

1977

An investigation of the consistency of judgments regarding successive approximations of /r/

Scott Robert Lane
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Speech Pathology and Audiology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Lane, Scott Robert, "An investigation of the consistency of judgments regarding successive approximations of /r/" (1977). *Dissertations and Theses*. Paper 2854.
<https://doi.org/10.15760/etd.2848>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.


AN ABSTRACT OF THE THESIS OF Scott Robert Lane for the Master of Science in Speech Communication: Emphasis in Speech Pathology/Audiology presented December 15, 1977.

Title: An Investigation of the Consistency of Judgements Regarding Successive Approximations of /r/.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:


Robert H. English


Mary E. Gordon


Jack Hegrenes

This investigation sought to determine the accuracy and consistency of judgements made by three groups of judges, relative to successive approximations of /r/. The three groups were made up of speech pathologists, student trainees, and untrained individuals, respectively. It was the task of these judges to rank order three /r/ productions into the following categories: correct; partially correct; and incorrect. This task is basically the same as reinforcing approxi-

mations of /r/ within the therapy situation, and appears not to require extensive training. Many authors (Curry et al., 1943; Perrin, 1954; Oyer, 1959; Siegel, 1962; Irwin, 1965; and Elbert et al., 1967) have found little difference between trained and untrained listeners in identifying correct versus incorrect articulation. An apparent need existed to investigate what the accuracy and consistency of judgements would be by introducing successive approximations as a controlled or independent variable.

Two specific questions were posed by this investigation:

What is the accuracy of each group's ratings in choosing successively closer approximations of /r/?

What is the consistency of accuracy with which judgements of successive approximations of /r/ are made from one occasion to another?

The results indicated no significant difference between the three groups in ability to accurately identify correct, partially correct, and incorrect productions of /r/. There does appear to be a significant difference ($P < .05$) in accuracy judgements when comparing the trained (combined scores of speech pathologists and student trainees) with the untrained individuals. The untrained judges did significantly better in identifying the stimuli than did the trained judges.

There was no significant difference between consistency scores of the three groups. This would seem to indicate the three groups were equal in ability to be consistent,

over time, in their judgements of correct, partially correct, and incorrect productions of /r/.

It is interesting to observe there was no significant difference, either in accuracy or consistency, between speech pathologists, student trainees, and untrained individuals in identifying successive approximations of the phoneme /r/. The small difference between groups (although insignificant) tended to imply the untrained judges were more successful in the above stated task.

It also was found the three groups, without exception, experienced more difficulty identifying incorrect productions than correct and more difficulty with partially correct than incorrect productions.

AN INVESTIGATION OF THE CONSISTENCY OF JUDGEMENTS
REGARDING SUCCESSIVE APPROXIMATIONS OF /r/

by

SCOTT ROBERT LANE

A thesis submitted in partial fulfillment of the
requirements for the degree of


MASTER OF SCIENCE
in
SPEECH COMMUNICATION:
EMPHASIS IN SPEECH PATHOLOGY/AUDIOLOGY

Portland State University

1978

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

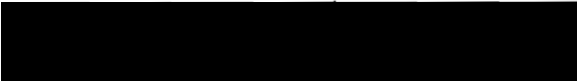
The members of the Committee approve the thesis of
Scott Robert Lane presented December 15, 1977.

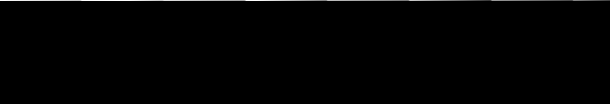

Robert H. English, Chairman


Mary E. Gordon


Jack Hegrenes

APPROVED:


Robert W. Vogelsang, Head, Department of Speech Communication


Stanley E. Rauch, Dean of Graduate Studies and Research

ACKNOWLEDGMENTS

I am deeply grateful and extremely happy to be able to thank those people whose assistance made the completion of this study possible. Their support, encouragement, and constructive criticism will always be appreciated.

I wish to thank Vickie Martinazzi who as my master clinician disagreed with me constantly regarding misarticulated /r/s, and consequently provided the inspiration and encouragement to initiate this study.

I would like to express my sincere appreciation to Dr. Jack Hegrenes and Mary Gordon. As the statistician for my thesis, Dr. Hegrenes provided the help which proved to be the cure for much panic and frustration regarding data and its analysis. I wish to thank Mary Gordon for proof-reading this thesis and providing accurate feedback concerning its progress.

I extend a very warm thank you to Dr. Robert H. English for his unselfish dedication and perseverance as the chairman of my thesis committee, and as my friend. His ability and knowledge of instrumentation, writing style and problem solving were an integral part throughout this study. Without his help this thesis would not have been possible.

To my family, who always knew I could do it, I wish to express my love and appreciation for helping when and how

they could. A special gratitude and respect are extended to my parents for enabling me to be more than I thought possible, and for teaching me to realize dreams can come true.

For my wife, Wendy, who has never known me without a thesis to do, I give all my love. Her support and understanding throughout this process is deeply appreciated.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS	iii
LIST OF TABLES	vii
CHAPTER	
I INTRODUCTION AND STATEMENT OF THE PROBLEM	1
INTRODUCTION.	1
STATEMENT OF THE PROBLEM.	2
SPECIFIC QUESTIONS TO BE ANSWERED	2
DEFINITIONS	3
II REVIEW OF THE LITERATURE.	5
STUDIES EVALUATING BINARY JUDGEMENTS OF ARTICULATION.	5
STUDIES EVALUATING JUDGEMENTS OF ARTICULATION SEVERITY	8
SUMMARY	11
III METHODS AND PROCEDURES.	13
SUBJECTS.	13
METHODS	13
Stimuli Instrumentation	
PROCEDURES.	16
ANALYSIS OF DATA.	17
IV RESULTS AND DISCUSSION.	18

CHAPTER	PAGE
RESULTS	18
Data Regarding Accuracy Scores of the Three Groups	
Data Regarding Consistency Scores of the Three Groups	
DISCUSSION.	27
Accuracy Scores of the Three Groups	
Consistency Scores of the Three Groups	
V SUMMARY AND IMPLICATIONS.	32
SUMMARY	32
IMPLICATIONS.	34
Clinical Research	
BIBLIOGRAPHY.	36
APPENDICES.	39
A INSTRUMENT SPECIFICATIONS	40
B PERMISSION REQUEST.	43
C JUDGE'S RESPONSE FORM	44
D INSTRUCTIONS TO JUDGES.	45

