

6-9-1994

# A Comparison Between Trained Ear Estimation and Orthographic Transcription When Measuring Speech Intelligibility of Young Children

Nancy Kay Sugarman  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Speech and Rhetorical Studies Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation


Sugarman, Nancy Kay, "A Comparison Between Trained Ear Estimation and Orthographic Transcription When Measuring Speech Intelligibility of Young Children" (1994). *Dissertations and Theses*. Paper 4775. <https://doi.org/10.15760/etd.6659>


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](mailto:pdxscholar@pdx.edu).


THESIS APPROVAL

The abstract and thesis of Nancy Kay Sugarman for the Master of Science in Speech Communication: Speech and Hearing Sciences were presented June 9, 1994, and accepted by the thesis committee and the department.


COMMITTEE APPROVALS:

  
Mary Gordon-Brannan, Chair

  
Ellen Reuler

  
Ruth Falco  
Representative of the Office  
of Graduate Studies


DEPARTMENT APPROVAL:

  
Stephen A. Kosokoff, Chair  
Department of Speech Communication

\* \* \* \* \*

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by

 on 26 August 1994

## ABSTRACT

An abstract of the thesis of Nancy Kay Sugarman for the Master of Science in Speech Communication: Speech and Hearing Sciences presented June 9, 1994.

Title: A Comparison Between Trained Ear Estimation and Orthographic Transcription When Measuring Speech Intelligibility of Young Children

When the primary mode of communication is speech, the crucial ingredient for successful communication is intelligible speech. The speech of children with disordered phonologies is often unintelligible. Accurate and reliable measurement of speech with compromised intelligibility is essential if appropriate treatment procedures are to be chosen and implemented.

The focus of this investigation was the measurement of speech intelligibility in young children. The purpose of this study was to investigate the relationship between the subjective method of trained ear estimation and the objective method of orthographic transcription when measuring the speech intelligibility of young speakers with a wide range of phonological proficiency. For this study, the standard measurement of intelligibility was operationally defined as the percentage of words understood in a continuous speech sample derived from orthographic transcription of the sample. The secondary purpose was to

investigate the accuracy of the speech-language pathologists' estimates as compared to the standard measure for each of the three groups: (a) the children with the most intelligibility, (b) with average intelligibility, and (c) with the least intelligibility.

Data were collected from 47 children, aged 4:0 to 5:6, who comprised three groups with varying levels of intelligibility. Two groups of listeners who were unfamiliar with the speakers, but familiar with the topic, rated the children's percentage of intelligibility from continuous speech samples via orthographic transcription or trained ear estimation.

The two methods of measuring speech intelligibility investigated in this study were found to correlate highly ( $r = .96$ ). However, there was a significant difference between the percentages derived from orthographic transcription and those derived from trained ear estimation for some speakers. The  $t$ -test analyses revealed significant differences between the two measures for the two most intelligible groups, and no significant difference for the least intelligible group. It appears that the subjective method of estimating speech intelligibility with trained ears correlates with the objective method of orthographic transcription, but yields a different percentage score for some speakers.

A COMPARISON BETWEEN TRAINED EAR ESTIMATION  
AND ORTHOGRAPHIC TRANSCRIPTION WHEN MEASURING  
SPEECH INTELLIGIBILITY OF YOUNG CHILDREN

by

NANCY KAY SUGARMAN

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

MASTER OF SCIENCE  
in  
SPEECH COMMUNICATION:  
SPEECH AND HEARING SCIENCES

Portland State University  
1994

## ACKNOWLEDGMENTS

Throughout the course of this challenging and time-consuming project, I've had the fortunate opportunity to work with a great group of people. Thanks to their encouragement and guidance, along with some raw determination, this task has been completed.

My advisor and mentor, **Mary Gordon-Brannan**, showed tremendous support throughout the writing of my thesis. Especially in the last month before its completion, she often critiqued and returned it in only a day's time. I sincerely appreciate the expert advise, encouragement, and valuable time that she afforded me. Another committee member, **Ellen Reuler**, invested much of her time and energy into revising and critiquing my rough drafts as well. She asked pertinent and probing questions which helped clarify my writing immensely. **Ruth Falco**, the final member on my committee, had helpful insights that added greatly to the finished product. Special thanks goes to all my committee members for accepting the task of reading my thesis in a matter of 4 short days in preparation for my defense. I'm grateful for their willingness to go above and beyond the expected call of duty to help me finish on time.

Four speech-language pathologists, **Karen Jenkins, Ann Lynn Laskey, Lou Ann McCoy, and Cathy Thompson,** contributed their time and "trained ears" and deserve special thanks for participating in my study.

My husband, **Ken,** was a loving and helpful support to me throughout. He was encouraging, understanding, and never lost faith in my ability to finish. He even searched the whole house when one of my computer discs was missing!

My thanks would not be complete without mentioning the other faculty that helped make my experience at PSU wonderful, as well as challenging. Special thanks go to **Allyson Goodwyn-Craine, Maria Monserrat-Hopple, Dr. Doug Martin, Joan McMahon, Dr. Rhea Paul, Dr. John Tetnowski, and Mary T. Withers.**

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS . . . . .	ii
LIST OF TABLES . . . . .	vii
CHAPTER	
I INTRODUCTION . . . . .	.1
Statement of Purpose . . . . .	.2
Definition of Terms . . . . .	.4
II REVIEW OF THE LITERATURE . . . . .	.8
Factors that Affect Intelligibility . . . . .	.8
General Factors	
Phonological Factors	
Speech Intelligibility Measures . . . . .	10
Objective Measures	
Subjective Measures	
Objective versus Subjective Measures	
Severity and Speech Intelligibility . . . . .	16
Summary . . . . .	18
III METHOD . . . . .	20
Participants . . . . .	20
Speakers	
Listeners	
Measures of Intelligibility . . . . .	24



	Procedures . . . . .	24
	Preliminary Procedures	
	Speech Samples	
	Judgements	
	Scoring	
	Reliability	
	Data Analysis . . . . .	29
IV	RESULTS AND DISCUSSION . . . . .	30
	Results . . . . .	30
	Reliability	
	Research Question I	
	Research Question II	
	Discussion . . . . .	37
	Listener Reliability	
	Degree of Intelligibility	
	Listener Estimations	
V	SUMMARY AND IMPLICATIONS . . . . .	47
	Summary . . . . .	47
	Implications . . . . .	49
	Clinical	
	Research	
	REFERENCES . . . . .	53
	APPENDIXES	
A	INDIVIDUAL SUBJECT DATA . . . . .	58
B	CHARACTERISTICS OF THE THREE SPEAKER GROUPS	59
C	INFORMED CONSENT-SLPs . . . . .	60
D	INFORMED CONSENT-PARENTS . . . . .	62
E	QUESTIONNAIRE . . . . .	63

F	LISTENER INSTRUCTIONS-GRADUATE STUDENTS . . .	64
G	LISTENER INSTRUCTIONS-SLPs . . . . .	65
H	RAW DATA BY LISTENER - PERCENTAGE-OF-WORDS UNDERSTOOD IN CONTINUOUS SPEECH/ ORTHOGRAPHIC TRANSCRIPTION - GRADUATE STUDENTS. . . . .	67
I	RAW DATA BY LISTENER - PERCENTAGE-OF-WORDS UNDERSTOOD IN CONTINUOUS SPEECH/ TRAINED EAR ESTIMATION - SLPs . . . . .	70
J	LISTENER RATINGS FOR SPEAKERS RATED TWICE-SLP . . . . .	73

LIST OF TABLES

TABLE		PAGE
1	Phonological Characteristics of The Three Speaker Groups (Ages 4 and 5 Years) . .	23
2	Orthographic Transcription Correlation Matrix for Intelligibility Measures . . . . .	31
3	Trained Ear Estimation Correlation Matrix for Intelligibility Measures . . . . .	31
4	Descriptive Statistics for Two Intelligibility Measures for Three Speaker Groups . . .	36

## CHAPTER I

### INTRODUCTION

Intelligible speech is crucial for effective communication between people. Among various researchers, Bernthal and Bankson (1988) considered the measure of intelligibility to be the primary indicator of communication competency.

The level of intelligibility in the speech of children with phonological deviancies can range from slightly to grossly unintelligible, therefore affecting their communicative effectiveness. Accurately assessing the speech of these children is very important when deciding if speech and language intervention should be provided, and if so, what type of intervention is needed.

A common method used today for assessing speech intelligibility is to elicit a continuous speech sample of a child's speech. From the speech sample, many speech-language pathologists (SLPs) use their trained ears to judge how intelligible a child is (Gordon-Brannan, 1993a). They estimate the number of words understood from a continuous speech sample in the form of a percentage. For example, they may phrase their estimate as, "Jimmy was judged to be approximately 75% intelligible when the

context was known, and approximately 35% intelligible when the context was unknown."

Orthographic transcription is an objective method used to measure percentage of speech intelligibility that involves writing down each intelligible word in an utterance. It is not possible to transcribe the unintelligible words. A percentage of intelligibility is derived by dividing the total number of words in the speech sample by the number of intelligible words. For example, if the speech sample had 100 words in it and only 65 were understandable, then the child's speech would be considered to be 65% intelligible.

A child's percentage of speech intelligibility is a major factor when determining what course intervention should take. It is very important, then, to make sure an accurate assessment is made of the child's speech. The question raised by this estimation procedure is: Is the estimate made by the trained ears of a SLP accurate enough to qualify a child for services, or should a more objective method be used to determine percentage of speech intelligibility?

#### Statement of Purpose

The purpose of this study was to investigate the relationship between the subjective method of trained ear

estimation and the objective method of orthographic transcription when measuring the speech intelligibility of preschool children with a wide range of phonological proficiency. For this study, intelligibility was defined as the percentage of words in connected-speech samples correctly understood by unfamiliar listeners. The secondary purpose was to investigate the accuracy of the speech-language pathologists' estimates in comparison to the intelligibility measure derived from orthographic transcription, for each of three speaker groups: (a) the children with the most intelligibility, (b) with average intelligibility, and (c) with the least intelligibility.

