

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

1987

Media and intelligibility influences on whole word broad phonetic transcription reliability

Alice Elizabeth Smith
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Smith, Alice Elizabeth, "Media and intelligibility influences on whole word broad phonetic transcription reliability" (1987). *Graduate Student Theses, Dissertations, & Professional Papers*. 7218.
<https://scholarworks.umt.edu/etd/7218>

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

COPYRIGHT ACT OF 1976

THIS IS AN UNPUBLISHED MANUSCRIPT IN WHICH COPYRIGHT
SUBSISTS. ANY FURTHER REPRINTING OF ITS CONTENTS MUST BE
APPROVED BY THE AUTHOR.

MANSFIELD LIBRARY
UNIVERSITY OF MONTANA
DATE: 1987

MEDIA AND INTELLIGIBILITY INFLUENCES
ON WHOLE WORD BROAD PHONETIC
TRANSCRIPTION RELIABILITY

by

Alice Elizabeth Smith

B.A., Mary Baldwin College, 1974

M.S., University of North Carolina, 1977

Presented in partial fulfillment of
the requirements for the degree of

Master of Arts

UNIVERSITY OF MONTANA

1987

Approved by:

Barbara Ben
Chairman, Board of Examiners

J. C. [Signature]
Dean, Graduate School

August 10, 1987
Date

UMI Number: EP38019

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI EP38019

Published by ProQuest LLC (2013). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Media and Intelligibility Influences on Whole Word Broad
Phonetic Transcription Reliability (134 pp.)

Director: Barbara A. Bain, Ph.D.

The purpose of this study was to explore the effects of audiotaped versus audio-videotaped recording conditions and speaker intelligibility on interobserver and intraobserver reliability calculated from whole word broad phonetic transcriptions. Descriptive analyses of the data were also conducted.

Two male children matched for mean length of utterance and differing in intelligibility were each audio-videotaped utilizing a high quality audio signal. The speech sample was the Goldman-Fristoe Test of Articulation (1969). Sixteen subjects transcribed each speech sample in an audio-only and an audio-visual condition. Each condition was transcribed two times for a total of eight transcriptions per subject. The order of transcription conditions was counterbalanced across subjects. Each transcription was scored on a point-by-point basis with a scoring key to calculate interobserver and intraobserver reliability and to analyze the target sounds transcribed incorrectly.

Results indicated a significant difference between the intelligibility conditions. No significant differences occurred between recording conditions or transcription times regardless of speaker intelligibility. Descriptive analyses indicated essentially no differences between recording conditions with regard to the percentages of target sound characteristics transcribed incorrectly. Voiced alveolar stop-plosive and fricative consonants and lax unrounded high and mid-front vowels were most frequently transcribed incorrectly. Variability existed in percentages of occurrence of transcription errors between intelligibility conditions but the relative frequency of occurrence of errors on phoneme characteristics was the same as found for recording conditions except for consonant voicing.

Conclusions indicated essentially no differences in reliability calculated from whole word broad phonetic transcriptions obtained from audio-only versus audio-videotaped recordings when speakers differed in degree rather than quality of intelligibility. Further research is indicated to explore calculated reliability differences which may occur due to qualitative intelligibility differences in speakers.

ACKNOWLEDGEMENTS

The completion of this thesis was made more bearable by the time, energy, and encouragement of many people. My love and thanks are extended to all of them. My subjects spent many hours in the recording lab. Janis Hayes-Strom, Twila Tabor, Raelene Hall, Mary Pat Schilly, and Carol Henighan provided suggestions, happy hour companions, and shoulders to lean on when things looked darkest. Mary Hardin never hung up on late night phone calls from Iowa and Montana. Mike Wynne loaned much valuable computer time and instruction. David Jones and Bonnie Edwards saw me through the perils and pleasures of learning Word Perfect while typing the document.

The members of my committee deserve special thanks. Barbara Bain's enthusiasm, patience, and red pen taught me to think and write more clearly and critically. Wes Shellen spent countless hours assisting me with the confusion of statistics and data organization which faced me. Donald Goldberg listened when I needed to think aloud and shared stacks of articles from his dissertation as I began my research.

Dana Smith was with me every step of the way.

TABLE OF CONTENTS

ABSTRACT.....		ii
LIST OF TABLES.....		vi
LIST OF FIGURES.....		vii
 Chapter		
I. INTRODUCTION.....		1
Media and Data Collection.....		3
Role of Visual Cues in Observation....		6
Observational Media and Phonetic Transcriptions.....		9
Influence of Media on Reliability.....		13
Statement of the Problem.....		19
II. METHODS.....		23
Subjects.....		23
Speakers.....		24
Stimulus Tapes.....		25
Reliability of Stimulus Tapes.....		27
Experimental Condition.....		29
Measurement, Design, and Analysis.....		32
Interobserver Reliability Design...		33
Interobserver Reliability Measurement.....		33
Interobserver Reliability Analysis.		34
Intraobserver Reliability Design...		34
Intraobserver Reliability Measurement.....		35
Intraobserver Reliability Analysis.		35
Descriptive Design.....		36
Descriptive Measurement.....		36
Descriptive Analysis.....		36
Investigator Reliability.....		37

Chapter

III.	RESULTS.....	38
	Audiotaped versus Audio-Videotaped Recordings.....	39
	Interobserver Reliability.....	39
	Intraobserver Reliability.....	42
	Descriptive Analyses.....	44
	Sound Errors.....	45
	Consonants.....	45
	Consonant Summary.....	55
	Vowels.....	56
	Vowel Summary.....	66
	Summary of Audiotaped versus Audio-Videotaped Recordings.....	67
	Speaker Intelligibility:	
	Intelligible versus Unintelligible.....	68
	Interobserver Reliability.....	68
	Intraobserver Reliability.....	70
	Descriptive Analyses.....	72
	Sound Errors.....	73
	Consonants.....	73
	Consonant Summary.....	83
	Vowels.....	84
	Vowel Summary.....	96
	Summary of Speaker Intelligibility...	97
IV.	DISCUSSION.....	98
	Clinical Impressions.....	100
	Audiotaped versus Audio-Videotaped Recordings.....	100
	Speaker Intelligibility.....	105
	Future Research.....	110
	ENDNOTES.....	115
	APPENDICES.....	116
	REFERENCES.....	127

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1	Results of analysis-of-variance (ANOVA) (time/intelligibility/media) for sixteen subjects..... 40
2	Intraobserver point-by-point percentages of agreement for each of 16 subjects in the audio-only and audio-visual recording conditions with percentages of difference between Time 1 (T1) and Time 2 (T2)..... 43
3	Percentages of all consonant substitution errors, omission errors, and addition errors analyzed according to place and manner of articulation in the audio-only and audio-visual recording conditions at Time 1 and Time 2..... 54
4	Percentages of all vowel substitution errors, omission errors, and addition errors analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 and Time 2..... 65
5	Intraobserver point-by-point percentages of agreement for each of 16 subjects in the intelligible and unintelligible conditions with percentages of difference between Time 1 (T1) and Time 2 (T2)..... 71
6	Percentages of all consonant substitution errors, omission errors, and addition errors analyzed according to place and manner of articulation in the intelligible and unintelligible conditions at Time 1 and Time 2..... 83
7	Percentages of all vowel substitution errors, omission errors, and addition errors analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height in the intelligible and unintelligible conditions at Time 1 and Time 2..... 95

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1	Point-by-point percentages of agreement among 16 subjects on all transcribed sounds in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	41
2	Percentages of all sounds transcribed incorrectly which occurred on consonants and vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	46
3	Percentages of all consonants transcribed incorrectly by place of production in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	48
4	Percentages of all consonants transcribed incorrectly by manner of production in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	49
5	Percentages of all consonants transcribed incorrectly on voiced and voiceless consonants in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	51
6	Percentages of all consonants transcribed incorrectly which were substitution, omission, or addition errors in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	53
7	Percentages of all vowels transcribed incorrectly which occurred on tense and lax vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....	57

8 Percentages of all vowels transcribed incorrectly which occurred on rounded and unrounded vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....59

9 Percentages of all vowels transcribed incorrectly by degree of tongue advancement in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2)....61

10 Percentages of all vowels transcribed incorrectly by degree of tongue height in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2)....62

11 Percentages of all vowels transcribed incorrectly which were substitution, omission, or addition errors in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2).....64

12 Point-by-point percentages of agreement among 16 subjects on all transcribed sounds in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2)...69

13 Percentages of all sounds transcribed incorrectly which occurred on consonants and vowels in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....74

14 Percentages of all consonants transcribed incorrectly by place of production in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....75

15 Percentages of all consonants transcribed incorrectly by manner of production in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....77

16 Percentages of all consonants transcribed incorrectly which occurred on voiced and voiceless consonants in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....79

17 Percentages of all consonants transcribed incorrectly which were substitution, omission, or addition errors in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....81

18 Percentages of all vowels transcribed incorrectly which occurred on tense and lax vowels in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....86

19 Percentages of all vowels transcribed incorrectly on rounded and unrounded vowels in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....88

20 Percentages of all vowels transcribed incorrectly by degree of tongue advancement in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....90

21 Percentages of all vowels transcribed incorrectly by degree of tongue height in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....91

22 Percentages of all vowels transcribed incorrectly which were substitution, omission, or addition errors in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2).....93

CHAPTER 1

INTRODUCTION

Reliability is defined as the stability, precision, and consistency in reporting behaviors and test results. Clinically, reliability provides confidence in diagnostic and treatment data. That is, for example, the accurate determination of the existence of a speech or language problem and the documentation of therapy progress depend in part on reliable measurements.

Interobserver reliability refers to agreement between two or more examiners on a particular set of obtained data and is an important indicator of objective and precise measurements of the behaviors of interest. Intraobserver reliability refers to agreement by one examiner on two or more observations of a particular set of obtained data and is an important indicator of objective and precise measurement of the behaviors of interest by the same examiner. The reliability obtained on whole word broad phonetic transcriptions of a child's speech by trained examiners is one issue of importance with regard to diagnostic speech assessments since phonetic transcriptions are a common basis for the determination of the existence of a speech

problem. More specifically, Henderson (1938) suggested the need for consistent and objective examiner judgments when obtaining and evaluating articulation test results. The present study is concerned with interobserver and intraobserver reliability obtained from whole word broad phonetic transcriptions of preschool children's speech.

Various factors have been identified which influence the reliability with which observers obtain data. Some of the identified factors have included the complexity of the behaviors being rated (Harris & Lahey, 1982; Kazdin, 1977; Mash & McElwee, 1974); examiner expectations about the behavior being assessed (Harris & Lahey, 1982; Kazdin, 1977; Lipinski & Nelson, 1974; Oller & Eilers, 1975; Redfield & Paul, 1976); changes in observer expectations about the behavior being assessed due to feedback (Harris & Lahey, 1982; Kazdin, 1977); knowledge of interobserver or intraobserver reliability assessments (Foster & Cone, 1980; Kazdin, 1977; Kent, Kanowitz, O'Leary, & Cheiken, 1977; Reid, 1970; Romanczyk, Kent, Diament, & O'Leary, 1973; Taplin & Reid, 1973); knowledge of who is assessing reliability (Foster & Cone, 1980; Kent, et al., 1977; Romanczyk, et al., 1973); prior training assessing the behavior of interest (Baer, 1977; Harris & Lahey, 1982; Mash & McElwee, 1974; Taplin & Reid, 1973); and the observational media from which the data are obtained (Amorosa, von Benda, Wagner, & Keck, 1985; Henderson, 1938; Hodson & Paden, 1983; Ingram, 1976; Irwin, 1970; Kent & Foster, 1977; Kent, O'Leary, Dietz, & Diament, 1979; Kipfmuehler & Prins, 1971; Lovaas, Koegel, Simmons, & Long, 1973; Mencke, Ochsner, & Testut, 1983; Moller & Starr, 1984;

Wright, 1954). This last factor, the influence of observational media on reliability, is a focus of the present study.

The remainder of this chapter will discuss the use of various media for obtaining behavioral data, the role visual cues play in speech and language observation and perception, the advantages and limitations of various media for obtaining reliable phonetic transcriptions, and comparisons of interobserver and intraobserver reliability utilizing various media for obtaining speech samples. The chapter concludes with a statement of the problem.

Media and Data Collection

Direct observations, audiotaped recordings, and audio-videotaped recordings are three prevalent methods of behavioral data collection. The type of medium chosen by an examiner depends in part on the complexity and type of behavior to be rated and the frequency and rate of occurrence of the behavior of interest (Cherry Wilkinson, Clevenger, & Dollaghan, 1981; Dollaghan & Miller, 1986; Holm, 1978; Rosenblum, 1978; Sackett, 1978).

Direct observations entail in vivo data collection while the behaviors of interest are occurring. Behaviors measured via direct observation are typically simple rather than complex and must be easily identified and recorded (e.g., out-of-seat behavior, occurrence/nonoccurrence of a specified behavior). Direct observa-

tions require the examiner to be able to easily view the subject in order to identify the behaviors to be measured. In addition, the examiner must be trained in the correct identification of the behavior of interest as well as in the measurement method being used since there is no second chance to verify the occurrence of a particular behavior.

Audiotaped recordings enable the examiner to identify verbal or vocal behaviors at some time after the behaviors have occurred. Simple or complex occurrences of verbal or vocal behavior may be tape recorded for later listening and the audiotape may be replayed as needed to verify the occurrence of a behavior. In addition, an audiotape provides a means by which further analysis beyond the original purpose of the recording may be conducted at any time. The examiner should be trained in listening to the behavior of interest on audiotape in order to easily identify and note the occurrence of the behavior without frequent replays (Shriberg & Kent, 1982). Care must be taken to obtain representative samples of the subject's verbal and/or vocal behaviors via appropriate recording equipment and control of the recording environment (e.g., constant speaker distance from the microphone, use of a unidirectional microphone, quiet recording location, appropriate signal-to-noise ratios and frequency response) (Code & Ball, 1984; Ingram, 1976; Shriberg & Kent, 1982).

Videotaped recordings also allow the examiner to observe and record the behaviors of interest at any time after the behaviors have

actually occurred. Unlike the direct observation method, the examiner may view the subject performing any type of simple or complex behavior since videotapes allow for replay of the behaviors of interest. In addition, the examiner may focus on verbal and/or vocal behaviors as well as visual cues which may facilitate identification of precipitating nonverbal behaviors not available on audiotaped recordings. As noted for audiotaped recordings, appropriate equipment and environmental controls should be utilized for video recording to obtain optimal videotaped samples of the behaviors of interest (e.g, good lighting, close-ups, and sharp focus) (Shriberg & Kent, 1982). The examiner should be trained in the correct identification and measurement methods for the behaviors of interest to decrease unnecessary replay time. As noted for audiotaped recording analysis, the examiner has the option to analyze the videotapes again at a later time for additional data not originally obtained in the initial analysis.

Three methods of obtaining behavioral data have been discussed. These include direct observation, audiotaped recordings, and videotaped recordings. Each method provides the examiner with a means by which behaviors of interest may be observed and measured, however, the latter two methods provide an added advantage since audio- and videotaped recordings may be replayed for data verification purposes.

All three observational media have been utilized alone or in combination with each other in the field of speech and language pathology for behavioral observation and data collection (Amorosa, et

al., 1985; Elphick, 1984; Henderson, 1938; Moller & Starr, 1984; Irwin, 1970; Mencke, et al., 1983; Monsen, 1983; Stephens & Daniloff, 1977; Wright, 1954). Comparisons of interobserver and intraobserver reliability measurements obtained within and between two or more of the above-mentioned media as addressed in the speech and language research will be discussed in a later section. The complexity and types of behaviors which may be measured differ slightly between media. Videotaped recordings provide an optimal medium for observation and measurement of the most complex speech and language behaviors and direct observations ususally provide data regarding simpler behaviors. Audiotaped recordings are typically reserved for data collection regarding only verbal or vocal behaviors since visual cues are unavailable on audiotape for measuring nonverbal behaviors. The influence of visual information on the observation and assessment of speech and language behaviors will be discussed next.

Role of Visual Cues in Observation

Visual cues provide information to an individual in a number of ways. Visual cues identify a speaker's nonverbal pragmatic behaviors and environmental contextual information such as items in a setting pertinent to conversational meaning. The identity of the speaker and listener are available via visual information if the interaction is taking place face-to-face. In addition, information about visual components of speech such as lip closure and lip postures may aid a

listener in interpreting a speaker's utterances which might otherwise be unintelligible (Shriberg & Kent, 1982).

A number of studies have cited the positive contribution of visual cues when accompanied by auditory input for increasing perceived intelligibility of a speaker (Hudgins & Numbers, 1942; Monsen, 1983; Subtelny, 1977) by increasing a listener's ability to make phonemic distinctions during speechreading tasks (Erber, 1979; Fisher, 1968; Walden, Prosek, & Worthington, 1974; Woodward & Barber, 1960), for identification of correct/incorrect consonants (Irwin, 1970; Mencke, et al., 1983; Irwin & Krafchick, 1965), and for speech perception during adverse listening conditions (Binnie, Montgomery, & Jackson, 1974; Miller & Nicely, 1955; Neely, 1956; O'Neill, 1954; Sanders & Goodrich, 1971; Steele, Binnie, & Cooper, 1978). Results of these studies suggested visual cues enhanced auditory perception. Intelligibility has been positively correlated with the number and type of articulation errors produced by a speaker (Andrews, Platt, & Young, 1977; Hudgins & Numbers, 1948; Jordan, 1960; Prins & Bloomer, 1968; Van Demark, 1964). Thus, the literature suggested visual cues enhanced the intelligibility of individuals, particularly those with a high number of articulation errors in their speech. However, the majority of the intelligibility studies reviewed which utilized speech samples obtained from speakers with varying degrees of articulation deficits have obtained data solely utilizing audiotaped recordings (Andrews, et al., 1977; Hoops & Curtis, 1971; Metz, Samar, Schiavetti, Sitler, & Whitehead, 1985; Monsen, 1978; Platt, Andrews, Young, &

Quinn, 1980; Prins & Bloomer, 1965; Prins & Bloomer, 1968; Tikofsky & Tikofsky, 1964; Yorkston & Beukelman, 1978; Yorkston & Beukelman, 1981) rather than from videotaped recordings (Irwin & Krafchick, 1965; Mencke, et al., 1983; Monsen, 1983). In addition, orthographic transcriptions of the speech samples were used in all of these studies. Orthographic transcription requires examiners to write down a word or words without regard to phonetic accuracy. An examiner could theoretically identify a word orthographically from a small number of phonemes which would allow the examiner to guess the rest of the word. Theoretically, an examiner could also guess at phonetic production from a small number of phonemes but without the accuracy which phonetic transcription requires. Of the studies cited above, only Yorkston and Beukelman (1978; 1981) utilized phonetic transcriptions obtained from speech samples but they used audiotaped recordings rather than videotaped recordings. Thus, past intelligibility studies have not addressed the influence of visual cues on whole word phonetic transcriptions. Since visual cues may enhance phoneme identification and intelligibility ratings one would expect the intelligibility studies to have made use of both visual information as well as phonetic transcriptions to accurately rate intelligibility and identify phonemic errors beyond the orthographic level.

Visual cues may play a significant role during speaker/listener interactions by also providing a means of identifying pragmatic intent of the language interaction, enhancing phoneme identification when necessary, and increasing speaker intelligibility. Although research has been directed toward the role of visual cues during speechreading,

adverse listening conditions, and identification of correct/incorrect consonant productions, a need exists for further research regarding the influence which visual cues may have during whole word phonetic transcriptions of the speech of unintelligible individuals. The following section will discuss the advantages and limitations of various media for obtaining reliable phonetic transcriptions.

Observational Media and Phonetic Transcriptions

Determining the existence of an articulation problem in a child has traditionally been obtained from live phonetic transcriptions of single word articulation tests. Some clinicians have suggested using in vivo transcriptions as their primary source for articulation test analyses in order to take advantage of the visual speech cues thought to play a role in speech intelligibility. Live transcriptions of connected speech samples are not typically utilized due to the inability of most clinicians to phonetically transcribe rapid speech on-line, particularly if a child's speech is unintelligible. Live transcriptions of single words may also prove difficult if a child is unintelligible since the clinician is unable to verify transcription accuracy in vivo. Requiring a child to repeat a word does not ensure identical word reproduction and, therefore, does not ensure transcription reliability. Two methods of assessing reliability for whole word transcriptions are the use of interobserver reliability and intraobserver reliability measures. Again, however, the lack of a recorded speech sample results in the inability of the examiners to readily

identify their sources of difference in order to obtain transcription agreement.

