

11-6-1996

A Comparison of Speech Intelligibility Measures Between Unsophisticated Listener Judgements and Orthographic Transcription

Carla J. Dukart
Portland State University

Follow this and additional works at: 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

Dukart, Carla J., "A Comparison of Speech Intelligibility Measures Between Unsophisticated Listener Judgements and Orthographic Transcription" (1996). *Dissertations and Theses*. Paper 5257.
<https://doi.org/10.15760/etd.7130>

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.

THESIS APPROVAL

The abstract and thesis of Carla J. Dukart for the Master of Science in Speech Communication: Speech and Hearing Sciences were presented November 6, 1996, and accepted by the thesis committee and the department.

COMMITTEE APPROVALS:

[Redacted]

Mary Gordon-Brannan, Chair

[Redacted]

John Tetnowski

[Redacted]

Cathleen Smith
Representative of the Office
of Graduate Studies

DEPARTMENT APPROVAL:

[Redacted]

Stephen A. Kosokoff, Chair
Department of Speech Communication

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by

[Redacted]

on 20 Nov 1996

ABSTRACT

An abstract of the thesis of Carla J. Dukart for the Master of Science in Speech Communication: Speech and Hearing Sciences presented November 6, 1996.

Title: A Comparison of Speech Intelligibility Measures Between Unsophisticated Listener Judgements and Orthographic Transcription.

Intelligible speech is a primary component for successful communication. However, the speech of children with disordered phonologies is often unintelligible. Therefore, when assessing the speech intelligibility of children in order to determine whether they qualify for intervention services, speech-language pathologists need reliable evaluation tools.

The focus of this investigation was the measurement of speech intelligibility. The purpose of this study was to examine the relationship between two methods for measuring speech intelligibility. The first, identification method, involves the listener transcribing a speech sample from which the percentage of words understood is calculated. The second, scaling procedure, involves the listener estimating the percentage of words understood from a continuous speech sample. The secondary purpose of this study was to examine the accuracy of the scaling method as measured by ear estimation compared to the identification method as measured through orthographic transcription for each of three groups of children with: (a) the most intelligibility, (b) average intelligibility, and (c) least intelligibility.

Four unsophisticated listeners rated the speech intelligibility of 48 speakers aged 4:0 to 5:6 who comprised three groups with varying levels of phonological proficiency. The listeners who were unfamiliar with the speakers, but familiar with the topic, rated the children's continuous speech samples using ear estimation. The data collected were then compared with intelligibility ratings as measured in a previous study (Gordon-Brannan, 1994) via orthographic transcription.

The two methods of measuring speech intelligibility examined in this study were found to be positively correlated ($r = .86$). However, the t -test analysis revealed significant differences between the two measures for the most and least intelligible groups, indicating discrepancies between the two methods when measuring the speech intelligibility of some children. Additional statistical analysis revealed poor intrajudge reliability which should be considered when interpreting the results presented. It does appear, however, that when measuring speech intelligibility, using ear estimation, is reflective of the orthographic transcription measure, although the actual estimated percentages of intelligibility appear to differ from the percentages derived from orthographic transcription.

A COMPARISON OF SPEECH INTELLIGIBILITY MEASURES BETWEEN
UNSOPHISTICATED LISTENER JUDGEMENTS AND ORTHOGRAPHIC
TRANSCRIPTION

by

Carla J. Dukart

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
1996

ACKNOWLEDGMENTS

This, my greatest accomplishment, to date, would not be complete without acknowledging the people who helped make it happen. To begin, I would like to thank my niece **Brittany Sherwood**, for inspiring me to become a Speech-Language Pathologist. Because of you I found my direction in life.

Thank you **Mary Gordon-Brannan** for letting me participate in this research project and for giving me the guidance to see the project through. I will always be amazed and grateful not only for your editing ability, but also for your personal dedication and flexibility in helping me complete this goal. Thank you also to **Dr. John Tetnowski** for advising me academically and providing much appreciated insight on this project and **Dr. Cathleen Smith** for not only serving as a committee member for this project, but also for your wonderful letter of recommendation that helped me get accepted to graduate school in the first place!

In addition, this project would not have been possible without the love and support of my family, especially my parents, **Jerome** and **Lorraine**. From you I learned the value of hard work and the rewards of obtaining a goal. You never once doubted that I could accomplish anything I set out to achieve. Thank you, mom and dad for teaching me to believe in myself!

I would also like to express a special thanks to my extended family and friends.

To **Ed** and **Edith Lippert** for your continuous words of encouragement and well timed humor, which helped me tick off the years. To **Beth** and **Tracey** for your unconditional friendship, which I cherish. To **Robin, John, Jeff,** and **Jenny** for volunteering to participate in a project out of friendship. I am truly grateful for your help and am amazed by your dedication; right up through the last tape.

Finally, my deepest thanks to my husband **Tim Lippert**. As my best friend, you stuck by me through the tough times and helped me celebrate the successes. Words cannot express how thankful I am for the amount of love and support you have shown me during these past years. As I close this chapter of my life, I will always remember the long conversations we had about how life beyond graduate school will be so wonderful. Well, it's here and it **IS** wonderful, thanks to you! ♡

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	vii
CHAPTER	
I INTRODUCTION.....	1
Statement of Purpose.....	2
Definition of Terms.....	3
II REVIEW OF THE LITERATURE.....	6
Factors Influencing Intelligibility.....	6
Speech Production Factors	
Speech Intelligibility Measures.....	8
Identification	
Closed-set Identification	
Open-set Identification	
Scaling	
Assessment Considerations.....	14
Sampling Modes	
Time Factors	
Levels of Listener Sophistication	
Summary.....	19
III METHOD.....	20

	Subjects.....	21
	Unsophisticated Listeners	
	Procedures.....	21
	Speaker	
	Speech Samples	
	Judgements	
	Scoring	
	Reliability	
	Data Analysis.....	25
IV	RESULTS AND DISCUSSION.....	26
	Results.....	26
	Interjudge and Intrajudge Reliability	
	Research Question I	
	Research Question II	
	Discussion.....	31
	Listener Reliability	
	Degree of Intelligibility	
	Listener Estimations	
V	SUMMARY AND IMPLICATIONS.....	41
	Summary.....	41
	Implications.....	43
	Clinical	
	Research	
	REFERENCES.....	47
	APPENDIXES	
	A INFORMED CONSENT-UNSOPHISTICATED LISTENERS.....	54

B	QUESTIONNAIRE.....	56
C	INDIVIDUAL SUBJECT DATA.....	57
D	CHARACTERISTICS OF THE THREE SPEAKER GROUPS.....	59
E	LISTENER INSTRUCTIONS-GRADUATE STUDENTS.....	60
F	LISTENER INSTRUCTIONS-UNSOPHISTICATED LISTENERS....	61
G	LISTENER RATINGS FOR SPEAKERS RATED TWICE- UNSOPHISTICATED LISTENERS.....	63
H	RAW DATA BY LISTENER-PERCENTAGE -OF-WORDS UNDERSTOOD IN CONTINUOUS SPEECH/ EAR ESTIMATION - UNSOPHISTICATED LISTENERS.....	64
I	RAW DATA BY LISTENER-PERCENTAGE-OF-WORDS UNDERSTOOD IN CONTINUOUS SPEECH/ ORTHOGRAPHIC TRANSCRIPTION - GRADUATE STUDENTS.....	66

LIST OF TABLES

TABLE		PAGE
1	Orthographic Transcription Correlation Matrix for Intelligibility Measures.....	27
2	Non-sophisticated Listener Ear Estimation Correlation Matrix for Intelligibility Measures.....	27
3	Descriptive Statistics for Two Intelligibility Measures for Three Speaker Groups.....	30

CHAPTER I

INTRODUCTION AND STATEMENT OF PURPOSE

A fundamental component for successful communication is intelligible speech. However, disordered phonologies often adversely affect the intelligibility of children's speech. Clinically, this reduced intelligibility is assessed through both transcription and scaling measures. While some researchers contend that transcription procedures are the only valid means of assessing speech intelligibility (Yorkston & Beukelman, 1981), others argue that scaling procedures are more efficient (Weiss & Lillywhite, 1981). According to Bernthal and Bankson (1993) and Schmidt (1984), the measurement of intelligibility provides essential information about intervention strategies for children with disordered phonologies. In today's world when speech-language pathologists need to document the efficacy of intervention, it is imperative that measurement of speech intelligibility be both accurate and reliable. Although one objective method that involves gathering a spontaneous speech sample, then orthographically transcribing each word, and calculating a percentage of words understood out of words spoken can provide a reliable assessment of a child's intelligibility (Gordon-Brannan, 1993, 1994; Kwiatkowski & Shriberg, 1992), the method is time consuming (Bacon, 1995). Perhaps a more efficient method, and one commonly in use by practicing speech-language pathologists to date, is making an estimate of the percentage of words understood, and then noting with a percentage how intelligible a child is (Gordon-Brannan, 1993, 1994; Kent, Miolo, & Blodel, 1994; Samar & Metz, 1988; Shriberg & Kwiatkowski, 1982;

Weiss, 1982; Yorkston & Beukelman, 1978).

Statement of Purpose

The focus of this study was to investigate measurement of speech intelligibility in young children. The purpose of this study was to analyze the relationship between two types of intelligibility measures for preschool children with varying degrees of phonological proficiency: the orthographic transcription method, and listener estimation. Orthographic transcription is defined as the percentage of words understood by the listener from a written, word-for-word transcript. Listener estimation is defined as the judgement of percentage of words understood from a continuous speech sample.