LIST OF TABLES

TABLE		PAGE
I	Means and Standard Deviations for Accuracy Scores of Speech Pathologists, Student Trainees, and Untrained Individuals in Judging Successive Approximations of the Phoneme /r/	19
II	Means, Standard Deviations, and <u>t</u> -Test Results Between the Accuracy Scores of Speech Pathologists and Student Trainees.	20
III	Means, Standard Deviations, and <u>t</u> -Test Results Between the Accuracy Scores of the Student Trainees and the Untrained Individuals	21
IV	Means, Standard Deviations, and <u>t</u> -Test Results Between the Accuracy Scores of Speech Pathologists and Untrained Individuals	21
V	Means, Standard Deviations, and <u>t</u> -Test Results Between the Accuracy Scores of Trained and Untrained Judges	22
VI	Consistency Scores Obtained by Speech Pathologists, Student Trainees, and	

	Untrained Individuals, from One Occasion to Another	23
VII	Means, Standard Deviations, and <u>t</u> -Test Results in Comparing the Consistency Scores of Speech Pathologists and Student Trainees.	24
VIII	Means, Standard Deviations, and <u>t</u> -Test Results in Comparing the Consistency Scores of Speech Pathologists and Untrained Individuals	24
IX	Means, Standard Deviations, and <u>t</u> -Test Results in Comparing the Consistency Scores of Student Trainees and Un- trained Individuals	25
X	Means, Standard Deviations, and <u>t</u> -Test Results in Comparing the Consistency Scores of the Trained and Untrained Judges.	26
XI	Group Percentages in Identifying Correct, Partially Correct, and Incorrect Productions	27

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

INTRODUCTION

The disorder of articulation is the most prevalent group of disorders with which the speech pathologist works in a public school setting. One frequently occurring, misarticulated sound is the /r/. Many clinicians report /r/ to be one of the sounds most resistant to treatment (Aungst and Frick, 1964).

Curtis and Hardy (1959) have reported correct production of /r/ does occur in children who misarticulate this phoneme. It has been found children's correct production of frequently misarticulated sounds does follow rules and patterns which influence the integrity of the production of those sounds (Buck, 1948). Given the possibility the child does occasionally produce a correct /r/, the clinician must be able to hear and reinforce this production if stabilization is to occur. If the child intermittently produces a correct /r/ sound, then the occurrence of approximations of a correct /r/ must abound. Progress toward the target sound, to a great extent, is dependent on the clinician. The clinician must be able to accurately identify and reinforce approximations of a correct /r/ in order to ensure progress.

In addition to identifying and reinforcing accurately, the clinician must be consistent. What is perceived as incorrect on one trial also must be considered incorrect on any other trial. What is a successively closer approximation of a correct /r/ must always be judged as successively closer. If the clinician is inconsistent, the client is likely to be not only inconsistent but confused. Therefore, no matter what treatment procedure is used, no matter what (if any) program is implemented, the determining factor for progress in sound production is the feedback the clinician provides to the client. Hence, the present study was designed to measure both the accuracy and consistency of judgements made by clinicians relative to the isolated production of /r/.

STATEMENT OF THE PROBLEM

The purpose of the present study was to compare judgements made by speech pathologists, student trainees, and untrained individuals relative to a series of successive approximations of /r/. More specifically, the investigation sought to determine the consistency with which judgements are made concerning the production of the /r/ sound in terms of correctness by degrees.

SPECIFIC QUESTIONS TO BE ANSWERED

This study sought to answer the following questions in regard to the problem as stated above:

- 1) What is the accuracy of each group's ratings in choosing successively closer approximations of /r/?
- 2) What is the consistency of accuracy with which judgements of successive approximations of /r/ are made from one occasion to another?

DEFINITIONS

The following are operational definitions of specific terms employed in the present investigation.

Accuracy Score

The subjects within this investigation made rankings of each of the ten sets of stimuli, relative to the degree of correctness of each stimulus. Accuracy score reflects how close their judgements were to the ideal. A perfect accuracy score would equal 40.

Consistency Score

Each subject made judgements, relative to 10 sets of /r/, on two separate occasions. Consistency score refers to the number of judgement changes from one occasion to the next. A perfect score would equal 0.

Correct Production

Correct production of a sound element is one that will be accepted by the listener as conforming to the standards of the general population of adult speakers and listeners of American English (Weybright, 1974).

Phonetic Symbolization of /r/

The phonetic symbol /r/ is used within this study to refer to the central vowel produced either by humping the central portion of the tongue higher in the mouth than for /ʌ/, or by raising (or retroflexing) the tongue-tip toward the alveolar ridge (Leutenegger, 1963). Symbolically it encompasses the phonetically written phonemes /ɹ/, and /ɻ/.

Successive Approximation

The operant term, successive approximation is used in reference to the principle of shaping which allows a clinician to obtain a response that has never been performed voluntarily insofar as the clinician is aware. Such responses are elicited by ". . . reinforcing crude approximations of the final topography instead of waiting for the correct response" (Skinner, 1959). By way of example, when toilet training a young child, parents would be advised to reward any efforts made in the direction of the bathroom, instead of waiting for the child to exhibit mature bowel and bladder control. As control grows, so too should parent's expectations of the child's ability to copy the adult model.

CHAPTER II

REVIEW OF THE LITERATURE

The accurate evaluation of articulation competency is critical for four reasons: 1) to effectively diagnose deviations; 2) to measure progress; 3) to evaluate the effectiveness of therapeutic procedures; and 4) to correctly reinforce successive approximations of the target sound. Barker (1960) has noted the problem of assessing progress and evaluating procedures in articulation treatment is essentially the same as correctly reinforcing successive approximations.

The process of evaluating articulation ability, either initially before treatment begins or for the purpose of determining progress, is subjective. Wood (1949) has noted, independent of the method used, the final judgement of articulation rests on subjective, auditory judgements. Any response to a stimulus conjures a judge's reference to an internalized "phonemic space" (Noll, 1970). This phonemic space becomes the clinician's "yardstick" in determining the correctness or partial correctness of the elicited sound.