The need to provide a record of a child's speech which could be replayed for verification of live transcription accuracy encouraged the use of audiotaped recordings by many clinicians. Audiotaped recordings may be used to transcribe connected speech samples as well as single word samples at any time by another clinician for reliability purposes and may also be used as a backup for in vivo transcriptions (Ingram, 1976; Shriberg & Kent, 1982). Audiotaped recordings are suggested as a backup for live whole word transcriptions since the audiotaped recordings do not provide visual information thought to aid in phoneme identification but do theoretically allow the clinician to identify auditory variations in the production of a word or sound. Audiotaped recordings are not typically used in place of live whole word transcriptions. Routine observations by this investigator of most commercial tape recorders in use clinically indicated the recorders typically provide poor audiotape quality which may not allow accurate whole word phonetic transcriptions. Some clinicians have stated the importance of using the best quality tape recorder possible for transcription purposes (Numbers & Hudgins, 1948; Shriberg & Kent, 1982). Those clinicians who do exclusively use audiotapes for transcription purposes typically keep notes containing the speech sample context, a gloss of the child's intended production, and remarks about the child's nonverbal behaviors to facilitate later transcription of the sample (Bernstein & Tiegerman, 1985; Dale, 1978).

Thus, the use of audiotaped recordings obtained on poor quality clinical equipment becomes a concern. More specifically, poor quality audiotaped recordings may confirm or replicate in vivo transcriptions but may not be any more accurate despite their replay advantage. The use of high quality audiotaped recordings appears, therefore, to be an important consideration for obtaining reliable and accurate interobserver and intraobserver results albeit without visual information.

Videotaped recordings have typically been used for interobserver and intraobserver reliability of language analyses due to the value of visual cues during language production, particularly with regard to pragmatic intent (Bernstein & Tiegerman, 1985). Given the importance of reliable transcription information for phonological analyses of a child's speech and, considering the role of visual cues for accurately identifying some speech sounds more easily when speech is unintelligible, the use of videotaped recordings appears to be a viable option for interobserver and intraobserver reliability of whole word phonetic transcriptions. The addition of high quality audio to the videotape would also appear to provide optimal recording conditions for both visual and auditory input from which to calculate reliability of whole word transcriptions, particularly since the audio quality found on most videorecorders is poor, as noted previously (Shriberg & Kent, 1982). Shriberg and Kent have identified, however, a number of other limitations to the use of videotape recordings for phonetic transcriptions. Among the limitations mentioned are the high cost of video equipment (although high quality audio equipment is also

expensive); the greater distraction factor while recording with video which may result in poorer performance by the child; more complicated operation of video recording usually requiring another operator besides the clinician; and anecdotal reports by transcribers of becoming more tired when transcribing from videotapes, using more playbacks, and looking at the videotape monitor less frequently as the transcription session progresses (Shriberg & Kent, 1982). Despite these limitations, the advantage which videotapes accompanied by high quality audio recordings might afford the clinician in reliably and accurately assessing a speech disorder which might not be reliably and accurately assessed by audiotape alone could well outweigh the limitations involved in the use and cost of video recordings.

A discussion of the advantages and limitations of live observations, audiotapes, and videotapes for obtaining whole word phonetic transcriptions suggest interobserver and intraobserver reliability would be better for high quality audiotaped recordings or videotaped recordings plus high quality audiotaped recordings than for live observations for two major reasons. First, recordings can be replayed any number of times to obtain interobserver and intraobserver agreement. Second, high quality recordings would provide optimal visual and/or auditory input for accuracy of transcription. The clinician's choice of which medium he or she prefers for transcription purposes should logically be based upon which medium provides the highest degree of interobserver and intraobserver reliability. A review of the literature revealed no research comparing reliability of whole

word phonetic transcriptions calculated from audiotaped recordings to reliability calculated from videotaped recordings. The next section will discuss the research which has addressed the issue of the influence of media on reliability of data obtained from speech samples.

Influence of Media on Reliability

The majority of research addressing reliability has not investigated the influence of more than one data collection medium on reliability. The published studies which have addressed media influences on reliable behavior assessment have either compared interobserver and/or intraobserver reliability for two or more media (Amorosa, et al., 1985; Eisler, Hersen, & Agras, 1973; Elphick, 1984; Henderson, 1938; Irwin & Krafchick, 1965; Kent, et al., 1979; Lovaas, et al., 1973; Mencke, et al., 1983; Moller & Starr, 1984; Monsen, 1983; Stephens & Daniloff, 1977; Wright, 1954) or the performance of subjects exposed to different interviewing media (Dinoff, Newmark, Barnhart, Holm, Stern, & Saunders, 1970; Dinoff, Stenmark, & Smith, 1970). Research which has addressed the influence of media on interobserver and intraobserver reliability will be discussed next.

Videotaped observations have been found to be as reliable as live observations for recording the occurrences and nonoccurrences of multiple behaviors of autistic children (Lovaas, et al., 1973) and nonverbal interactions of married couples (Eisler, et al., 1973).

Kent, et al. (1979) compared the use of live observation, observation through a one-way mirror, and closed circuit television observations for recording the occurrences of nine categories of behaviors in first and second grade children. Reported results indicated insignificant differences between media for recording behaviors. Elphick (1984) found no significant differences between live and video presentations of a speechreading task to children, although scores were higher in the live mode of presentation for both single word and sentence identification. No research was found which compared the use of videotaped recordings to live observation recordings for observing complex oral speech and language behaviors. The studies cited above utilized behavior occurrence/nonoccurrence or correct/incorrect scoring methods and did not address the issue of specifically identifying oral speech and language behaviors. One would expect videotaped recordings accompanied by high quality auditory input to produce more reliable results than live observations for identifying oral speech and language behaviors since videotaped recordings may be replayed any number of times and would provide optimal visual and/or auditory input.

Eight studies addressed the influence of media on the reliability of assessing speech behaviors. These studies will be reviewed below.

Audiotaped recordings have been found to be slightly more reliable than live observations for rating types of articulation behaviors on a seven point scale (72% exact agreement interobserver ratings in live

observations as compared to 76% exact agreement interobserver ratings for audiotaped recordings) (Wright, 1954). Wright also reported intraobserver agreement ratings between media (79%, 82%, and 75%, respectively, for each of three examiners). Rank difference correlation coefficients (ρ) were calculated on interobserver and intraobserver rating scores. Results indicated greater interobserver and intraobserver reliability occurred during tape recorded conditions than in live observation.

Audiotaped recordings which subjects were allowed to listen to more than once were found to be more reliable for phonetic transcriptions of initial consonants for the speech of two unintelligible subjects (96% and 92%, respectively) than were live observations (89% and 76%, respectively) (Amorosa, et al., 1985). When whole word narrow transcriptions were analyzed, interobserver agreements were higher for the audio-only condition (73% and 70%, respectively, for each subject) than for the live condition (57% and 55%, respectively). Again, subjects listened to the audio-only productions more than once. Statistical analyses were not completed to determine statistical differences between media.

Stephens and Daniloff (1977) found judges tended to be more stringent when determining the acceptability of /s/ productions from audio-visual recordings than from audio-only recordings although overall interjudge agreement was similar between the two conditions (47%, audio-only; 52%, audio-visual). In a second experiment in the

same study, the investigators found more reliable /s/ ratings for a live condition than an audio-only condition when normal and "defective" /s/ speakers were rated. Results indicated 98% to 100% agreement in the live condition that the normal/s/ productions were acceptable and the defective /s/ productions were not acceptable. However, when the same productions were heard on audiotape by the same judges, more normal and "defective" /s/ productions were judged to be unacceptable. The investigators concluded live recordings are more reliable for appropriately judging /s/ productions and audio-visual recordings are slightly more reliable than audio-only recordings. Statistical analyses were not conducted to determine significance levels.

Moller and Starr (1984) compared speech ratings obtained from live observations, obtained via a mirror while listening to a loudspeaker system, and obtained from listening to an audiotaped recording. Judges rated speech samples of cleft palate speakers on the variables of resonance, articulation, voice, and overall speech acceptability on a subjective scale of 0 (no problem) to 7 (severe problem). They also rated speech intelligibility as a percentage. The authors found judges rated articulation, resonance, overall speech acceptability, and intelligibility similarly across all three media. Live voice ratings were significantly better than ratings of voice from the audio-visual and tape recorded media. One-way analyses-of-variance (ANOVAS) computed across conditions showed only ratings of voice reached statistical significance in any condition. The mean intel-

ligibility was 88% and the mean articulation rating was 2.4. Since intelligibility may be directly related to the number and type of articulation errors and most speakers in this study were highly intelligible and exhibited only mild articulation deficits, the lack of a statistically significant difference in ratings between audio and visual media may not reflect an adequate sampling of intelligibility and articulation deficits. Further research is needed to compare the effects of media on whole word phonetic transcriptions of subjects with varying degrees of intelligibility before conclusions can be drawn as to differences between media.

Henderson (1938) found two judges were slightly more reliable when identifying correct/incorrect consonants from audiotaped recordings (90%) than for live observation (89%) or when listening to a loudspeaker (78%). Phonetic transcriptions were also slightly more reliable when obtained from audiotaped recordings (85%) than from live observation (80%) or when listening to a loudspeaker (75%). No statistical analyses were conducted to determine significance levels. In addition, different speakers with different articulatory abilities were used in each condition, possible confounding interpretation of the data. A replication of the study using the same speakers in each condition and including the application of statistical analyses appears warranted.

Videotaped recordings have been shown to be more reliable than audiotaped recordings for interobserver agreement on identification of

initial and final consonants (Mencke, et al., 1983), for identifying phonetic errors (Irwin & Krafchick, 1965), and for identifying intelligible sentences (Monsen, 1983). Irwin and Krafchick (1965) reported statistically significant differences for interobserver correct and incorrect consonant identifications. Monsen (1983) reported a statistically significant difference for interobserver agreement of intelligibility between the media. As noted earlier, visual cues appear to play a facilitative role for speechreading and intelligibility by enhancing auditory information. None of the studies cited above compared reliability ratings for whole word phonetic transcriptions obtained from videotaped recordings to audiotaped recordings thus suggesting a need for investigation in this area.

All of the eight studies reviewed reported media effects on interobserver reliability. Only Wright (1954) also reported intraobserver reliability. Four of the seven studies performed statistical analyses of their data to determine the significance of results (Irwin & Krafchick, 1965; Moller & Starr, 1984; Monsen, 1983; Wright, 1954). A viable need exists for research addressing the effects of media on both interobserver and intraobserver reliability as well as statistical analyses of the data to determine whether differences between media are significant. Results of the literature review will be summarized in the next section.

Statement of the Problem

Research which has compared the effects of media on reliability of whole word phonetic transcriptions is inconclusive. Two studies attempted to show media effects on transcription reliability (Amorosa, et al., 1985; Henderson, 1938). However, Amorosa et al. (1985) only provided evidence of a need to review audiotaped recordings more than once for reliable results and left unanswered the role visual cues might play in a live observation medium versus a simulated live medium. Henderson's (1938) results were confounded partially due to her use of different speakers in each medium. In addition, the quality of audiotaped recordings at the time of her study cannot be compared to the quality of audiotapes today. Furthermore, statistical analyses were not used to determine the level of significance of media differences.

Only one of the eight studies which compared the effects of media on reliability reported intraobserver reliability as well as interobserver reliability (Wright, 1954). Interobserver reliability, as noted previously, is the agreement between two or more judges on the same set of data obtained within a medium. Intraobserver reliability, again, is the agreement reached by one judge on the same set of data obtained in the same or in different media (Shriberg & Kent, 1982). Both types of data are useful for providing confidence in data interpretation. None of the research comparing the effects of media

on reliability addressed both interobserver and intraobserver reliability with regard to whole word phonetic transcriptions.

Intelligibility has been shown to have a direct relationship with the number and types of articulation errors produced by a speaker. Thus, the degree of intelligibility of a speech sample may directly effect the reliability calculated from a whole word phonetic transcription of the sample. Visual cues have been cited as increasing perceived intelligibility. However, the effect of visual cues on whole word phonetic transcriptions obtained from speech samples with different degrees of intelligibility has not been addressed in the literature.

There is increased pressure in speech and language settings to provide services on a "prescribed basis" to children with articulation disorders, i.e., to identify and provide services to those children who demonstrate the greatest need for services. Therefore, a need exists to obtain the most reliable and valid diagnostic and treatment data possible to determine the existence of a problem, the need for therapy, and criteria for dismissal. Clinicians frequently utilize audiotapes or videotapes for analysis purposes either as a backup to live observations or, perhaps, in place of live observations. A clinician working in the field often may not have access to another clinician to obtain reliability checks. Therefore, clinicians should consider purchasing equipment which has been shown to provide the most reliable data. If both audiotape and videotape equipment provide

equivalent reliability, the clinician should logically purchase whichever equipment is less expensive and best suits his or her needs. If videotape equipment is equivalent to audiotape equipment in reliability, the videotape equipment provides the added advantage of being available for nonverbal language behavior analyses. Thus, this study is designed to address the issue of what type of recording medium best serves the clinician's needs.

In summary, the influence of media on reliability suggested videotaped observations were equivalent to live observations for recording the occurrence of nonverbal behaviors, audiotaped recordings were more reliable than live observations for judging speech behaviors, and videotaped recordings were more reliable than audiotaped recordings for speechreading and general intelligibility tasks. Obviously, a need exists to further investigate the effects of videotaped recordings and audiotaped recordings on the interobserver and intraobserver reliability calculated on whole word phonetic transcriptions obtained from speech samples representative of different levels of intelligibility to determine if differences exist between the media. The present study is designed to answer the following research questions:

- 1) What is the effect of high quality audiotaped versus high quality videotaped recordings on the interobserver and intraobserver reliability calculated on whole word broad phonetic transcriptions?

2) Does speaker intelligibility affect the interobserver and intraobserver reliability of whole word phonetic transcriptions obtained from audiotaped and videotaped recordings?

CHAPTER 2

METHODS

Subjects

The subjects were nine female graduate students in the Communication Sciences and Disorders Department at the University of Montana and seven female speech/language pathologists practicing in the Missoula, Montana area for a total of sixteen subjects. Each subject was selected according to the following criteria:

- 1) successful completion of a class in phonetic transcription methods with a course grade of B or better to ensure basic knowledge and skills in phonetic transcription procedures;
- 2) responding to a pure tone hearing screening administered by the investigator at 20 dB HL at .5, 1, 2, and 4kHz (ASHA, 1984) according to ANSI (1978) standards to ensure minimal hearing acuity for the perceptual transcription task in this study;
- 3) obtaining a word recognition score of 96% or better on an audiotaped modified 25-word list (Griffiths, 1967) presented at 50 dB HL through headphones in an audiological sound booth by the investigator at the University of Montana (Appendix A) to ensure adequate speech recognition skills at conversational loudness for the perceptual transcription task in this study;
- 4) reportedly normal or corrected vision to ensure adequate

visual acuity for the perceptual transcription task in this study;

5) obtaining correct/incorrect consonant agreement of 75% or better (Amorosa, et al., 1985; Henderson, 1938; Norris, Harden, & Bell, 1980; Shriberg, Kwiatkowski, & Hoffman, 1984) on the Grand Quiz on Phonetics Tape 4-A from Shriberg and Kent's (1982) training program to ensure current minimal competency for performing the phonetic transcription task in this study (Appendix B);

6) no previous contact with either of the two children selected as speakers for the study to decrease the possibility of transcription bias related to speaker familiarity.

Each subject completed a survey addressing previous transcription experience utilizing audiotaped and videotaped recordings as well as previous clinical experience. A copy of the survey and a summary of results may be found in Appendix C.

Speakers

Two male children ages 4:3 and 3:8 years were chosen as speakers for the stimulus tapes. The children were selected according to the following criteria:

1) enrollment in speech therapy locally due to the presence of an articulation problem;

- 2) responding to a pure tone hearing screening administered by the investigator at 20 dB HL at .5, 1, 2, and 4kHz according to ANSI (1978) methods in an audiological sound booth at the University of Montana;
- 3) the absence of any physical or mental handicapping conditions as reported by their parents;
- 4) a match for mean length of utterance (MLU) (Chapman, 1981) such that both children had an MLU of 4.4;
- 5) a Percent Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982) modified intelligibility rating on single word production (rather than connected speech) such that one child received a score of 85% in the mild-moderate category and one child received a score of 50% in the moderate-severe category (Appendix D).

Stimulus Tapes

Two videotapes were prepared as stimuli for the present study. The videotapes were recorded using an Hitachi VHS portable VTR (Model VT-7A), a Panasonic video camera (Model WV3160), and two T-120 Scotch VHS videocassette tapes. The audio portion of the videotapes was enhanced utilizing a Nakamichi cassette deck (Model BX-1), a TOA-FM wireless microphone, and a TOA tuner (Model WT-02). A schematic of the recording system may be found in Appendix E. The videotape and audiotape recordings were prepared in the recording laboratory of the Communication Sciences and Disorders Department at the University of

Montana. A head and shoulders view of each child was recorded on each videotape.

Each recording consisted of the investigator administering the 44-item Sounds-in-Words subtest of the Goldman-Fristoe Test of Articulation (GFTA) (1969) to a child. The GFTA is a single word articulation test and was chosen as the experimental stimulus for this study since practicing clinicians typically use single word articulation tests for diagnosis and assessment of children with articulation errors. Administration time of the GFTA is short and the pictures on the Sounds-in-Words subtest represent vocabulary familiar to most children (Goldman & Fristoe, 1969). The stimulus plates each contain one color picture of an object which allowed the investigator to provide a minimal fifteen-second period after a speaker's production of a word before introducing the next stimulus plate. This fifteen-second time period allowed adequate time for subjects to listen to the speaker's production of a word and transcribe the whole word on the scoring form before attending to the next word. Additional transcription time was provided by pausing the recording if requested by a subject.

The investigator presented each stimulus plate on the Sounds-in-Words subtest of the GFTA to the speaker according to the protocol outlined in the test manual (Goldman & Fristoe, 1969). If the speaker could not remember the name of a picture the investigator provided an indirect model to obtain a spontaneous response (e.g., "This is a house. What is this?"). If the speaker turned his head away from the

house. What is this?"). If the speaker turned his head away from the camera or became distracted (e.g. coughed) during the production of a word, he was reminded to look at the camera and was asked to repeat the word. These procedures were followed to ensure the visual and auditory clarity of the production of all 44 words.

Reliability of Stimulus Tapes

Interobserver reliability of the whole word phonetic transcriptions was established by two non-subject observers using the same equipment and conditions described in the experimental condition below. Both non-subject observers were selected according to the subject selection criteria previously described (see Appendices A-C). The two observers were not chosen as subjects because they were unable to make a long-term time commitment to the present study. The observers individually listened to and phonetically transcribed each videotape under two media conditions, audio-only and audio-visual. The order of condition presentation was counterbalanced between the observers. The same speaker was not listened to in both conditions on the same day. A minimal half-hour break was offered between the two condition presentations listened to on the same day to reduce fatigue. A minimal one-week break was provided before the other two conditions were transcribed (see Appendix F). Whole word phonetic transcriptions were obtained for each of the 44 test items for each speaker. There were a total of 127 consonants and 69 vowels (total=196 sounds) for the intelligible speaker and 99 consonants and 67 vowels (total=166

sounds) for the unintelligible speaker (see Appendix G). A percentage of agreement was obtained on a point-by-point comparison. Disagreements were discussed and the tapes replayed as often as necessary until differences were resolved and one transcription key was agreed upon for each speaker (see Appendix H). These scoring keys were used as the basis of comparison of the experimental subject data for obtaining interobserver and intraobserver reliability measures.