The hypotheses tested include:

1. There is a significant correlation between the orthographic transcription measure of speech intelligibility and the ear estimation measure of speech intelligibility.
2. There is no significant difference in intelligibility percentages between ear estimation and orthographic transcription for each of three groups of preschool children with: (a) most intelligibility, (b) average intelligibility, and (c) least intelligibility.

Definition of Terms

The following terms were defined as follows for this study:

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, e.g., /dɒdi/ for doggie, and /tæt/ for cat (Hodson & Paden, 1991).

Backing - Replacing an anterior consonant with a posterior one, e.g. /go/ for toe and /kʌb/ for tub (Hodson & Paden, 1991).

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

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).

Gloss - The listener's perceived interpretation of unintelligible word(s).

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 & Padden, 1986).

Morphology - The formation of words such as plurals, past tense, or possessives (Weiss et al., 1987).

Orthographic - Refers to the accurate or accepted spelling of words using the symbols of an alphabet. Most orthographic systems relate written language to spoken language, using alphabet letters as phonograms (Calvert, 1980).

Phonology - The study of the sounds that comprise language and the rules for using them (Weiss et al., 1987).

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

Pitch - Quality of sound caused by its frequency; proceeding on a scale from low to high (Shames & Wigg, 1990).

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

Prosody - Aspects of language that convey meaning and mood, and give melody to the speech act by changes in rate, rhythm, or stress (Shames & Wiig, 1990).

Semantics - Aspects of language concerned with rules governing the meaning or content of words, and grammatical rules (Shames & Wiig, 1990).

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

Stopping - Substitution of stops for other consonants, e.g., /kIt/ for kiss, /dUt/ for juice,

and /tʌni/ for funny (Hodson & Paden, 1991).

Suprasegmental - Characteristics greater than the linguistic segments of an utterance, relating to junctural or prosodic features (Shames & Wiig, 1990).

Syntax - Organizational rules for ordering words in a sentence, specifying word order, sentence organization, and word relationships.

CHAPTER II

REVIEW OF THE LITERATURE

Because the fundamental purpose of communication is to be understood, assessing intelligibility is of utmost concern when evaluating communication disorders in children. According to Gordon-Brannan (1993), a critical component for the assessment of children with disordered phonologies is accurate measurement of speech intelligibility.

The focus of this literature review is to explore factors that influence intelligibility in speech and methods routinely used to assess intelligibility of speakers. Within the scope of this review, consideration is given to speech samples used, time considerations, and levels of listener sophistication.

Factors Influencing Intelligibility

Research indicates that many factors can influence the intelligibility of speech (Bernthal & Bankson, 1993; Kent, 1992; Kent et al., 1994; Shriberg & Kwiatkowski, 1982; Weiss, 1980; Weston & Shriberg, 1992). Among these are suprasegmental , contextual, linguistic, and speech sound production factors. Suprasegmental factors that influence the intelligibility of a speech sample include prosodic features such as pitch, rate, stress, and phrasing. In addition, the quality, loudness, and resonance of a speakers' voice or the rhythm of the speech sample may negatively affect intelligibility. Contextual factors such as listener familiarity with the speaker, context of the message

being sent, and clarity of visual and acoustic signals of speech can alter the speaker's intended communication. However, since communication relies on both a speaker and a listener, Kent (1992) stated that lighting and distance from speaker to listener can also influence the intelligibility of the speech signal. According to Gordon-Brannan (1994), syntax, mean length of utterance, morphology, morphophonemics, and semantic encompass the linguistic features of speech that must be taken into consideration when determining speech intelligibility.

In 1992, Weston and Shriberg conducted two studies to determine the influence of contextual and linguistic variables on the intelligibility of speech. Results of the first study indicated that the position of the word within an utterance and its relation to other words has a significant effect on how well the intended message is understood. In addition to position of the word, word length and fluency of speech were also considered factors relating to intelligibility. The second study suggested that the degree to which an utterance is intelligible is dependent on the syllable structure of the word, its grammatical form, and phonological complexity. Weiss (1982) also identified 22 elements that can be classified as linguistic, suprasegmental, and articulatory factors that influence the individual's capability to be understood. However, this list did not include speech sound production errors used by children which also leads to unintelligibility.

Speech Sound Production Factors

According to a study conducted by Hodson and Paden (1981), 60 three-to-eight year old children who were unintelligible demonstrated frequent use of the following

phonological deviations: cluster reduction, stridency deletion, stopping, liquid deviation, and assimilation. A subsequent study conducted by Billman (1986) revealed that the two deviations that had the most adverse affect on a child's speech intelligibility were backing and prevocalic singleton omission. Billman also suggested that these two deviations should be given priority when selecting remediation targets.

Therefore, given the wide variety of divergent influences that can alter a speaker's intelligibility, quantifying and analyzing intelligibility can prove to be a formidable task. The goal of intelligibility measurements is to provide objective data regarding speech production deficits, quantify change in speech production skills, and present an overall picture of functional communication (Ansel & Kent, 1992). The most commonly used techniques to assess speech intelligibility can be separated into two categories, that is, transcription procedures and rating scales.

Speech Intelligibility Measures

Speech intelligibility is generally measured in two ways. One method is identification that involves the listener specifying the words a speaker says (Schiavetti, 1992). Another method for measuring intelligibility is the scaling procedure. This involves a listener listening to a speech sample and assigning an overall rating of speech intelligibility based on a scale (Ansel & Kent, 1992).

Identification Method

One of the most common identification procedures is orthographic transcription of a speech sample. Typically speech intelligibility is assessed by eliciting a speech sample through picture stimulus or repetition tasks. Measuring speech intelligibility

using the identification procedure is conducted either by using a closed-set word format or an open-set word format. Closed-set word identification involves identifying words spoken from a word list, while open-set word identification involves specifying the words understood from a continuous speech sample (Gordon-Brannan, 1994). In a review of speech intelligibility measures, Kent et al. (1994) discussed identification methods as quantitative in nature because speech intelligibility is determined by computing a percentage of words understood in a speech sample. For this reason, identification assessments have more face validity than estimation measures because the listener must objectively understand the stimulus (Ansel & Kent, 1992; Metz, Schiavetti, & Sitler, 1980; Yorkston & Beukelman, 1980). The following is a review of identification measures for assessing speech intelligibility in children with disordered phonologies.

Closed-set Word Identification

The SPEech INtelligibility Evaluation (SPINE), developed by Monsen (1981), is used to measure phonetic contrast and the percentage of words correctly identified from a total set. The speaker is instructed to say consonant-vowel-consonant (CVC) words that are contrastive in a phonetic feature; the examiner then selects the word believed to be spoken from possible choices on an answer sheet. Results are expressed as the percentage of words correctly identified by the examiner. The percentage is then used as the metric of intelligibility.

Children's Speech Intelligibility Test (CSIT), developed by Kent et al. (1994), is a word-identification test in which the stimuli are grouped into phonological subsets

associated with developmental levels of vowel and consonant mastery. The developmentally based construction also gives the examiner immediate opportunity to judge the child's performance relative to age-typical patterns. The stimuli are judged by listeners using a closed-set response where the listeners select the item heard from a multiple-choice format. The listeners responses are analyzed to yield an overall score (percentage of words correctly transmitted), a number of phonological contrast scores (based on error rates for each subset indicated above), and a composite score (the means of the error rates for a group of subset scores).

The Preschool Speech Intelligibility Measure (P-SIM) was designed by Morris, Wilcox, & Schooling (1995) as an objective, clinical measure of speech intelligibility in preschool children. For this assessment, the child repeats 50 target words modeled by the examiner. These words are randomly selected from 50 sets of 12 phonetically similar forms. Responses are audiotaped and the listener identifies, from taped responses, each of the 50 words from a forced choice listening paradigm. The percentage of correctly identified words serves as a metric of speech intelligibility.

One of the drawbacks of closed-set testing for young children is word familiarity. In order to compose an articulation or intelligibility test, it is difficult to ensure that children are familiar with the words that are selected and used to elicit speech. However, the risk of unfamiliarity with target words may be reduced when repetition tasks are used rather than identification tasks (Kent et al., 1994).

Open-set Word Identification

Identification assessments that rely on the listener writing down exactly what the

child says and then analyzing the written transcriptions by calculating a percentage of words understood out of words spoken are open-set word identification (Gordon-Brannan, 1994). One example of a published open-set word identification assessment is the Weiss Intelligibility Test (WIT; Weiss, 1982).

The WIT, which emphasizes word identification without phonetic or phonological deviation analysis, provides measures of intelligibility in single words, contextual speech, and overall intelligibility (Weiss, 1982). In the single word assessment, the child is asked to name pictures as they are presented. This is audiotaped to be scored later. Only words that are completely understood are written down. The examiner calculates the number of correctly identified words and multiplies by four. From this score, the Intelligibility of Isolated Words is determined (Weiss, 1982). In order to determine intelligibility of contextual speech, the child is instructed to describe a set of pictures while being audiotaped. The listener transcribes 200 words from the sample. The examiner counts the number of understood words and divides the number by two to ascertain the Intelligibility of Contextual Speech score (Weiss, 1982). For both areas, the examiner calculates the number of correctly identified words from an audio-taped speech sample. An overall intelligibility score is obtained by calculating the mean of the two previous scores.