The following research questions were investigated:

1. What is the correlation between trained ear estimations of the percentage of words understood in continuous speech samples and the percentage of words understood derived from the method of orthographic transcription?

2. Are there significant differences between trained ear estimations of the percentage of intelligibility and the percentages derived from orthographic transcription of the three groups: (a) preschool children with the most intelligibility, (b) with average intelligibility, and (c) with the least intelligibility?

### Definition of Terms

The following terms were defined as follows for this study:

AAPS - Arizona Articulation Proficiency Scale (Fudala & Reynolds, 1986).

ACI - Articulation Competence Index (Shriberg, 1993).

APD - Average percentage of occurrence of phonological deviations (Hodson, 1986).

APP-R - Assessment of Phonological Processes-Revised (Hodson, 1986).

Assimilation - Influence of one sound in a word or phrase upon another sound to make it the same as the influencing sound in one or more features (i.e., voicing, place, or manner). Example: /dadi/ for doggie, and /t et/ for cat (Hodson & Paden, 1991).

Backing - Replacing an anterior consonant with a posterior one. Example: /go/ for toe, and /k b/ for tub (Hodson & Paden, 1991).

Cluster Reduction - One or more of the consonants of a consonant cluster is omitted. Example: /mok/ for smoke, and /bek/ for brake (Hodson & Paden, 1991).

Consonant Sequence Omission - The omission of one or more sound segments from two or more contiguous consonants. Example: /tIn/ for string (Hodson & Paden, 1991).

DME - Direct magnitude estimation (subjective

intelligibility measure) (Schiavetti, 1992).

Dysarthria - Motor speech disorder caused by weakness, paralysis, slowness, incoordination, or sensory loss in the muscle groups responsible for speech (Brookshire, 1992).

Dysfluency - Occurs when the forward flow of speech is interrupted abnormally by repetitions or prolongations of a sound, syllable, or articulatory posture, or by avoidance and struggle behaviors (Van Riper & Emerick, 1990).

Dyspraxia - Neuromuscular speech problem characterized by inability or difficulty in performing speech acts voluntarily (Weiss, Gordon, & Lillywhite, 1987).

Final Consonant Deletion - The final singleton consonant in a word is omitted. Example: /bo/ for boat, and /ma/ for mop (Hodson & Paden, 1991).

Fronting - Replacing a posterior consonant with an anterior one. Example: /to/ for go, and /t b/ for cub (Hodson & Paden, 1991).

Glide Deviation - Omission of the glide feature by substitution of a non-glide sound or by totally omitting the target glide sound. Example: /In/ for win, and /g s/ for yes (Hodson & Paden, 1991).

Gliding - A glide (i.e., /w/ or /j/) is substituted for a sound in another class. Example: /w d/ for red or



/jaIt/ for light (Hodson & Paden, 1991).

Glottal Replacement - Substituting a glottal stop for a consonant. A glottal stop is not a distinctive sound in most dialects of the English language. Example: /fI?In/ for fishing, and /b e?/ for bath (Hodson & Paden, 1991).

Initial Consonant Deletion - A singleton consonant in a syllable is omitted. Example: / et/ for hat (Hodson & Paden, 1991).

Laryngectomy - The surgical removal of the larynx (Van Riper & Emerick, 1990).

Liquid Deviation - A liquid (i.e., /l/ and /r/) is omitted entirely or is replaced by a non-liquid (Hodson 1986).

Nasal deviation - Omission of the nasal feature by substitution of a non-nasal sound or by totally omitting the target nasal sound. Example: /bi/ for me, and /bek/ for make (Hodson & Paden, 1991).

Nasalization - Nasal emission during the production of typically non-nasal sounds (Hodson & Paden, 1991).

NTID - National Technical Institute for the Deaf (Johnson, 1975).

Phonological Process - A regularly occurring deviation in an individual's utterances, usually one that simplifies an adult phonological pattern (Hodson & Paden, 1991).

Postvocalic Singleton Omission - Final singleton consonant in a word is omitted. Same as final consonant deletion (Hodson & Paden, 1991).

Prevocalic Singelton Omission - A singleton consonant in a syllable is omitted. Same as initial consonant deletion (Hodson & Paden, 1991).

Prevocalic Voicing - Voicing an unvoiced consonant when it precedes a vowel. Example: /bIɡ/ for pig (Hodson & Paden, 1991).

SLP - Speech-language pathologist

Stridency Deletion or Stridency Deviation - Omission of the strident feature by substitution of a nonstrident sound or by totally omitting the target strident sound. Example: /tar/ for star, and /bIp / for zipper (Hodson & Paden, 1991).

Stopping - Substitution of stops for other consonants.

Example: /kIt/ for kiss, /dut/ for juice, and /t ni/ for funny (Hodson & Paden, 1991).

Velar Deviation - A velar (i.e., /k/, /g/, and /y/) is omitted entirely or is replaced by a non-velar (Hodson, 1986).

Weak Syllable Deletion - Omission of an unstressed syllable in a multisyllabic word. Example: /teto/ for potato, and /n en / for banana (Hodson & Paden, 1991).

## CHAPTER II

### REVIEW OF THE LITERATURE

Intelligible speech is considered one of the most fundamental aspects necessary for communication (Connolly, 1986). Numerous factors influence an individual's intelligibility level including any of the following conditions: dysarthria, hearing impairment, aphasia, dyspraxia, and laryngectomy. In addition to etiology, a number of other factors affect intelligibility and are not associated with the aforementioned disorders. Examples include familiarity of the listener with the speaker, length of utterance, phonological proficiency, and word pronunciation.

General factors that affect intelligibility will be discussed in the first part of this literature review followed by factors specific to phonologically disordered speech. In the second part of this review, some objective and subjective measurements commonly used to evaluate the speech intelligibility of persons with phonological disorders will be described, followed by a short section about severity and speech intelligibility.

#### Factors That Affect Intelligibility

##### General Factors

A number of factors unrelated to a phonological deficiency affect intelligibility (Gordon-Brannan, 1993a, Kent, Miolo, & Bloedel, 1994; Shriberg & Kwiatkowski, 1982; Weiss & Lillywhite, 1981). Contextual factors such as listener familiarity with the speaker and the spoken material, context of the message being sent (i.e., a known vs. an unknown topic), quality and clarity of the acoustic and visual signals, and linguistic factors such as sentence structure and length of utterance are related to degree of intelligibility. Speech-related characteristics also influence intelligibility, including word pronunciation, speech-sound additions, articulation errors, communicative dysfluency, and suprasegmental features such as phrasing, rate, stress, loudness, pitch, and quality.

#### Phonological Factors

From their research results, Hodson and Paden (1981) specified 11 processes that occur most frequently in speakers with phonological deficiencies and that most adversely affect intelligibility. In their study of 60 three-to-eight year old children who were unintelligible, the 5 most commonly occurring phonological processes that had the highest correlation with intelligibility were cluster reduction, stridency deletion, stopping, liquid deviation, and assimilation. The 6 remaining processes occurred less frequently, but were associated with

decreased intelligibility, including velar deviation, final consonant deletion, weak syllable deletion in 3- and 4-syllable words, prevocalic voicing, glottal replacement, and backing.

In a later study of the relationship between intelligibility and phonological process usage, Billman (1986) found that the 2 processes that had the most adverse effect on a child's speech intelligibility and probably should be given priority in selecting remediation targets were backing and prevocalic singleton omission. A third finding revealed that a high percentage of liquid deviation occurred in the children's speech, but probably should not be considered a top priority when determining remediation targets since liquid deviations "were not significantly correlated with intelligibility ratings" (p. 41).

#### Speech Intelligibility Measures

In the following section, quantitative and qualitative measures of speech intelligibility will be discussed. Quantitative, or objective, methods for measuring intelligibility involve calculating an actual percentage of words understood in a speech sample. Qualitative, or subjective, methods are more impressionistic and rely on what the listener perceives as intelligible.

## Objective Measures

Speech intelligibility has been quantified primarily through word identification tests whereby the listener writes down what the speaker says (Schiavetti, 1992). Some of the most common methods of eliciting speech is through word repetition or picture identification tasks, and contextual or conversational speech samples (Kent et al., 1994; Weiss, 1982; Weston & Shriberg, 1992; Wilcox, Schooling, & Morris, 1991).

Many of the word identification tests are designed to measure intelligibility in terms of the percentage of words understood. For example, the Weiss Intelligibility Test (Weiss, 1982) is divided into two subtests:

(a) Isolated Words and (b) Contextual Speech. The listener transcribes 25 words from the first subtest and derives a percentage of intelligibility. A continuous speech sample is elicited for the second subtest, from which the listener transcribes 200 words and calculates the number of words understood and those not understood. The total number of words understood from the two subtests are averaged together to yield an overall intelligibility score. A severity level of normal, mild, moderate, severe, or profound is applied to this score after it is compared with intelligibility norms.

In the Preschool Speech Intelligibility Measure (PSIM)

(Wilcox et al., 1991), the intelligibility of preschool children is evaluated through a word repetition task. The child is required to repeat a set of 50 randomly chosen words after a model by the clinician. A percentage of intelligibility is then calculated from the number of words correctly understood versus the number of words spoken.

Another instrument for measuring percentage of intelligibility is the Children's Speech Intelligibility Test (CSIT) (Kent et al., 1994). The CSIT was developed to test children with limited expressive speech due to any delay, disability, or limitation that prevents the use of phrase length, sentence length, or conversational speech. Single words are elicited through word repetition or picture identification tasks. In addition to determining phonological and composite contrast scores, an overall score of the percentage of words correctly produced is also derived by dividing correct productions by total productions (Kent et al., 1994).

A more informal approach to calculating the percentage of words understood involves orthographically transcribing each word from a conversational speech or reading sample. The percentage of intelligibility score is then derived by dividing the number of words uttered by the number of words understood (Kent et al., 1994).

### Subjective Measures

Scaling procedures are available specifically for the

measurement of speech intelligibility in communication disorders. According to Schiavetti (1992), the most widely used methods are equal-appearing interval scaling and direct magnitude estimation (DME). Both of these methods use a listener's perception of a speech sample to assign numbers that reflect that speaker's percentage of intelligibility and severity level.