Appendix G provides a summary of the consonant and vowel characteristics presented by each speaker on the stimulus tapes and agreed upon for the transcription scoring keys. Few differences existed between the intelligible and unintelligible speakers when percentages of consonant characteristics were compared. The intelligible speaker used higher percentages of liquids, rhotic/palatals, velars, and voiced consonants and the unintelligible speaker used higher percentages of stop-plosives, nasals, bilabials, and voiceless consonants. Most percentages of difference in consonant use between the two speakers did not exceed 6% except for liquids (11%). When vowel usage percentages were compared, the unintelligible speaker used higher percentages of high, front, back, tense, and unrounded vowels. The intelligible speaker used higher percentages of mid, central, lax, and rounded vowels. Percentages of difference in vowel use tended to be approximately 10% when tongue height and advancement were compared and approximately 4% when tenseness of production and degree of lip rounding were considered. Therefore, the two speakers used ap-

proximately the same percentages of consonant and vowel characteristics in the speech samples presented to the subjects in this study.

Experimental Condition

Videotapes were presented to the subjects utilizing an Hitachi VHS portable VTR (Model VT-7A), an Hitachi random access tuner (model VT-T67A), and a JVC Telstar color monitor (Model C-2073US). The audio portion of the videotape was presented utilizing a Pioneer stereo amplifier (Model A-5) and one Boston Acoustics (Model A40) speaker. A schematic of the recording playback may be found in Appendix E.

Subjects individually observed the videotapes with the investigator in the recording laboratory of the Communication Sciences and Disorders Department at the University of Montana. Each subject was seated at a table directly facing the color monitor not more than five feet away (Erber, 1971). Tape loudness presentation levels were set at 55 dB HL as measured by a General Radio Company sound level meter (Model 1565-A) reading obtained by the investigator next to the subject's chair.

Tape stimuli were presented in two media conditions, audio-only and audio-visual, for each speaker. This was accomplished by use of the color monitor during audio-visual presentations and by turning off the color monitor during audio-only presentations. Therefore, each subject listened to each speaker on two different occasions in each of the two media for a total of four experimental conditions:

- 1) Audio-intelligible;
- 2) Audio-unintelligible;
- 3) Audiovisual-intelligible; and
- 4) Audiovisual-unintelligible.

The four experimental conditions were utilized to provide a means to obtain subject data for comparison of interobserver and intraobserver reliability of whole word broad phonetic transcriptions obtained from audiotaped recordings and videotaped recordings and to assess the influence of intelligibility on whole word broad phonetic transcriptions obtained from the two media. In addition, the use of the same audiotaped recording in both the audio-only and audio-visual conditions for each speaker insured a consistent auditory input for the purpose of determining the effects of visual cues on interobserver and intraobserver transcription reliability.

The presentation of tapes was counterbalanced across all conditions such that no two subjects listened to the four conditions in the same order or heard the same speaker on the same day (see Appendix F) to control for possible interactions between order of presentation and learning effects. Each subject listened to each experimental condition two times during the study. First, each experimental condition was presented to each subject in the assigned counterbalanced order presented in Appendix F. On Day 1, the first two experimental conditions were presented to each subject. A half-hour break was offered between conditions on each day to reduce fatigue (Shriberg & Kent, 1982). On Day 2, which took place at least one week after Day

1, the second two experimental conditions were presented to each subject. On Day 3, which took place at least one week after Day 2, each subject listened to the first two experimental conditions in reverse order. On Day 4, which took place at least one week after Day 3, each subject listened to the second two experimental conditions in reverse order. The subject data obtained from Day 1, Day 2, Day 3, and Day 4 were used to determine interobserver reliability (Day 1 versus Day 2, Day 3 versus Day 4) and intraobserver reliability (Day 1 versus Day 3, Day 2 versus Day 4).

Each subject was provided with a transcription sheet which included a list of the stimulus words to be transcribed and a space next to each word for the whole word broad phonetic transcription. Passage familiarity has been cited as influencing intelligibility estimates by listeners (Beukelman & Yorkston, 1980) and as a possible influence decreasing validity of transcription (Shriberg & Kent, 1982). However, a written gloss of intended production for whole word phonetic transcription is typically recommended (Shriberg & Kent, 1982). Since clinicians typically utilize an articulation test scoring form which provides a written gloss of intended production and this study is designed in part to address issues of clinical relevance to practicing clinicians, a written gloss was provided for the subjects. In addition, if a stimulus word had been repeated more than once by the speaker, a notation was placed beside that word on the transcription sheet indicating which production the subject was to transcribe. A key was also provided on each transcription sheet to

Appendix H. In addition, a phoneme transcription key was provided for each subject to decrease the possibility of inaccurate transcriptions due to subject inability to recall a phonemic symbol.

Each subject was allowed to listen or look and listen to each stimulus word one time. Shriberg and Kent (1982) recommended listening to a word a maximum of two times when phonetically transcribing. Sufficient controls were provided to enhance whole word broad phonetic transcription following only one presentation by controlling the loudness of stimulus presentations, and by providing optimal audio-only and audio-visual clarity, a phoneme transcription key, and a minimal fifteen-second time period in which to transcribe. Each subject was instructed to enhance visual input during the audio-visual condition for each speaker by attending to the color monitor. Each subject was instructed not to discuss transcription findings outside of the recording laboratory.

Measurement, Design, and Analysis

Sixteen subjects performed whole word broad phonetic transcriptions of the Sounds-in-Words subtest of the Goldman-Fristoe Test of Articulation (GFTA) (1969) in four conditions:

- 1) Audio-intelligible;
- 2) Audio-unintelligible;
- 3) Audiovisual-intelligible; and
- 4) Audiovisual-unintelligible.

4) Audiovisual-unintelligible.

Each subject listened to each condition two times during the experimental study for a total of eight transcriptions per subject.

Interobserver Reliability Design

A 2x2x2 repeated measures split-plot factorial design (time/intelligibility/media) was utilized for the present study with repeated measures over the three factors. This design was used to answer the interobserver component of both research questions. The use of Pearson product-moment correlations was originally planned to partially answer the interobserver reliability components of both research questions but was found to be inappropriate for the data analysis (see Endnote 1).

Interobserver Reliability Measurement

The eight whole word broad phonetic transcriptions obtained from each subject were scored according to the transcriptions keys on a point-by-point basis. The percentage of agreements ($[\text{total \# of agreements} / \text{total \# of agreements} + \text{disagreements}] \times 100$) obtained by each subject was recorded for each phonetic transcription in each medium condition for a total of eight scores per subject (two scores per condition):

- 1) Audio-intelligible;
- 2) Audio-unintelligible;

- 3) Audiovisual-intelligible; and
- 4) Audiovisual-unintelligible.

Interobserver Reliability Analysis

A 2x2x2 (time/intelligibility/media) analysis-of-variance (ANOVA) was performed to determine whether significant main effects were present for any of the factors or interactions between factors which might have affected the interobserver reliability obtained on the whole word phonetic transcriptions. There were two levels per factor (Time 1, Time 2, Intelligible, Unintelligible, Audio-only, Audio-visual) so no follow-up tests were required. The ANOVA was obtained utilizing the Ullrich-Pitz program on the University of Montana DECSYSTEM-2065 (DECA), TOPS 20 mainframe computer. In addition, percentage agreement data were collapsed at both transcription times such that interobserver reliability was first compared between the two recording conditions (audio and audio-visual) and then compared between the two intelligibility conditions (intelligible and unintelligible).

Intraobserver Reliability Design

A percentage of agreement ($[\text{total \# of agreements} / \text{total \# of agreements} + \text{disagreements}] \times 100$) design was utilized for the present study. This design was used to answer the intraobserver reliability component of both research questions. The use of Pearson product-moment correlations was originally planned to answer the intraobserver

reliability components of both research questions but was found to be inappropriate for the data analysis (see Endnote 1).

Intraobserver Reliability Measurement

The two whole word broad phonetic transcriptions obtained from each subject in each condition were compared on a point-by-point basis with the transcription keys (see Interobserver Reliability Measurement). Data were collapsed such that the percentage of agreement in each of the two transcription times in each experimental condition was recorded for each subject for a total of eight scores per subject (two scores per condition):

- 1) Audio-only;
- 2) Audio-visual;
- 3) Intelligible; and
- 4) Unintelligible.

Intraobserver Reliability Analysis

The two percentages of agreement obtained in each condition for each subject were compared to determine the effect of time between transcriptions on intraobserver reliability. Percentages of difference between the two transcriptions in each condition were calculated for each subject.

Descriptive Design

A percentage of agreement design was utilized to describe the overall agreements and disagreements on all sounds transcribed correctly and incorrectly. This design was utilized to describe the types of errors made in each of the experimental conditions.

Descriptive Measurement

All phonemes transcribed were tallied in each experimental condition at both transcription times. Percentages of agreement were calculated for all sounds, for consonants only, and for vowels only.

Descriptive Analysis

Data were collapsed such that descriptive results were analyzed in four conditions: audio-only, audio-visual, intelligible, and unintelligible. The percentage of all sounds transcribed correctly and incorrectly in each condition were reported. For purposes of this study errors were analyzed and percentages were reported relative to total transcription errors rather than to the frequency of occurrence of various consonant and vowel characteristics on the scoring key. Results should be interpreted accordingly. The percentages of all sounds transcribed incorrectly which were errors on consonants and vowels were reported. The percentages of all consonants which were

transcribed incorrectly were analyzed and reported for place, manner, and voicing characteristics. The percentages of all consonants transcribed incorrectly which were substitution, omission, or addition transcription errors were also reported. The percentages of all vowels which were transcribed incorrectly were analyzed and reported for the characteristics of tenseness of production, lip configuration during production, and tongue position during production. The percentages of all vowels transcribed incorrectly which were substitution, omission, or addition transcription errors were also reported.

Investigator Reliability

Interobserver agreement was calculated on the scoring of subject transcriptions for the interobserver and intraobserver analyses between the investigator and one of the non-subject observers described earlier. Eight subject transcriptions were independently scored utilizing the scoring key. In addition, eight subject transcriptions were independently scored for intraobserver analysis. Agreement between the investigator and the non-subject observer was 97% for each type of scoring procedure.

CHAPTER 3

RESULTS

Results are presented in two sections which correspond with the two research questions posed for this study. The first section describes the effects of high quality audiotaped recordings versus high quality audio-videotaped recordings on the interobserver and intraobserver reliability calculated from whole word broad phonetic transcriptions. The second section describes the effect of speaker intelligibility on interobserver and intraobserver reliability of whole word broad phonetic transcriptions calculated from audiotaped and audio-videotaped recordings. Each section is divided into three subsections: interobserver reliability, intraobserver reliability, and a descriptive analysis of the data.

Audiotaped versus Audio-Videotaped Recordings

Interobserver Reliability

In order to determine the relative effects of audiotaped versus audio-videotaped recordings on the interobserver reliability calculated from whole word broad phonetic transcriptions, a 2x2x2 (time/intelligibility/media) analysis-of-variance (ANOVA) was utilized. Results are shown in Table 1 and indicate no main effect occurred for the recording media [$F(1,15)=.31$; NS] or time of presentation [$F(1,15)=1.40$; NS]. In addition, Table 1 shows there were no significant interactions between the recording media and time of transcription [$F(1,15)=.916$, NS] or between the recording media and the degree of intelligibility of the speaker [$F(1,15)=.05$; NS]. Thus, interobserver reliability calculated from whole word broad phonetic transcriptions was not significantly affected by the audiotaped versus audio-videotaped recordings regardless of the intelligibility of the speaker or elapsed time between transcriptions.

Point-by-point percentages of agreement were calculated on all transcribed sounds for all subjects at Time 1 (T1) and Time 2 (T2) in both recording conditions with the data collapsed across the intelligibility factor. Subject agreement on all transcribed sounds was 84% at T1,T2 for the audio-only condition and 83% at T1 and 84% at T2 for the audio-visual condition. Results, Figure 1, indicated

TABLE 1. Results of analysis-of-variance (ANOVA)
(time/intelligibility/media) for 16 subjects.

Sources of Variance	MS	df	F	p
Time	.000847	1	1.40	.25
Intelligibility	.182053	1	113.13	.00*
Time/Intelligibility	.000124	1	0.31	.59
Media	.000065	1	0.31	.59
Time/Media	.000261	1	0.92	.64
Intelligibility/Media	.000012	1	0.05	.82

*Significant at $p < .01$

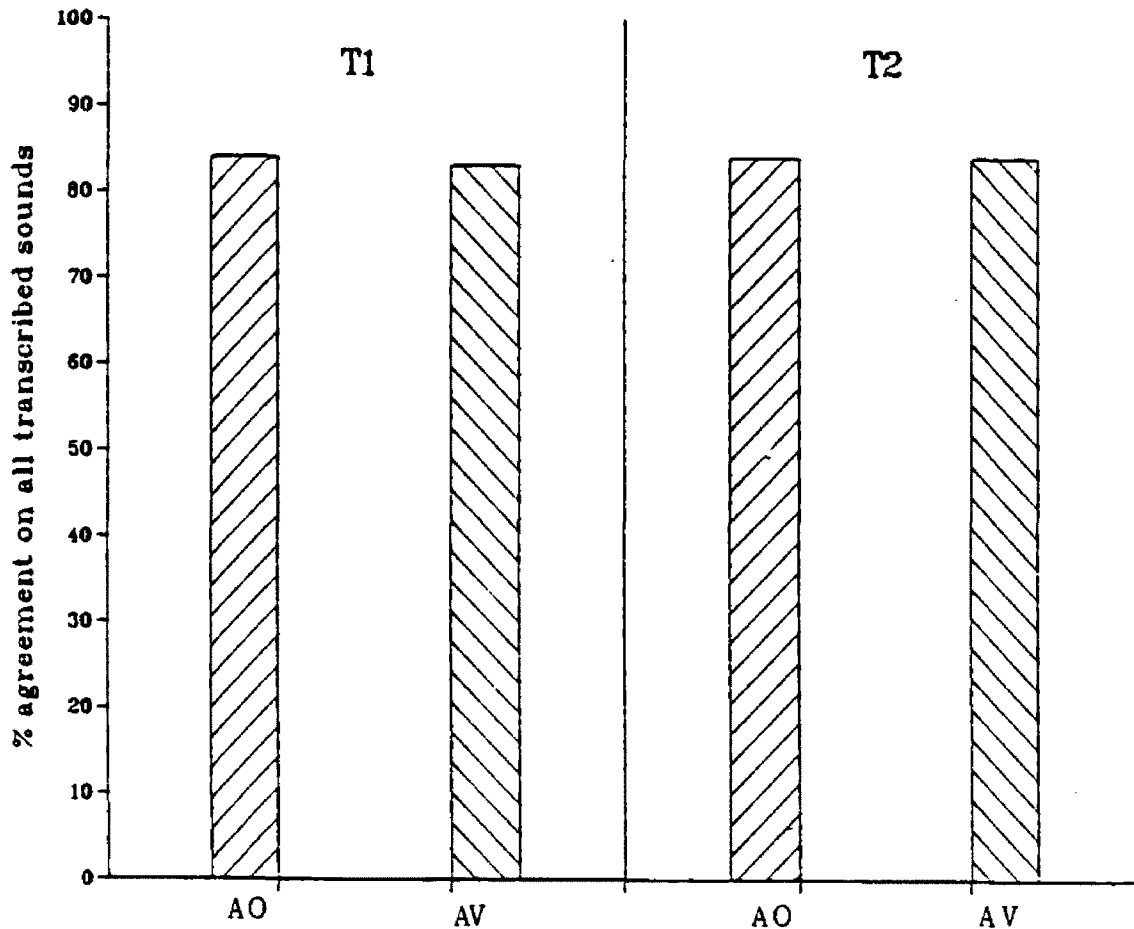


FIGURE 1. Point-by-point percentages of agreement among sixteen subjects on all transcribed sounds in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

point-by-point interobserver percentages of agreement calculated on whole word broad phonetic transcriptions were equivalent at both T1 and T2 for audiotaped and audio-videotaped recordings when the data were collapsed across the factor of intelligibility.

Intraobserver Reliability

In order to determine the effects of the two recording media on intraobserver reliability calculated from whole word broad phonetic transcriptions, point by point percentages of agreement with reference to the transcription scoring keys were calculated from all transcribed sounds for each subject in each recording condition at both transcription times with the data collapsed across the factor of intelligibility. Results, Table 2, showed intraobserver agreement ranged from 80% to 87% at T1 and from 79% to 89% at T2 in the audio-only condition and from 80% to 89% at T1 and from 80% to 91% at T2 in the audio-visual condition. Intraobserver variability between Time 1 and Time 2 transcriptions ranged from 0% to 3% (Mean=1%) in the audio-only condition and from 0% to 7% (Mean=2%) in the audio-visual condition. Results in Table 2 indicated subjects tended to be equally consistent on whole word broad phonetic transcriptions between T1 and T2, with reference to the scoring keys, in both recording conditions regardless of recording medium.

As Table 2 also indicated, the seven subjects with more clinical experience (indicated by "**") had intraobserver agreement percentages

TABLE 2. Intraobserver point-by-point percentages of agreement for each of 16 subjects in the audio-only and audio-visual recording conditions with percentages of difference between Time 1 (T1) and Time 2 (T2). Data are collapsed across the intelligibility factor.

*=speech/language pathologist +=student

Subject	Audio-only			Audio-Visual		
	T1 %	T2 %	%dif	T1 %	T2 %	%dif
1+	85	83	2	85	85	0
2*	84	82	2	88	81	7
3+	82	84	2	83	82	1
4+	82	84	2	81	84	3
5*	83	84	1	83	84	1
6+	87	89	2	89	91	2
7+	85	85	0	84	87	3
8+	82	83	1	83	82	1
9+	81	82	1	80	82	2
10+	80	80	0	81	82	1
11+	86	84	2	86	85	1
12*	81	82	1	80	80	0
13*	80	82	2	79	84	5
14*	87	87	0	83	87	4
15*	86	83	3	84	87	3
16*	84	84	0	84	84	0
Range=	80-87	79-89	0-3	80-89	80-91	0-7
Mean=	83	84	1	83	84	2

equivalent to those of the nine subjects with less clinical experience during both recording conditions. Variability between the two transcription times in the audio-only condition ranged from 0% to 3% (mean=1%) for the experienced subjects and from 0% to 2% (mean=1%) for the less experienced subjects. Variability between the two transcription times in the audio-visual condition ranged from 0% to 7% (mean=3%) for the experienced subjects and from 0% to 3% (mean=2%) for the less experienced subjects. Experienced subjects were slightly more variable in the audio-visual condition. These results indicated essentially no differences in intraobserver agreement on broad whole word phonetic transcriptions in both recording conditions regardless of the amount of clinical transcription experience.

Descriptive Analyses

Analyses of the data were performed in order to describe all targeted sounds transcribed by all subjects at Time 1 and Time 2 during the audio-only and audio-visual recording conditions. The data were collapsed across the intelligibility factor.

Descriptive analyses are reported in three sections. The first section reports the percentage of all target sounds transcribed incorrectly (sound errors=SEs). The second section describes consonants transcribed incorrectly (consonant errors=CEs) as a percentage of total SEs, as percentages of total CEs by place, manner, and voicing characteristics, and as percentages of total CEs which were

substitution, omission, or addition errors. The third section describes vowels transcribed incorrectly (vowel errors=VEs) as a percentage of total SEs, as percentages of total VEs according to the characteristics of tenseness/laxness, lip configuration, and tongue position (front, central, back, high, mid, low), and as percentages of total VEs which were substitution, omission, or addition errors. Each analysis was performed on both recording conditions, audio-only (AO) and audio-visual (AV), at both transcription times. As noted previously, consonant and vowel characteristics were analyzed relative to total transcription errors rather than to the frequency of occurrence of each characteristic on the scoring keys. Thus results should be interpreted accordingly.

Sound Errors (SEs)

The percentage of SEs which occurred in the AO condition was 16% at both Time 1 (T1) and Time 2 (T2). The percentages of SEs which occurred on all transcribed sounds in the AV condition were 17% at T1 and 16% at T2. Thus, the percentages of all sounds transcribed incorrectly were equivalent for the two recording conditions, audio-taped and audio-videotaped, at both transcription times.

Consonants

The percentages of total SEs which were transcription errors on consonants (CEs) in the AO condition were 49% at T1 and 48% at T2.