In 1982, Shriberg and Kwiatkowski developed a severity measure called the Percentage of Consonants Correct (PCC). In this assessment, the listener audiotapes a continuous speech sample and determines the percentage of consonants correct as the total number of consonants judged correct by the listener divided by the total number of

intended consonants expressed. Although the PCC was not designed for measuring intelligibility, it has been shown to correlate highly with intelligibility (Osberger, 1992). For this reason, researchers often cite it as a metric for intelligibility. Rather than an intelligibility measure, Shriberg and Kiawtkowski (1982) consider the PCC to be a severity measure that encompasses intelligibility, disability, and handicap. The values gleaned from the PCC are meant to reflect 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%.

To summarize, when measuring the degree of speech intelligibility of young children, one common procedure is identification, or calculating an actual percentage of words understood from a speech sample. This can be done using closed-set word format or open-set format. Closed-set word identification methods allow the examiner a closer look at specific sound production or phonemic errors, but are limited in that word or sentence repetition tasks are rarely present in everyday speaking situations. On the other hand, open-set word identification methods give the examiner a better overall picture of the speaker's daily conversational speaking abilities. Yet, this method requires more time for both administration and scoring.

Scaling Procedures

A second method for measuring speech intelligibility involves having the listener assign a numerical rating of overall intelligibility on either an equal-appearing interval scale or a direct magnitude scale. In equal-appearing interval scaling procedures, the speaker's intelligibility is assigned a numerical value along a continuum.

The continuum is commonly a scale numbered from 1, to 5, 7, or 9. These numbers represent the degree of intelligibility ranging from unintelligible to intelligible.

According to Schiavetti (1992), an odd numbered scale is used to provide a middle value and two end points.

Estimating the percentage of speech intelligibility is an equal-appearing scaling procedure (Kent, 1992). With this method, the listener assigns a numerical value based on a continuum ranging from 0% (unintelligible) to 100% (fully intelligible) rather than a scale of 1 to 5, 7, or 9. For example, if the listener understands half of the speakers' utterances, the speech sample is considered 50% intelligible.

In 1992, Osberger developed the Meaningful Use of Speech Scale (MUSS), which is a scaling procedure that involves surveying parents and teachers who rate a children's ability to communicate effectively. The parents or teachers assign a numerical value as to whether the children adapt their speech to the listeners' familiarity, or provide clarification and repair techniques when speech is not understood. The answers to the probes are then scored according to the frequency of occurrence (e.g., 0 = never occurs; 4 = always occurs). The goal of this assessment is to provide global information of a child's spoken language. The MUSS is limited, however, in that it does not offer detailed information of error patterns present in speech that contributes to reduced intelligibility. Another scaling procedure discussed by Schiavetti in 1992 is the National Technical Institute for the Deaf (NTID) rating scale, in which a value on a 5-point equal interval scale is assigned based on the listener's judgement of intelligibility.

A second type of scaling procedure is direct magnitude estimation. In this method, the listener judges the intelligibility of a speech sample and assigns a value relative to either a standard stimulus (Schiavetti, 1984) or to the first sample heard. This value is the "perceived magnitude" of intelligibility based on the standard stimulus or the initial assessment. Direct magnitude differs from equal-appearing interval scaling procedures in that it is not constrained by a fixed maximum or minimum level.

Although scaling procedures as a measure of speech intelligibility are limited in that they do not provide data on specific error patterns, they do offer a more realistic picture of the speaker's communication abilities in everyday speaking situations (Beukelman & Yorkston, 1980). Since these assessments are based on perceptual judgement, they are considered qualitative measures, thus the rating assigned reflects the degree of difficulty the listener has in understanding the context of the speech sample (Platt, Andrews, Young, & Nielson, 1978; Platt, Andrews, Young, & Quinn, 1980; Yorkston & Beukelman, 1978). Further, scaling procedures provide the examiner with an efficient tool for measuring speech intelligibility of young children.

Assessment Considerations

Assessing speech intelligibility involves more than selecting among the various methods available. In addition to selecting an approach to use, the examiner must also consider variables that influence the assessment environment. These other variables to consider when selecting a method of speech intelligibility assessment are mode of speech sample, time, and level of listener sophistication.

Sampling Modes

According to Kwiatkowski and Shriberg (1992), the major problem in accounting for variable intelligibility measures in clinical populations is not only selecting which method of assessment to use, but also selecting which speech mode to sample. For purposes of intelligibility assessment, three standard speech modes are considered. These include word, sentence, and conversational speech levels.

Ansel and Kent (1992) stated that single word and sentence level transcription tasks lend themselves to quantification and have been shown to provide the most stringent and valid measure of speech intelligibility because they provide the opportunity for phonemic or word analysis of the speaker's intended utterance. In support of word-level measures, Kent et al. (1989, p. 495) stated that the advantage of single word intelligibility testing includes "(a) quantification in terms of percentage correct, (b) potential for a phonetic feature analysis of the errors, (c) significant elimination of syntactic, prosodic, and other variables that effect sentence production or conversation, and (d) a simple response from subjects."

In contrast, Kwiatkowski and Shriberg (1992) stated that word- or sentence-level procedures are not valid for the purposes of intelligibility assessment. In fact, the authors stated that conversational speech sampling provides the only valid approach for assessing intelligibility in children with phonological disorders of unknown origin, whether obtained spontaneously or by imitation. Assessment of conversational speech samples is also considered a better indicator of functional performance because it tends to be more similar to daily speaking situations as opposed to production of words in isolation (Beukelman & Yorkston, 1980).

Time Factors

Scaling procedures that rely on the evaluation of the spoken utterance as judged by listeners produce data much more rapidly. A speech passage may be sampled and judged in 30 seconds as opposed to several minutes required to obtain a score with an objective transcription test (Cox, Alexander, & Rivera, 1991). This method also allows ready quantification of intelligibility of speech messages that are similar to everyday connected speech, whereas most objective tests use a type of speech (e.g., single monosyllables) not often found in everyday listening.

Other researchers agree that the reason scaling procedures are more often the procedure of choice among practicing speech-language pathologists is because they require less time to administer and few listeners are needed to analyze the speech sample (Kelly, Dancer, & Bradley, 1986; Schiavetti, 1981). In addition, Samar and Metz (1988) stated that although the transcription method of intelligibility assessment has a greater face validity than scaling methods, rating scales are generally regarded as being more "clinically tractable" (p. 307) or requiring less time and effort for administration.

Levels of Listener Sophistication

A measurement of a speech disorder is primarily a perceptual event, and the observer's response necessarily represents the final validation for any measurement (Young, 1969). However, each observer uses a certain amount of background knowledge or experiences when judging the intelligibility of speech samples. For example, there are three connotations for level of listener experience. The first

classification includes those listeners who are familiar with normal and deviant speech development. The second classification consists of those listeners closely associated with the speakers who have disordered speech (e.g., parents/caregivers). The third classification includes listeners with little or no knowledge of speech development and/or limited exposure to young children. This background knowledge determines the listener's level of sophistication ranging from experienced to unsophisticated. Several studies that focused on speech samples from speakers who are deaf, dysarthric, or alaryngeal have shown that individuals who have experience working with such speakers (usually speech-language pathologists) tend to judge speech samples of these speakers as more intelligible than do listeners without such previous experience (Beukelman & Yorkston, 1980; Managan, 1961; Markides, 1983; McCroskey & Mulligan, 1963; McGarr, 1983; Monsen, 1978; Thomas, 1963; Williams & Watson, 1985).

Research that focused on analyzing the influence of listener sophistication on intelligibility measures has revealed various classifications for experienced listeners. In their study, Kwiatkowski and Shriberg (1992) defined experienced listeners as graduate speech-language clinicians in their first or second year of school with experience providing intervention services for one or two children. Ellis and Fucci (1992) defined experienced listeners as those who held masters degrees in speech-language pathology, received certificates of clinical competence (CCC) from the American Speech-Language Hearing Association (ASHA), had recent ongoing experiences judging the intelligibility of taped speech samples, and had experience with phonetic transcription

of speech.

Even though sophisticated listeners had experience judging speech samples from unintelligible speakers, Kwaitkowski and Shriberg (1992) found that a significant problem using conversational speech samples to examine moments of unintelligibility is that the intended targets of some unintelligible strings cannot be ascertained. For this reason, caregivers and others familiar with the speaker are routinely enlisted in clinical and research environments to provide glosses for unintelligible utterances. The authors defined caregiver as "those who spent more than 40 hours a week with the child" (p. 1097). In 1987, Goehl and Martin observed three mother-child pairs using a sentence repetition task and concluded that caregivers were as much as 30% better than experienced speech clinicians at recognizing words spoken by their child.

At the other end of the continuum, Ellis and Fucci (1992) defined unsophisticated listeners as those who held masters degrees in a variety of other fields, had no previous experience judging the intelligibility of taped speech samples, and had no previous experience with phonetic transcription. In their study, Ellis and Fucci examined the effects of listener experience on two measures of intelligibility. For their study, 10 experienced listeners and 10 inexperienced listeners judged the intelligibility of nine audiotaped speech sample. The speech samples consisted of a male speaker who maintained the same stress and intonation pattern for all nine tapes, but varied the number of correct phonemes produced from 0 to 24. The major finding of the study was that experienced and inexperienced listeners did not differ significantly in their speech intelligibility judgements when using both transcription and scaling procedures.