Interval scaling procedures involve placing a number on a continuum that represents the intelligibility of a speaker's speech sample. The continuum is usually a scale numbered from 1 to 5, 7, or 9 with the numbers representing degree of intelligibility. For example, in the National Technical Institute for the Deaf (NTID) rating scale, the beginning of the continuum, number 1, is assigned to speech that is completely unintelligible, and the end of the continuum, number 5, is assigned to speech that is completely intelligible (Johnson, 1975). Some rating scales are not labeled with descriptors, only numbered intervals. It appears that there is no significant difference between using numerical or descriptive scales when determining intelligibility (Guilford, 1954; Stevens, 1975).

DME does not require that the rating fall upon a scaled continuum. The only requirement is that the rating of perceived intelligibility is proportional to the rest of



the speech samples (Schiavetti, 1992; Toner & Emanuel, 1989). With DME, a standard/modulus may be used. One of the speech samples is rated for intelligibility and becomes the standard speech sample. A modulus number, usually either 10 or 100, is given to this standard sample and represents a subjective value of intelligibility. This standard sample and modulus are used as reference points for scoring the intelligibility of the remaining samples (Schiavetti, 1992).

When a standard and modulus are not used, the listener may assign any number to the first speech sample. The samples that follow are given numbers "that correspond to the ratios of the perceived magnitudes of the intelligibility of the various speech samples" (Schiavetti, 1992, p. 21).

#### Objective versus Subjective Measures

There are advantages and disadvantages to either method of measuring intelligibility. According to some researchers (Beukelman & Yorkston, 1979; Metz, Samar, Schiavetti, Sitler & Whitehead, 1985; Samar & Metz, 1988; Schiavetti, 1992; Yorkston & Beukelman, 1978), word identification measures are more advantageous than scaling procedures for the following reasons:

1. The results are easier to use and understand for the lay person and other professionals.

2. There is good criterion validity for both isolated word and contextual speech intelligibility measures.

3. Administration and scoring time is essentially equal to that of the NTID rating scale, a scaling procedure.

4. Word identification tests are reliable measures of speech intelligibility.

5. In addition to assigning severity levels to speech disorders, a description of the disorder and documentation of any therapeutic change may be provided.

These reasons were generated from research primarily with persons with hearing impairments and with dysarthria.

Despite these reasons, Schiavetti (1992) indicated it is important to be able to recognize when a scaling procedure might be a more direct and appropriate method than any available quantitative measure. Young (1969) indicated that subjective measurement should always be recognized as a more appropriate and valid method than quantitative measures when drawing final conclusions about intelligibility. He stated:

... a measurement of a speech disorder is primarily a perceptual event, and the observer's response necessarily represents the "final" validation for any measurements. (p. 135)

Scaling procedures are considered advantageous by some because of their convenience, cost, efficiency, and ease

(Metz, Schiavetti, & Sitler, 1980). Quantitative measures appear to have better face validity than scaling measures, but can become very time-consuming and costly. Scaling procedures provide the SLP with a short and inexpensive alternative to assessing a client's speech.

### Severity and Speech Intelligibility

Level of severity and degree of intelligibility are constructs that are very similar to each other. Bernthal and Bankson (1988) regarded speech intelligibility as the factor used most often when judging severity level. Shriberg and Kwiatkowski (1982) proposed that severity encompasses three constructs, including intelligibility, disability, and handicap. Research by Billman (1986) revealed a significant positive correlation ( $r = +.79$ ) between intelligibility and severity measures of children with phonological disorders. Billman also stated that severity level is the perceived degree of impairment, and is correlated to the amount of difference between a person's utterances and that of adults in the linguistic community. The terms normal, mild, moderate, severe, and profound are often used to describe level of severity (Hodson, 1986; Shriberg & Kwiatkowski, 1982; Weiss, Gordon, & Lillywhite, 1987).

Severity level has been measured by methods such as

the Percentage of Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982), Articulation Competence Index (ACI) (Shriberg, 1993), the average percentage of occurrence of phonological deviations (APD) from the Assessment of Phonological Processes-Revised (APP-R) (Hodson, 1986), and the Arizona Articulation Proficiency Scale (AAPS) (Fudala & Reynolds, 1986).

The PCC, that is, the total number of consonants correct divided by the total number of intended consonants in a speech sample, was devised by Shriberg and Kwiatkowski (1982) as a severity measurement. It is translated into one of the following severity levels: (a) mild, 85-100%; (b) mild-moderate, 65-85%; (c) moderate-severe, 50-65%; and (d) severe, less than 50%. Shriberg (1993) later developed the ACI to account for the lesser influence of articulatory distortions as compared to omissions and substitutions, upon severity level. In the ACI, a conversational speech sample is used to determine the percentage of consonants correctly produced (PCC) and the percentage of consonants produced incorrectly because of articulatory disorders. The latter percentage is interpreted as the relative percentage of distortion errors (RDI), that is, the total number of distortion errors divided by the total number of articulation errors in a speech sample. To determine a child's "articulation

competence," the PCC and RDI of the child's speech sample are added together and divided by 2 to yield an ACI score between 0-100%. Such a measure gives less weight to distortions as compared to substitutions and omissions.

In the APP-R, a phonological deviation score is calculated from the percentage scores of 10 frequently occurring phonological processes. The phonological deviation score is converted into a level of severity (i.e., mild, moderate, severe, or profound), based upon the average percentage of occurrence of phonological deviations and the age of the child (Gordon-Brannan, 1993a; Kent et al., 1994).

In the AAPS, consonant production in the initial and final positions of words are assessed. Each consonant is numerically weighted according to its frequency of occurrence in the English language. The idea behind this test is that the more errors made on the frequently occurring consonants, the more unintelligible a child's speech. An estimated percentage of speech intelligibility can be derived from the total weighted value of misarticulated sounds. This percentage can be interpreted to show the severity level of a child's speech (Fudala & Reynolds, 1986).

#### Summary

Speech intelligibility within the phonologically

deficient population needs to be assessed with reliable, valid methods. There are advantages and disadvantages to both subjective and objective methods of measuring intelligibility. Perhaps a combination of both methods is the best solution. When time is limited in the schools for assessing children with disorders, however, it is often the quickest method of evaluating speech intelligibility that is implemented, rather than the most reliable or valid method. The best interest of the child must be of top priority which means that accurate, valid, and reliable assessments must be used. In this study, the accuracy of a relatively quick method of assessing intelligibility, that is, estimation by tracking from continuous speech samples, was explored.

## CHAPTER III

### METHOD

The purpose of this study was to investigate the relationship between the subjective method of trained ear estimation and the objective method of orthographic transcription when measuring the speech intelligibility of preschool children with varying levels of phonological proficiency. Forty-seven continuous speech samples were analyzed for percentage of intelligibility via orthographic transcription and estimation. In an earlier study (Gordon-Brannan, 1993b), unfamiliar, yet trained listeners orthographically transcribed the speech samples and a percentage of intelligibility score was computed for each sample. In this study, the speech samples were judged again, this time by experienced speech-language pathologists who estimated the children's intelligibility.

#### Participants

##### Speakers

The speakers consisted of 47 preschoolers, 20 females and 27 males, ranging in age from 4:0 to 5:6 (mean = 4:7). They comprised three groups of children who ranged phonologically from no articulation errors to numerous

omission errors. These speakers were selected for an earlier study on intelligibility (Gordon-Brannan, 1993b), according to the following criteria:

1. No known neurological, motor, or physical impairment that might affect speech production noted according to parent interview and questionnaire, and investigator observation.

2. No clinically significant laryngeal or resonance deviancy noted at the time of testing according to investigator observation.

3. Age-appropriate receptive language according to the results from the Test of Auditory Comprehension-Revised (TACL-R) (Carrow-Woolfolk, 1985).

4. Hearing sensitivity within the normal/mild loss range according to a pure tone hearing screening set at 35 dB at 500K, 1K, and 2K Hz.

The speakers who qualified for inclusion in the original study were then dispersed into 4 subgroups according to the scores they received on the 1-Minute Measure of Homonymy and Intelligibility (Hodson, 1992). For this study, only 47 of the original 48 subjects were included due to "white noise" on the recording of one speech sample. The sample was omitted since the child's voice was difficult to hear. The subjects in this study were regrouped into two groups of 16, and one group of 15



children, based on intelligibility. Group I was comprised of those demonstrating the most intelligibility, group II was made up of those with average intelligibility, and group III was composed of those with the least intelligibility. Table 1 shows the phonological characteristics of each of the three groups of children. Characteristics of individual speakers and the three speaker groups are provided in Appendixes A and B.

### Listeners

The speech samples were judged by two groups of listeners. The first group was involved in the earlier Gordon-Brannan (1993b) study, and was comprised of four graduate students in the Speech and Hearing Sciences Program at Portland State University (PSU). This group had completed coursework in phonology, had clinical experience with clients with speech and language disorders, and were unfamiliar with the speakers. A percentage of words understood for each speech sample was calculated from their orthographic transcriptions.

The second group of listeners was comprised of four SLPs from the greater Portland, Oregon, metropolitan area. Each had a minimum of 10 years experience (range = 10 to 17 years; mean = 14 years) working in the schools with caseloads including students with articulation/phonological disorders, and signed an informed consent form (Appendix C).

Table 1

Phonological Characteristics of the Three Speaker Groups  
(Ages 4 and 5 Years)

Group	Phonological Proficiency/Deficiency	Speech-Sound Characteristics on the 1-Minute Measurement of Homonymy & Intelligibility
Most	Adult-like	No more than 1 phonemic substitution, addition, or metathesis and 1 non-phonemic variation (e.g., lisp), no omissions/glottal replacements
Average	Mild/Mod deficient	3-12 phonemic substitutions, additions, and metathesis; 0-9 omissions/glottal replacements
Least	Mod/Severely deficient	At least 12 phonemic substitutions, additions, and metathesis; 1-9 omissions/glottal replacements (moderate); at least 11 omissions and/or glottal replacement (severe)

These listeners were unfamiliar with the speakers and estimated the percentage of words understood to each speech sample after listening to them via audiotape.