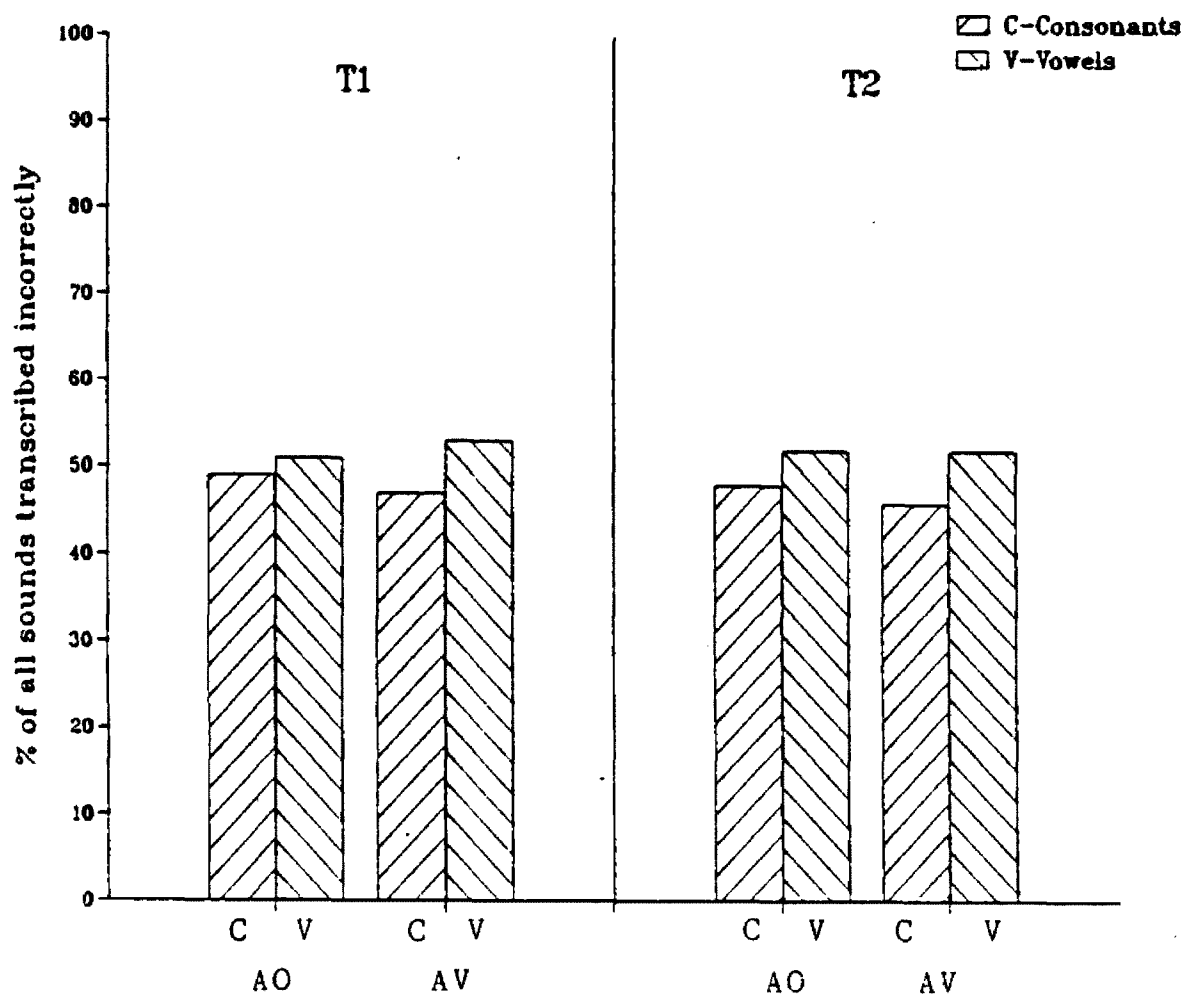


FIGURE 2. Percentages of all sounds transcribed incorrectly which occurred on consonants and vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

The percentages of total SEs which were CEs in the AV condition were 47% at T1 and 46% at T2. Results, Figure 2, indicated approximately half of all transcription errors were errors on consonants in both recording conditions at both transcription times.

Total CEs were analyzed according to place of production. Percentages of total CEs were calculated for the following places of articulation: bilabial, labiodental, linguadental, alveolar, palatal, labial/velar, velar, rhotic/palatal, and glottal (Figure 3). The percentages of CEs which occurred on alveolar sounds were 62% at T1 and 61% at T2 during the AO condition and 63% at T1, T2 during the AV condition. Transcription errors occurred on consonants in all other places of production with similar percentages between T1 and T2 less than or equal to 13% during both recording conditions. Thus, almost two-thirds of the CEs occurred on alveolar consonants during both recording conditions at both transcription times. The remainder of CEs during both recording conditions at Time 1 and Time 2 occurred in approximately equal percentages on consonants in the other eight places of production. Thus, essentially no differences existed between the audio-only and audio-visual recording conditions at both transcription times when place of production consonant transcription errors were analyzed.

Total CEs were also analyzed according to manner of production. Percentages of total CEs were calculated for the following manners of production: stop-plosives, fricatives, affricates, nasals, liquids,

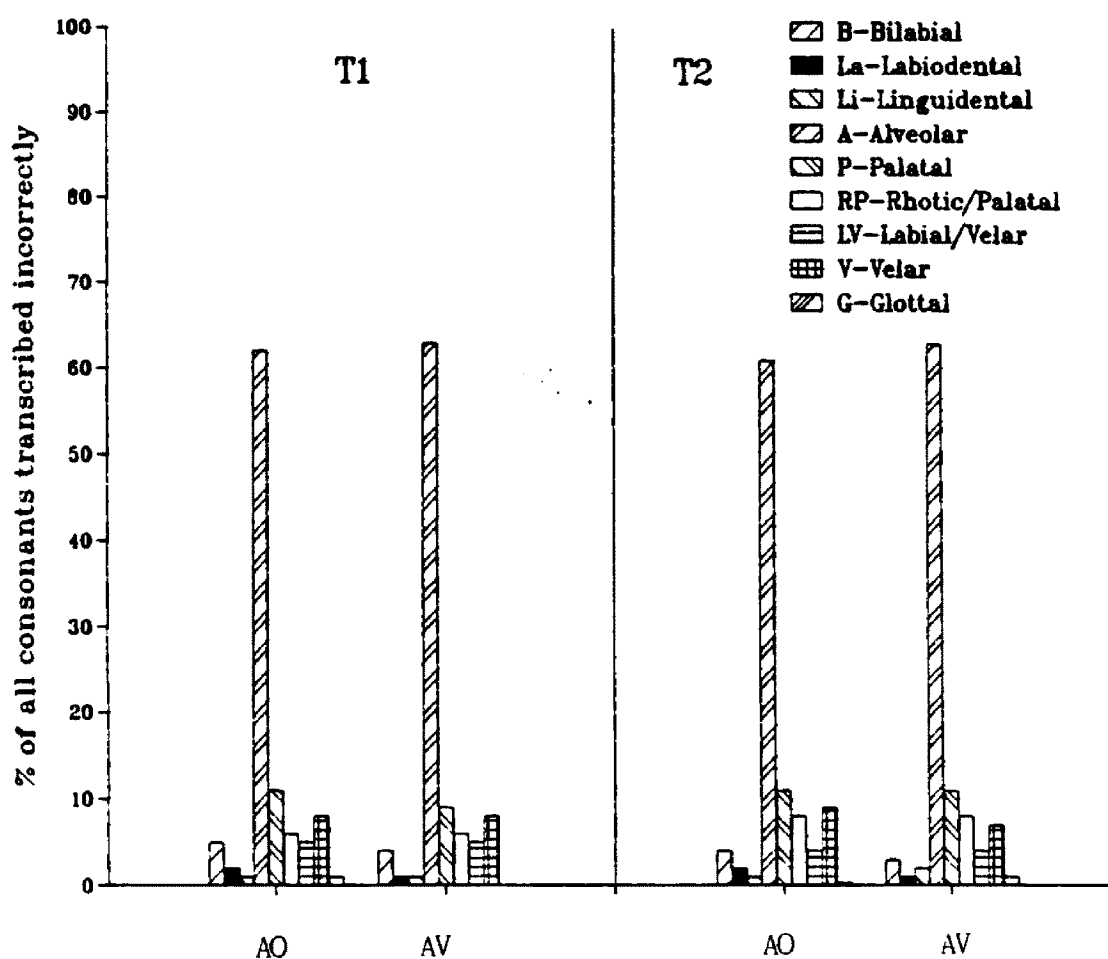


FIGURE 3. Percentages of all consonants transcribed incorrectly by place of production in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

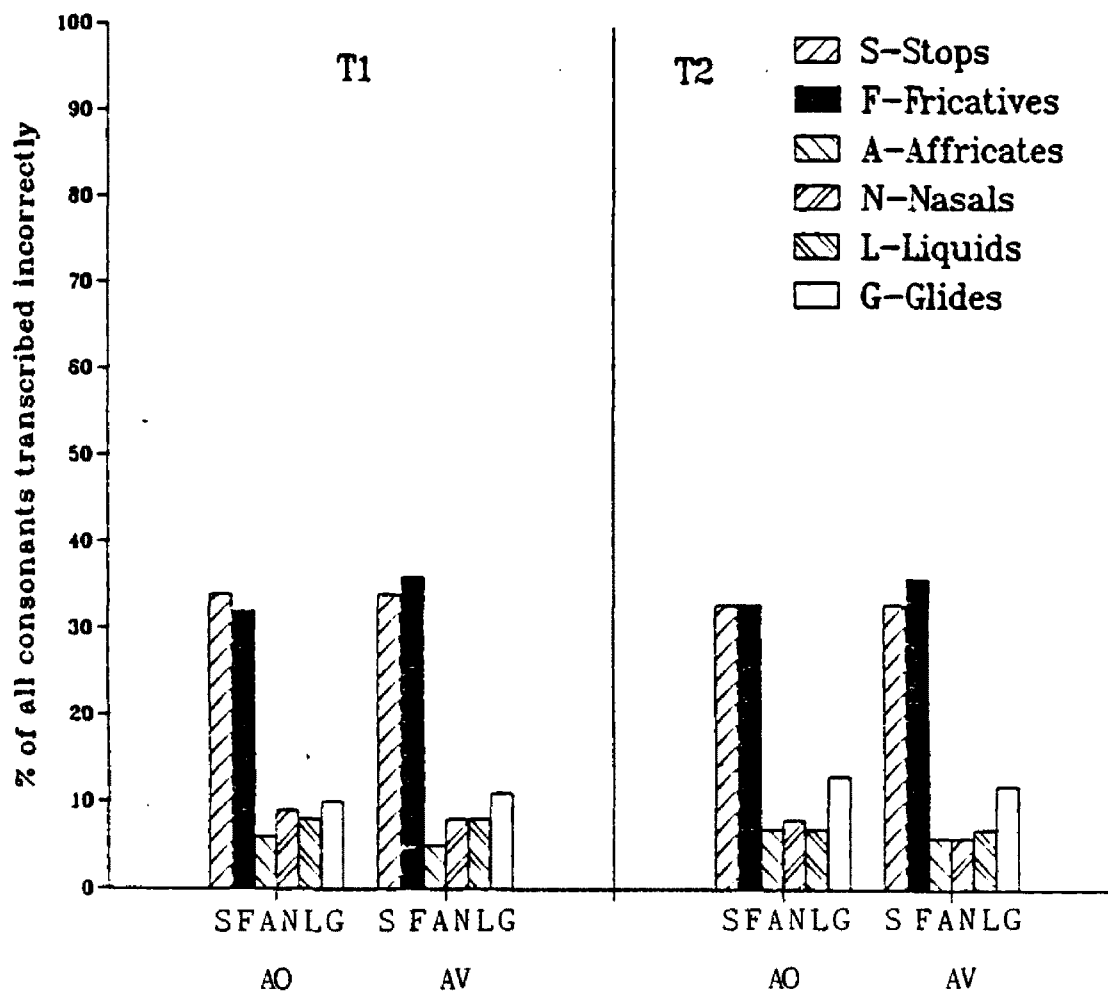


FIGURE 4. Percentages of all consonants transcribed incorrectly by manner of production in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

and glides (Figure 4). Errors occurred on stop-plosives with percentages of 34% at T1 and 33% at T2 in both recording conditions. Errors occurred on fricatives with percentages of 32% at T1 and 33% at T2 in the AO condition and 36% at T1, T2 in the AV condition. Errors occurred on affricates, nasals, liquids, and glides with equivalent percentages between Time 1 and Time 2 less than or equal to 13% in the AO condition and less than or equal to 12% in the AV condition. These results indicated approximately one-third of all CEs occurred on stop-plosives and approximately one-third of all CEs occurred on fricatives during both recording conditions at both transcription times. The remaining CEs occurred with similar percentages on affricates, nasals, liquids, and glides in both recording conditions at both transcription times. Thus, essentially no differences existed between the audio-only and audio-visual recording conditions at both transcription times when manner of production consonant transcription errors were analyzed.

Total CEs were analyzed according to voiced and voiceless production (Figure 5). Transcription errors occurred on voiceless consonants with percentages of 42% at T1 and 43% at T2 in the AO condition and 43% at T1 and 44% at T2 in the AV condition. Transcription errors occurred on voiced consonants with percentages of 58% at T1 and 57% at T2 in the AO condition and 57% at T1 and 56% at T2 in the AV condition. Results indicated slightly more than one-half of the consonant transcription errors occurred on voiced consonants and slightly less than one-half of the CEs occurred on voiceless con-

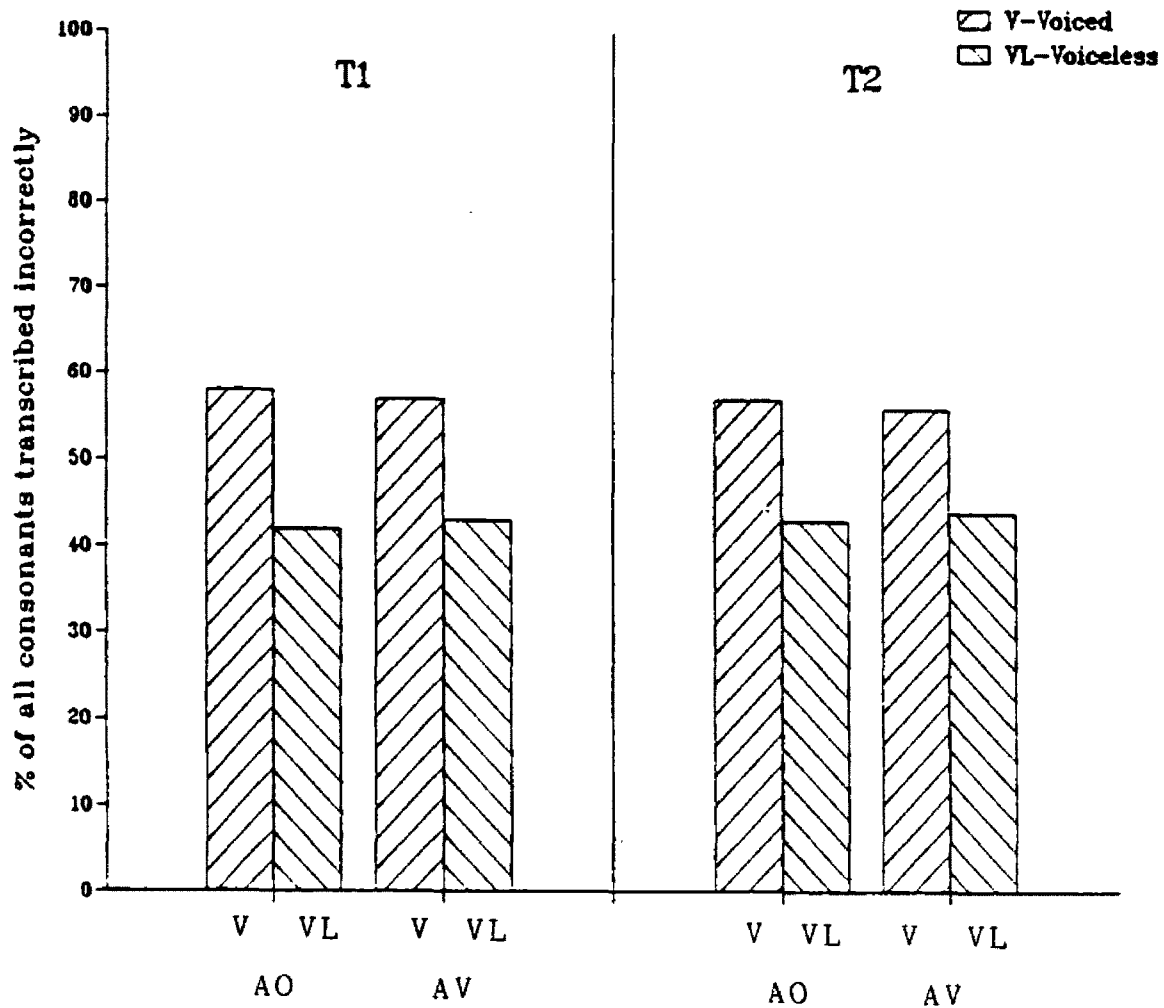


FIGURE 5. Percentages of all consonants transcribed incorrectly on voiced and voiceless consonants in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

sonants during both recording conditions at both transcription times. Thus, essentially no differences existed between the audio-only and the audio-visual recording conditions at both transcription times when voiced/voiceless consonant transcription errors were analyzed.

Total CEs were further analyzed to determine the percentages of occurrence of substitution, omission, and addition transcription errors in the audio-only and audio-visual recording conditions at both transcription times (Figure 6). The percentages of total CEs which were phoneme substitution errors were 60% at T1 and 59% at T2 in the AO condition and 59% at T1, T2 in the AV condition. The percentages of total CEs which were consonant omission errors were 20% at T1 and 21% at T2 in both recording conditions. The percentages of total CEs which were addition errors were 20% at T1, T2 in the AO condition and 21% at T1, T2 in the AV condition. Thus, during both recording conditions at both transcription times the majority of transcription errors were substitution errors. Approximately the same percentages of omission errors as addition errors were present in both recording conditions at both transcription times. Essentially no differences existed between the two recording conditions during both transcription times when transcription error types (substitutions, omissions, additions) were analyzed.

Total consonant transcription substitution errors, omission errors, and addition errors were each analyzed according to place and manner of articulation (Table 3). Results indicated the majority of

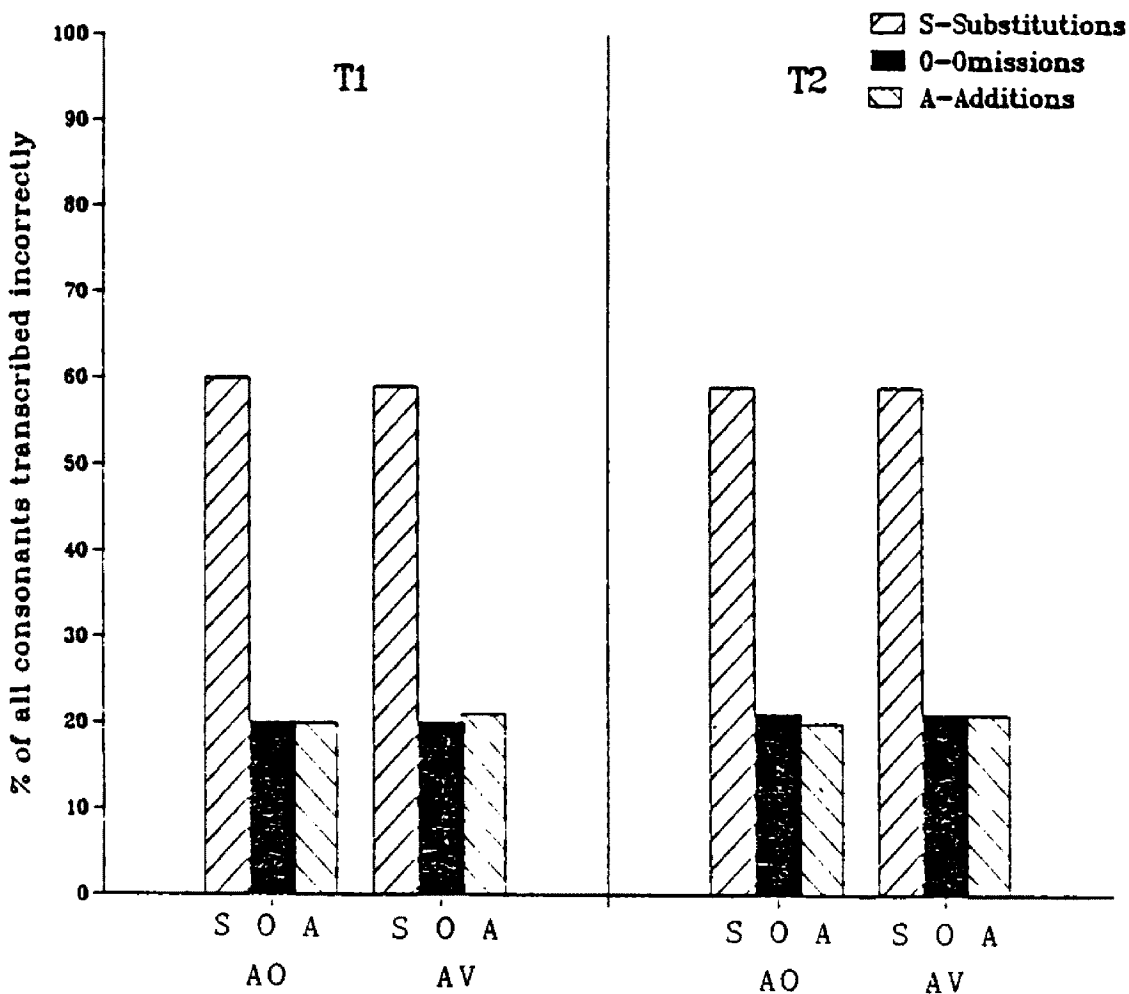


FIGURE 6. Percentages of all consonants transcribed incorrectly which were substitution, omission, or addition errors in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

TABLE 3. Percentages of all consonant substitution errors, omission errors, and addition errors each analyzed according to place and manner of articulation in the audio-only and audio-visual recording conditions at Time 1 and Time 2 with the data collapsed across the intelligibility factor.