STUDIES EVALUATING BINARY JUDGEMENTS OF ARTICULATION

In the studies reviewed, judges, both trained and untrained, demonstrated high correlations in their judgements of

correct versus incorrect articulation ability. Henderson (1938) did a study of inter-judge reliability, using three judges in several different settings, in order to assess articulation ability. The judges attained 80 percent agreement when making binary evaluations of correct or incorrect within a live test situation. Agreement decreased to 72 percent when required to state the type of error. The author concluded a binary judgement of correct versus incorrect can be more accurately made than judgements involving type of error.

It has been reported by Siegel (1962) that inexperienced examiners can be trained to make judgements regarding correct or incorrect articulations with a minimum of training. In the study done by Siegel, training involved four hours of listening to correct and incorrect speech sounds and took place between the first and second test occasions with no training between the second and third test occasions. The correlations of two inexperienced examiners on the three testing occasions were: .92 before training; and .97 and .96 after training. Within the same study Siegel employed two experienced examiners on the same task excluding the four training hours between the first and second test occasions. They attained correlation coefficients of .86, .94, and .91. Although there was high reliability, the author cautioned that examiners tended to differ significantly in absolute scores they assigned to the subjects.

In a study done by Aungst and Frick (1964), speech pathologists used the McDonald Deep Test of Articulation to assess the articulation of /r/ as produced by children. They demonstrated high inter- and intra-judge agreement when making correct versus incorrect judgements. The authors chose to use the /r/ phoneme in their study because of its clinical frequency and because of clinician's reports that it is one of the most difficult sounds to correct. Inter-judge reliability was 77.38 percent for spontaneous productions and 90.62 percent for the imitated productions. Here again, it is to be noted the judges were making only binary judgements of "right" or "wrong", and were not required to make a third, more precise judgement of a possible approximation.

Irwin (1970) conducted an investigation concerned with the consistency of judgements regarding articulation using students enrolled in a clinical methods class. She found consistency of agreements to be 66 percent when identifying misarticulations and 84 percent when identifying correct productions. Irwin concluded:

Since the consistency of agreements was generally much poorer in identification of misarticulation than for correct productions of consonants, it would appear that special attention should be given to the study of incorrect productions of sound. It may be that uncertain standards for correct productions as well as inadequate perceptual abilities may be contributing to the low consistency of agreement for the identification of misarticulations.

It would appear necessary, to this writer, to give special

attention not only to the incorrect productions of sound, but also to the degrees of deviation. This type of attention would help develop two major elements of auditory perception, discrimination and memory, thus, enhancing the clinician's ability to identify successive approximations.

In another study Irwin (1965) found experienced clinicians, graduating seniors in speech pathology, and classroom teachers differed very little in their ability to identify misarticulations. The first group correctly identified misarticulations with an accuracy percentage of 72.1, while the other two groups had 72.5 percent and 64.2 percent, respectively. Here again, the judges were required only to make a binary judgement, and were more accurate in identifying correct articulation than they were in recognizing incorrect articulation.

Other studies (Curry et al., 1943; Perrin, 1954; Oyer, 1959; and Elbert et al., 1967) have demonstrated, when making binary judgements of articulation ability, there is no significant difference between sophisticated listeners and unsophisticated listeners. Within these investigations, no reasons were given to explain why this might be true.

STUDIES EVALUATING JUDGEMENTS OF ARTICULATION SEVERITY

Sherman and Morrison (1955) conducted a study to determine if an observer could be reliable on judgements of articulation severity. A nine point scale was employed, with one

representing least defective articulation and nine representing most defective articulation. The intra-class reliability of a group of trained judges was reported to be 94 percent using this nine point scale. It was concluded mean-scale values of severity of defective articulation can be obtained from the responses of trained individual observers, but absolute values of severity measures of defective articulation are not necessarily comparable from one individual observer to another.

Another investigation using a similar nine-point scale was conducted by Barker (1960). Ten judges were used to rate tape recorded examples of defective articulation along a continuum from one to nine. One represented least defective articulation and nine represented most defective articulation. The judge's ratings were then correlated to an Articulation Score, which was a measure of relative occurrence of all speech sounds. The Articulation Score indicated the amount of speech which probably can be understood in ordinary conversation. The correlation between judge's ratings and Articulation Scores was .94. This indicates the judge's severity ratings correlate highly with the amount of speech which can be understood in ordinary conversation.

Wright (1954) reports the use of a seven-point scale of articulation, beginning with correct, followed by four levels of distortion, then substitutions, and omissions. It was found evaluations using this scale can be reliable.

Intra-judge consistency was reported as 87.8 percent while inter-judge consistency was approximately 90 percent.

Shames (1952) investigated the agreement between evaluations made by two clinicians. The clinicians were to indicate whether the speech on the second recording presented to them was better than, the same as, or worse than the speech on the first recording. The correlation between the judgments of the two clinicians was .57, which was moderately significant. This study involved making a trinary choice using two stimulus sounds; the correlation score indicates such a judgement to be a more difficult task than a simple binary judgement; as reflected by the high correlations reported in the Siegel (1962) and Aungst and Frick (1964) investigations.

A five-point scale was used by Burgi and Matthews (1960) to compare one speech sample with a second one and then rate the second one. In this scale, 0 represented a great deal worse, 1 represented a little worse, 2 represented the same as, 3 represented a little better, and 4 represented a great deal better. Four groups of judges were used. The first group consisted of untrained individuals, the second was composed of classroom teachers, the third was made up of speech pathology students in their first term of practicum, and the fourth consisted of trained speech pathology graduates. The intra-group correlations were as follows: group one, untrained individuals, attained a correlation coeffi-

cient of .80; group two, classroom teachers, had a correlation of .88; group three, beginning speech pathology students, achieved a correlation of .68; and group four, speech pathology graduate students, had a correlation of .75. It can be seen the two groups with the least training had the highest intra-group agreement and the two groups with the most training had less intra-group agreement.

SUMMARY

The foregoing data would seem to suggest very little difference between trained and untrained judges in ability to identify correct and incorrect articulation. It also has been suggested correct articulations are more accurately identified than incorrect articulations and more accurately identified than type and degree of articulation error.

The literature further implies the identification of degrees of severity is a much more difficult task than that of making a binary judgement relative to articulation. Both trained and untrained judges can have high intra- and inter-judge correlations, however, judges with less training attain higher correlation coefficients.

