening @ 25 dB

Left Ear ___ 500 ___ 1000 ___ 2000 ___ 4000

Right Ear ___ 500 ___ 1000 ___ 2000 ___ 4000

Pass ___ Fail ___

Speech Screening

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling point of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Pass ___ Fail ___

Speech Production Perception Test

Target sound /r/ Error /w/ Control Sound /l/

Instructions: I am going to show you a picture and say the name of it. I will say it different ways and I want you to tell me if I say it the right way or the wrong way. You will listen to 18 trails. Listen: Is this a _____?

1. /w/ Error	**yes/No**
2. /l/ Control	**yes/No**
3. /r/ Target	**Yes/no**
4. /r/ Target	**Yes/no**
5. /w/ Error	**yes/No**
6. /l/ Control	**yes/No**
7. /r/ Target	**Yes/no**
8. /l/ Control	**no/Yes**
9. /w/ Error	**yes/No**
10. /r/ Target	**Yes/no**
11. /w/ Error	**yes/No**
12. /l/ Control	**yes/No**
13. /w/ Error	**yes/No**
14. /r/ Target	**Yes/no**
15. /w/ Error	**yes/No**
16. /l/ Control	**yes/No**
17. /r/ Target	**Yes/no**
18. /l/ Control	**yes/No**

Appendix C

Subject ID Number _____ Date _____ Group Assignment R1 ___ W2 ___ DR3 ___

MARK EACH ITEM AS RIGHT (√) OR WRONG (X)

1. ___ 2. ___ 3. ___ 4. ___ 5. ___ 6. ___ 7. ___ 8. ___ 9. ___ 10. ___

11. ___ 12. ___ 13. ___ 14. ___ 15. ___ 16. ___ 17. ___ 18. ___ 19. ___ 20. ___

21. ___ 22. ___ 23. ___ 24. ___ 25. ___ 26. ___ 27. ___ 28. ___ 29. ___ 30. ___

31. ___ 32. ___ 33. ___ 34. ___ 35. ___ 36. ___ 37. ___ 38. ___ 39. ___ 40. ___

41. ___ 42. ___ 43. ___ 44. ___ 45. ___ 46. ___ 47. ___ 48. ___ 49. ___ 50. ___

51. ___ 52. ___ 53. ___ 54. ___ 55. ___ 56. ___ 57. ___ 58. ___ 59. ___ 60. ___

61. ___ 62. ___ 63. ___ 64. ___ 65. ___ 66. ___ 67. ___ 68. ___ 69. ___ 70. ___

71. ___ 72. ___ 73. ___ 74. ___ 75. ___ 76. ___ 77. ___ 78. ___ 79. ___ 80. ___

81. ___ 82. ___ 83. ___ 84. ___ 85. ___ 86. ___ 87. ___ 88. ___ 89. ___ 90. ___