	TIME 1						TIME 2					
	Audio-only			Audio-Visual			Audio-only			Audio-Visual		
PLACE	S	O	A	S	O	A	S	O	A	S	O	A
Bilabial	5	11	0	4	7	0	4	6	0	2	9	1
Labiodental	2	1	0	1	0	1	2	1	1	2	1	0
Linguadental	1	0	1	2	0	0	2	0	0	2	0	1
Alveolar	65	68	47	67	69	54	62	71	50	64	73	49
Palatal	13	7	10	13	6	4	14	6	9	14	7	8
Rhotic/Palatal	3	0	20	3	1	21	6	0	23	4	0	27
Labial/Velar	2	2	16	3	3	13	2	4	7	2	4	8
Velar	9	12	3	7	13	7	8	13	9	9	6	2
Glottal	0	0	3	0	0	0	0	0	2	0	0	4
MANNER	S	O	A	S	O	A	S	O	A	S	O	A
Stop-Plosives	18	74	42	17	72	46	17	69	44	18	69	52
Fricatives	49	6	9	53	5	15	49	7	13	54	7	13
Affricates	10	0	0	8	0	0	12	0	0	10	0	0
Nasals	13	9	0	10	9	0	9	10	1	8	10	2
Liquids	8	2	24	9	4	22	11	4	27	9	3	31
Glides	2	9	25	3	9	17	2	10	15	2	11	16

S=Substitution
O=Omission
A=Addition

transcription substitution errors occurred on alveolar fricative consonants in both recording conditions at both transcription times. The majority of transcription omission and addition errors occurred on alveolar stop-plosive consonants in both recording conditions at both transcription times. Thus, essentially no differences existed between the audio-only and audio-visual recording conditions when the place and manner characteristics of consonant error types were analyzed.

Consonant Summary

Results indicated approximately half of all sounds transcribed incorrectly were consonants during both the audio-only and audio-visual recording conditions at both transcription times. Consonants which were transcribed incorrectly (CEs) with reference to the scoring key were analyzed according to place, manner, and voicing characteristics and according to transcription error types (substitutions, omissions, additions) with the data collapsed across the factor of intelligibility. Descriptive analyses revealed voiced alveolar fricatives and stop-plosives were most frequently transcribed incorrectly at both transcription times during both recording conditions. The majority of transcription errors on consonants according to error type were substitutions of other sounds for alveolar fricatives during both recording conditions at both transcription times. Omission and addition errors occurred most frequently on alveolar stop-plosives during both recording conditions at both transcription times. Thus, essentially no differences were present between the audio-only and

audio-visual recording conditions at both transcription times when whole word broad phonetic transcriptions of consonants were analyzed.

Vowels

The percentages of all sounds transcribed incorrectly (SEs) which were transcription errors on vowels (VEs) in the AO condition were 51% at T1 and 52% at T2. The percentages of SEs which were VEs in the AV condition were 53% at T1 and 54% at T2. Results, Figure 2, indicated approximately half of all transcription errors were errors on vowels in both recording conditions at both transcription times.

The percentages of total VEs which were errors on diphthongs were 1% at T1 and 2% at T2 in the AO condition and 3% at T1 and 2% at T2 in the AV condition. Thus, there were essentially no differences between the audio-only and audio-visual recording conditions at both transcription times when diphthong transcription errors were analyzed. Diphthong errors were included in the total count of VEs during vowel characteristic analyses for this study. Therefore, vowel error percentages do not total 100% for the results reported in the remainder of this section.

Total VEs were analyzed according to the characteristics of tenseness and laxness of production. The percentages of VEs which occurred on tense vowels were 40% at T1 and 38% at T2 in the AO condition and 41% at T1 and 38% at T2 in the AV condition. The

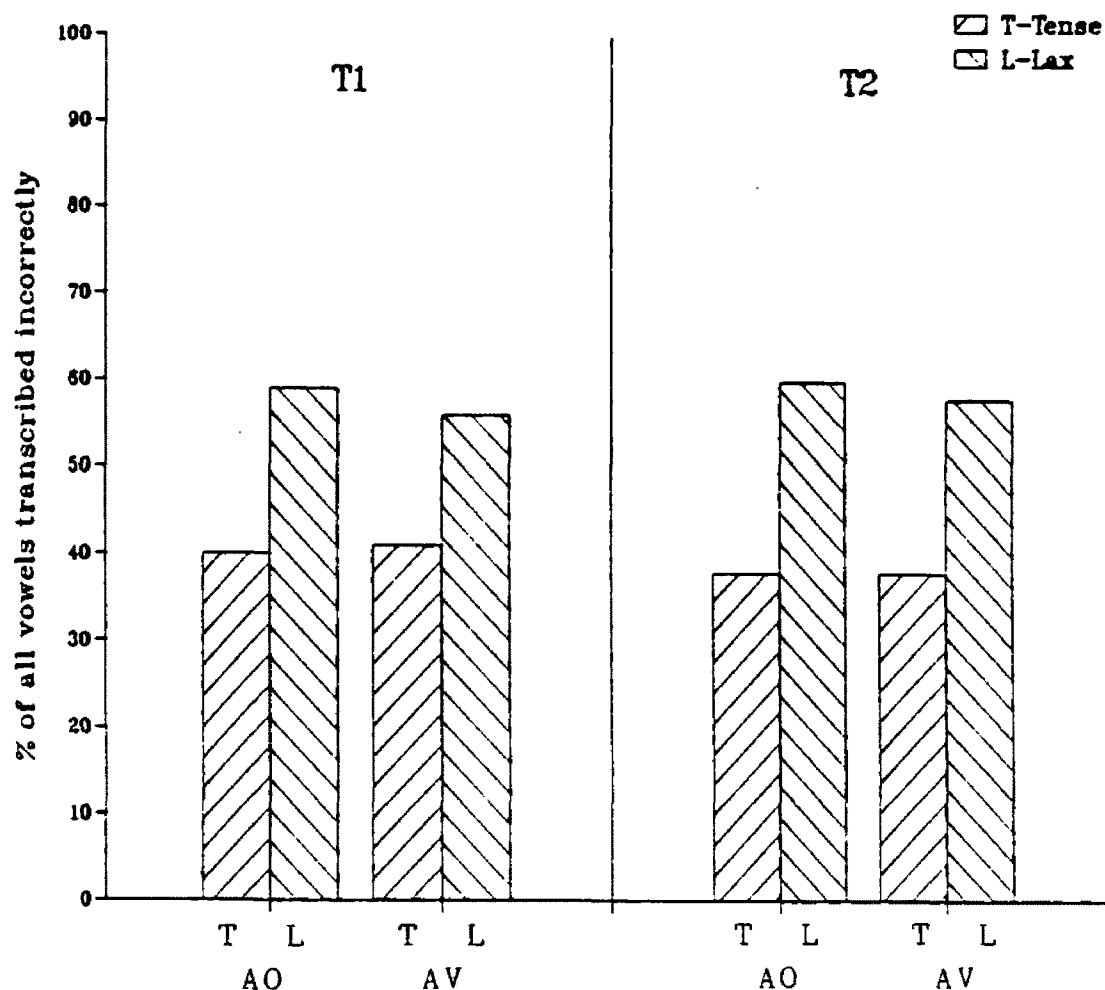


FIGURE 7. Percentages of all vowels transcribed incorrectly which occurred on tense and lax vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

percentages of VEs which occurred on lax vowels were 59% at T1 and 60% at T2 in the AO condition and 56% at T1 and 58% at T2 in the AV condition. Results, Figure 7, indicated more transcription errors occurred on lax vowels than on tense vowels during the two recording conditions at both transcription times. Thus, essentially no differences were present between the audio-only and audio-visual recording conditions at both transcription times when transcription errors on tense and lax vowels were analyzed.

Total VEs were also analyzed according to the characteristic of lip configuration (i.e. rounded and unrounded) during production. The percentages of VEs which occurred on rounded vowels were 28% at T1 and 26% at T2 in the AO condition and 27% at T1 and 24% at T2 in the AV condition. The percentages of VEs which occurred on unrounded vowels were 70% at T1 and 72% at T2 in the AO condition and 70% at T1 and 74% at T2 in the AV condition. Results, Figure 8, indicated almost three-fourths of all VEs occurred on unrounded vowels during both recording conditions at both transcription times. Essentially no differences were present between the audio-only and audio-visual recording conditions at both transcription times when transcription errors on rounded and unrounded vowels were analyzed.

Total VEs were further analyzed according to tongue position (i.e. tongue advancement and tongue height) during production. The characteristic of tongue advancement (i.e. front, central back) was analyzed first. The percentages of total VEs which occurred on front

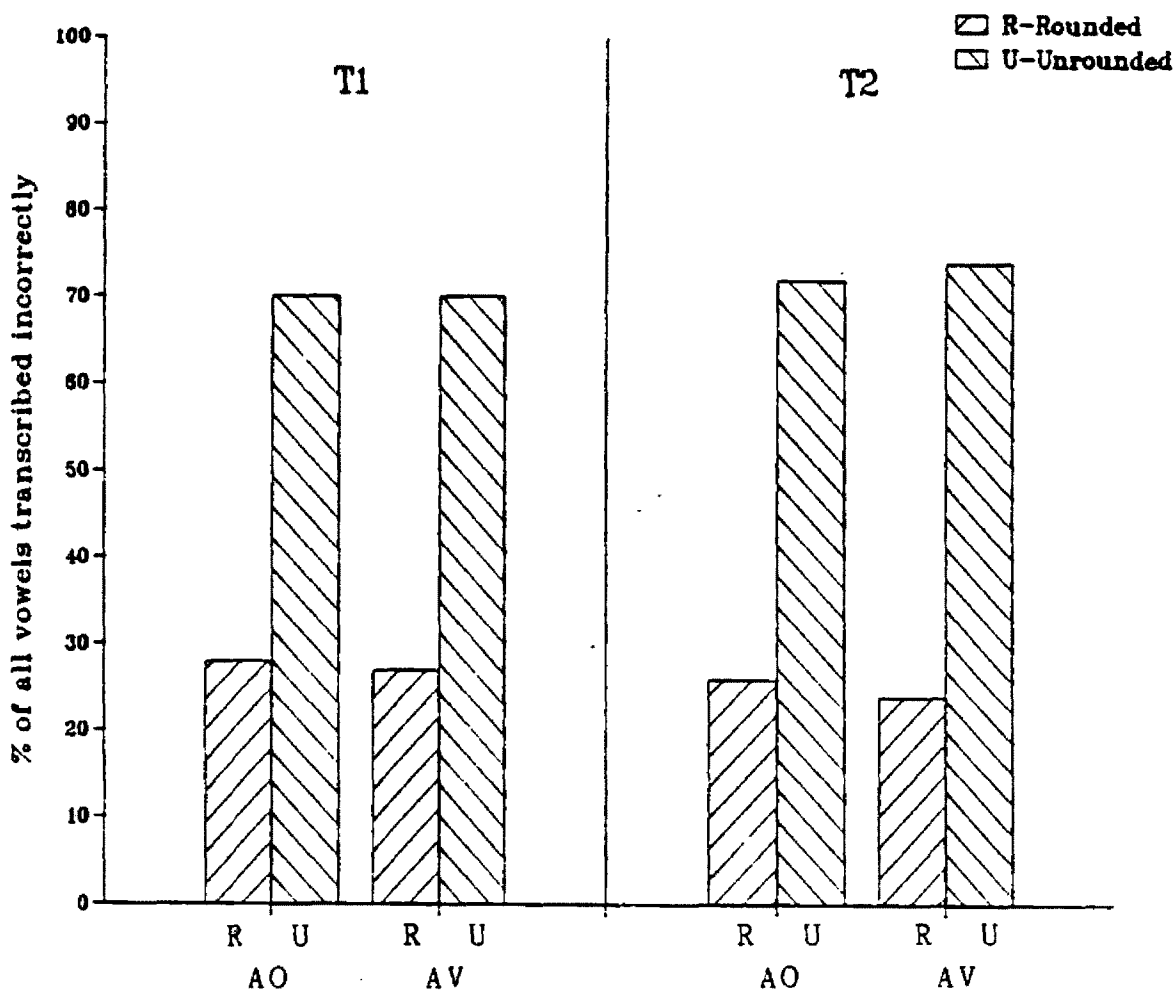


FIGURE 8. Percentages of all vowels transcribed incorrectly which occurred on rounded and unrounded vowels in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

vowels were 55% at T1, T2 in the AO condition and 52% at T1 and 55% at T2 in the AV condition. The percentages of VEs which occurred on central vowels were 19% at T1 and 21% at T2 in the AO condition and 21% at T1, T2 in the AV condition. The percentages of total VEs which occurred on back vowels were 25% at T1 and 22% at T2 in the AO condition and 24% at T1 and 21% at T2 in the AV condition. Results, Figure 9, indicated about half of the vowels transcribed incorrectly were front vowels during both recording conditions at both transcription times. Percentages of error were similar for central and back vowels for both recording conditions at both transcription times. Thus, essentially no differences existed between the audio-only and audio-visual recording conditions at both transcription times when vowel transcription errors were analyzed according to the degree of tongue advancement.

Total VEs were also analyzed according to tongue height (i.e. high, mid, low) during production. The percentages of VEs which occurred on high vowels were 40% at T1, T2 in the AO condition and 39% at T1 and 42% at T2 in the AV condition. The percentages of VEs which occurred on mid vowels were 51% at T1, T2 in the AO condition and 52% at T1 and 49% at T2 in the AV condition. The percentages of VEs which occurred on low vowels were 8% at T1, T2 in the AO condition and 7% at T1, T2 in the AV condition. Results, Figure 10, indicated the majority of VEs occurred on high and mid vowels during both recording conditions at both transcription times. Essentially no differences were present between the audio-only and audio-visual recording conditions

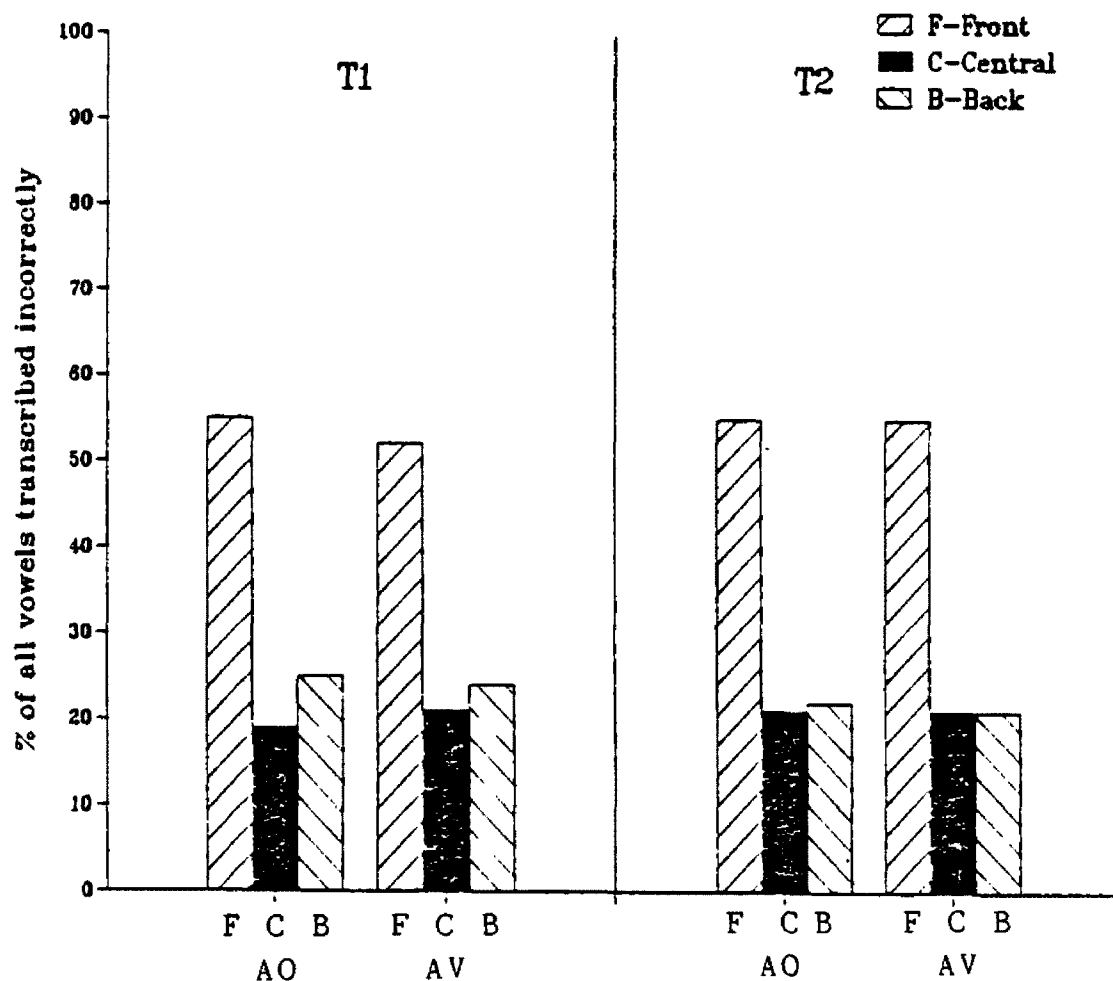


FIGURE 9. Percentages of all vowels transcribed incorrectly by degree of tongue advancement in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

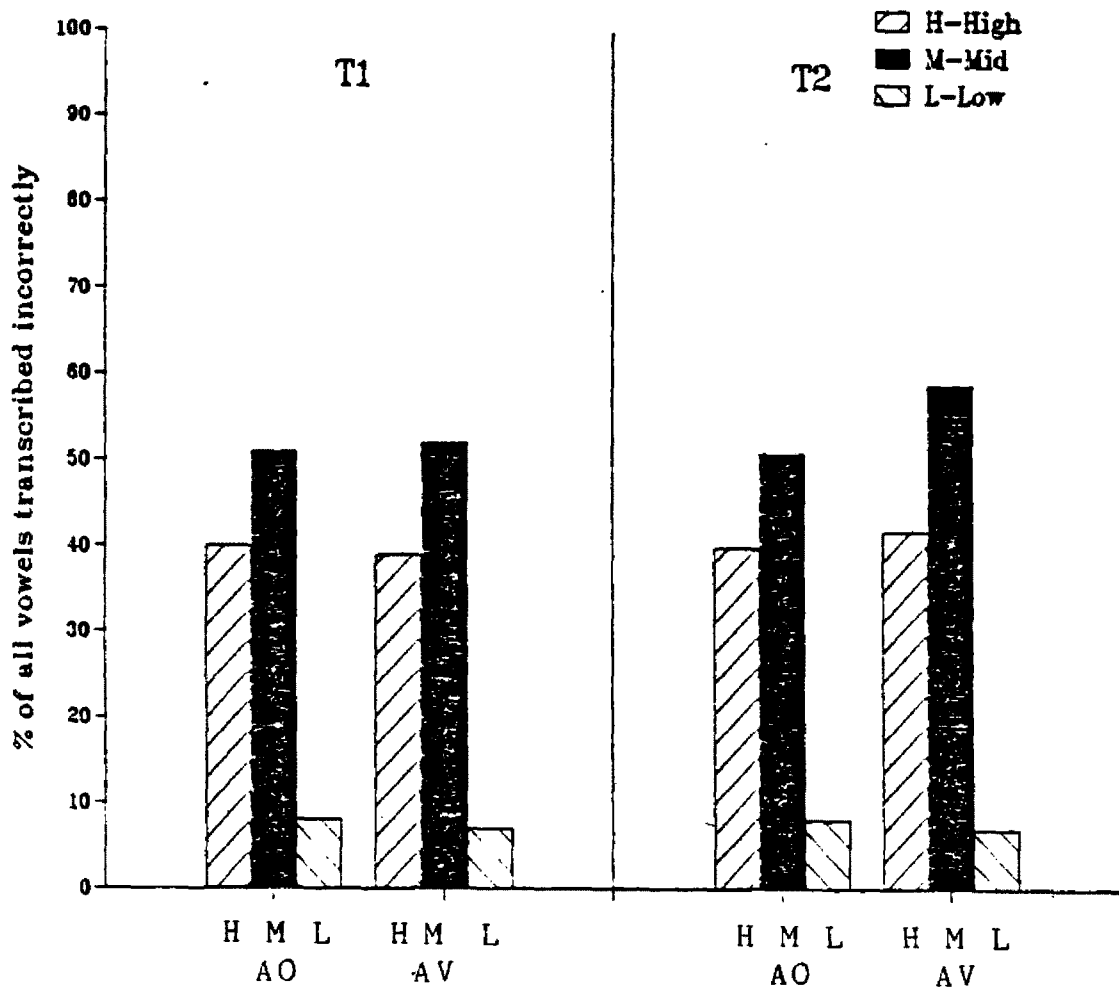


FIGURE 10. Percentages of all vowels transcribed incorrectly by degree of tongue height in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

at both transcription times when vowel transcription errors were analyzed according to tongue height.