It would appear to this investigator the ability to accurately and consistently identify degrees of severity (successive approximations) would facilitate the monitoring of progress made in treatment. In the opinion of this investigator, it is unrealistic to be so stringent in

articulation treatment as to accept only correct or "normal" productions. To enhance positive movement in articulation treatment, from an incorrect to a correct production, it is advantageous to accept and reinforce closer and closer approximations of the target sound.

CHAPTER III

METHODS AND PROCEDURES

SUBJECTS

Twenty-one subjects, divided into three groups of seven subjects each, comprised the experimental sample. The first group consisted of seven speech pathologists. Each subject within this group was required to have a minimum of five years experience in the public schools, and was selected from the Portland Public Schools. The second group consisted of seven graduate students majoring in speech pathology who had completed their student teaching requirement and were enrolled in the Speech and Hearing Sciences Program at Portland State University. The third group was composed of seven students from a speech 100 series class at Portland State University, thus, representing the inexperienced or untrained group of judges. All twenty-one subjects were randomly selected.

METHODS

To avoid contamination of results by a possible hearing loss, only evaluators with normal hearing were included in the study. A puretone audiometric screening test was administered to all candidates, using a Beletone Model 10C

portable audiometer, with Aural-dome headphones, Model AR 100 (see Appendix A, 1 and 2 for specifications). The speech frequencies tested were 500, 1000, and 2000 Hz, at 20 dB (ISO).

Stimuli

Ten children, judged by the speech pathologist in their school to have normal or correct production of the /r/ phoneme, were used in collection of stimuli. Adults were not utilized for stimuli because the investigation was aimed primarily at public school or public school-bound populations; hence, the use of children for the collection of stimuli. The children consisted of five males and five females, all ten years of age. Each child phonated one "normal" /r/, which was recorded once and duplicated two times. (This procedure provided for ten sets of stimuli with three /r/s' within each set or a total of thirty stimuli.) See Appendix B for the parental consent form.

Identification of a normal /r/ sound, according to Lehiste (1964), is dependent on the position of the third formant. O'Connor et al. (1957) have reported the distinguishing factors of the /r/ phoneme as being the second and third formants. In other words, the starting point of the second formant transition and the rising third formant of the /r/, distinguish it from other phonemes. Each child's production of /r/ was analyzed, using the Kay Sona-Graph, enabling the investigator to accurately identify and filter

critical frequency bandwidths. After this filtering process there was one sound within each set which was unfiltered or left in the "natural" state as recorded. The remaining two were filtered, the first /r/ had the complete third formant filtered out, and the second /r/ had both the second and third formants filtered out. The final result was that for each of the ten children there was one set of /r/ stimuli which was presented to the judges. Within each set there were three stimuli, randomly ordered, one of which was left unfiltered, another with only the third formant filtered out which approximated a partially correct /r/ sound, and another with the second and third formants filtered out and which sounded similar to /ʌ/, or for the purpose of this study, an incorrect /r/ sound.

Instrumentation

The stimuli were recorded using an Ar-Tik Speech and Hearing Recorder with an Electro-Voice microphone, Model RE 15 (see Appendix A, 3 and 4 for specifications). From the Ar-Tik the stimuli were analyzed, using a Kay Sona-Graph, Model 6061-B, sound spectrograph (Appendix A, 5) to identify the second and third formants by visually inspecting the formant positions in relation to frequency, for the purpose of filtering. This filtering process was accomplished through the use of a Krohn-Hite Filter Model 3202 (see Appendix A, 6, for specifications). From the Ar-Tik the stimuli were passed through the filters and dubbed to a

second tape using an Ampex, Model AG 500, tape recorder (see Appendix A, 7, for specifications) which was used in presenting the stimuli to the judges. In an attempt to reduce a judgement in frequency and quality of /r/ being made on the basis of loudness, fluctuations in intensity were controlled by monitoring the stimuli recorded on this second tape utilizing the VU-meter on the Ampex recorder, making the appropriate loudness-level adjustments as necessary during the dubbing process.

Presentation of the re-recorded, filtered stimuli to the judges was done on the Ampex using an Ampex Model 622 speaker (see Appendix A, 8, for specifications).

The end result was a tape of ten sets (or thirty stimuli). The order within each set was randomized, so there was no consistent pattern in presentation.

PROCEDURES

The task of the judges was to discriminate among the three types of stimuli of /r/ sounds. There were ten sets in all with three stimuli per set. Each judge was required to evaluate each of the ten sets, arranging the stimuli in rank order relative to the degree of correctness. The categories under which the stimuli were placed included: correct; partially correct; and incorrect (see Appendix C). Each set was presented to each judge three times during each ranking session. There were two presentations of the ten

sets with the second presentation following the first by two weeks.

ANALYSIS OF DATA

Comparisons in both accuracy and consistency scores were made among the three groups of judges. Mean scores and standard deviations were determined for each group and a one tailed t-test was utilized to determine if there were any significant differences from one group to the next. Mean scores for accuracy and consistency were obtained by adding each groups individual scores together and then dividing by the N. The range of scoring possibilities for consistency was 30-0, with 0 representing no judgement changes from one occasion to the next. The range for accuracy was 0-40, with 40 representing a perfect score.

CHAPTER IV

RESULTS AND DISCUSSION

RESULTS

The purpose of this study was to compare judgements made by speech pathologists, student trainees, and untrained individuals relative to a series of successive approximations of the phoneme /r/. More specifically, this investigation sought to determine the consistency with which judgements are made concerning the production of the /r/ sound in terms of correctness by degrees. This comparison was accomplished by having subjects within the three groups make judgements, on two separate occasions, concerning the relative degree of correctness of each /r/ phoneme. There were 10 sets with 3 stimuli to a set for a total of 30 stimuli. All subjects were verbally given the same instructions (see Appendix D) for each of the two trials which were two weeks apart. A copy of the recording form can be found in Appendix C.

Data obtained from the above mentioned experiment were compiled and analyzed to answer the following questions:

- 1) What is the accuracy of each group's ratings in choosing successively closer approximations of /r/?

- 2) What is the consistency of accuracy with which judgements of successive approximations of /r/ are made from one occasion to another?

DATA REGARDING ACCURACY SCORES OF THE THREE GROUPS

Tables I through V deal specifically with question number 1 regarding the accuracy of judgements of the three groups.