Summary

When determining eligibility for intervention services, speech-language pathologists must assess speech intelligibility in the most accurate and reliable method. In clinical settings, however, time constraints often dictate the procedure for intelligibility assessment. Therefore, a method that is both efficient and accurate is essential when fulfilling the requirements for eligibility, while decreasing the amount of time spent on assessment.

The literature maintains that a combination of both identification methods and scaling procedures is necessary in order to present the best overall picture of the client's intervention needs (Ansel & Kent, 1992; Cox et al., 1991; Samar & Metz, 1988). However, as caseloads in the school system soar, the need for a quick and reliable method of assessing speech intelligibility is critical. Therefore, the focus of this study was to provide data supporting the reliability of scaling procedures, specifically ear estimation, by comparing speech intelligibility percentages as measured by orthographic transcription with speech intelligibility percentages as measured by ear estimation.

CHAPTER III

METHOD

The purpose of this study was to investigate the relationship between two types of measures for ascertaining the percentage of speech intelligibility for preschool children with varying degrees of phonological proficiency: (a) the orthographic transcription method and (b) the ear estimation method. Continuous speech samples from 48 children were analyzed for the percentage of words understood. The listeners used ear estimation methods to judge the percentage of speech understood from continuous samples. The estimations were compared to orthographic transcription percentages from a previous study.

In the original study (Gordon-Brannan, 1993), speech-language pathology (SLP) graduate students orthographically transcribed continuous speech samples from 48 speakers. The investigator computed the percentage of words understood from orthographic transcriptions of each speech sample. In this study, the speech samples were judged by novice listeners who rated the intelligibility of the children's speech by estimating the percentage of words understood. The ratings of these unsophisticated listeners were compared with ratings by graduate student SLP listeners.

Subjects

Unsophisticated Listeners

The student SLP listeners, from a previous study (Gordon-Brannan, 1993), included four experienced SLP graduate students from Portland State University. In addition to participation in articulation and language clinic, these students had completed a minimum of 3 credit hours of articulation and phonology course work. In the original study, the trained listeners orthographically transcribed each speech sample, from which a percentage of words understood by each listener was computed for each speaker, and a mean percentage of the 4 listeners was calculated for each speaker.

The listeners for this study included 4 unsophisticated listeners from the greater Portland area. Each unsophisticated listener had earned a bachelor's degree in a program other than speech-language pathology and had not completed any courses in phonology or articulation. In addition, the listeners for this study passed a hearing screening at 20dB HL for the frequencies of 500Hz, 1K Hz, 2K Hz, and 4K Hz and signed an informed consent form (Appendix A). After listening to an audiotaped speech sample, the listeners estimated the percentage of words understood.

Procedures

Speakers

The speakers included 48 preschoolers selected from Portland area preschools and speech-language pathology caseloads. Before testing was done, parents/caregivers signed informed consent forms. The children ranged in age from 4:0 to 5:6 (mean =

4:7). The speakers comprised three groups of children who ranged in level of phonological proficiency from no articulation errors to multiple omission errors. These speakers were selected for an earlier study on intelligibility (Gordon-Brannan, 1993).

In order to participate in the previous study (Gordon-Brannan, 1993), the children passed a bilateral hearing screening at 35dB for the frequencies 500, 1K, and 2K Hz, scored at or above the 10th percentile on the Test of Auditory Comprehension-Revised (TACL-R) (Carrow-Woolfolk, 1985), and demonstrated normal laryngeal and resonance function as determined by investigator observation. The children also had to have no known physical, neurologic, or motor impairments as reported through parent interview and questionnaire, and investigator observation (Appendix B).

For this study the groups were determined by ranking the speakers in numerical order from 1-48 according to their intelligibility percentage, as measured by orthographic transcription. The speakers were then divided into three groups of 16 (i.e. #1-#16, #17-#31, and #32-#48). Group I consisted of the 16 children demonstrating the most intelligibility, group II was comprised of the middle 16 children exhibiting average intelligibility, and group III included the 16 children with the least intelligibility. Appendixes C and D show the age, gender, and TACL-R scores, as well as the characteristics of each speaker group.

Speech Samples

For the earlier study by Gordon-Brannan (1993), speech samples were emitted by telling a story using pictures in The Relatives Came (Rylant & Gammell, 1985). Each

100-word speech sample was both audiotaped and videotaped in a sound treated room at Portland State University. A sharp SX D200 digital audiotape recorder, VHS recorder, AG-100, and Panasonic camcorder was used to record the speech samples. Each of the children spoke into an AKG (Model C451) capacitor flat microphone which was placed approximately 6" away from their mouths in either a microphone stand, on a cloth covered table, or on foam.

In the original study, transcripts were made from the initial orthographic transcriptions for the investigator to verify. Then the parent/caregiver reviewed the content of the transcriptions and either verified or corrected the listener's interpretation of his/her child's speech by identifying words that were unintelligible or misunderstood by the investigator.

In the earlier study, both speaker and investigator utterances were dubbed onto listener tapes in random order. These were used later for orthographic transcription and rating. To determine interjudge reliability, five speech samples with at least one sample from each of the three speaker groups were presented twice. These repeated samples were presented at the end of the tape.

Judgements

For the original study, the 4 listeners were familiarized with the story used to elicit the speech samples. While listening to the samples up to three times on their personal analogue audiotape recorders, they transcribed each utterance. Verbal and written instructions for this task were given (Appendix E).

For this study, 4 unsophisticated listeners were also familiarized with the story used to elicit the speech samples. After they were familiarized with the story, each speech sample was presented to the group during one 4-hour session. The equipment used for the presentations was a Denon digital audiotape recorder (Model DTR-80) connected to a Sony table-top speaker (Model SRS-150). The equipment specified is identical to equipment used in the previous Gordon-Brannan study (1993) because the use of different tape players could alter the output signal of the speech sample. The listeners estimated the percentage of words understood after each sample was played once. The directions for this task were delivered both orally and in writing (Appendix F).

Scoring

In the previous study (Gordon-Brannan, 1993), orthographic transcriptions were compared with pre-made score keys, to determine the accuracy of listener interpretation. Correct scores were given for words that were identified by the listener but not by the score key, or differed only in morphological form. Incorrect scores were given for words that were not identified by the listener but were on the score key, or were not identified by either the listener or score key. The percentage of words understood from each continuous speech sample was computed. For this study, the medians of estimated percentages by unsophisticated listeners and identification percentages by SLP student listeners were determined for each speaker.

Reliability

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

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

Data Analysis

To determine whether or not the listener estimation method for measuring speech intelligibility correlates with speech intelligibility as measured by the identification method, a Pearson r was used. To determine whether there is a significant difference in intelligibility percentages between unsophisticated listener judgment and orthographic transcription by student SLPs for each of the three groups of speech samples, preschool children with: (a) most intelligibility, (b) average intelligibility, and (c) least intelligibility, a two-tailed t -test was applied. The statistical significance level was set at the .05 level of confidence for all data analysis.

CHAPTER IV

RESULTS AND DISCUSSION

Results

The following section includes results from the study comparing speech intelligibility measures between unsophisticated listener judgements and orthographic transcription. Further, interjudge and intrajudge reliability was examined for the two listener groups, non-sophisticated listeners using ear estimation and graduate students using orthographic transcription. Finally, results presented include a comparison of speech measures for three groups of children divided into most, average, and least intelligible.

Interjudge and Intrajudge Reliability

Prior to reporting results for the two research hypotheses, interjudge and intrajudge reliability for the two methods of measuring intelligibility are presented. To determine interjudge reliability between nonsophisticated listener judgement and orthographic transcription, a Pearson r correlation matrix was constructed. Pearson r correlation matrices were provided through the SYSTAT computer program. Reliability matrices for each listener group are provided in Tables 1 and 2. Pearson r correlations for percentage of words understood in a continuous speech sample ranged from .86 to .95 for six pairs of listeners using the orthographic transcription method and .75 to .90 for six listeners pairs using nonsophisticated ear estimation. The statistics

Table 1
Orthographic Transcription Correlation Matrix for Intelligibility Measures

Listener	Listeners			
	1	2	3	4
1	1.00			
2	.89	1.00		
3	.86	.92	1.00	
4	.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 2

Non-sophisticated Listener Ear Estimation Correlation Matrix for Intelligibility Measures

Listener	Listeners			
	A	B	C	D
A	1.00			
B	.89	1.00		
C	.85	.90	1.00	
D	.75	.83	.84	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.

from these listener groups indicate that the listeners were in general agreement when assessing speech intelligibility from continuous speech samples.