Total VEs were analyzed to determine the percentages of occurrence of transcription substitution, omission, and addition errors in the two recording conditions at Time 1 and Time 2. The percentages of VEs which were phoneme substitution errors were 74% at T1, T2 in both the AO condition and the AV condition. The percentages of VEs which were vowel addition errors were 12% at T1, T2 in both recording conditions. The percentages of VEs which were vowel omission errors were 14% at T1 and 16% at T2 in the AO condition and 13% at T1, T2 in the AV condition. Results, Figure 11, indicated the majority of transcription errors on vowels in both recording conditions at both transcription times were transcription substitution errors. Approximately the same percentages of addition errors were present as omission errors during both recording conditions at both transcription times. Essentially no differences existed between the audio-only and audio-visual recording conditions at both transcription times when vowel transcription errors according to error types (substitutions, omissions, additions) were analyzed.

Total vowel transcription substitution errors, omission errors, and addition errors were each analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height. Results, Table 4, indicated the majority of transcription substitution errors occurred on unrounded high-front and mid-front vowels in both

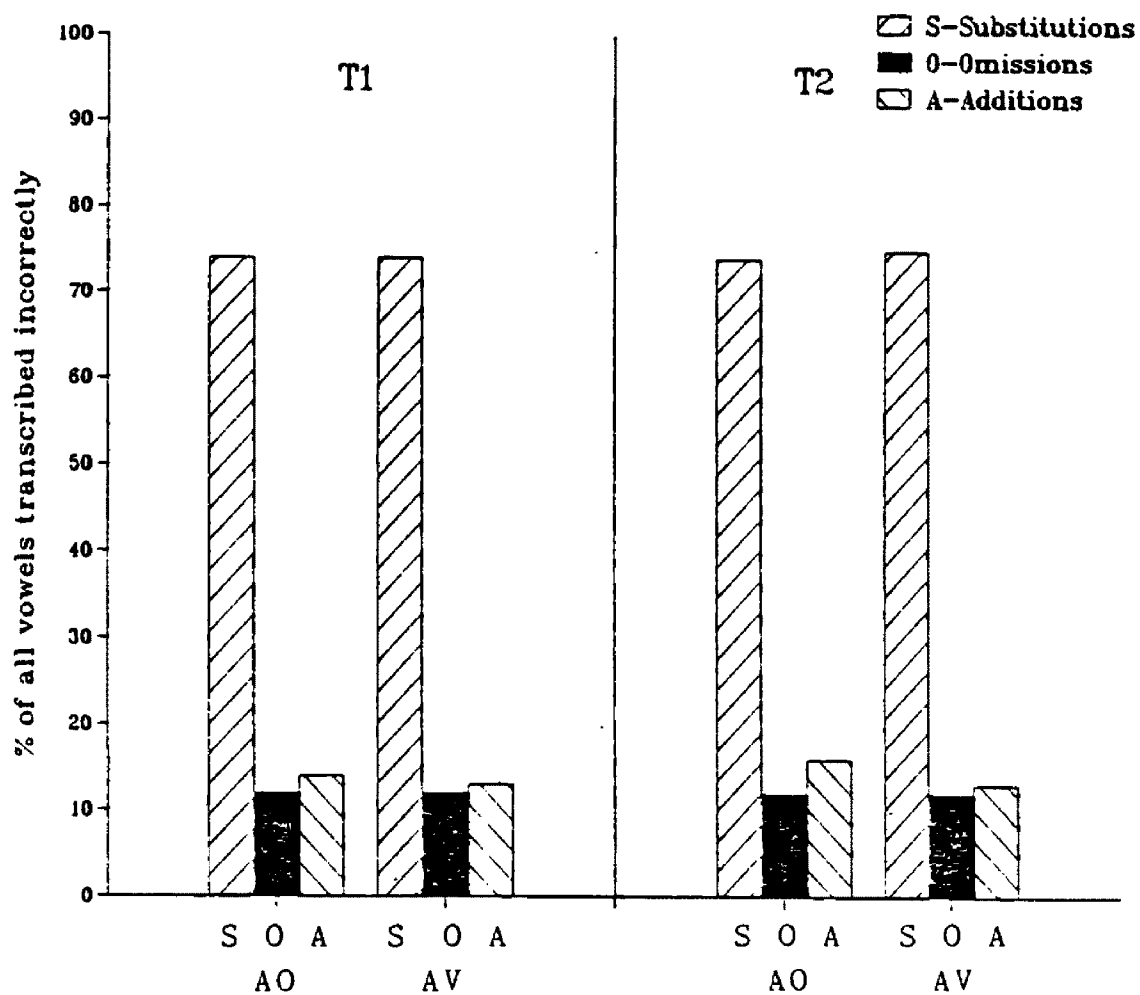


FIGURE 11. Percentages of all vowels transcribed incorrectly which were substitution, omission, or addition errors in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the intelligibility factor.

TABLE 4. Percentages of all vowel substitution errors, omission errors, and addition errors analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height in the audio-only (AO) and audio-visual (AV) recording conditions at Time 1 and Time 2 with the data collapsed across the intelligibility factor.

	TIME 1						TIME 2					
	AO			AV			AO			AV		
TENSENESS	S	O	A	S	O	A	S	O	A	S	O	A
Tense	47	17	21	48	21	21	45	21	14	45	20	14
Lax	51	81	77	49	79	78	53	67	84	52	80	85
LIP	S	O	A	S	O	A	S	O	A	S	O	A
Rounded	33	1	28	31	1	30	30	6	20	28	1	23
Unrounded	66	97	70	66	98	69	68	94	79	70	99	76
ADVANCEMENT	S	O	A	S	O	A	S	O	A	S	O	A
Front	57	82	19	53	79	24	58	81	23	58	79	21
Central	11	15	64	15	19	53	13	13	63	15	20	58
Back	30	1	16	28	3	22	27	4	12	25	1	19
HEIGHT	S	O	A	S	O	A	S	O	A	S	O	A
High	41	53	19	41	53	21	43	56	14	45	53	19
Mid	46	46	78	49	49	71	46	41	79	45	46	75
Low	11	0	2	8	1	7	9	3	6	8	1	5

S=Substitution
O=Omission
A=Addition

recording conditions at both transcription times. The percentage of substitution errors was equivalent between tense and lax vowels in both recording conditions at both transcription times. The majority of transcription omission errors occurred on unrounded lax high-front and mid-front vowels in both recording conditions at both transcription times. The majority of transcription addition errors occurred on unrounded lax mid-central vowels in both recording conditions at both transcription times. Thus, essentially no differences were present between the audio-only and audio-visual recording conditions at both transcription times when the characteristics of the transcription error types were analyzed.

Vowel Summary

Results indicated approximately one-half of all sounds transcribed incorrectly were vowels during both recording conditions at both transcription times. Vowels which were transcribed incorrectly, VEs, were analyzed according to the characteristics of tenseness of production, lip configuration, and tongue position, as well as error types (substitutions, omissions, and additions) with the data collapsed across the factor of intelligibility. Results revealed lax unrounded high-front and mid-front vowels were most frequently transcribed incorrectly. The majority of errors according to error type were phoneme substitutions in both recording conditions at both transcription times. Thus, essentially no differences were present

between the audio-only and audio-visual recording conditions when whole word broad phonetic transcriptions of vowels were analyzed.

Summary of Audiotaped versus Audio-Videotaped Recordings

Analysis-of-variance (2x2x2, time by intelligibility by media) results indicated no significant main effects involving the two recording conditions, audio-only and audio-visual. Further analyses were performed with the data collapsed across the factor of intelligibility. Results indicated essentially no differences between the two recording conditions at both transcription times when interobserver agreement, intraobserver agreement, and errors which occurred on vowels and consonants were analyzed. That is, subjects did not appear to perform whole word broad phonetic transcriptions better from audiotaped recordings than from audio-videotaped recordings.

The next section will describe the effect of speaker intelligibility on the interobserver and intraobserver reliability of whole word broad phonetic transcriptions. Descriptive results are also reported.

Speaker Intelligibility: Intelligible versus Unintelligible

Interobserver Reliability

In order to determine the relative effects of speaker intelligibility on the interobserver reliability calculated from broad whole word phonetic transcriptions, a 2x2x2 (time/intelligibility/-media) analysis-of-variance (ANOVA) was utilized. Results, Table 1, indicated a main effect was found for the factor of intelligibility [$F(1,15)=113.13$; $p<.01$]. As noted previously, there was no significant interaction between intelligibility and the recording media [$F(1,15)=.05$; NS]. In addition, there was no significant interaction between the time of transcription and speaker intelligibility [$F(1,15)=.31$; NS]. These results indicated a significant difference between the interobserver reliability percentages of agreement calculated from whole word broad phonetic transcriptions obtained from the intelligible and the unintelligible speaker.

Point-by-point percentages of agreement were calculated on all transcribed sounds for all subjects at Time 1 (T1) and Time 2 (T2) in both recording conditions with data collapsed across the media factor. Subject agreement on all transcribed sounds was 87% at T1 and 88% at T2 in the intelligible condition and 80% at T1, T2 in the unintelligible condition. Results, Figure 12, indicated point-by-point interobserver percentages of agreement calculated on whole word broad phonetic transcriptions were significantly higher for the intelligible

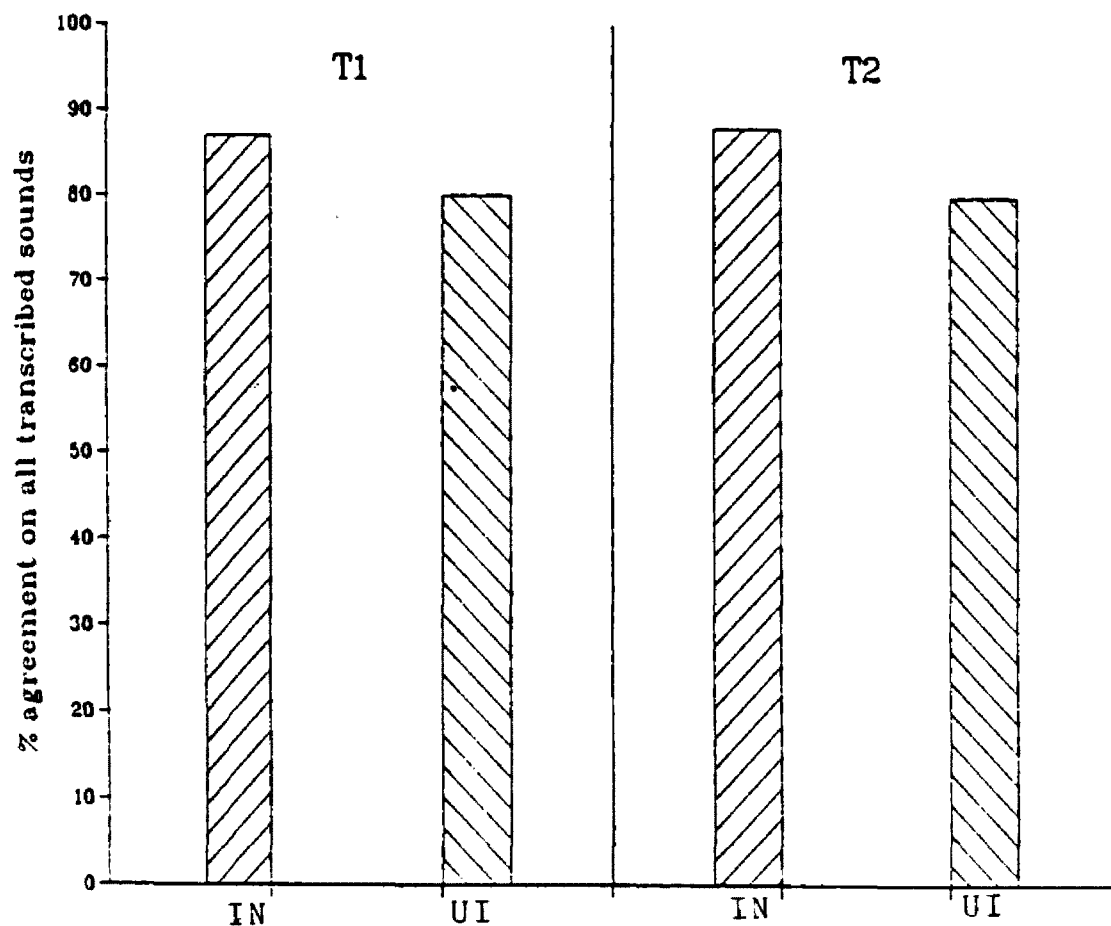


FIGURE 12. Point-by-point percentages of agreement among sixteen subjects on all transcribed sounds in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

condition than for the unintelligible condition when data were collapsed across the media factor. Results were approximately equivalent between the two transcription times, however, suggesting the subjects were consistent across time regardless of speaker intelligibility.

Intraobserver Reliability

In order to determine the effects of intelligibility on intraobserver reliability calculated from whole word broad phonetic transcriptions, point-by-point percentages of agreement with reference to the transcription scoring keys were calculated from all transcribed sounds for each subject at both transcription times with data collapsed across the media factor (Endnote 2). Results, Table 5, showed intraobserver agreement ranged from 80% to 90% at Time 1 and 85% to 93% at Time 2 in the intelligible condition and from 75% to 83% at Time 1 and 75% to 87% at Time 2 in the unintelligible condition. Intraobserver variability ranged from 0% to 5% (mean=2%) in the intelligible condition and from 0% to 6% (mean=2%) in the unintelligible condition. Results, shown in Table 5, indicated subjects tended to be consistent, with reference to the scoring keys, between Time 1 and Time 2 in both intelligibility conditions regardless of the degree of intelligibility.

As Table 5 also indicated, the seven subjects with more clinical experience (indicated by "**") had intraobserver agreement percentages

TABLE 5. Intraobserver point-by-point percentages of agreement for each of 16 subjects in the intelligible and unintelligible conditions with percentages of difference between Time 1 (T1) and Time 2 (T2). Data are collapsed across the media factor.

*=speech/language pathologist +=student

Subject	Intelligible			Unintelligible		
	T1 %	T2 %	%dif	T1 %	T2 %	%dif
1+	88	85	3	81	84	3
2*	89	86	3	83	77	6
3+	88	89	1	78	78	0
4+	87	87	0	77	80	3
5*	86	88	2	80	79	1
6+	90	93	3	86	87	1
7+	88	88	0	81	84	3
8+	87	39	2	79	76	3
9+	83	85	2	78	80	2
10+	87	87	0	75	75	0
11+	89	87	2	83	82	1
12*	87	86	1	75	77	2
13*	80	85	5	79	80	2
14*	89	91	2	81	83	2
15*	90	90	0	80	81	1
16*	89	88	1	79	81	1
Range=	80-90	85-93	0-5	75-83	75-87	0-6
Mean=	87	88	2	80	80	2

equivalent to the nine subjects with less clinical experience for both intelligibility conditions. Variability between the two transcription times in the intelligible condition ranged from 0% to 5% (mean=2%) for the experienced subjects and from 0% to 3% (mean=1%) for the less experienced subjects. Variability between the two transcription times in the unintelligible condition ranged from 1% to 6% (mean=2%) for the experienced subjects and from 0% to 3% (mean=2%) for the less experienced subjects. These results indicated essentially no differences in intraobserver agreement on whole word broad phonetic transcriptions regardless of the amount of clinical experience.

Descriptive Analyses

Analyses of the data were performed in order to describe all targeted sounds transcribed by all subjects at Time 1 and Time 2 during the intelligible (IN) and the unintelligible (UI) conditions. The data were collapsed across the media factor. Descriptive analyses were performed and are reported as described in Section 1 (Audiotaped versus Audio-videotaped) for total transcription errors on sounds (SEs), on consonants (CEs), and on vowels (VEs) except analyses were performed on both intelligibility conditions. As noted previously, consonant and vowel characteristics were analyzed relative to total transcription errors rather than to the frequency of each characteristic on the scoring keys. Thus, results should be interpreted accordingly.

Sound Errors

The percentages of sounds transcribed incorrectly (SEs) in the intelligible (IN) condition were 13% at Time 1 (T1) and 12% at Time 2 (T2). The percentages of SEs in the unintelligible (UI) condition were 20% at both Time 1 and Time 2. Results indicated more transcription errors occurred in the unintelligible condition than in the intelligible condition at both transcription times.

Consonants

The percentages of total SEs which were errors on consonants (CEs) were 52% at T1 and 49% at T2 in the intelligible (IN) condition and 45% at T1, T2 in the unintelligible (UI) condition. Results, Figure 13, indicated slightly more errors were made on consonants in the intelligible than in the unintelligible condition at both transcription times. Overall, however, approximately half of all transcription errors on sounds were CEs in both intelligibility conditions at both transcription times.