Group means and standard deviations for the accuracy scores of the three groups are presented in Table I. Examination of data reveal the untrained group mean (34.86) exceeded that of the speech pathologists (31.86) by 3 points and the student trainees (32.43) by 2.43 points. Additionally, it is to be noted standard deviations grow larger as accuracy scores diminish. This tends to indicate the untrained group had less fluctuation and greater agreement in judgements as a group than did either the student trainees or the speech pathologists. Speech pathologists, as a group, showed the greatest

TABLE I

MEANS AND STANDARD DEVIATIONS FOR ACCURACY SCORES OF SPEECH PATHOLOGISTS, STUDENT TRAINEES, AND UNTRAINED INDIVIDUALS IN JUDGING SUCCESSIVE APPROXIMATIONS OF THE PHONEME /r/*

	Speech Pathologists	Student Trainees	Untrained Individuals
Mean	31.86	32.43	34.86
S.D.	4.099	3.690	2.545

*Highest score possible was 40.

deviation from the mean with a S.D. of 4.099 indicating less agreement among them with greater fluctuation.

A comparison of group accuracy scores was made between speech pathologists and student trainees. In Table II are presented the means, S.D.s and t-test results. It is to be noted there is no statistically significant difference between the two groups, although there was a small difference between the means (.57) in favor of the student group. To be significant at the .05 level of confidence P would have to equal or exceed 1.796.

Table III contains the means, standard deviations, and the t-test value between the accuracy scores of the student trainees and the untrained individuals. Although a t-test score of 1.4378 reveals no significant difference at the .05 level of confidence (to be significant P would have to equal or exceed 1.796), it can be seen the untrained group did slightly better (2.43 points) than did the student trainee group on mean scores.

TABLE II

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS BETWEEN THE ACCURACY SCORES OF SPEECH PATHOLOGISTS AND STUDENT TRAINEES

	Speech Pathologists	Student Trainees	<u>t</u>
Mean	31.86	32.43	.2740*
S.D.	4.099	3.690	

* $P > .05$, d.f. 12.

TABLE III

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS BETWEEN THE ACCURACY SCORES OF THE STUDENT TRAINEES AND THE UNTRAINED INDIVIDUALS

	Student Trainees	Untrained Individuals	t
Mean	32.43	34.86	1.4378*
S.D.	3.690	2.545	

*P > .05, d.f. 12.

Accuracy scores between speech pathologists and untrained individuals also were compared. Resultant data between these two groups of judges may be found in Table IV. Here it is to be observed there is no statistically significant difference between the means; to be significant P would have to equal or exceed 1.796. There was, however, a difference of 3.0 points between the accuracy scores in favor of the untrained individuals.

TABLE IV

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS BETWEEN THE ACCURACY SCORES OF SPEECH PATHOLOGISTS AND UNTRAINED INDIVIDUALS

	Speech Pathologists	Untrained Individuals	t
Mean	31.86	34.86	1.645*
S.D.	4.099	2.545	

*P > .05, d.f. 12.

Table V presents a comparison of trained and untrained judge's accuracy scores. The trained judges included seven speech pathologists and seven student trainees, while the untrained group consisted of seven speech 100 students. Inspection of the data reveals a significant difference between the two groups with a one-tailed t score of 1.95, d.f. 19, with $P < .05$. This difference indicates the untrained judges were more successful in identifying correct, partially correct, and incorrect productions of /r/ than were the trained judges.

DATA REGARDING CONSISTENCY SCORES OF THE THREE GROUPS

Tables VI through X deal specifically with question number 2 regarding the consistency of judgements of the three groups.

Table VI provides means and standard deviations for the consistency scores (defined as the number of changes from one occasion to the next) of the three groups. It can be readily

TABLE V
MEANS, STANDARD DEVIATIONS, AND t -TEST
RESULTS BETWEEN THE ACCURACY SCORES
OF TRAINED AND UNTRAINED JUDGES

	Trained	Untrained	t
Mean	32.143	34.86	1.95*
S.D.	3.76	2.545	

* $P < .05$, d.f. 19.

TABLE VI
 CONSISTENCY SCORES OBTAINED BY SPEECH
 PATHOLOGISTS, STUDENT TRAINEES, AND
 UNTRAINED INDIVIDUALS, FROM ONE
 OCCASION TO ANOTHER*

	Speech Pathologists	Student Trainees	Untrained Individuals
Mean	6.143	7.286	5.714
S.D.	3.288	4.536	2.360

*A perfect score equals 0,
denoting no judgement changes.

seen by visual inspection the untrained group had a lower mean consistency score and standard deviation value than did either the student trainees or the speech pathologists, reflecting the greatest consistency in judgements from one occasion to the next. The student trainees had the greatest standard deviation value, which indicated the least consistency from one occasion to the next. Having obtained a mean consistency score of 6.143 and a standard deviation value of 3.288, the speech pathologists fall in a medial position between the two student groups, relative to consistency of judgements from one occasion to another.

Observation of Table VII shows a t-test value of .0539 which indicates no significant difference between the mean consistency scores of the speech pathology group and the student trainee group. At a .05 level of confidence it would take a t value of 1.796, with 12 degrees of freedom

TABLE VII

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS IN COMPARING THE CONSISTENCY SCORES OF SPEECH PATHOLOGISTS AND STUDENT TRAINEES

	Speech Pathologists	Student Trainees	<u>t</u>
Mean	6.143	7.286	.0539*
S.D.	3.288	4.536	

*P > .05, d.f. 12.

to indicate a significant difference. These two groups appear to be congruent in their consistency of judgements from one occasion to another.

The number of changes made by the speech pathology group and the untrained individuals are presented in Table VIII in terms of means, standard deviations, and t-test value. In comparing the two mean consistency scores a t-test value of 0.280 is obtained which reflects no signifi-

TABLE VIII

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS IN COMPARING THE CONSISTENCY SCORES OF SPEECH PATHOLOGISTS AND UNTRAINED INDIVIDUALS

	Speech Pathologists	Untrained Individuals	<u>t</u>
Mean	6.143	5.714	0.280*
S.D.	3.288	2.360	

*P > .05, d.f. 12.

cant difference between the number of judgement changes made by the two groups, from one occasion to the next.

Comparisons between the student trainees and the untrained individuals can be found in Table IX. Although there is no significance between the mean consistency scores (t-test value of 0.8132) of the two groups, it is interesting to note there is a 1.572 point difference between the means, with the untrained group having less changes from one occasion to another.