### Measures of Intelligibility

The original study used orthographic transcriptions as the standard for measuring speech intelligibility. Speech-language pathology students wrote down each identifiable word in the speech samples. An X was used to indicate any syllable that was not understood. From these transcriptions, the mean percentage of words correctly understood was calculated for each speaker.

This study employed the use of trained ear estimations to measure speech intelligibility. Experienced speech-language pathologists used their trained ears to estimate percentage of intelligibility (i.e., estimate of percentage of words understood) of each child's speech sample.

### Procedures

#### Preliminary Procedures

For the Gordon-Brannan (1993b) study, speakers were selected from greater Portland, Oregon, metropolitan area preschools and speech-language pathology caseloads. An informed consent form was signed by the parents/caregivers before any testing on their children was done (Appendix D).

In addition to the informed consent form, the parents/caregivers filled out questionnaires on their children's speech, hearing, developmental history, and socioeconomic level (Appendix E). With one exception, all subjects met the selection criteria specified above. This child (Speaker 47) did not meet the criterion regarding receptive vocabulary, that is, obtaining a score at the 10th percentile or above on the TACL-R. However, this child was admitted to the study due to the professional judgement of the original investigator and the child's SLP that his receptive language skills were appropriate for his age level (Gordon-Brannan, 1993b).

#### Speech Samples

The continuous speech samples from the earlier study were elicited in a sound-treated recording room at PSU, and were both audio-taped and video-recorded. The video was viewed later by parents of the children with moderate/severe phonological deficiencies. The equipment used to audiotape and videotape the samples included a Panasonic camcorder, VHS Reporter, Ag-100, and a Sharp SX D200 digital audiotape recorder. An AKG, Model C451, capacitor flat microphone was used to record the speech samples and was placed approximately 6" from the speaker's mouth. The microphone sat on foam or in a microphone stand on a table covered with cloth. The 100-word speech sample was

elicited through retelling the story, The Relatives Came (Rylant & Gammell, 1985), and five pictures of children engaged in everyday activities, if needed.

In the original study, scoring keys for each continuous speech sample were prepared. Transcripts were made from the initial orthographic transcriptions for the investigator to verify. Then the parent/caregiver of the children with moderate/severe deviciencies reviewed the content of the transcriptions. They either verified or corrected the listener's interpretations of their child's speech by identifying words that were unintelligible or misunderstood by the investigator.

In the previous study, the speech samples, including both the speakers' and investigator's utterances, were dubbed onto listener tapes. These audiotapes were dubbed onto the listener tape in random speaker order, and were used for later transcription and rating. A total of five speech samples, at least one from each of the original speaker groups, were presented twice on the listener tapes for the purpose of determining intrajudge reliability. The repeated samples were presented at the end of the listener tape.

### Judgements

The graduate students from the original study were familiarized with the pictures and book used to elicit the

continuous speech samples before they listened to the samples. They were instructed to listen to each utterance a maximum of three times on their own analogue audiotape recorders, and then orthographically transcribed each utterance. Written directions for this task were given (Appendix F).

The four SLPs for this study were also familiarized with the pictures and book used to elicit the continuous speech samples. After the stimulus materials were shown and described to them, they listened to the continuous speech samples as a group on two different occasions via digital audiotape equipment. The digital listener tapes were presented to the listeners through a Denon digital audiotape recorder (Model DTR-80P) connected to a Sony table-top speaker (Model SRS-150). To test interrater reliability, 5 of the 47 speech samples were played twice without the listeners' knowledge, for a total of 52 samples. Each sample was played once for the SLPs during which they estimated the percentages of words understood for each sample. The directions for this task were given both orally and in writing (Appendix G).

### Scoring

In the original Gordon-Brannan (1993b) study, the listeners' orthographic transcriptions were compared with the transcription keys prepared by the investigator and

then verified by the parent. If the listener identified a word that differed from the one identified by the parent/investigator transcript, the listener's transcribed word was considered incorrect. Words with differences in morphological form only were not considered incorrect. Words that were not identified by either the listeners or the parent/investigator were considered incorrect. Words identified by the listener, but not by the parent/investigator, were considered correct. The percentage of words understood by each listener for each continuous speech sample was computed. Additionally, the means of the SLP estimated percentages and the graduate student transcription percentages were determined for each speaker.

### Reliability

The Pearson product-moment correlation (Pearson  $r$ ) was used to determine interjudge reliability between each pair of graduate student listeners and each pair of SLP listeners. Six comparisons for each listener group were made for a total of 12.

To determine intrajudge reliability within each experienced SLP, the percentages assigned to the five speech samples that were recorded twice were compared by using the Pearson  $r$ . Discrepancy scores were also computed for each listener to examine intrajudge reliability further.

### Data Analysis

A Pearson  $r$  was used to address the first research question of the correlation between the trained ear estimation method for measuring speech intelligibility and the method of orthographic transcription. The results indicated whether or not the subjective method (estimation) correlates with speech intelligibility as measured by the objective standard method (orthographic transcription). To address the second research question of whether there is a significant difference between trained ear estimation and orthographic transcription intelligibility percentages for the three groups of children: (a) with the most intelligibility, (b) with average intelligibility, and (c) with the least intelligibility, two-tailed  $t$ -tests were used. Statistical significance was set at the .05 level of confidence for all data analyses.



## CHAPTER IV

### RESULTS AND DISCUSSION

#### Results

The results of a comparison between orthographic transcription and trained ear estimation when measuring speech intelligibility of preschool children with varying levels of phonological proficiency are presented in the following sections. In addition, mean scores from four listeners for three groups of children divided into most, average, and least intelligible, are compared.

#### Reliability

Preliminary to comparing the two measures of intelligibility, interjudge and intrajudge reliability was investigated for the two listener groups. The percentage data for each speaker sample by each listener appear in Appendixes H and I. Tables 2 and 3 provide the reliability coefficients for interjudge reliability between each pair of listeners within the two groups of listeners, those from the Gordon-Brannan study (1993) who used orthographic transcription and those who used their trained ears to estimate speech intelligibility. Pearson  $r$  correlation matrices were provided through the SYSTAT computer program for all listener pairs in each listener group. Pearson  $r$

Table 2

Orthographic Transcription Correlation Matrix for  
Intelligibility Measures

Listener	Listeners			
	A	B	C	D
A	1.00			
B	.89	1.00		
C	.86	.92	1.00	
D	.88	.95	.94	1.00

Note: Each correlation represents 2 listeners. The critical value for a 1-tailed  $r(3)$  at the .01 level is .930; at the .05 level, .805.

Table 3

Trained Ear Estimation Correlation Matrix for  
Intelligibility Measures

Listener	Listeners			
	1	2	3	4
1	1.00			
2	.94	1.00		
3	.88	.92	1.00	
4	.92	.95	.98	1.00

Note: Each correlation represents 2 listeners. The critical value for a 1-tailed  $r(3)$  at the .01 level is .930; at the .05 level, .805.

correlations for the percentage-of-words understood in continuous speech ranged from .86 to .95 for the six pairs of listeners using orthographic transcription, and from .88 to .98 for the six pairs of listeners using trained ear estimation. These correlations indicate the four listeners in each group were in general agreement in assessing the speech intelligibility of the continuous speech samples.

Pearson  $r$  correlations were used to determine intrajudge reliability within each listener of the second group. This was achieved by comparing the percentages assigned to five speaker samples that were judged twice. The  $r$ -values for the listeners in the group that used trained ear estimation to measure intelligibility are as follows: Listener 1, .99; Listener 2, .99; Listener 3, .96; and Listener 4, .95. These results indicate that each of the listeners who used trained ear estimation to rate the same five speakers twice were highly reliable. The group of listeners from the initial study used a rating scale, not orthographic transcription, to rate the five speakers twice. Those data will not be included in this investigation since it was not for the orthographic transcription data.

To examine intrajudge reliability further, a discrepancy model was employed (Appendix J). For three

of the five speakers listened to twice, Listener 1 assigned estimated percentages of intelligibility to within 1% of the other continuous speech sample presentations. She rated the second presentation of the last two speakers within 7% and 10% of their first assigned percentage. Thus, the discrepancy scores of Listener 1 were -1, +10, +7, +1, and +1, with a discrepancy score mean of 4 percentage points. Listener 2 assigned the same estimated percentage on two presentations to one of the five speakers listened to twice. She assigned percentages within 2, 5, 7, and 12 percentage points of the first assigned percentage for the four remaining speakers. Her discrepancy score mean was 5.2 percentage points. Listener 3 assigned the same percentage to two of the five continuous speech samples. She estimated the intelligibility of the second presentation for the remaining three speakers to within 2, 5, and 24 percentage points of their first presentation. Her discrepancy score mean was 6.2%. Listener 4 estimated the same percentage of intelligibility for one of the five speakers, and assigned percentages of within 1, 2, 5, and 28 points of the other for the remaining four speakers. Her discrepancy score mean was 7.2 percentage points. Based on these results from both methods of investigating intrajudge reliability, the four SLP listeners were consistent in their judgement

of speech intelligibility.

#### Research Question I

The first research question investigated was: What is the correlation between trained ear estimations of the percentage of words understood in continuous speech samples, and the percentage of words understood derived from the method of orthographic transcription? A Pearson  $r$  correlation was used to determine the correlation between the two measures. The mean percentage of intelligibility for each speaker for both measures are provided in Appendix A. The resultant Pearson  $r$  correlation was .96 ( $N = 47$ ;  $p = .012$ ). These results indicate speech intelligibility derived by trained ear estimation is highly correlated with the percentage of intelligibility derived from orthographic transcription.