Total CEs were analyzed according to place of production. Percentages of total CEs were calculated for the following places of articulation: bilabial, labiodental, linguadental, alveolar, palatal, labial/velar, velar, rhotic/palatal, and glottal (Figure 14). The percentages of CEs which occurred on alveolar sounds were 65% at T1 and 63% at T2 in the IN condition and 61% at T1 and 62% at T2 in the

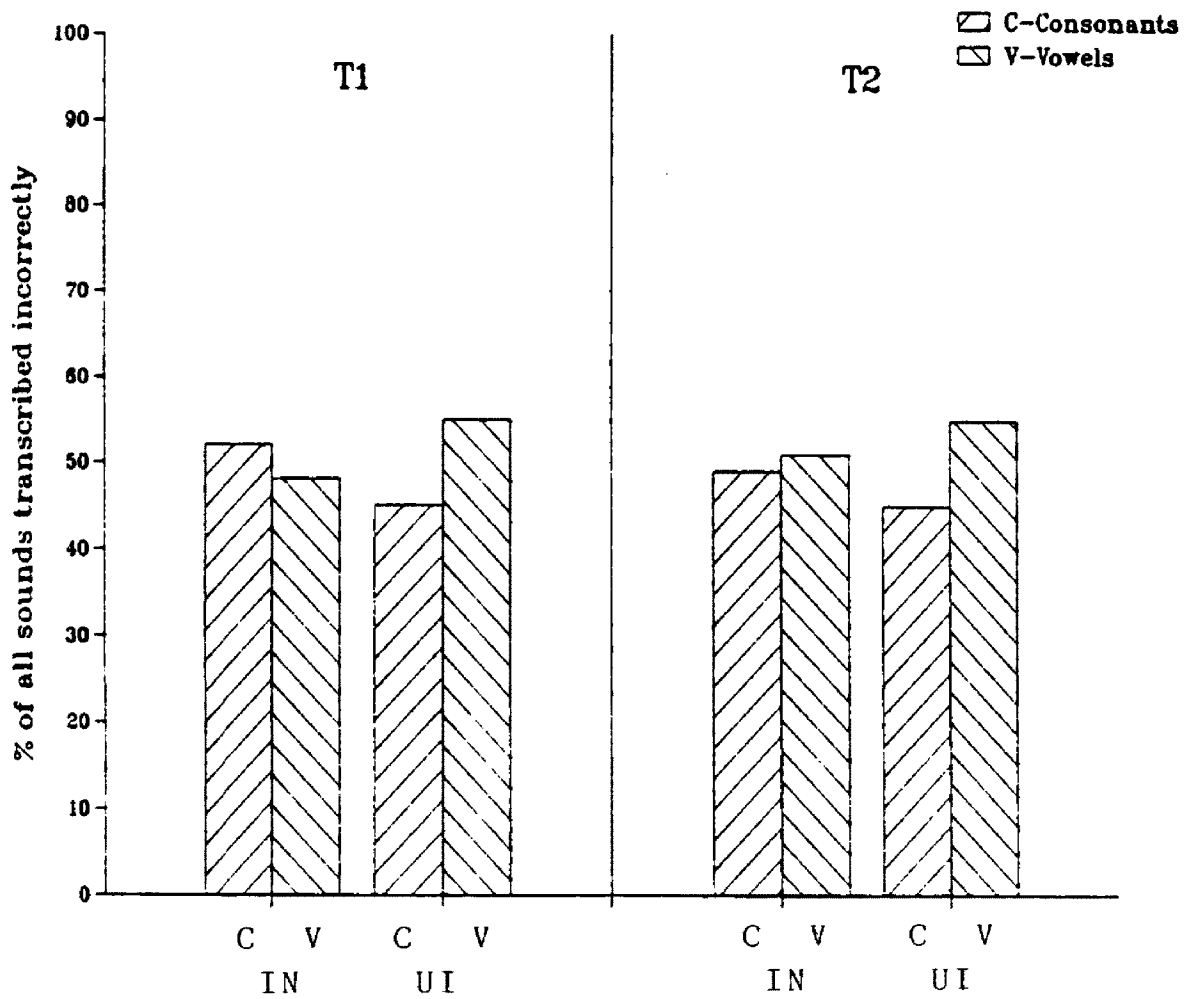


FIGURE 13. Percentages of all sounds transcribed incorrectly which occurred on consonants and vowels in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

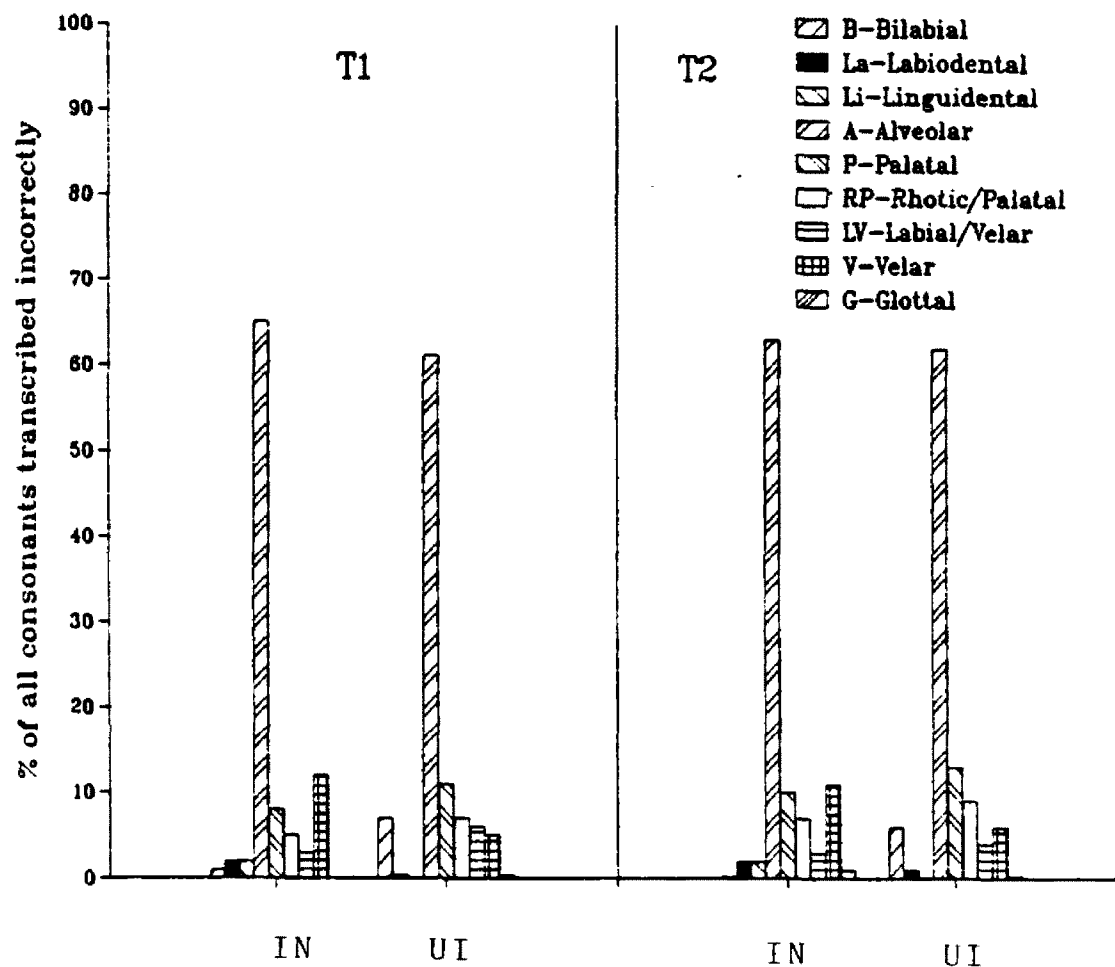


FIGURE 14. Percentages of all consonants transcribed incorrectly by place of production in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

UI condition. Errors occurred on bilabials with percentages of less than or equal to 1% at T1, T2 in the IN condition and 7% at T1 and 6% at T2 in the UI condition. Errors occurred on velar consonants with percentages of 12% at T1 and 11% at T2 in the IN condition and 5% at T1 and 6% at T2 in the UI condition. Errors occurred on all other places of production with equivalent percentages between T1 and T2 of less than or equal to 13% in both intelligibility conditions. Thus, the majority of CEs occurred on alveolar consonants in both intelligibility conditions at both transcription times. Error percentages for most places of production were similar regardless of intelligibility condition except for velars, which were transcribed incorrectly more frequently in the IN condition at both transcription times, and bilabials, which were incorrectly transcribed more frequently in the UI condition at both transcription times. Results indicated slight percentage differences between the intelligibility conditions at both transcription times for two places of production. In general no differences occurred between the two intelligibility conditions at both transcription times when place of production consonant errors were analyzed.

Total CEs were also analyzed according to manner of production. Percentages of total CEs were calculated for the following manners of production: stop-plosives, fricatives, affricates, nasals, liquids, and glides (Figure 15). Errors occurred on stop-plosives with percentages of 32% at T1 and 30% at T2 in the IN condition and 36% at T1 and 35% at T2 in the UI condition. Errors occurred on fricatives

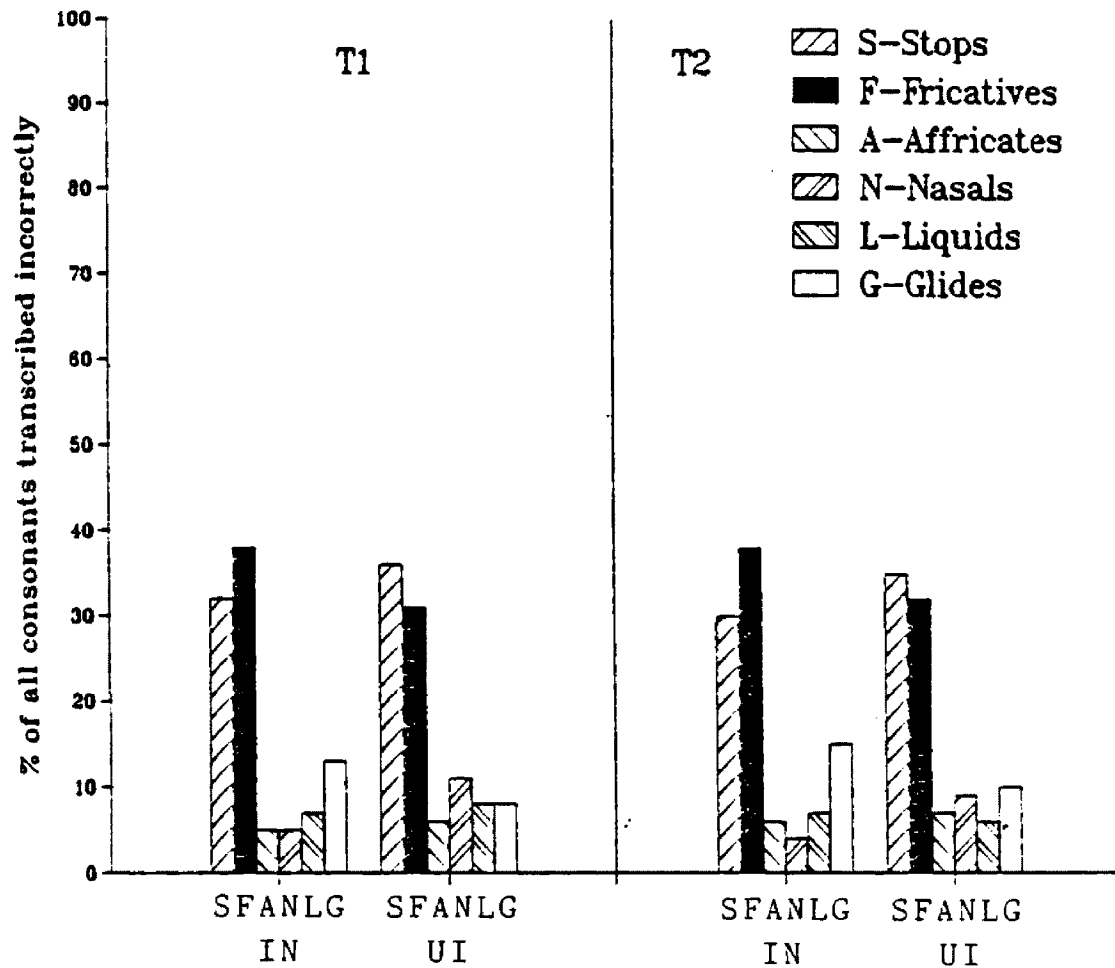


FIGURE 15. Percentages of all consonants transcribed incorrectly by manner of production in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

with percentages of 38% at T1, T2 in the IN condition and 31% at T1 and 32% at T2 in the UI condition. Errors occurred on nasals with percentages of 5% at T1 and 4% at T2 in the IN condition and 11% at T1 and 9% at T2 in the UI condition. Errors occurred on glides with percentages of 13% at T1 and 15% at T2 in the IN condition and 8% at T1 and 10% at T2 in the UI condition. Errors occurred on affricates and liquids with equivalent percentages between T1 and T2 less than or equal to 8% in both intelligibility conditions. These results indicated approximately one-third each of all CEs occurred on stop-plosives and fricatives in both intelligibility conditions at both transcription times. Slightly higher percentages of error occurred on stop-plosives and nasals in the unintelligible condition than in the intelligible condition at both transcription times. Slightly higher percentages of error occurred on fricatives and glides in the intelligible condition than in the unintelligible condition at both transcription times. Results indicated variations existed between the two intelligibility conditions at both transcription times when manner of production consonant transcription errors were analyzed.

Total CEs were also analyzed according to voiced and voiceless production (Figure 16). The percentages of transcription errors on voiceless consonants were 32% at T1, T2 in the IN condition and 51% at T1, T2 in the UI condition. The percentages of transcription errors on voiced consonants were 68% at T1, T2 in the IN condition and 49% at T1, T2 in the UI condition. Results indicated a higher percentage of consonant transcription errors occurred on voiced consonants in the IN

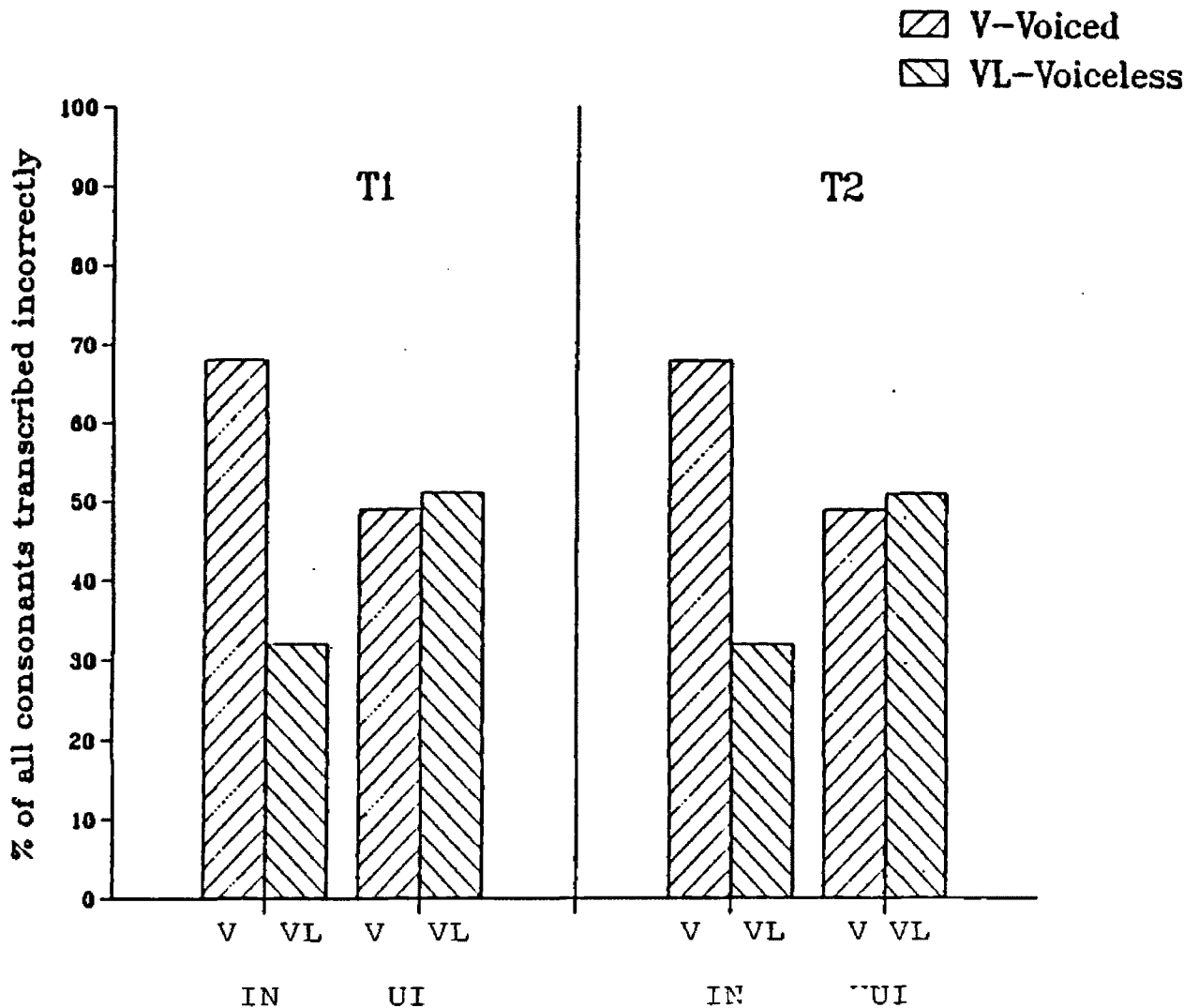


FIGURE 16. Percentages of all consonants transcribed incorrectly which occurred on voiced and voiceless consonants in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media condition.

condition than in the UI condition and a higher percentage of errors occurred on voiceless consonants in the UI condition than in the IN condition. Results indicated differences between the two intelligibility conditions when voiced/voiceless consonant transcription errors were analyzed.

Total CEs were further analyzed to determine the percentages of occurrence of substitution, omission, and addition errors in the IN and UI conditions at both Time 1 and Time 2 (Figure 17). The percentages of total CEs which were phoneme substitution errors were 60% at T1, T2 in the IN condition and 59% at T1, T2 in the UI condition. The percentages of total CEs which were consonant omission errors were 30% at T1 and 31% at T2 in the IN condition and 13% at T1 and 14% at T2 in the UI condition. The percentages of total CEs which were consonant addition errors were 10% at T1 and 9% at T2 in the IN condition and 28% at T1 and 27% at T2 in the UI condition. Thus, during both intelligibility conditions at both transcription times, the majority of transcription errors on consonants were substitution errors. A higher percentage of CEs were omission errors at both transcription times in the IN condition than in the UI condition. A higher percentage of CEs were addition errors at both transcription times in the UI condition than in the IN condition. Results indicated differences between the two intelligibility conditions which were equivalent between Time 1 and Time 2 when consonant transcription error types (substitutions, omissions, additions) were analyzed.

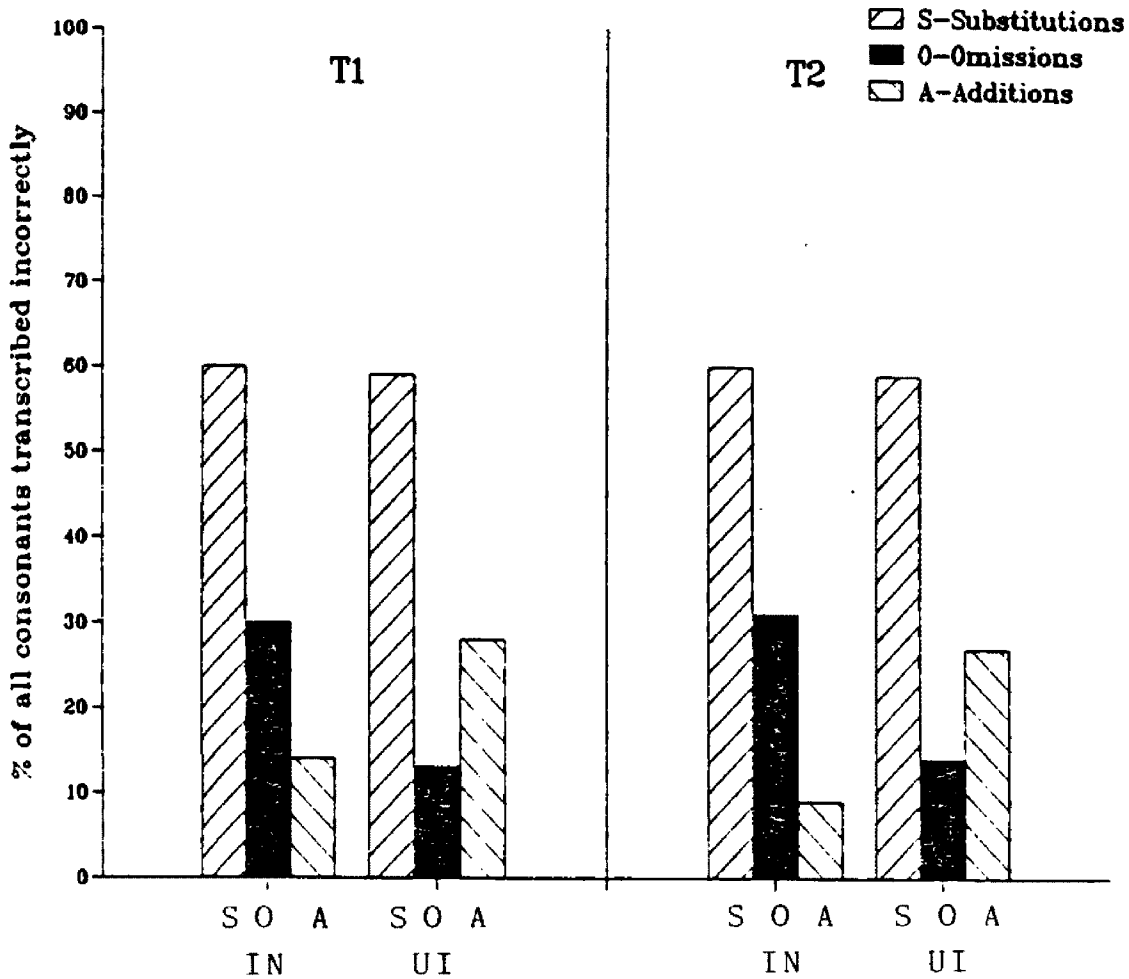


FIGURE 17. Percentages of all consonants transcribed incorrectly which were substitution, omission, or addition errors in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

Total consonant transcription substitution errors, omission errors, and addition errors were each analyzed according to place and manner of articulation (Table 6). Results indicated substitution errors occurred most frequently on alveolar fricative consonants in both intelligibility conditions at both transcription times. Omission and addition errors occurred most frequently on alveolar stop-plosive consonants in both intelligibility conditions at both transcription times. Omission errors occurred more frequently in the IN condition than in the UI condition. Addition errors tended to occur more frequently in the UI than in the IN condition. Results indicated although differences existed between the intelligibility conditions on the percentages of occurrence of error types, the place and manner characteristics of the error types were the same for the two intelligibility conditions at both transcription times.