The Pearson r was used to determine intrajudge reliability among nonsophisticated listeners. Reliability scores were determined by comparing each nonsophisticated listener judgement of five speech samples rated twice (Appendix G). The r value for Listener A was .60; Listener B, .74; Listener C, .58; and Listener D, .27. These results indicate poor reliability for nonsophisticated listeners using ear estimation to judge the speech intelligibility for the same five speakers twice. A discrepancy model (Appendix G) was used to examine intrajudge reliability further. When assigning speech intelligibility percentages of the 5 speakers presented twice, Listener A was within 2 percentage points for 1 speaker, 20 and 45 percentage points for 2 of the 4 speakers, and 60 percentage points for the remaining two speakers. Thus, Listener A's discrepancy scores were -60, -45, -20, +2, and +60, with a mean discrepancy score of 37.4 percentage points. Listener B assigned the same intelligibility percentage to 1 speaker on both listening opportunities, and was within 5 percentage points for another speaker. On 3 of the 5 continuous speech samples, Listener B assigned percentages within 40, 45, and 61 points of the first assigned percentage points. Her discrepancy score mean was 30.2%. Listener C judged 1 of the 5 speakers the same intelligibility percentage following both presentations. She was within 9, 32, 51 and 56 percentage points from the first presentation for the remaining four speakers, giving her a discrepancy score mean of 29.6%. Listener D judged 2 of the 5 speakers within 1% of the first presentation rating. He was within 12, 14, and 76 percentage points for other

three speakers. His discrepancy score mean was 20.6%. Based on these results, the four nonsophisticated listeners were inconsistent in their judgements of speech intelligibility.

Research Question I

The first research hypothesis tested was: There is a significant correlation between the orthographic transcription measure of intelligibility and the subjective ear estimation measure of intelligibility. To determine the correlation between the two methods, a Pearson r correlation was used. Appendix C provides data on mean percentage of speech intelligibility for each speaker measured by listener estimation and orthographic transcription. The Pearson r correlation was .85, indicating that the percentage of speech intelligibility as judged by ear estimation is positively correlated with the percentage of speech intelligibility as judged by orthographic transcription; however, intrajudge reliability for the ear estimation procedure was poor.

Research Question II

The second research hypothesis tested was: There is no significant difference in intelligibility percentages as judged by non-sophisticated ear estimation and orthographic transcription for each of three speaker groups of preschool children with: (a) most intelligibility, (b) average intelligibility, and (c) least intelligibility. Non-sophisticated listener ear estimation and orthographic transcription percentages for each speaker group were compared using a two-tailed t -test. Table 3 shows means, standard deviations, and ranges for each of the three speaker groups as determined by the two intelligibility measures. The most intelligible speaker group received the highest mean

Table 3

Descriptive Statistics for Two Intelligibility Measures for Three Speaker Groups

Measure: transcription		Groups		
		I Most Intelligible	II Average Intelligible	III Least Intelligible
%-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.

Measure: estimation		Groups		
		I Most Intelligible	II Average Intelligible	III Least Intelligible
%-of Words	Mean	83%	76%	28%
Understood	SD	18.3	20.2	20.9
in Speech	Range	30-100%	30-100%	0-88%

Note: Listener group - Non-sophisticated listener measurement technique - ear estimation; all numbers are rounded up to the nearest percent.

scores from both non-sophisticated listener ear estimation (\underline{M} = 83%) and orthographic transcription (\underline{M} = 93%). The speakers with average intelligibility received moderate mean scores from non-sophisticated listener ear estimation (\underline{M} = 76%) and orthographic transcription (\underline{M} = 81%), while the least intelligible speaker group received the lowest mean scores from both listener groups; non-sophisticated listener ear estimation (\underline{M} = 28%) and orthographic transcription (\underline{M} = 52%). The resulting t -values indicate a significant difference between the judgements of intelligibility as measured by non-sophisticated listeners' ear estimation compared to orthographic transcription for the most intelligible group ($t = -2.97$, $df = 15$, $p = .010$) and for the least intelligible group ($t = -7.94$, $df = 15$, $p = .000$). In these two groups, the mean percentage of words understood was higher for the orthographic transcription measure. There was no significant difference between non-sophisticated listener ear estimation and orthographic transcription intelligibility percentages for the average speaker group ($t = -.99$, $df = 15$, $p = .337$).

In summary, the non-sophisticated listener ear estimation percentages of intelligibility were significantly different from the orthographic transcription percentages for the young speakers in both the most intelligible and least intelligible speaker groups (adult-like and moderate/severe), but were not significantly different for the average intelligible group (mild/moderate).

Discussion

For this study, non-sophisticated listeners estimated the percentage of words understood from a continuous speech sample via auditory input only. These scores

were then compared to intelligibility percentages as measured by orthographically transcribing what each speaker said and calculating a percentage of words understood from the speech sample. The accuracy of the orthographic transcription method was enhanced by using parent/caregiver verification of the children with moderately and severely deficient phonologies. For the original Gordon-Brannan study (1993), parents/caregivers reviewed the content of the transcriptions. Then they either verified or corrected the listener's interpretation of their child's speech by identifying words that were unintelligible or misunderstood by the investigator. Comparisons made in this study included investigating the correlation between the two methods of measuring speech intelligibility. These results suggested that the two measures were correlated, although the actual percentages of the two measures differed significantly for children in the most intelligible group and least intelligible group. However, the listener's poor intrajudge reliability should be considered when interpreting these results.

Listener Reliability

When rating the five speech samples twice, the non-sophisticated listeners demonstrated poor reliability. Although the reliability scores for 2 of the 5 speakers were within 10-20 percentage points from the first rating, the reliability scores for the remaining three speakers showed great variability and therefore decreased listener reliability. For example, speakers in the average intelligibility group received scores with the largest variability. Specifically, Listener A estimated the intelligibility of Subject 38 as 30% for the first presentation and 90% after the second listening; Listener B, 38% and 99%; Listener C, 44% and 100%; and Listener D, 68 % and 100%. The

largest discrepancies between estimations were with Listener D, who assigned percentages to the same speech sample a total of 76 percentage points apart. In addition, the remaining three listeners also had high discrepancy ranges (58, 61, and 56) especially when estimating the intelligibility of speakers in the average or least intelligible groups. However, listener scores were more reliable for the most intelligible speaker and one of the least intelligible speakers, indicating that judgements made for speakers on either end of the continuum (most/least) are more consistent than judgements made on the average speaker group.

An investigation of interjudge reliability indicates a wide range of estimated percentages from the nonsophisticated listeners (Appendix H). In this study, one speaker (#38) was judged identically by all 4 listeners to be 100% intelligible. At the other end of the continuum, Subject 33 was judged 100% intelligible by Listener D, but only 30% intelligible by Listener A, for a difference of 70%. Notable similarities (10% or less) occurred for six speakers (8, 9, 13, 25, 36, and 45) in the most intelligible group. In addition, one other speaker (#47) in the average intelligible group received scores from all 4 listeners with a difference of 10% or less. On the other end of the continuum, notable differences (50% or more) occurred for two speakers (#14 and #33) in the most intelligible group, one speaker (#38) in the average intelligible group, and two speakers (#2 and #16) in the least intelligible group. The remaining 37 speakers received intelligibility scores that differed among the four listeners from 11% to 49%. These results indicate that, although the listeners were not in general agreement on the majority of intelligibility estimations, their estimations for individual speakers in the

most intelligible group were within 10% of each other. Thus, visual inspection of the data shows that there was more variability among nonsophisticated listeners for speakers with most intelligibility and least intelligibility.

A closer look at the graduate students who used orthographic transcription to assess intelligibility indicates that, although they were in general agreement with each other, the differences in the percentages for some of the speakers also reveal a wide range of percentages (Appendix I). For example, only 1 speaker, Subject 9, was judged 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. The largest percentage difference (42%) among the 4 listeners occurred on Subject 6, who received a percentage score of 70% from Listener 3, while Listener 1 assigned a score of 28%. Differences of over 10% occurred for 3 speakers (5, 25, and 37) in the most intelligible group, for 8 speakers (1, 12, 17, 28, 34, 35, 42, and 44) in the average intelligible group, and for 13 speakers (2, 3, 6, 11, 16, 20, 23, 24, 31, 32, 43 and 46) in the least intelligible group. These results demonstrate high variability among 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, the SLP graduate student listener's measures, using orthographic transcription, follow a pattern of increased variability as intelligibility decreases, suggesting that there are listener differences in understanding preschool speakers who are not essentially intelligible. On the other hand, nonsophisticated listener ear estimation measures indicated the highest

variability among listeners was within the average intelligible group, with the least intelligible group showing the lowest variability. These data suggest that nonsophisticated listeners using the ear estimation method were in higher agreement when determining most and least intelligible speakers, but differed when judging speakers with average intelligibility.

Degree of Intelligibility

The intelligibility scores ascertained through orthographic transcription procedures ranged from 19% to 100% for 48 children with varying degrees of phonological proficiency. The scores derived via non-sophisticated listener ear estimation ranged from 4% to 100% for the same 48 children.

When separated into groups according to intelligibility level, children in the most intelligible group were understood at least 80% of the time by the graduate students using the orthographic transcription method. Children with average intelligibility were understood at least 70% of the time, while the least intelligible group was understood 70% or less, with 6 of the members understood less than 50% of the time.

Comparison of the two measures shows that the percentage of words understood in the most intelligible group ranged from 100% to 88% for the orthographic transcription method and from 100% to 46% for the non-sophisticated listener ear estimation. The range of percentage points differed by an average of 24 points from orthographic transcription to ear estimation, indicating a significant difference between the two measures for assessing speech intelligibility of speakers with adult-like speech.

The group with average intelligibility received percentages from orthographic transcription ranging from 86% to 71% and from 96% to 45% from ear estimation. Although the highest nonsophisticated ear estimation was 10% higher than the highest orthographic transcription percentage, the *t*-test results revealed no significant difference between the two methods when assessing speech intelligibility of children with mild/moderate phonological proficiency.