#### Research Question II

The second research question investigated was: Are there significant differences between estimations of the percentage of intelligibility and the percentages derived from orthographic transcriptions of the three groups: (a) young children with the most intelligibility, (b) with average intelligibility, (c) with the least intelligibility? Prior to addressing the group comparisons, the two intelligibility measures were compared for all the speakers using a two-tailed  $t$ -test. The

overall mean score for the listeners who used orthographic transcription to measure speech intelligibility was 76%, while the overall mean score for those who used trained ear estimation was 78%,  $t(46) = -2.61$ ,  $p = .012$ . These results indicate a statistically significant difference for the preset level of confidence (i.e.,  $p = .05$ ).

Means, standard deviations, and ranges for each of the three speaker groups are shown in Table 4. The most intelligible group received the highest mean scores for the estimation measure, as well as for the standard measure. The mean scores were lower for each succeeding speaker group, with the least intelligible group receiving the lowest mean percentage. To determine if there was a significant difference between the percentages of intelligibility derived from orthographic transcription and those from listener estimations, the means for each speaker group were compared using paired sample  $t$ -tests. When judging the most intelligible speaker group, orthographic transcription ( $M = 93\%$ ) was significantly different from trained ear estimation ( $M = 96\%$ ;  $t(15) = -3.731$ ,  $p = .002$ ). The orthographic transcription mean score of the speaker group with average intelligibility was 81%, while the trained ear estimation mean score for this group was 85% ( $t(15) = -3.564$ ,  $p = .003$ ). These results indicate that the orthographic transcription mean

Table 4

Descriptive Statistics for Two IntelligibilityMeasures for Three Speaker Groups

Measure (transcription)		Groups		
		I	II	III
		Adult-like	Mild/Mod	Mod/Severe
%-of-Words	Mean	93%	81%	52%
Understood	SD	3.5	5.6	13.9
in Speech	Range	88-100%	71-86%	19-71%

Note: Listener group - graduate students; Measurement technique - orthographic transcription; all numbers have been rounded up to the nearest percent; SD = standard deviation; Mod = moderate.

Measure (estimation)		Groups		
		I	II	III
		Adult-like	Mild/Mod	Mod/Severe
%-of-Words	Mean	96%	85%	53%
Understood	SD	4.2	9.1	21.7
in Speech	Range	85-100%	61-95%	9-87%

Note: Listener group - Speech-language pathologists; measurement technique - trained ear estimation; all numbers are rounded up to the nearest percent; Mod = moderate.

for the group with average intelligibility is significantly different from the trained ear estimations for this group. In these two groups, the mean percentage-of-words understood was higher for the estimations. A significant difference was not found between the orthographic transcription measure ( $\bar{M} = 52\%$ ) and trained ear estimation ( $\bar{M} = 53\%$ ) when judging intelligibility of the least intelligible speaker group ( $t(14) = -0.159, p = .876$ ).

In summary, the estimated percentages of intelligibility were significantly different from the orthographic transcription percentages for the young speakers in the two most intelligible subgroups (adult-like and mild/moderate), but were not different for the least intelligible group (moderate/severe).

### Discussion

Two methods for measuring speech intelligibility were investigated in this study, that is, orthographic transcription and trained ear estimation. For this study, the objective method of orthographic transcription is considered the standard measurement because it includes writing down each word the speaker utters in a continuous speech sample. The initial study employed this method of measuring speech intelligibility and enhanced the accuracy



of the scoring of the transcriptions even more by verifying the data with a parent/care-giver of the children with moderately and severely deficient phonologies. In this study, the subjective method of trained ear estimation was compared with the orthographic transcription method to determine the correlation between the two measures, as well as to compare the actual percentages derived from the two methods. The results indicated the two measures were highly correlated, although the actual percentages of the two measures differed significantly for the adult-like and mild/moderate groups.

#### Listener Reliability

The SLPs demonstrated high reliability within themselves when rating five samples twice (Appendix J). Even though their estimations were highly correlated, large discrepancies were noted between percentages assigned to some speakers. For example, one of the speakers from the least intelligible group received the score with the largest variability. Specifically, Listener 1 estimated the intelligibility of Subject 40 as 58% for the first presentation and 68% for the second listening; Listener 2, 42% and 49%; Listener 3, 49% and 73%; and Listener 4, 42% and 70%. The largest discrepancies between estimations were within Listeners 3 and 4, who assigned percentages to the same speech sample a total of 24 and 28 percentage

points apart, respectively. One other speaker (#47) received a discrepant score of 12% by Listener 2. All other discrepancies were 7% or less.

Regarding interjudge reliability, some speakers received a wide range of estimated percentages for the four SLPs (Appendix I). Three speakers, 2, 5, and 7, were judged identically by all four SLPs to be 100% intelligible. At the other end of the continuum, Subject 41 was judged 65% intelligible by Listener 3, but only 20% intelligible by Listener 1, for a difference of 45%. Other notable differences (over 10%) occurred for three speakers (10, 14, and 16) in the adult-like group, seven speakers (25, 26, 28, 29, 30, 31, and 32) in the mild/moderate group, and for 12 speakers (35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, and 46) in the moderate/severe group. These results indicate that, even though the SLPs were in general agreement with each other, their estimations for some individual speakers varied greatly. There was more variability among expert listeners as intelligibility decreased.

The graduate students who used orthographic transcription to assess intelligibility were in general agreement with each other as well. The differences in the percentages for some of the speakers reveal a wide range of orthographically transcribed percentages assigned by the

graduate student listeners (Appendix H). Only one speaker, Subject 1, was shown to be 100% intelligible by two of the four graduate student listeners. No other speakers were found to be 100% intelligible by the transcription method. Subject 41 received the largest percentage difference (42%), as Listener C's transcription was 70%, and Listener A's transcription was 28%. The differences over 10% in the adult-like group occurred for three speakers (7, 15, and 16); in the mild/moderate group for eight speakers (17, 18, 19, 23, 27, 28, 29, and 30); and in the moderate/severe group for thirteen speakers (33, 34, 35, 36, 37, 38, 41, 42, 43, 44, 45, 46, and 47). These results demonstrate high variability between listeners even though their transcriptions were in general agreement with each other.

The two methods of measuring intelligibility discussed in this section yield a wide range of percentages for some of the speakers. Consequently, they both follow a pattern of increased variability as intelligibility decreases.

#### Degree of Intelligibility

The intelligibility scores derived from orthographic transcription ranged from 19% to 100% for the 48 children in the initial study. The scores derived via trained ear estimation ranged from 8% to 100% for the 47 (of the original 48) children in this study. According to the orthographic transcription data, all of the children in the

most intelligible group were understood at least 80% of the time. The group with average intelligibility was understood at least 70% of the time, while the least intelligible group was understood 70% of the time or less, with 6 of its 15 members understood less than 50% of the time.

The percentage of words understood in the most intelligible group ranged from 100% to 88% for orthographic transcription, and from 100% to 85% for the trained ear estimation group. The range of percentages was quite similar for the two measures; however, the statistical analysis indicated a significant difference between estimation and transcription in the most intelligible speaker group.

The group with average intelligibility received percentages from the orthographic transcribers ranging from 86% to 71%, and from 94% to 61% from the trained ear estimators. Thus, the highest estimated score was 8% higher and the lowest estimated score was 10% lower than the standard measure. Again, the t-test results indicated a significant difference between the intelligibility ratings of trained ear estimation and orthographic transcription for this group.

The range of percentages in the least intelligible group were from 70% to 19% by the orthographic

transcription measure, and from 86% to 9% by the estimation measure. Thus, the range difference in this group was 16% between the highest estimated percentage and the standard measure, and 10% between the lowest estimation and the standard measure. The least intelligible group had the greatest amount of variability between the estimated and transcribed percentages. Even though the more intelligible speakers of the least intelligible group received the scores with the largest range between the estimated and orthographic percentages, statistical analysis indicated no significant difference between the two methods of measuring speech intelligibility for this group. In fact, the mean percentages derived from estimation ( $\bar{M} = 53\%$ ) and transcription ( $\bar{M} = 52\%$ ) were virtually identical.

These findings suggest that as intelligibility decreases, the accuracy of estimated intelligibility increases when compared to the standard measure. Also, the variability between orthographic and estimated percentages is greater with decreasing intelligibility.

Visual inspection of the raw data shows that some of the continuous speech samples yielded percentages that differed by 5% or more for the two measurement techniques (Appendix A). In the most intelligible group, 5 of the 16 mean percentage scores derived by trained ear estimation, differed from the orthographic transcription mean by 5% or

more, with the largest difference being 8%. Eleven of the 16 samples in the mild/moderate intelligibility group differed from the orthographically transcribed mean by 5% or more, with 2 of them differing by 11 percentage points. Eleven of the 15 samples in the least intelligible group differed from the orthographically transcribed score by 5% or more, with 3 of them differing by at least 15%.

Appendix A shows the difference of the SLP estimates in plus or minus amounts from the percentages derived from the standard measure. The estimated percentages assigned by the SLPs were generally higher than the orthographic percentages in all speaker groups except for the least intelligible group. For instance, the SLPs assessed 13 of the 16 speakers in the most intelligible group at a higher percentage of intelligibility than did the graduate students using orthographic transcription. Subjects 1, 3, and 15 were the only speakers to receive mean percentages below their standard measure percentage of intelligibility. In the group with average intelligibility, Subject 31 received the only mean estimate below the orthographic transcription mean. The difference between the two percentages was -10%. The only speaker to receive identical means from estimation and orthographic transcription was Subject 26, who was judged to be 80% intelligible by both listener groups. Nine of the 15 least

intelligible speakers received mean estimated percentages below the mean orthographic percentages. Speakers 38 and 44 differed the most between estimated and transcribed percentages with discrepancies of +22.00 and -19.00 percentage points, respectively.

In summary, these results indicate that estimated intelligibility is somewhat higher than actual intelligibility derived from orthographic transcription for children whose speech is 70% or more intelligible. Conversely, estimated percentage of intelligibility tends to be lower than actual intelligibility for children whose speech is less than 70% intelligible.

#### Listener Estimations

The subjectivity of trained ear estimation often raises questions regarding accuracy and reliability (Gordon-Brannan, 1993a; Kent et al., 1994). According to the results found here, there is a high correlation between this measure and the standard, objective method of orthographic transcription. What factors may have affected the listeners' ability to understand the speech samples in this study? The four SLPs in this study were required to have at least 5 years of experience in the schools. In fact, three of the four had 15+ years, and one had 10 years experience as a SLP in the public schools.