Table X presents a comparison of trained and untrained judge's consistency scores. Seven speech pathologists and seven student trainees made up the trained group while the untrained judges consisted of seven speech 100 students. Observation of Table X reveals the untrained group had less change than did the trained group and a lower standard deviation value. There was no significant difference, however, between groups as can be seen by the t-test score of 0.734.

TABLE IX

MEANS, STANDARD DEVIATIONS, AND t-TEST RESULTS IN COMPARING THE CONSISTENCY SCORES OF STUDENT TRAINEES AND UNTRAINED INDIVIDUALS

	Student Trainees	Untrained Individuals	<u>t</u>
Mean	7.286	5.714	0.8132*
S.D.	4.536	2.360	

* $P > .05$, d.f. 12.

TABLE X
 MEANS, STANDARD DEVIATIONS, AND t-TEST
 RESULTS IN COMPARING THE CONSISTENCY
 SCORES OF THE TRAINED AND
 UNTRAINED JUDGES

	Trained	Untrained	<u>t</u>
Mean	6.714	5.714	0.734*
S.D.	3.85	2.36	

*P > .05, d.f. 19.

It would take a t value of 1.796 with 19 degrees of freedom, to make a significant difference at the .05 level of confidence.

In Table XI the percentage of accurate judgements made by the three groups is presented. It can be noted here the correct /r/ productions were most accurately identified in relation to partially correct and incorrect productions. The most difficult /r/ productions to identify were those which were partially correct. Looking at the three groups, it can be seen the untrained group attained the highest percentage in identifying correct, partially correct, and incorrect productions than did the other two groups. By inspection there appears to be a marked difference using percentages in the ability of judges to identify the correct versus the incorrect productions, as well as the correct versus the partially correct productions of /r/.

TABLE XI
 GROUP PERCENTAGES IN IDENTIFYING CORRECT,
 PARTIALLY CORRECT, AND INCORRECT
 PRODUCTIONS OF /r/

	Correct	Partially Correct	Incorrect
Speech Pathologists	.91	.64	.67
Student Trainees	.89	.66	.74
Untrained Individuals	.94	.79	.89

DISCUSSION

It was the intent of this investigation to compare judgements made by three groups of listeners, relative to successive approximations of the /r/ phoneme. The three groups consisted of speech pathologists, student trainees, and untrained individuals. The judgements made by the three groups were analyzed for the accuracy of judgements made by each group as well as the consistency of each group's judgements from one occasion to another. A discussion of results reflecting inter-group accuracy and consistency scores follows.

Accuracy Scores of the Three Groups

What is the accuracy of each group's ratings in choosing successively closer approximations of /r/?

The accuracy scores reported in this study refer to the number of correct judgements each group of judges made

in ranking successive approximations of /r/ into three categories: correct, partially correct, and incorrect. The results indicate the three groups differed very little in their ability to accurately identify approximations of /r/. The speech pathologists attained a mean accuracy score of 31.86, the student trainees a mean accuracy score of 32.43, and the untrained individuals a mean accuracy score of 34.86. These findings are compatible with those of Irwin (1965) who found little difference between experienced judges, student trainees, and inexperienced judges in their ability to identify misarticulations. Irwin gave no explanation for her findings. Other studies (Curry et al., 1943; Perrin, 1954; Oyer, 1959; and Elbert et al., 1967) have demonstrated no significant difference between judgements made by sophisticated and unsophisticated listeners, supportive of the findings of this investigation.

Standard deviations for the three group's accuracy scores were as follows: speech pathologists, 4.099; student trainees, 3.690; and untrained individuals, 2.545. These S.D. values are indicative of the amount of intra-group agreement. Although the difference in accuracy scores was insignificant, it is interesting to note, the group with the least training had the greatest amount of agreement, and vice versa. This finding is corroborated by Siegel (1962) and Burgi and Matthews (1960). They found inexperienced judges to have higher inter-judge reliability than did

experienced judges. No explanation of this finding was given. Within the present study it is possible that mechanically filtered sound (/r/), used to approximate misarticulated sound (incorrect /r/), does not adequately suffice. The question "Do filtered /r/ phonemes sound like naturally misarticulated /r/ phonemes?" must be considered. It also is possible the "mental set" of each group of judges may have differed to the extent that results were influenced. For example, the speech pathologist group may have been satiated with auditory /r/ stimuli, while at the other extreme the untrained individuals were "fresh" in terms of the /r/ phoneme.

When the two groups with training in auditory perceptual discrimination were combined and compared with the group with no training, a significant difference resulted. The untrained group was significantly more accurate in identifying correct, partially correct, and incorrect /r/ phonemes than the trained group. The highest accuracy score possible was 40 with the trained group mean accuracy score of 32.143 and the untrained group mean accuracy score of 34.86. This finding is consistent with that of Burgi and Matthews (1960) who found groups with more training had less intra-group agreement than did groups with less training.

The findings involving identification of correct versus incorrect articulation are congruent with the findings of Irwin (1970) relative to the difficulty factor. Irwin found,

as did this investigator, that all judges, regardless of sophistication, have more success identifying correct productions than incorrect productions. In the present study it also was found the addition of an approximation or partially correct sound was more difficult to identify than either the correct or the incorrect production. Irwin hypothesized ". . . it may be that uncertain standards for correct productions as well as inadequate perceptual abilities may be contributing to the low consistency of agreement . . ." in the identification of incorrect sound productions.

To summarize the answer to question number one, there was no significant difference in the accuracy ratings of the three groups in choosing successively closer approximations of /r/. It is interesting to observe, however, the group with the least training had the highest accuracy ratings and the groups with the most training had the lowest accuracy ratings. This, we have noted, reached significance ($P < .05$) when the scores of the two trained groups were combined and compared with the untrained group.

Consistency Scores of the Three Groups

What is the consistency of accuracy with which judgements of successive approximations of /r/ are made from one occasion to another?

Consistency scores refer to the consistency of judgements from one occasion to another. More specifically, consistency scores are the mean number of changes each group made from one judging occasion to the next; hence, the

smaller the score the more consistent the group. The mean number of judgement changes for each group were as follows: speech pathologists, 6.143; student trainees, 7.286; and untrained individuals, 5.714. The untrained group had the highest consistency of accuracy in judgements made from one occasion to the next. The student trainees had the greatest number of changes, therefore they were the least consistent from one occasion to the next. The difference between the group scores, however, are not significant. The findings of the present investigation corroborate those of Wright (1954) in that judges tend to be consistent with themselves whether sophisticated or unsophisticated. Motivation may have influenced the outcome of this study. The speech pathology group may have been satiated in terms of articulation discrimination, which would have had a direct affect on their judgements. It also might be hypothesized the experienced group was accustomed to attending to other cues besides auditory, such as visual, etcetera.