The range of intelligibility percentages in the least intelligible group were from 70% to 19% for orthographic transcription and from 55% to 4% for ear estimation. The difference between methods of measuring speech intelligibility of children with moderate/severe phonological deficiency was 41%.

For this study, the criterion for accuracy is considered to be orthographic transcription. These findings suggest that accuracy of non-sophisticated ear estimation is greater in children with average intelligibility and there is greater variability between orthographic transcription and nonsophisticated ear estimation percentages when assessing the speech intelligibility of children at either end of the continuum. In other words, non-sophisticated listeners demonstrate difficulty estimating and differentiating average speech from the speech of children with either most intelligibility or least intelligibility.

Visual inspection of the raw data shows that over half of the continuous speech samples yielded percentages that differed by 20% or more between the two measurement techniques (Appendix C). In the most intelligible group, 4 out of 16 speaker percentages differed by 20% or more, with the largest difference being 44%.

Three out of 16 speakers in the average intelligible group received percentages from ear estimation that differed from orthographic transcription percentages by 20%. The largest discrepancy was on Subject 30 who was rated a mean difference of 43% between the two methods. In the least intelligible group, 7 out of 16 speakers were judged 20 percentage points apart from the orthographic transcription method to the ear estimation method.

According to the data (Appendix C), estimations were generally lower for non-sophisticated listener ear estimation across all speaker groups. In a similar study, researchers found that individuals who have experience working with speakers who are deaf, dysarthric, or alaryngeal tend to judge speech samples of these speakers as more intelligible than persons without such experience (Beukelman & Yorkston, 1980; Managan, 1961; Markides, 1983; McCroskey & Mulligan, 1963; McGarr, 1983; Monsen, 1978; Thomas, 1963; Williams & Watson, 1985). In fact, of the 48 speakers, only Subject 25 received mean percentages of 88.5 from both orthographic transcription and ear estimation. The largest discrepancy between methods for measuring speech intelligibility occurred in the least intelligible group. In this group, 15 out of the 16 speakers were given a lower mean percentage by ear estimation method as compared with the orthographic transcription method. In fact, of the 16 speakers from the least intelligible group, only one (Subject 27) received a higher mean percentage from ear estimation (54%) than from orthographic transcription (49.5%), for a difference of +4.25%. The largest discrepancies overall occurred on Subjects 14, 30, 23, and 46, with discrepancy scores of -44,

-43, -41.75, and -40.75, respectively.

In summary, these results indicate that, overall, speech intelligibility percentages measured by nonsophisticated listener ear estimation are lower than those derived by orthographic transcription. Further, the range of percentages between the two methods dramatically increased as intelligibility decreased, thus suggesting a greater difficulty for nonsophisticated listeners to understand intelligible speech in children with moderate/severe phonological deficiencies. In addition to level of listener sophistication, another explanation for high discrepancy scores among the nonsophisticated listeners is the type of scale used. For instance, in this study, listeners were instructed to make a speech intelligibility judgement based on a scale from 0% to 100%, thus leaving much room for variability among listeners; whereas other scales offer ranges on a continuum from 1, to 5, 7, or 9, thereby restricting the margin of error.

Listener Estimations

The practice of measuring speech intelligibility using ear estimation raises questions of accuracy and reliability (Gordon-Brannan, 1993; Kent et al., 1994). According to the results found in this study, there is a positive correlation between ear estimation as measured by nonsophisticated listeners and orthographic transcription. Yet, due to the poor intrajudge reliability, it is essential that these results be interpreted carefully. These findings are similar to the results from a study conducted by Ellis and Fucci (1992). In their study, experienced listeners and inexperienced listeners did not differ significantly in their speech intelligibility judgements when using both transcription and scaling procedures. There are, however, certain factors that may have

affected the listeners' ability to understand the speech samples. Since the purpose of this study was to investigate whether nonsophisticated listeners are able to judge speech intelligibility with the same accuracy as orthographic transcription procedures, one of the main factors influencing these results was level of listener experience or sophistication. The four listeners for this study were required to have earned degrees in other fields and have had no prior experience with phonological disorders in children. In fact, three out of the four had no experience with children at all while one had experience with a person who was hearing impaired.

Other factors that may have decreased the accuracy of the listeners' estimations include the quality of some of the speech sample recordings which made interpretation difficult, fatigue, and desensitization. Listeners complained that some of the speech sample recordings were extremely loud while others were barely audible. In addition, the listeners noted that throughout two of the speech samples they were very distracted by a continuous "buzzing" sound which made comprehension extremely difficult. All 4 listeners mentioned feeling fatigued after 4 hours of listening to speech samples. Even though the listeners were given opportunities to rest or reconvene on another night, they all chose to complete the study in one evening. After 2 hours of listening to children speak about a story with which the listeners were familiar, all 4 reported being somewhat desensitized to the task. The more they listened to the children talking about the same story, the more they anticipated or interpreted what the child said. In fact, it was noted that although the listeners were shown the materials prior to conducting the study and were given the opportunity to score three practice speech samples, the speech

intelligibility judgements tended to increase throughout the session. Generally, listening to the same story, spoken by children with disordered speech, for an extended period of time would tend to decrease the ability to listen objectively to each individual speech sample. Indeed, for the listeners in this study, the familiarity and anticipation of words may have actually increased their understanding of the children's speech. In fact, three out of four listeners reported rating the speakers intelligibility according to how much they understood from the context of the speech sample rather than specific words, which may have contributed to the poor intrajudge reliability. For this reason, future research projects involving these continuous speech samples should counterbalance the order of speaker presentation.

A factor that may have increased the accuracy of listener estimations was the examiner's utterances which could be heard on the taped speech samples. Although an attempt was made to control for this by not giving cues, the listeners stated that the examiner's utterances occasionally cued the listener to what word or phrase the child was saying, thus aiding their interpretation of the children's utterances.

In conclusion, the factors that influence the accuracy of listener estimation include listener experience, quality of stimulus materials, fatigue, desensitization, and examiner utterances present on taped speech samples. These factors should be considered when reading and interpreting the results presented in this study.

CHAPTER V

SUMMARY AND IMPLICATIONS

Summary

Intelligible speech is a primary component for successful communication. However, the speech of children with disordered phonologies is often unintelligible. Therefore, when assessing the speech intelligibility of children in order to determine whether they qualify for intervention services, speech-language pathologists need reliable evaluation tools.

The focus of this investigation was the measurement of speech intelligibility. The purpose of this study was to examine the relationship between two methods for measuring speech intelligibility. The first, identification method, involves the listener transcribing a speech sample from which the percentage of words understood is calculated. The second, scaling procedure, involves the listener estimating the percentage of words understood from a continuous speech sample. The secondary purpose of this study was to examine the accuracy of the scaling method as measured by ear estimation compared to the identification method as measured through orthographic transcription for each of three groups of children with: (a) the most intelligibility, (b) average intelligibility, and (c) least intelligibility.

Four unsophisticated listeners rated the speech intelligibility of 48 speakers aged

4:0 to 5:6 who comprised three groups with varying levels of phonological proficiency. The listeners who were unfamiliar with the speakers, but familiar with the topic, rated the children's continuous speech samples using ear estimation. The data collected were then compared with intelligibility ratings as measured in a previous study (Gordon-Brannan, 1994) via orthographic transcription.

The two methods of measuring speech intelligibility examined in this study were found to be positively correlated ($r = .86$). However, the t -test analysis revealed significant differences between the two measures for the most and least intelligible groups, indicating discrepancies between the two methods when measuring the speech intelligibility of some children. Additional statistical analysis revealed poor intrajudge reliability which should be considered when interpreting the results presented. It does appear, however, that when measuring speech intelligibility, using the ear estimation method, is reflective of the orthographic transcription measure, although the actual estimated percentages of intelligibility appear to differ from the percentages derived from orthographic transcription.

Implications

Clinical

The t -test results indicate a significant difference between the two measures investigated in this study when rating speech intelligibility of young speakers. These results show that the scaling method of ear estimation is not a reliable tool for measuring speech intelligibility when the listener has little or no experience with phonological disorders. The t -test results regarding the most, average, and least intelligible groups revealed significant differences between ear estimation and orthographic transcription in the groups with the most and least intelligibility. The average difference between the two measures for the most and least intelligible group was 24% and 41%, respectively. These measures may be clinically significant for setting intervention goals and criteria. In other words, when nonsophisticated listeners judge the speech intelligibility of young children, statistics from this study show that they have increased difficulty distinguishing intelligible speech from average speech.

On the other hand, there was no significant difference between the two methods when measuring the speech intelligibility of children with average intelligibility. Notably, 5 of the 16 speakers for this group differed by 15% or more on the two measures, which is probably clinically significant. While the statistical results of this study support the method of ear estimation when measuring speech intelligibility of young children who are between 70% to 87% intelligible, the differences between estimated and actual percentages for the majority of the children appear to be clinically relevant, as it could affect a child's eligibility for services. In addition to using the ear

estimation for assessing speech intelligibility, statistical results from this study also indicate that children whose intelligibility is below 70% should be evaluated by an objective method such as orthographic transcription to obtain accurate results.