All four preferred to estimate the intelligibility

of the speech samples by tracking understood versus not-understood words with slash marks, plus/minus signs, or other symbols. This form of tracking is not a "Gestalt" way of estimating speech intelligibility, but was allowed for this study since all four reported using it when evaluating children in the schools. Amount of experience working with speech/language disordered children and the tracking method may have influenced the accuracy and reliability of the SLP estimations.

Another factor that may have increased the accuracy of their estimations, was the examiner's utterances on the sample tapes. The sample tapes included the conversation between the examiner and the child. The SLPs reported that the content of what the child said was occasionally evident by what the examiner said, and that this may have aided their interpretation of the children's utterances.

Two factors that may have hindered the listeners' understanding were fatigue and desensitization. The SLPs mentioned feeling fatigued after 2½ hours of listening to speech samples. Even though they had a 10-15 minute break, they reported that they became tired, and somewhat desensitized to the task. The more they listened to the children talking about the same story, the more familiar they became with what might be chosen to talk about. Also, listening to disordered speech for such an extended period



of time may have affected the SLPs' ability to listen objectively to each individual speech sample.

Even though the two measures are highly correlated, the time in which it took to use each of the methods greatly differed. The SLPs estimated speech intelligibility of the 47 speech samples on two separate occasions totaling approximately 4½ hours. The graduate students spent anywhere from 12 to 30 hours orthographically transcribing the 48 speech samples from the original study. These results reveal that estimating intelligibility takes considerably less time than orthographic transcription.

In conclusion, the factors that may have influenced the results of this study include the experience of the SLPs, the tracking method of estimating intelligibility, the examiner's utterances on the listener tapes, and listener fatigue or desensitization. These factors should be considered when interpreting the results of this study.

## CHAPTER V

### SUMMARY AND IMPLICATIONS

#### Summary

When the primary mode of communication is speech, the crucial ingredient for successful communication is intelligible speech. The speech of children with disordered phonologies is often unintelligible. Accurate and reliable measurement of speech with compromised intelligibility is essential if appropriate treatment procedures are to be chosen and implemented.

The focus of this investigation was the measurement of speech intelligibility in young children. The purpose of this study was to investigate the relationship between the subjective method of trained ear estimation and the objective method of orthographic transcription when measuring the speech intelligibility of young speakers with a wide range of phonological proficiency. For this study, the standard measurement of intelligibility was operationally defined as the percentage of words understood in a continuous speech sample derived from orthographic transcription of the sample. The secondary purpose was to investigate the accuracy of the speech-language pathologists' estimates as compared to the standard measure for each of three groups: (a) the children with the most

intelligibility, (b) with average intelligibility, and (c) with the least intelligibility.

Data were collected from 47 children, aged 4:0 to 5:6, who comprised three groups with varying levels of intelligibility. Two groups of listeners who were unfamiliar with the speakers, but familiar with the topic, rated the children's percentage of intelligibility from continuous speech samples via orthographic transcription or trained ear estimation.

The two methods of measuring speech intelligibility investigated in this study were found to correlate highly ( $r = .96$ ). However, there was a significant difference between the percentages derived from orthographic transcription and those derived from trained ear estimation for some speakers. The  $t$ -test analyses revealed significant differences between the two measures for the two most intelligible groups, and no significant difference for the least intelligible group. It appears that the subjective method of estimating speech intelligibility with trained ears correlates with the objective method of orthographic transcription, but yields a different percentage score for some speakers.

## Implications

Clinical

The mean score for trained ear estimation was 79%, and the mean score for orthographic transcription was 76%. The t-test results indicate a significant difference between the two measures, whereas the Pearson r correlation (.96) indicated the two measures were highly correlated when rating speech intelligibility of young speakers. Furthermore, the t-test results regarding the most, average, and least intelligible groups revealed significant differences between estimation and transcription in the groups with the most and average intelligibility. However, the differences between the two measures were 8% or less for the adult-like group and thus seem not to be clinically significant. The differences between the measures for 2 of the 16 children in the middle group were 11% and thus may be important differences clinically for these two children, but probably not for the others in this group. .

On the other hand, there was no significant difference between the two methods when measuring the speech intelligibility of the least intelligible group. Notably, 3 of the 15 speakers for this group differed by 16% or more on the two measures, which is probably clinically significant. While the statistical results of this study support the method of trained ear estimation when measuring

speech intelligibility of young children who are 70% intelligible or less, the differences between estimated and actual percentages for some children appear to be clinically relevant, as it could affect a child's eligibility for services. The statistical results of this study also indicate that children whose intelligibility is above 70% should be evaluated by an objective method such as orthographic transcription to obtain accurate results; however, the differences don't appear to be significant enough to jeopardize a child's eligibility for services.

If time is a major consideration when evaluating a child's speech, using estimation by tracking to assess speech intelligibility is acceptable since the measures derived in this way do correlate with percentages derived from orthographic transcription. Clinicians need to be aware that differing methods for determining percentage of intelligibility yield differing results. For example, using a tracking method, rather than a "Gestalt" assignment of percentage, to record intelligible and unintelligible utterances may influence accurate estimation of speech intelligibility, but this is yet to be determined.

Monitoring speech competency involves periodic evaluation of the skill. This is usually accomplished by pre- and post-testing, and then comparing the results to

assess progress. Since level of intelligibility is often a determining factor regarding priority for services, occasionally evaluating it enables the clinician to monitor a client's speech competency, and determine at what priority level the child qualifies. Accurately assessing speech intelligibility, then, is important for determining which children should receive services.

### Research

SLPs need to be held accountable for the methods chosen to evaluate and treat clients. It is important that they choose appropriate, research-based instruments. This study focused on a method of measuring speech intelligibility that has been used by SLPs for a number of years, but has rarely been challenged as to whether or not it is an accurate, reliable, and valid way of evaluating speech. The results of this study give some validity to trained ear estimation. However, further research in this area is warranted. Should this study be duplicated in any way, it is suggested that either the examiner's voice be deleted from the audiotapes, or be kept away from the microphone. It is also suggested that an audio-video tape be used in order to provide more realistic information to the listeners, since speech samples are usually collected with the examiner/listener and child interacting together. Besides being able to hear the child's speech, the examiner

can usually see the child's mouth when speaking, and can note any accompanying gestures as well. Another area in need of further investigation includes comparing the estimations of less experienced SLPs with orthographic transcription to determine if experience has any influence on estimated percentage of intelligibility. It would also be interesting to investigate the correlation between orthographic transcription and assigning a percentage based on a "Gestalt" impression. This would not include tracking of any kind, but would require the listener to provide an estimated percentage of intelligibility derived purely from listening to the speech sample, an approach used by many practicing SLPs (Gordon-Brannan, 1993a; Kent et al., 1994).

Assuring that the tools used for assessment are the best, most appropriate ones available for each client is a large part of a practitioner's responsibility. The results of this study should help practicing SLPs continue to make wise, appropriate decisions when choosing evaluation tools. Since different methods for determining percentage of intelligibility yield different results, the method used to measure a client's intelligibility should be indicated within the diagnostic report. With additional research on procedures of estimating intelligibility, those who choose estimation as a diagnostic tool will be better informed about whether or not it is accurate and reliable.

## References

- Bernthal, J. E., & Bankson, N. W. (1988). Articulation and phonological disorders (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Beukelman, D. R., & Yorkston, K. M. (1979). The relationship between information transfer and speech intelligibility of dysarthric speakers. Journal of Communication Disorders, 12, 189-196.
- Billman, K. S. (1986). Phonological processes and intelligibility of spontaneous utterances in young children. Unpublished master's thesis, San Diego State University, San Diego, CA.
- Brookshire, R. H. (1992). An introduction to neurogenic communication disorders (4th ed.). St. Louis, MO: Mosby-Year Book, Inc.
- Carrow-Woolfolk, E. (1985). Test for auditory comprehension of language-revised. Allen, TX: DLM.
- Connolly, J. H. (1986). Intelligibility: A linguistic view. British Journal of Communication Disorders, 21, 371-376.
- Fudala, J. B., & Reynolds, W. M. (1986). Arizona articulation proficiency scale (2nd ed.). Los Angeles: Western Psychological Services.



- Gordon-Brannan, M. (1993a). Assessing intelligibility: Children's expressive phonologies. Topics in Language Disorders, 14, 17-25.
- Gordon-Brannan, M. (1993b). Speech intelligibility assessment of young children with varying levels of phonological proficiency/deficiency. Unpublished doctoral dissertation, The Wichita State University, Wichita, KS.
- Guilford, J. P. (1954). Psychometric methods. New York: McGraw-Hill.
- Hodson, B. W. (1992). 1-minute measure of homonymy and intelligibility. Unpublished manuscript.
- Hodson, B. W., & Paden, E. P. (1991). Targeting intelligible speech (2nd ed.). Austin, TX: ProEd.
- Hodson, B. W. (1986). The assessment of phonological processes-revised. Austin, TX: Pro-Ed.
- Hodson, B. W., & Paden, E. P. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. Journal of Speech and Hearing Disorders, 46, 369-373.
- Johnson, D. P. (1975). Communication characteristics of NTID students. Journal of the Academy of Rehabilitative Audiology, 8, 17-32.

Kent, R. D., Miolo, G., & Bloedel, S. (1994).

Intelligibility of children's speech: A review of evaluation procedures. American Journal of Speech-Language Pathology, 4, 81-95.

Metz, D. E., Samar, V. J., Schiavetti, N., Sitler, R. W., & Whitehead, R. L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. Journal of Speech and Hearing Research, 28, 345-355.

Metz, D. E., Schiavetti, N., & Sitler, R. W. (1980). Toward an objective description of the dependent and independent variables associated with intelligibility assessments of hearing-impaired adults. In J. D. Subtelny, Speech assessment and speech improvement for the hearing impaired (pp.72-81). Washington, D. C.: The Alexander Graham Bell Association for the Deaf, Inc.