In summary to question number two, there was no significant difference in the three groups' consistency scores from one occasion to another.

CHAPTER V

SUMMARY AND IMPLICATIONS

SUMMARY

This investigation sought to determine the accuracy and consistency of judgements made by three groups of judges, relative to successive approximations of /r/. The three groups were made up of speech pathologists, student trainees, and untrained individuals, respectively. It was the task of these judges to rank order three /r/ productions into the following categories: correct; partially correct; and incorrect. This task is basically the same as reinforcing approximations of /r/ within the therapy situation, and appears not to require extensive training. Many authors (Curry et al., 1943; Perrin, 1954; Oyer, 1959; Siegel, 1962; Irwin, 1965; and Elbert et al., 1967) have found little difference between trained and untrained listeners in identifying correct versus incorrect articulation. An apparent need existed to investigate what the accuracy and consistency of judgements would be by introducing successive approximations as a controlled or independent variable.

Two specific questions were posed by this investigation:

What is the accuracy of each group's ratings in choosing successively closer approximations of /r/?

What is the consistency of accuracy with which judgements of successive approximations of /r/ are made from one occasion to another?

The results indicated no significant difference between the three groups in ability to accurately identify correct, partially correct, and incorrect productions of /r/. There does appear to be a significant difference ($P < .05$) in accuracy judgements when comparing the trained (combined scores of speech pathologists and student trainees) with the untrained individuals. The untrained judges did significantly better in identifying the stimuli than did the trained judges.

There was no significant difference between consistency scores of the three groups. This would seem to indicate the three groups were equal in ability to be consistent, over time, in their judgements of correct, partially correct, and incorrect productions of /r/.

It is interesting to observe there was no significant difference, either in accuracy or consistency, between speech pathologists, student trainees, and untrained individuals in identifying successive approximations of the phoneme /r/. The small difference between groups (although insignificant) tended to imply the untrained judges were more successful in the above stated task.

It also was found the three groups, without exception, experienced more difficulty identifying incorrect productions

than correct and more difficulty with partially correct than incorrect productions.

IMPLICATIONS

Clinical

The results of this investigation indicate a wide variance in ability to identify correct, partially correct, and incorrect productions of /r/; therefore, it would seem advantageous to place more emphasis on training students to recognize incorrect and partially correct productions of sound. Along with this is the idea of enhancing two major areas of auditory perception, discrimination, and memory, thus improving speech pathologist's ability to accurately and consistently identify successive approximations of a target sound.

One other major clinical implication within this investigation has to do with the use of aides in articulation therapy. Trained and untrained judges demonstrated no significant difference in ability to identify sound productions, within this study, therefore, it seems reasonable to assume the same relationship would carry-over to the therapy room.

Research

The finding that there was no significant difference between trained and untrained judges is, in itself, significant. More research needs to be done at this point in regard to the identification of the specific cues used by listeners in judging articulation and determine if training influences

attending to those cues. Curtis (1954) has observed the importance of identifying the acoustic characteristics of speech sounds, but emphasizes the identification of the specific cues i.e., auditory, auditory-visual, visual, etcetera used by listeners to judge those speech sounds.

It would be interesting to duplicate this study, using naturally produced, correct, partially correct, and incorrect productions of sound. Such a study would determine if the mechanically filtered sounds within this investigation influenced the outcome. Another possible aspect could entail using visual, as well as auditory cues, by way of videotape. This would determine which cues, visual or auditory, were more valuable to the listener.

B I B L I O G R A P H Y

BIBLIOGRAPHY

- AUNGST, L.F. and FRICK, J.V., Auditory discrimination ability and consistency of articulation of /r/. J. Speech and Hearing Dis., 29, 76-85 (1964).
- BARKER, J.O., A numerical measure of articulation. J. Speech and Hearing Dis., 25, 79-88 (1960).
- BUCK, M.W., A Study of the Misarticulation of /r/ in Children from Kindergarten Through Third Grade. M.A. Thesis, U. of Iowa, 1948.
- BURGI, E.J. and MATTHEWS, J., Effects of listener sophistication upon global ratings of speech behavior. J. Speech and Hearing Res., 3, 348-353 (1960).
- CURRY, R., KENNEDY, L., WAGNER, L., and WILKIE, W., Phonographic scale for the measurement of defective articulation. J. of Speech Dis., 8, 123-126 (1943).
- CURTIS, J.F., Systematic research in experimental phonetics: 3. the case for dynamic analysis in acoustic phonetics. J. Speech Hearing Dis., 19, 147-157 (1954).
- CURTIS, J.F. and HARDY, J.C., A phonetic study of misarticulation of /r/. J. Speech and Hearing Res., 2, 244-257 (1959).
- ELBERT, M., SHELTON, R.L., and ARNDT, W.B., A task for the evaluation of articulation change: 1. development of methodology. J. Speech and Hearing Res., 10, 281-288 (1967).
- HENDERSON, F.M., Accuracy in testing the articulation of speech sounds. J. Educational Res., 31, 348-356 (1938).
- IRWIN, R.B., Consistency of judgements of articulatory productions. J. Speech and Hearing Res., 3, 548-555 (1970).
- IRWIN, R.B. and KRAFCHICK, I.P., An audio-visual test for evaluating the ability to recognize phonetic errors. J. Speech and Hearing Res., 8, 281-290 (1965).