Finally, SLPs using ear estimation methods to set intervention goals and outcome criteria must devise a procedure to monitor whether the client has acquired speech competency. This is normally achieved by pre- and post-testing and comparing the results in order to determine progress. Since remediation of unintelligible speech is a high priority, SLPs may choose to conduct periodic intelligibility assessments throughout the course of intervention. By conducting periodic assessments, SLPs may be able to determine whether the child still qualifies for intervention and/or at what priority level the child qualifies. Therefore, accurately assessing speech intelligibility is an integral component in the evaluation process because the outcome may influence the child's eligibility for services. In an ideal clinical setting, measuring speech intelligibility by incorporating both scaling and identification procedures gives the best overall picture of the speaker's communication ability. However, as caseloads increase, SLPs may turn to speech intelligibility assessments methods that are less time consuming as long as the measure is reliable.

As this study also indicates, different methods for determining percentage of intelligibility yield different results; therefore, the method used to measure a child's intelligibility should be indicated in the diagnostic report. Furthermore, the diagnostic report should also include characteristics of the listener or discuss who is doing the interpreting. In other words, the examiner should list his/her level of sophistication,

from knowledgeable about normal and deviant speech development to familiar with the speaker or unfamiliar with the speaker.

Research

As the speech-language profession expands and caseloads increase, SLPs need to incorporate methods for assessing children with varying degrees of speech intelligibility that are efficient, reliable and focus on the child's functional communication goal. Therefore, it is imperative that SLPs choose research-based evaluation tools. This study compared two methods for measuring speech intelligibility, identification and scaling.

The results of this study did lend some support for the validity of using ear estimation as a tool for measuring speech intelligibility in young children, regardless of listener experience level. However, further research in this area is warranted. Should this study be duplicated in any way, it is suggested that the examiner's voice be either eliminated from the speech samples or kept away from the microphone. In addition, future examination of these data should include rating the speech samples from videotaped sessions. Since speech samples are usually collected with the examiner/listener and child interacting together, the speaker's intelligibility is influenced by what the listener/examiner sees on the child's face when speaking, which is more like everyday speaking situations.

When considering the range of listener sophistication from little or no experience to experienced SLPs, another area in need of further investigation includes comparing the estimations of parents, educators, multi-disciplinary team members, or

SLP fellows with orthographic transcription to determine what influence level of experience has on estimating speech intelligibility.

In addition to choosing assessment tools that are efficient, it is the SLP's responsibility to assure that all methods used for evaluation are the most appropriate ones available for the child. The results of this study should help practicing SLPs make wise decisions regarding evaluation tools, intervention goals, and outcome criteria. As further research is conducted regarding ear estimation as a diagnostic tool, practitioners will be better informed about whether or not it is valid and reliable.

References

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. Journal of Speech and Hearing Research, 35, 296-308.

Bacon, V. J. (1995). Validity and efficiency of the check-slash transcription method for measuring intelligibility. Unpublished master's thesis, Portland State University, Portland, OR.

Bernthal, J. E., & Bankson, N. W. (1993). Articulation and phonological disorders (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Beukelman, D. R., & Yorkston, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. Journal of Communication Disorders 13, 33-41.

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.

Calvert, D. R. (1980). Descriptive phonetics. New York: Division Theime-Stratton, Inc.

Carrow-Woolfolk, E. (1985). Test for auditory comprehension of language-revised. Allen, TX: DLM.

Cox, R. M., Alexander, G. C., Rivera, I. M. (1991). Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. Journal of Speech and Hearing Research, 34, 904-915.

Ellis, L. W., Fucci, D. J. (1992). Effects of listeners' experience on two measures of intelligibility. Perceptual and Motor Skills, 74, 1099-1104.

Geohl, H., & Martin, P. (1987, November). Intelligibility and familiarity: Is experience enough? Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Gordon-Brannan, M. (1993). Speech intelligibility assessment of young children with varying levels of phonological proficiency/deficiency. Doctoral dissertation, The Wichita State University, Wichita, KS.

Gordon-Brannan, M. (1994). Assessing intelligibility: Children's expressive phonologies. Topics in Language Disorders, 14, 17-25.

Hodson, B. W., & Paden, E. P. (1991). Targeting intelligible speech (2nd ed.). Austin, TX: Pro-Ed.

Hodson, B. W., & Paden, E. P. (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.

Kelly, C., Dancer, J., & Bradley, R. (1986). Correlation of the SPINE test scores to judges' ratings of speech intelligibility in hearing-impaired children.

The Volta Review, 88, 145-150.

Kent, R. D. (1992). Speech intelligibility and communicative competence in children. In A.P. Kaiser & D.B. Gray (Eds.), The social use of language: Research foundations for early language intervention. Baltimore, MD: Brooks.

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.

Kwiatkowski, J., & Shriberg, L. S. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. Journal of Speech and Hearing Research, 35, 1,095-1,104.

Managan, K. (1961). Speech improvement through articulation testing. American Annals of the Deaf, 106, 391-396.

Markides, A. (1983). The speech of hearing impaired children. Oxford, UK: Manchester University Press.

McCroskey, R. L., & Mulligan, M. (1963). The relative intelligibility of esophageal speech and artificial larynx speech. Journal of Speech and Hearing

Disorders, 28, 37-41.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. Journal of Speech and Hearing Research, 26, 451-458.

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.

Monsen, R. B. (1978). Towards measuring low well hearing impaired children speak. Journal of Speech and Hearing Research, 21, 197-219.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. American Annals of the Deaf, 126, 845-852.

Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The preschool intelligibility measure. American Journal of Speech Language Pathology, 4, (4), 24-31.

Osberger, M. J. (1992). Speech intelligibility in the hearing impaired: Research and clinical implications. In R.D. Kent (Ed.), Speech intelligibility in speech disorders: Theory, measurement and management (pp. 233-264).

Amsterdam: John Benjamins.

Platt, L. J., Andrews, G., Young, M., & Nielson, P. D. (1978). The

measurement of speech impairment of adults with cerebral palsy. Folia Phoniatica, 30, 50-58.

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.

Schmidt, M. (1984). Intelligibility and the child with multiple articulation deviations. In H. Winitz (Ed.), Treating articulation disorders. Baltimore, MD: University Park Press.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), Intelligibility in speech disorders. Philadelphia, PA: John Benjamins Publishing.

Schiavetti, N., Metz, D. E., & Sitler, R. W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired. Journal of Speech and Hearing Research, 24, 441-445.

Schriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. Journal of Speech and Hearing Disorders, 47, 256-270.

Shames, G. H., & Wiig, E. H. (1990). Human communication disorders (3rd ed.). New York: Macmillan Publishing.

Thomas, W. (1963). Intelligibility of the speech of deaf children. Procedures of the International Congress on Education of the Deaf, Document No. 106.

Van Riper, C., & Emerick, L. (1990). Speech correction: An introduction to speech pathology and audiology (8th ed.). Englewood Cliffs, NJ: Prentice Hall.

Weiss, C. (1980). Weiss comprehensive articulation test. Allen, TX: DLM Teaching Resources.

Weiss, C. E. (1982). Weiss intelligibility test. Tigard, OR: C. C. 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.

Williams, S. E., & Watson, J. B. (1985). Differences in speaking proficiencies in three laryngectomee groups. Archives of Otolaryngology, 111, 216-219.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dyarthric speech. Journal of Communication Disorders, 11, 499-512.

Yorkston, K. M., & Beukelman, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single word intelligibility. Journal of Communication Disorders, 13, 15-31.

Yorkston, K. M., & Beukelman, D. R. (1981). Assessment of intelligibility of dysarthric speech. Tigard, OR: C. C. Publications.

Young, M. A. (1969). Observer agreement: Commulative effects of rating many samples. Journal of Speech and Hearing Research, 12, 135-143.

Appendix A

Informed Consent - Unsophisticated Listeners

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 provide supporting data that the method of 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 it may benefit others in the future.

Carla J. Dukart 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 B

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 _____
If so, how many times? _____		
When was the last time? _____		
had medical treatment for ear infections	yes _____	no _____
If 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 C

Individual Subject Data

Group Diff. of	Subject	Age	Sex	TACL	Intell. %	Intell. %	NS%-OT%
	#			%tile	OT (M)	NS (M)	
Most	9	4:11	F	88	99.50	99.25	-0.25
Most	36	4:8	F	95	98.25	99.00	+0.75
Most	45	5:2	F	61	97.50	98.25	+0.75
Most	8	4:5	F	89	96.25	97.00	+0.75
Most	13	4:3	M	67	95.50	90.50	-5.00
Most	48	4:11	M	91	95.25	74.00	-21.25
Most	5	4:1	F	98	93.50	78.00	-15.50
Most	39	5:5	M	59	93.25	100.00	+6.75
Most	4	5:5	M	76	93.25	88.00	-5.25
Most	29	4:11	F	66	92.00	73.75	18.25
Most	21	4:3	M	79	91.50	93.75	+2.25
Most	15	4:7	F	84	91.25	66.75	24.50
Most	14	4:0	F	64	90.00	46.00	44.00
Most	33	4:1	M	17	89.50	69.00	20.50
Most	25	4:5	M	94	88.50	88.50	+/-0.00
Most	37	4:6	F	39	87.75	68.75	19.00
Average	30	4:0	F	17	86.25	47.25	43.00
Average	35	5:6	F	73	86.00	94.00	+8.00
Average	44	5:0	M	46	86.00	85.75	-0.25
Average	10	4:10	M	50	86.00	83.25	-2.75
Average	26	4:1	M	57	85.75	79.75	-6.00
Average	47	4:8	M	31	85.50	95.75	+10.25
Average	34	5:0	M	57	84.50	87.75	+3.25
Average	7	4:11	F	97	83.00	45.00	-38.00
Average	40	4:0	M	35	81.25	96.25	+15.00
Average	18	5:0	M	76	80.00	78.50	-1.50
Average	17	5:0	M	94	78.75	74.25	-4.50
Average	42	4:1	F	97	77.50	77.25	-0.25
Average	28	4:7	F	85	75.50	57.50	-18.00
Average	12	4:8	M	11	73.00	91.25	+18.25
Average	22	4:2	M	14	71.25	80.25	+9.00
Average	38	4:9	M	57	70.75	49.50	21.25
Least	2	4:3	M	47	70.50	50.75	19.75
Least	11	5:11	F	35	69.50	55.00	14.50

Individual Subject Data (cont.)