Rylant, C., & Gammell, S. (1985). The relatives came. New York: Bradbury Press.

Samar, V. J., & Metz, D. E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. Journal of Speech and Hearing Research, 31, 307-316.

- Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent, Intelligibility in speech disorders: Theory, measurement and management (pp. 11-34). Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Shriberg, L. D. (1993). Four new speech and prosody measures for genetics research and other studies in developmental phonological disorders. Journal of Speech and Hearing Research, 36, 105-140.
- Shriberg, L., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. Journal of Speech and Hearing Disorders, 47, 256-270.
- Stevens, S. S. (1975). Psychophysics. New York: Wiley.
- Toner, M. A., & Emanuel, F. W. (1989). Direct magnitude estimation and equal appearing interval scaling of vowel roughness. Journal of Speech and Hearing Research, 32, 78-82.
- Van Riper, C., & Emerick, L. (1990). Speech correction: An introduction to speech pathology and audiology (8th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C. E. (1982). Weiss intelligibility test. Tigard, OR: CC Publications.

- Weiss, C. E., Gordon, M. E., & Lillywhite, H. S. (1987). Clinical management of articulatory and phonologic disorders. Baltimore: Williams & Wilkins.
- Weiss, C. E., & Lillywhite, H. S. (1981). Communicative disorders (2nd ed.). St. Louis: C. V. Mosby Company.
- Weston, A. D., & Shriberg, L. D. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. Journal of Speech and Hearing Research, 35, 1316-1332.
- Wilcox, K. S., Schooling, T. L., & Morris, S. R. (1991). The preschool intelligibility measure (PSIM). Paper presented at the annual meeting of the American Speech-Language-Hearing Association, Atlanta, GA.
- Yorkston, K. M., & Beukelman, D. R., (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. Journal of Communication Disorders, 11, 499-512.
- Young, M. A. (1969). Observer agreement: Cumulative effects of rating many samples. Journal of Speech and Hearing Research, 12, 135-143.

Appendix A  
Individual Subject Data

Group	Subject #	Age	Sex	TACL %ile	Intell.%-OT (mean)	Intell.%-SLP (mean)	Difference of SLP % from OT %
Most	1	4:11	F	88	99.50	99.25	-.25
Most	2	4:8	F	95	98.25	100.00	+1.75
Most	3	5:2	F	61	97.50	97.00	-.50
Most	4	4:5	F	89	96.25	97.50	+1.25
Most	5	4:3	M	67	95.50	100.00	+4.50
Most	6	4:11	M	91	95.25	97.75	+2.50
Most	7	4:10	F	98	93.50	100.00	+6.50
Most	8	5:5	M	59	93.25	99.00	+5.75
Most	9	5:5	M	76	93.25	96.50	+3.25
Most	10	4:11	F	66	92.00	92.50	+.50
Most	11	4:3	M	79	91.50	97.25	+5.75
Most	12	4:7	F	84	91.25	99.25	+8.00
Most	13	4:0	F	64	90.00	94.75	+4.75
Most	14	4:1	M	17	89.50	89.75	+.25
Most	15	4:5	M	94	88.50	85.25	-3.25
Most	16	4:6	F	39	87.75	93.75	+6.00
Avg.	17	4:0	F	17	86.25	92.50	+6.25
Avg.	18	5:6	F	73	86.00	94.50	+8.50
Avg.	19	5:0	M	46	86.00	93.00	+7.00
Avg.	20	4:1	M	50	86.00	86.25	+.25
Avg.	21	4:1	M	57	85.75	93.00	+7.25
Avg.	22	4:8	M	31	85.50	93.00	+7.50
Avg.	23	5:0	M	57	84.50	87.75	+3.25
Avg.	24	4:11	F	97	83.00	94.00	+11.00
Avg.	25	4:0	M	35	81.25	89.75	+8.50
Avg.	26	5:0	M	76	80.00	80.00	+/-0.00
Avg.	27	5:0	M	94	78.75	79.25	+.50
Avg.	28	4:10	F	97	77.50	82.75	+5.25
Avg.	29	4:7	F	85	75.50	86.75	+11.25
Avg.	30	4:8	M	11	73.00	76.25	+3.25
Avg.	31	4:2	M	14	71.25	61.00	-10.25
Avg.	32	4:9	M	57	70.75	77.00	+6.25
Least	33	4:3	M	47	70.50	86.50	+16.00
Least	34	5:1	F	35	69.50	75.00	+5.50
Least	35	4:2	F	12	67.00	73.25	+6.25
Least	36	4:9	M	29	63.50	63.25	-.25
Least	37	4:4	M	43	61.00	71.00	+10.00
Least	38	4:2	F	38	59.25	81.25	+22.00
Least	39	4:1	F	35	55.75	49.50	-6.25
Least	40	4:8	M	31	50.25	47.75	-2.50
Least	41	4:10	F	27	50.00	40.00	-10.00
Least	42	4:8	F	16	49.50	41.50	-8.00
Least	43	4:9	M	27	47.00	43.00	-4.00
Least	44	4:0	M	41	46.00	27.00	-19.00
Least	45	4:0	M	37	41.75	39.75	-2.00
Least	46	4:2	M	44	35.75	45.00	+9.25
Least	47	4:5	M	02	18.75	08.50	-10.25

Note: Intell.%-OT (mean) = Average percentage of words understood in continuous speech sample by graduate students' orthographic transcription; Intell.%-SLP (mean) = Average percentage of words understood in continuous speech sample by speech-language pathologists' trained ear estimation; Most = most intelligible group; Avg. = group with average intelligibility; Least = least intelligible group.

## Appendix B

Characteristics of the Three Speaker Groups

Group	Mean Age (Age Range)	# of Females	# of Males	TACL-R %ile (Range)
Adult- like	4:8 (4:0-5:5)	9	7	72 (17-98)
Mild/ Moderately Deficient	4:9 (4:0-5:6)	5	11	56 (11-97)
Moderately/ Severly Deficient	4:5 (4:0-5:1)	6	9	31 (2-47)

## Appendix C

Informed Consent-SLPs

I, \_\_\_\_\_, consent to serve as a listener in this research project on measuring speech intelligibility of children.

I understand that the study involves giving an estimate of speech intelligibility in the form of a percentage to 48 speech samples. It will take a total of approximately 4 hours to listen to all 48 tapes.

I understand that participation in this study will present no physical, social, economic, or other risks except for the possible inconvenience of coming to the PSU campus to participate in the study. All data obtained during the course of the study will remain confidential. Published data and public records will not reveal my name.

It has been explained to me that the purpose of the study is to learn if the method of trained ear estimation is an accurate and reliable measurement of speech intelligibility. I may not receive any direct benefit from taking part in this study, but my participation may help to increase knowledge which may benefit others in the future.

Nancy Sugarman has offered to answer any questions I may have about the study and what is expected of me in the

study. I understand that I am free to withdraw from participation in this study at any time without jeopardizing my relationship with Portland State University.

I have read and understand the foregoing information and agree to participate in this study.

Date: \_\_\_\_\_ Signature: \_\_\_\_\_

If you experience problems that are the result of your participation in this study, please contact the Chair of the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 105 Neuberger Hall, Portland State University, (503) 725-3417.



## Appendix D

### Informed Consent

I, \_\_\_\_\_, give my consent for my child \_\_\_\_\_ to serve as a subject in the research project on measuring understandability of speech in children conducted by Mary Gordon-Brannan, Program Director, Speech & Hearing Sciences Program at Portland State University. I understand that the study involves the administration of hearing, speech, and receptive language tests to children.

My child would respond to hearing tests, point to pictures, repeat words and sentences, and talk with the examiner. I would fill out a questionnaire form regarding speech, language, and hearing case history, as well as parental occupation information. Also I would help interpret the speech my child uses during the testing. I understand that participation in this study will present no physical or psychological risks. All data obtained during the course of the study will remain confidential. Published data will not reveal the name of my child.

It has been explained to me that the purpose of the study is to learn how to measure how much of children's speech can be understood. My child may not receive any direct benefit from participation in this study, but my participation may help to increase knowledge which may benefit others in the future.

Mary Gordon-Brannan has offered to answer any questions I may have about the study and what is expected of me in the study. I understand that I am free to withdraw from participation in this study at any time without jeopardizing my relationship with Portland State University.

I have read and understand the foregoing information and agree to participate in this study.

Date: \_\_\_\_\_ Signature: \_\_\_\_\_

*If you experience problems that are the result of your participation in this study, please contact the Chair of the Human Subjects Research Review Committee, Office of Graduate Studies, 303 Cramer Hall, Portland State University, 503/725-5262.*

## Appendix E

Questionnaire

Child's Name: \_\_\_\_\_ Birthdate: \_\_\_\_\_

Parent(s): \_\_\_\_\_

Address: \_\_\_\_\_ Phone #: \_\_\_\_\_

Father's Occupation: \_\_\_\_\_

Mother's Occupation: \_\_\_\_\_

Relationship of person completing the questionnaire \_\_\_\_\_

1. Has your child ever been diagnosed as demonstrating any of the following:

neurological impairment    yes \_\_\_\_\_    no \_\_\_\_\_

orthopedic or physical handicap    yes \_\_\_\_\_    no \_\_\_\_\_

motor or movement impairment    yes \_\_\_\_\_    no \_\_\_\_\_

2. Has your child had a history of ear infections as indicated by the following:

complained of ear aches    yes \_\_\_\_\_    no \_\_\_\_\_

had ear aches or infections    yes \_\_\_\_\_    no \_\_\_\_\_

Is so, how many times? \_\_\_\_\_

When was the last time? \_\_\_\_\_

had medical treatment for ear infections    yes \_\_\_\_\_    no \_\_\_\_\_

Is so, how many times? \_\_\_\_\_

When? \_\_\_\_\_

had ventilation tubes inserted    yes \_\_\_\_\_    no \_\_\_\_\_

If so, when? \_\_\_\_\_

Are tubes currently one or both ears? \_\_\_\_\_

3. Provide information about speech development:

When did your child say his/her first word? \_\_\_\_\_

What was the first word? \_\_\_\_\_

When did your child begin to put 2 words together? \_\_\_\_\_

Do family members have difficulty understanding your child's speech?    yes \_\_\_\_\_    no \_\_\_\_\_

Do persons outside the family have difficulty understanding your child's speech?    yes \_\_\_\_\_    no \_\_\_\_\_

## Appendix F

Instructions to Graduate Student Listeners

You will hear 48 children, 4-6 years of age, in conversational speech samples. You are not to transcribe the last 5 samples, samples 49-53. Write down the sample # and subject # in the blanks at the top of each sample form. Then write down or type the words you think the child said on the response forms numbered from 1 to 50 or use a computer for your orthographic transcriptions. Write down one utterance per numbered line. You will not use all the lines for each child. Do not be concerned about how you divide the utterances. Use an X to indicate each syllable that you do not understand. If you understand part of a word, write down the part of the word you understand along with an X, e.g., Xing. You do not need to write down fillers such as um, mm-mm, uh huh, etc. You are encouraged to guess the words said. While transcribing the sample, **you may listen to each utterance a maximum of three times**. When finished with the transcription, you may listen to the whole sample once to check your transcription. Turn in individual transcriptions to M. Gordon-Brannan as you complete them. Do you have any questions about what you are to do? If so, ask me now or call me at 725-3143 (W) or 227-3356(H).