- LEHISTE, I., Acoustical Characteristics of Selected English Consonants. Bloomington, Ind.: Indiana University (1964).
- LEUTENEGGER, R.R., The Sounds of American English. Milwaukee, Wis.: University of Wisconsin, Scott, Foresman and Company (1963).
- NOLL, J.D., Articulation Assessment. In J. Fricke (Ed.), Speech and the Dentofacial Complex: The State of the Art, ASHA Reports Number 5. Washington, D.C.: American Speech and Hearing Association, 1970 as cited in Shriberg, L.D., Articulation judgements: some perceptual considerations, J. Speech and Hearing Res., 4, 876-882 (1972).
- O'CONNOR, S.D., GESTMAN, L.I., LIBERMAN, A.M., DELATTRE, P.C., and COOPER, F.S., Acoustic cues for the perception of initial /w, j, r, l/ in english. Word, 13, 24-43 (1957).
- OYER, H.J., Speech error recognition ability. J. Speech Hearing Dis., 24, 391-394 (1959).
- PERRIN, E.L., The rating of defective speech by trained and untrained observers. J. Speech and Hearing Dis., 19, 48-51 (1954).
- SHAMES, G., An investigation of prognosis and evaluation in speech therapy. J. Speech and Hearing Dis., 17, 386-392 (1952).
- SHERMAN, D. and MORRISON, S., Reliability of individual ratings of severity of defective articulation. J. Speech and Hearing Dis., 20, 352-358 (1955).
- SIEGEL, G.M., Experienced and inexperienced articulation examiners. J. Speech and Hearing Dis., 27, 28-35 (1962).
- SKINNER, B.F., Cumulative Record. New York: Appleton-Century-Crofts, 1959.
- WEYBRIGHT, G.D., An Investigation of the Development of the Phonemes /t/ and /k/ in the Speech of Preschool Children. M.S. Thesis, Portland State University, 1974.
- WOOD, K.S., Measurement of progress in the correction of articulatory defects. J. Speech and Hearing Dis., 14, 171-174 (1949).
- WRIGHT, H.N., Reliability of evaluations during basic articulation and stimulation testing. J. Speech and Hearing Dis., (1954) Monograph Supplement, 4, 19-27.

A P P E N D I C E S

APPENDIX A

INSTRUMENT SPECIFICATIONS

1. Beltone Model 10C Portable Audiometer; equipped with a calibrated set of TDH 39 Earphones, mounted in an Aural Dome, circum-aural headset.
2. Aural Dome headphones, Model AR 100.
3. Ar-Tik Speech and Hearing Recorder.
Frequency Response: ± 1.5 dB, 50 to 12K Hz for both record and playback.
Flutter and Wow: .2%.
Signal to Noise Response: Undetermined at this time.
4. Electro-Voice Microphone Model RE-15; Dynamic Cardioid, 150 OHMs.
5. Kay Sona-Graph, Model 6061-B Sound Spectrograph.
Frequency Range: 85-16000 Hz in two ranges.
Analysis Time: 1.3 minutes.
Effective Resolution: 80-8000 Hz 45 and 300 Hz
160-16000 Hz 90 and 600 Hz
AGC Range: Variable 20 to 40 dB down to 10.
Frequency Calibration: Switchable at 50, 500, or 1000 Hz intervals
Response: ± 2 dB over entire range.
Recording Time: 80-8000 Hz 2.4 sec.
160-16000 Hz 1.2 sec.
Amplifier Characteristics: Flat or 13 dB high-frequency pre-emphasis

Input Impedance: 200, 600, or 10,000 , Switchable.

Recording Medium: Nickle-cobalt plated turntable.

Microphone Supplied: Altec-Lansing 681A dynamic.

Power Supply: 117V, 50/60 CPS, 100W., Regulated.

6. Krohn-Hite Filter Model 3202; high-pass and low-pass cutoff frequencies continuously adjustable from 20 Hz to 2 MHz in five bands.

Attenuation Slope: Nominal 24 dB per octave per channel in high-pass or low-pass.

Maximum Attenuation: Greater than 80 dB.

Insertion Loss: Zero + 1/2 dB to 2 MHz; 3 dB at approximately 10 MHz. 6 dB in Band-Reject operation.

Input Characteristics: Maximum Input Amplitude - 3 v rms up to 2 MHz, decreasing to 1 v rms at 10 MHz.
Maximum DC Component - Low-Pass Mode: Combined ac plus dc should not exceed 4.2 v, peak. High-Pass Mode: 200 v.
Impedance - 100 k ohms in parallel with 50 pf.

Output Characteristics: Maximum Voltage - 3 v, rms, to 2 MHz (1.5 v, rms, in Band-Reject operation).
Maximum Current - 10 ma (less in Band-Reject operation).
Internal Impedance - 50 ohms, approximately (higher in Band-Reject operation).

Power Requirements: 105-125 or 210-250 volts, single-phase, 50-400 Hz, 15 watts.

7. Ampex, Model AG 500, Tape Recorder.

Input: 100,000 ohms, unbalanced. Will accept input signal levels as low as -18 dbm for normal operating level.

Output: Will feed a 600 ohm line, balanced or unbalanced, at a nominal +4 dbm output level. Maximum reproduce output level before clipping is +24 dbm.

Overall Frequency Response: 15 ips: +2 dB, 30 to 18,000 Hz
 (500 Hz reference)
 7-1/2 ips: +2 dB, 60 to 10,000 Hz, +2-4 dB, 30 to 15,000 Hz
 3-3/4 ips: +2 dB, 50 to 7,500 Hz
 1-7/8 ips: +3 dB, 100 to 6,000 Hz
 15/16 ips: +3 dB, 100 to 3,000 Hz

8. Ampex Model 622 Speaker-Amplifier (Serial No. 3664524).

Power: Volts A.C. 117.

Cycles: 50/60.

Amperes: 0.5.

APPENDIX B

PERMISSION REQUEST

Dear Parent:

I am a graduate student in speech pathology at Portland State University and I am studying the /r/ sound as in bird and father. I have the approval of the Lake Oswego School District and with your permission, I would like to record your child making the /r/ sound.

This is not a test, and in no way will your child's name be used in this study. Will you please help me by signing this slip and returning it to your child's teacher. The teacher will give it to me.

Parent's Signature

Thank you,

Scott R. Lane
Graduate Student - PSU

APPENDIX C

JUDGE'S RESPONSE FORM

NAME: _____

DATE: _____

SEX: M F

BIRTHDATE: _____

PHONE: _____

SET NUMBER	CORRECT	PARTIALLY CORRECT	INCORRECT
Sample A			
Sample B			
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			

APPENDIX D

INSTRUCTIONS TO JUDGES

There are three categories with which to classify stimuli in this study. They are, correct, partially correct, and incorrect. Your task is to decide which stimulus is compatible with each category and to appropriately record your response. For example, if Stimulus "A" is incorrect write "A" in that column. You will hear 10 sets of stimuli, with 3 stimuli in each set. The first stimulus you hear in each set will be labeled A, the second, B, and the third, C. Each set will be repeated 3 times. You must put one stimulus in each category. Are there any questions?