Group of	Subject #	Age	Sex	TACL %tile	Intell.% OT (x)	Intell.% NS (x)	Diff. NS%-OT%
Least	1	4:6	M	62	68.75	32.25	36.50
Least	32	4:2	F	12	67.00	27.00	40.00
Least	23	4:9	M	29	63.50	21.75	41.75
Least	16	4:4	M	43	61.00	54.75	-6.25
Least	46	4:2	F	38	59.25	18.50	40.75
Least	31	4:1	F	35	55.75	30.25	25.25
Least	19	4:8	M	31	50.25	16.25	34.00
Least	6	4:10	F	27	50.00	23.50	26.50
Least	27	4:8	F	16	49.50	54.00	+4.50
Least	43	4:9	M	27	47.00	19.75	30.25
Least	20	4:0	M	41	46.00	14.00	32.00
Least	3	4:0	M	37	41.75	17.25	24.50
Least	41	4:2	M	44	35.75	6.25	29.50
Least	24	4:5	M	2	18.75	4.00	14.75

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

Appendix D

Characteristics of the Three Speaker Groups

Group	Mean Age (Age Range)	# of Females	# of Males	TACL-R %ile (Range)
Most intelligible	4:8 (4:0-5:5)	9	7	73%ile (17-98)
Average Intelligible	4:8 (4:0-5:6)	5	11	56%ile (11-97)
Least intelligible	4:6 (4:0-5:11)	6	10	33%ile (2-47)

Appendix E

Instructions to Graduate Student Listeners

You will hear 48 children, 4 to 6 years of age, in conversational speech samples. You are not to transcribe the last 5 samples, samples 49-53. Write down the sample # 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 F

Listening Instructions to Unsophisticated Listeners

Listener name: _____ Listener #: _____

You will be listening to a total of 53 audiotapes with connected speech samples. This will be split into two sessions. During this session, you will listen to 53 connected 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. However, you I ask that you not take notes on any of the samples you listen to. When each 3 minute sample has been completed, you will have approximately 15 seconds to select a percentage of intelligibility between 0-100% for that speaker. Put your estimated percentage of intelligibility in the space provided which corresponds with the speech sample presented.

Your name will not be used in any way with the findings of this study. You will be referred to by listener number only. Do you have any questions?

Data Entry Sheet

Sample 1 _____ %	Sample 19 _____ %	Sample 37 _____ %
Sample 2 _____ %	Sample 20 _____ %	Sample 38 _____ %
Sample 3 _____ %	Sample 21 _____ %	Sample 39 _____ %
Sample 4 _____ %	Sample 22 _____ %	Sample 40 _____ %

Sample 5 _____ %	Sample 23 _____ %	Sample 41 _____ %
Sample 6 _____ %	Sample 24 _____ %	Sample 42 _____ %
Sample 7 _____ %	Sample 25 _____ %	Sample 43 _____ %
Sample 8 _____ %	Sample 26 _____ %	Sample 44 _____ %
Sample 9 _____ %	Sample 27 _____ %	Sample 45 _____ %
Sample 10 _____ %	Sample 28 _____ %	Sample 46 _____ %
Sample 11 _____ %	Sample 29 _____ %	Sample 47 _____ %
Sample 12 _____ %	Sample 30 _____ %	Sample 48 _____ %
Sample 13 _____ %	Sample 31 _____ %	Sample 49 _____ %
Sample 14 _____ %	Sample 32 _____ %	Sample 50 _____ %
Sample 15 _____ %	Sample 33 _____ %	Sample 51 _____ %
Sample 16 _____ %	Sample 34 _____ %	Sample 52 _____ %
Sample 17 _____ %	Sample 35 _____ %	Sample 53 _____ %
Sample 18 _____ %	Sample 36 _____ %	

Appendix G

Listener Estimations for Speakers Rated Twice

Group	Sub. #	Listener A		Listener B		Listener C		Listener D	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd
		Difference		Difference		Difference		Difference	
Most	4	98	100	80	85	86	95	88	100
		2		5		9		12	
Average	28	60	0	40	0	51	0	79	3
		60		40		51		76	
Average	38	30	90	38	99	44	100	86	100
		60		61		56		14	
Least	1	30	10	50	50	38	38	11	11
		20		0		0		0	
Least	11	60	15	60	15	61	29	39	40
		45		45		32		1	

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; Average = group with average intelligibility; Least = least intelligible group.

Appendix H

Raw Data by ListenerPercentage-of -Words Understood in Continuous Speech
Ear Estimation -- Non-sophisticated Listeners

Sub.#	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
Group:						
Most						
9	99	98	100	100	99.25	2
36	99	99	98	100	99.00	2
45	99	99	95	100	98.25	5
8	95	95	98	100	97.00	5
13	95	85	89	93	90.50	10
48	95	65	63	73	74.00	32
5	90	60	84	78	78.00	30
39	100	100	100	100	100.00	0
4	98	80	86	88	88.00	18
29	70	70	73	82	73.75	12
21	90	93	92	100	93.75	10
15	40	60	81	86	66.75	41
14	25	50	30	79	46.00	54
33	30	50	96	100	69.00	70
25	95	85	86	88	88.50	10
37	70	55	75	75	68.75	20
						M = 20

Group:						
AVG						
30	50	39	54	46	47.25	15
35	90	99	89	98	94.00	10
44	95	73	86	89	85.75	22
10	55	93	85	100	83.25	45
26	90	80	87	62	79.75	28
47	99	90	94	100	95.75	10
34	90	80	98	83	87.75	18
7	30	40	34	76	45.00	46
40	99	89	97	100	96.25	11
17	75	75	63	84	74.50	21
42	90	75	58	86	77.25	32
28	60	40	51	79	57.50	39

Sub.#	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
12	90	99	98	78	91.25	21
22	65	75	81	100	80.25	35
38	30	38	44	86	49.50	<u>56</u>

M = 26

Least

2	12	39	74	78	50.75	66
11	60	60	61	39	55.00	22
1	30	50	38	11	32.25	39
32	15	18	42	33	27.00	27
23	10	5	37	35	21.75	32
16	60	39	32	88	54.75	56
46	15	8	16	35	18.50	27
31	20	35	27	40	30.50	20
19	20	10	11	24	16.25	14
6	30	20	18	26	23.50	12
27	30	75	48	63	54.00	45
43	10	20	9	28	16.75	19
20	5	8	13	30	14.00	25
3	5	25	4	35	17.25	31
41	2	1	6	16	6.25	15
24	2	0	3	11	4.00	<u>11</u>

M = 17

Appendix I

Raw Data by Listener

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

Sub.#	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
Group: Most						
9	99	99	100	100	99.50	1
36	98	99	99	97	98.25	2
45	98	98	96	98	97.50	2
8	95	97	96	97	96.25	2
13	92	98	96	96	95.50	6
48	98	97	92	94	95.25	6
5	92	86	98	98	93.50	12
39	91	92	97	93	93.25	6
4	92	95	97	89	93.25	8
29	96	91	95	86	92.00	10
21	96	89	91	90	91.50	7
15	90	92	92	91	91.25	2
14	91	89	92	88	90.00	4
33	92	91	89	86	89.50	6
25	92	95	82	85	88.50	13
37	96	82	88	85	87.75	<u>14</u>
						M=6

**Group:
AVG.**

30	75	91	92	87	86.25	17
35	92	79	95	78	86.00	17
44	92	78	84	90	86.00	14
10	87	80	87	90	86.00	10
26	85	88	83	87	85.75	5
47	89	85	81	87	85.50	8

Sub.#	Listener 1	Listener 2	Listener 3	Listener 4	Mean Score	Range
7	83	83	83	83	83.00	0
40	86	79	76	84	81.25	10
18	74	84	84	78	80.00	10
17	70	82	83	80	78.75	13
42	85	71	82	72	77.50	14
28	82	74	76	70	75.50	12
12	65	72	78	77	73.00	13
22	70	71	68	76	71.25	8
38	68	77	71	67	70.75	<u>10</u>
						M=10

**Group:
Least**

2	61	72	74	75	70.50	14
11	59	73	64	82	69.50	23
1	60	71	65	70	68.75	6
32	70	69	59	70	67.00	11
23	72	59	58	65	63.50	14
16	49	69	62	64	61.00	20
46	68	67	46	56	59.25	22
31	58	55	58	52	55.75	6
19	53	51	53	44	50.25	9
6	28	42	70	60	50.00	42
27	56	53	36	53	49.50	20
43	51	45	40	52	47.00	12
20	43	48	40	32	46.00	16
3	37	46	35	49	41.75	14
41	42	34	28	39	35.75	14
24	35	10	8	22	18.75	<u>27</u>
						M=17