## Appendix G

Listening Instructions to SLPs

Listener name: \_\_\_\_\_ Listener #: \_\_\_\_\_  
 Number of years experience as SLP in public schools \_\_\_\_\_

You will be listening to 27 connected speech samples tonight, and 26 speech samples at a second meeting still to be arranged, for a total of 53 speech samples. They will be presented in random order according to severity of intelligibility and age.

Please use as much objectivity as possible while listening. You may take notes on the provided sheet of paper if you wish. When each 3 minute sample has been completed, you will have about 1 minute to evaluate what you've heard, and select an arbitrary percentage of intelligibility between 0-100% for that speaker. Put your estimated percentage of intelligibility in the space provided which correlates with the speech sample presented.

Your name will not be used in any publications or public files related to this study. You will be referred to by listener number only. Do you have any questions?

Sample 1 _____%	Sample 12 _____%
Sample 2 _____%	Sample 13 _____%
Sample 3 _____%	Sample 14 _____%
Sample 4 _____%	Sample 15 _____%
Sample 5 _____%	Sample 16 _____%
Sample 6 _____%	Sample 17 _____%
Sample 7 _____%	Sample 18 _____%
Sample 8 _____%	Sample 19 _____%
Sample 9 _____%	Sample 20 _____%
Sample 10 _____%	Sample 21 _____%
Sample 11 _____%	Sample 22 _____%

Sample 23 \_\_\_\_\_%  
Sample 24 \_\_\_\_\_%  
Sample 25 \_\_\_\_\_%  
Sample 26 \_\_\_\_\_%  
Sample 27 \_\_\_\_\_%  
Sample 28 \_\_\_\_\_%  
Sample 29 \_\_\_\_\_%  
Sample 30 \_\_\_\_\_%  
Sample 31 \_\_\_\_\_%  
Sample 32 \_\_\_\_\_%  
Sample 33 \_\_\_\_\_%  
Sample 34 \_\_\_\_\_%  
Sample 35 \_\_\_\_\_%  
Sample 36 \_\_\_\_\_%  
Sample 37 \_\_\_\_\_%  
Sample 38 \_\_\_\_\_%

Sample 39 \_\_\_\_\_%  
Sample 40 \_\_\_\_\_%  
Sample 41 \_\_\_\_\_%  
Sample 42 \_\_\_\_\_%  
Sample 43 \_\_\_\_\_%  
Sample 44 \_\_\_\_\_%  
Sample 45 \_\_\_\_\_%  
Sample 46 \_\_\_\_\_%  
Sample 47 \_\_\_\_\_%  
Sample 48 \_\_\_\_\_%  
Sample 49 \_\_\_\_\_%  
Sample 50 \_\_\_\_\_%  
Sample 51 \_\_\_\_\_%  
Sample 52 \_\_\_\_\_%  
Sample 53 \_\_\_\_\_%

Listener #: \_\_\_\_\_

## Appendix H

Raw Data by Listener

Percentage-of-Words Understood in Continuous Speech  
Orthographic Transcription - Graduate Students

Sub. #	Listener A	Listener B	Listener C	Listener D	Mean Score	Range
<b>Group:</b>						
<b>MOST</b>						
1	99	99	100	100	99.50	1
2	98	99	99	97	98.25	2
3	98	98	96	98	97.50	2
4	95	97	96	97	96.25	2
5	92	98	96	96	95.50	6
6	98	97	92	94	95.25	6
7	92	86	98	98	93.50	12
8	91	92	97	93	93.25	6
9	92	95	97	89	93.25	8
10	96	91	95	86	92.00	10
11	96	89	91	90	91.50	7
12	90	92	92	91	91.25	2
13	91	89	92	88	90.00	4
14	92	91	89	86	89.50	6
15	92	95	82	85	88.50	13
16	96	82	88	85	87.75	<u>14</u>
					<u>M=</u>	6
<b>Group:</b>						
<b>AVG.</b>						
17	75	91	92	87	86.25	17
18	92	79	95	78	86.00	17
19	92	78	84	90	86.00	14

Sub. #	Listener A	Listener B	Listener C	Listener D	Mean Score	Range
20	87	80	87	90	86.00	10
21	85	88	83	87	85.75	5
22	89	85	81	87	85.50	8
23	88	77	84	89	84.50	12
24	83	83	83	83	83.00	0
25	86	79	76	84	81.25	10
26	74	84	84	78	80.00	10
27	70	82	83	80	78.75	13
28	85	71	82	72	77.50	14
29	82	74	76	70	75.50	12
30	65	72	78	77	73.00	13
31	70	71	68	76	71.25	8
32	68	77	71	67	70.75	10
						<u>M=</u> 10

Group:

**LEAST**

33	61	72	74	75	70.50	14
34	59	73	64	82	69.50	23
35	70	69	59	70	67.00	11
36	72	59	58	65	63.50	14
37	49	69	62	64	61.00	20
38	68	67	46	56	59.25	22
39	58	55	58	52	55.75	6
40	53	51	53	44	50.25	9
41	28	42	70	60	50.00	42
42	56	53	36	53	49.50	20

---

Sub. #	Listener A	Listener B	Listener C	Listener D	Mean Score	Range
43	51	45	40	52	47.00	12
44	43	48	40	32	46.00	16
45	37	46	35	49	41.75	14
46	42	34	28	39	35.75	14
47	35	10	8	22	18.75	<u>27</u>

---

M = 18

---

---



## Appendix I

Raw Data by Listener

Percentage-of-Words Understood in Continuous Speech  
Trained Ear Estimation - SLPs

Sub. #	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
Group:						
<b>MOST</b>						
1	100	98	100	99	99.25	2
2	100	100	100	100	100.00	0
3	98	91	100	99	97.00	9
4	98	92	100	100	97.50	8
5	100	100	100	100	100.00	0
6	97	96	100	98	97.75	4
7	100	100	100	100	100.00	0
8	99	97	100	100	99.00	3
9	98	92	98	98	96.50	6
10	93	87	98	92	92.50	11
11	98	93	100	98	97.25	7
12	100	99	99	99	99.25	1
13	96	94	94	95	94.75	2
14	82	85	98	94	89.75	16
15	90	82	84	85	85.25	8
16	92	87	98	98	93.75	11
					<u>M= 5.5</u>	
Group:						
<b>AVG.</b>						
17	92	95	93	90	92.50	5
18	93	89	98	98	94.50	9
19	91	94	95	92	93.00	4

Sub. #	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
20	91	91	81	82	86.25	10
21	93	90	95	94	93.00	5
22	93	90	96	93	93.00	6
23	88	81	91	91	87.75	10
24	95	90	96	95	94.00	6
25	93	82	92	92	89.75	11
26	88	82	72	78	80.00	16
27	82	76	80	79	79.25	6
28	86	68	90	87	82.75	22
29	90	79	91	87	86.75	12
30	50	75	90	90	76.25	40
31	62	69	58	55	61.00	14
32	74	78	72	84	77.00	12
						<u>M= 12</u>
-----						
Group:						
<b>LEAST</b>						
33	84	84	89	89	86.50	5
34	77	75	74	74	75.00	3
35	65	71	82	75	73.25	17
36	67	56	61	69	63.25	13
37	80	60	71	73	71.00	20
38	74	77	87	87	81.25	13
39	31	43	64	60	49.50	29
40	58	42	49	42	47.75	16
41	20	30	65	45	40.00	45
42	34	31	54	47	41.50	23

Sub. #	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
43	48	25	55	44	43.00	30
44	25	17	42	24	27.00	25
45	31	28	52	48	39.75	24
46	47	25	57	51	45.00	32
47	8	8	10	8	8.50	2
					<u>M=</u>	20

Note: The groups are divided into most, average, and least intelligible based on the percentage-of-words understood via orthographic transcription, not trained ear estimation. Thus, the percentages listed here are not in numerical order within the categories of most, average, and least intelligible.

## Appendix J

Listener Estimations for Speakers Rated Twice

Group	Sub.#	Listener 1		Listener 2		Listener 3		Listener 4	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd
		Difference		Difference		Difference		Difference	
Most	4	98	99	92	97	100	100	100	99
		1		5		0		1	
Most	8	99	98	97	97	100	100	100	100
		1		0		0		0	
Avg.	33	84	91	84	82	89	94	89	94
		7		2		5		5	
Least	40	58	68	42	49	49	73	42	70
		10		7		24		28	
Least	47	8	9	8	20	10	12	8	10
		1		12		2		2	

Note: 1st = the first presentation of the continuous speech samples for estimating speech intelligibility; 2nd = the second presentation of the continuous speech samples for estimating speech intelligibility; Difference = difference between the 1st estimation and the 2nd estimation; Most = most intelligible group; Avg. = group with average intelligibility; Least = least intelligible group.