Consonant Summary

Results indicated approximately half of all transcription errors on sounds (SEs) were errors on consonants for both intelligibility conditions at both transcription times. The overall percentage of occurrence of consonants transcribed incorrectly (CEs) was slightly higher for the intelligible condition than the unintelligible condition at both transcription times. CEs were analyzed according to place, manner, and voicing characteristics and according to error type (substitutions, omissions, additions) with data collapsed across the media factor. Descriptive analyses revealed alveolar stop-plosives

TABLE 6. Percentages of all consonant substitution errors, omission errors, and addition errors each analyzed according to place and manner of articulation in the intelligible and the unintelligible conditions at Time 1 and Time 2 with the data collapsed across the media factor.

	TIME 1						TIME 2					
	Intelligible			Unintelligible			Intelligible			Unintelligible		
PLACE	S	O	A	S	O	A	S	O	A	S	O	A
Bilabial	1	1	0	7	22	0	0	0	2	5	19	0
Labiodental	4	0	0	1	1	1	3	1	0	1	1	1
Linguadental	4	0	2	0	0	0	4	0	2	1	0	0
Alveolar	68	65	49	64	77	51	65	69	40	62	77	53
Palatal	8	10	4	16	0	8	9	11	9	18	0	8
Rhotic/Palatal	3	1	29	3	0	19	6	0	30	5	0	23
Labial/Velar	1	4	16	4	0	14	2	4	5	2	3	8
Velar	11	20	0	6	0	6	11	16	2	7	0	6
Glottal	0	0	0	0	0	2	0	0	9	0	0	1
MANNER	S	O	A	S	O	A	S	O	A	S	O	A
Stop-Plosives	11	73	38	22	73	45	9	68	35	23	71	42
Fricatives	60	2	10	45	12	13	59	4	16	46	12	12
Affricates	8	0	0	10	0	0	9	0	0	12	0	0
Nasals	5	5	0	16	15	0	3	8	0	12	13	2
Liquids	15	5	31	3	0	21	17	6	35	5	0	28
Glides	1	14	20	4	0	21	2	15	14	2	3	16

S=Substitution
O=Omission
A=Addition

and fricatives were most frequently transcribed incorrectly in both intelligibility conditions at both transcription times. Voiced consonants were most frequently transcribed incorrectly in the intelligible condition and voiceless consonants were most frequently transcribed incorrectly in the unintelligible condition at both transcription times. The majority of transcription errors on consonants according to error type were phoneme substitutions for alveolar fricatives during both intelligibility conditions at both transcription times. Omission and addition transcription errors occurred most frequently on alveolar stop-plosives during both intelligibility conditions at both transcription times. Results indicated percentage of occurrence differences between the two intelligibility conditions, particularly on voicing, manner of production, and error type. Essentially no differences occurred between the intelligibility conditions when place of articulation was analyzed although differences on velar and bilabial sounds existed. However, the relative frequency of occurrence of errors analyzed according to place, manner, and error type (substitutions, omissions, additions) were similar between the two intelligibility conditions at both transcription times.

Vowels

The percentages of all sounds transcribed incorrectly (SEs) which were errors on vowels (VEs) in the intelligible (IN) condition were 48% at T1 and 51% at T2. The percentages of SEs which were VEs in the

unintelligible (UI) condition were 55% at T1,T2. Results, Figure 13, indicated slightly more transcription errors were made on vowels in the UI condition than in the IN condition at both transcription times. Overall, however, approximately half of all transcription errors on sounds were VEs in both intelligibility conditions at both transcription times.

The percentages of VEs which were transcription errors on diphthongs were less than 1% at T1,T2 in the IN condition and were 3% at T1,T2 in the UI condition. Thus, essentially no differences existed between the two intelligibility conditions when diphthong transcription errors were analyzed. Diphthong errors were included in the total count of VEs during vowel characteristic analyses for this study. Therefore, vowel error percentages do not total exactly 100% for the results reported in the remainder of this section.

Total VEs were analyzed according to the characteristics of tenseness and laxness of production. The percentages of VEs which occurred on tense vowels were 38% at T1,T2 in the IN condition and 42% at T1 and 38% at T2 in the UI condition. The percentages of VEs which occurred on lax vowels were 62% at T1,T2 in the IN condition and 55% at T1 and 59% at T2 in the UI condition. Results, Figure 18, indicated more transcription errors occurred on lax vowels than tense vowels during both intelligibility conditions at both transcription times. More transcription errors occurred on lax vowels in the unintelligible condition than in the intelligible condition at both

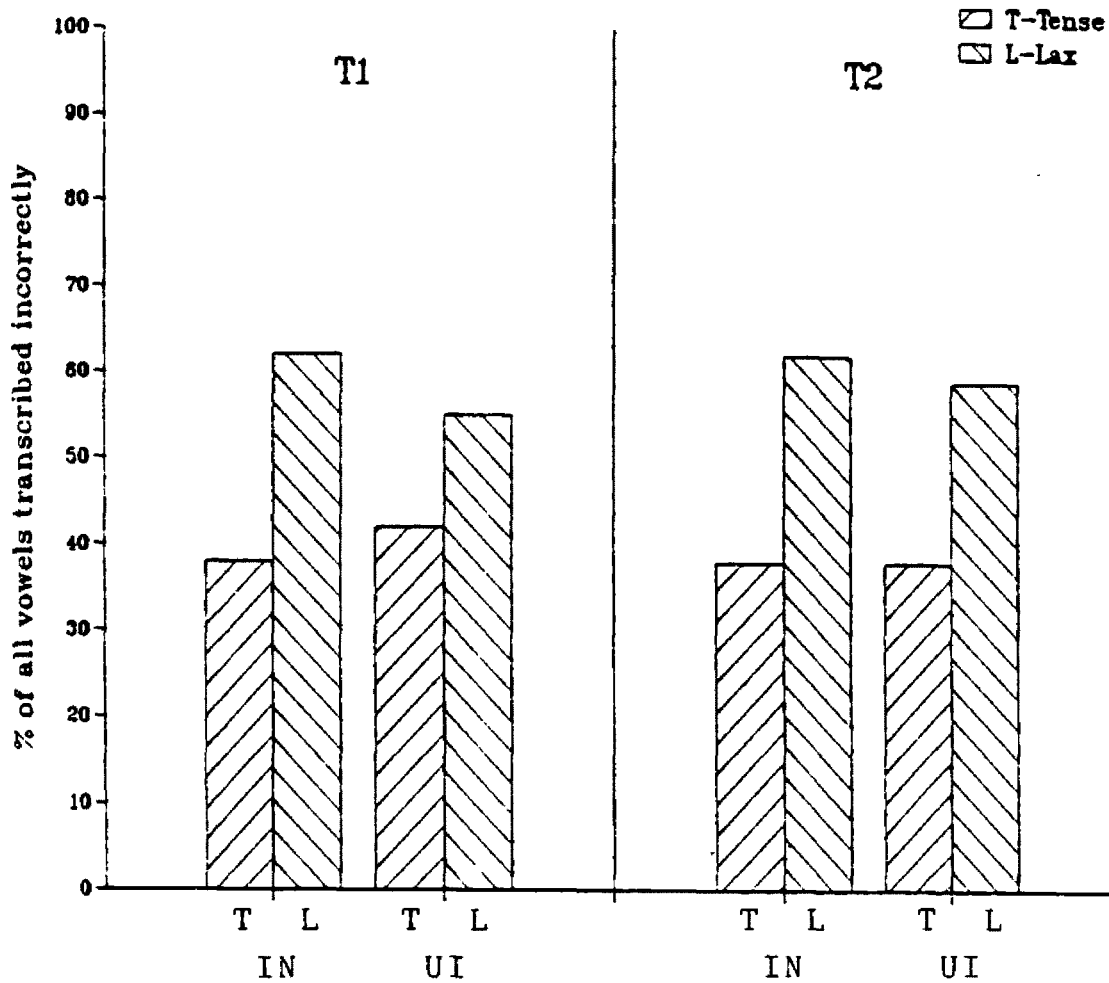


FIGURE 18. Percentages of all vowels transcribed incorrectly which occurred on tense and lax vowels in the intelligible (IN) and unintelligible (UI) recording conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

transcription times. Essentially no differences existed between the intelligible and unintelligible conditions when the relative frequency of occurrence of transcription errors on tense and lax vowels were analyzed.

Total VEs were also analyzed according to the characteristic of lip configuration (i.e. rounded and unrounded) during production. The percentages of VEs which occurred on rounded vowels were 24% at T1 and 22% at T2 in the IN condition and 30% at T1 and 27% at T2 in the UI condition. The percentages of VEs which occurred on unrounded vowels were 75% at T1 and 78% at T2 in the IN condition and 67% at T1 and 70% at T2 in the UI condition. Results, Figure 19, indicated more transcription errors occurred on unrounded vowels in the UI condition and more errors occurred on rounded vowels in the IN condition at both transcription times. Errors occurred more frequently on unrounded vowels in both intelligibility conditions at both transcription times. Results indicated percentage differences existed between the intelligible and unintelligible conditions at both transcription times when transcription errors on rounded and unrounded vowels were analyzed. However, the majority of errors occurred on unrounded vowels in both conditions at both transcription times.

Total VEs were further analyzed according to tongue position (i.e. tongue advancement and tongue height) during production. The characteristic of tongue advancement was analyzed first. The percentages of VEs which occurred on front vowels were 52% at T1 and 55% at T2 in the

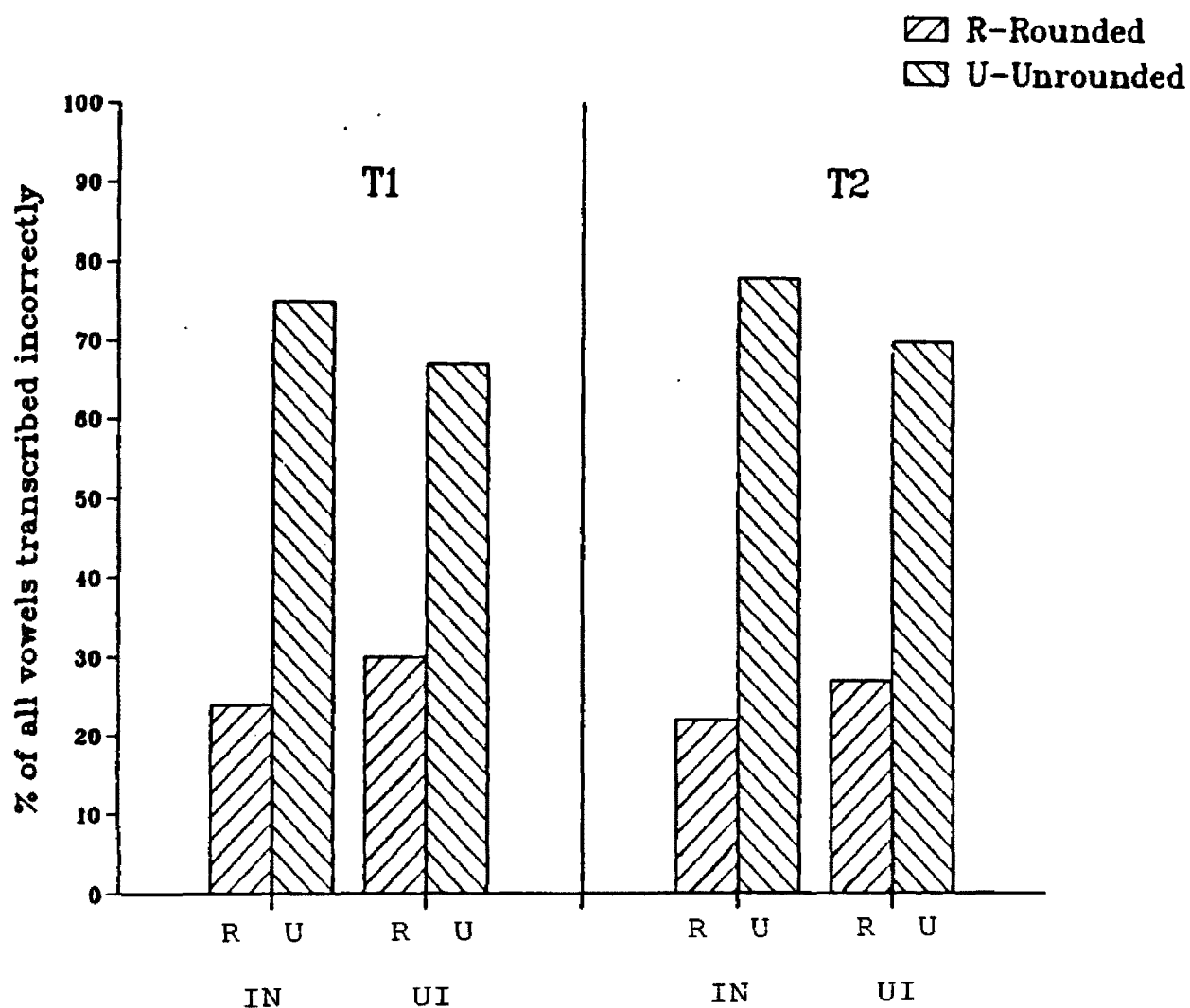


FIGURE 19. Percentages of all vowels transcribed incorrectly on rounded and unrounded vowels in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

IN condition and 54% at T1 and 55% at T2 in the UI condition. The percentages of VEs which occurred on central vowels were 25% at T1, T2 in the IN condition and 17% at T1 and 19% at T2 in the UI condition. The percentages of VEs which occurred on back vowels were 24% at T1 and 20% at T2 in the IN condition and 25% at T1 and 23% at T2 in the UI condition. Results, Figure 20, indicated more than half of the VEs analyzed according to tongue advancement were front vowels during both intelligibility conditions at both transcription times. Transcription errors occurred on front vowels slightly more in the IN condition than the UI condition at both transcription times. Results indicate slight differences occurred between the intelligible and unintelligible conditions at both transcription times when central vowels were analyzed. Essentially no differences existed between the two intelligibility conditions at both transcription times when front and back vowels were analyzed.

Total VEs were also analyzed according to tongue height (i.e. high, mid, low) during production. The percentages of VEs which occurred on high vowels were 36% at T1 and 40% at T2 in the IN condition and 41% at T1, T2 in the UI condition. The percentages of VEs which occurred on mid vowels were 62% at T1 and 58% at T2 in the IN condition and 46% at T1 and 45% at T2 in the UI condition. The percentages of VEs which occurred on low vowels were 2% at T1 and 3% at T2 in the IN condition and 11% at T1 and 10% at T2 in the UI condition. Results, Figure 21, indicated the majority of VEs analyzed according to tongue height occurred on high and mid vowels during both intelligibility

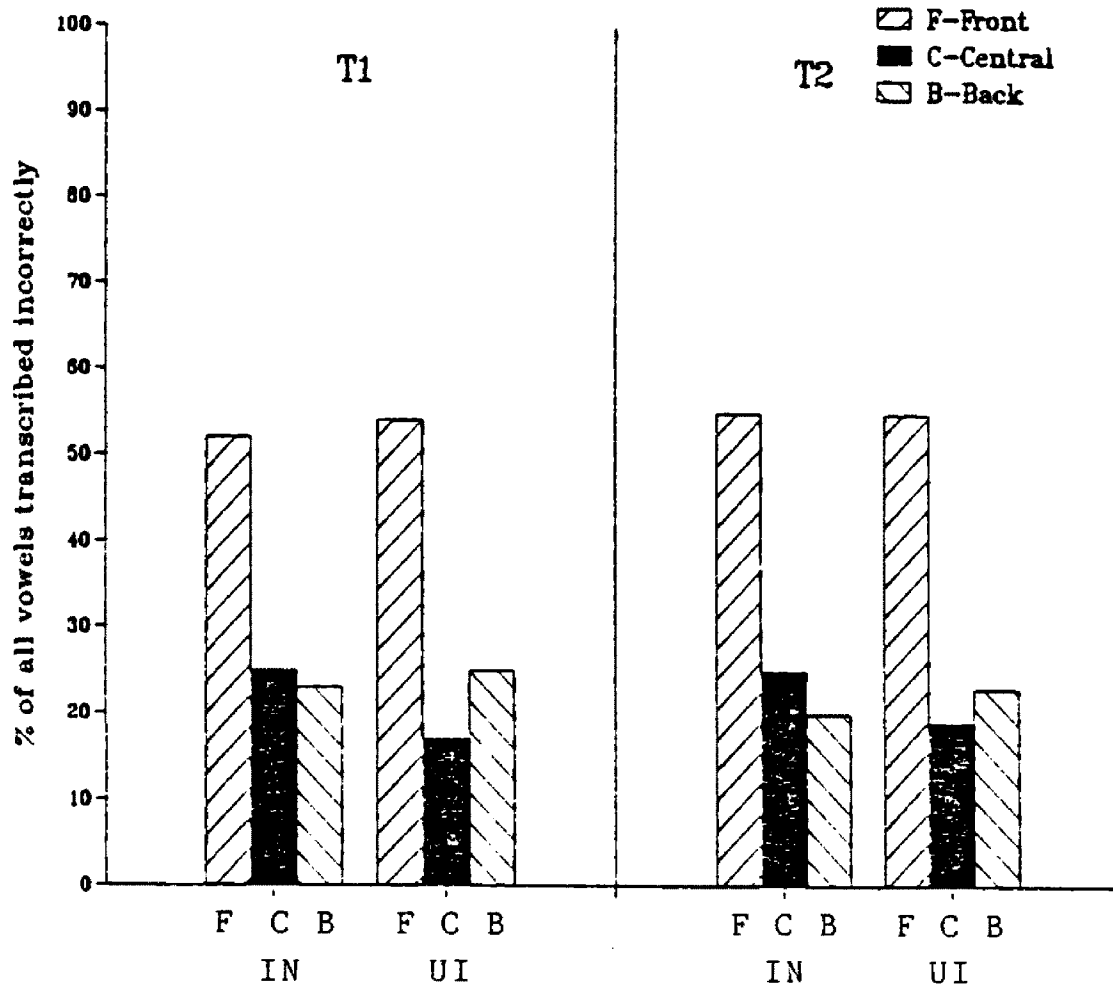


FIGURE 20. Percentages of all vowels transcribed incorrectly by degree of tongue advancement in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

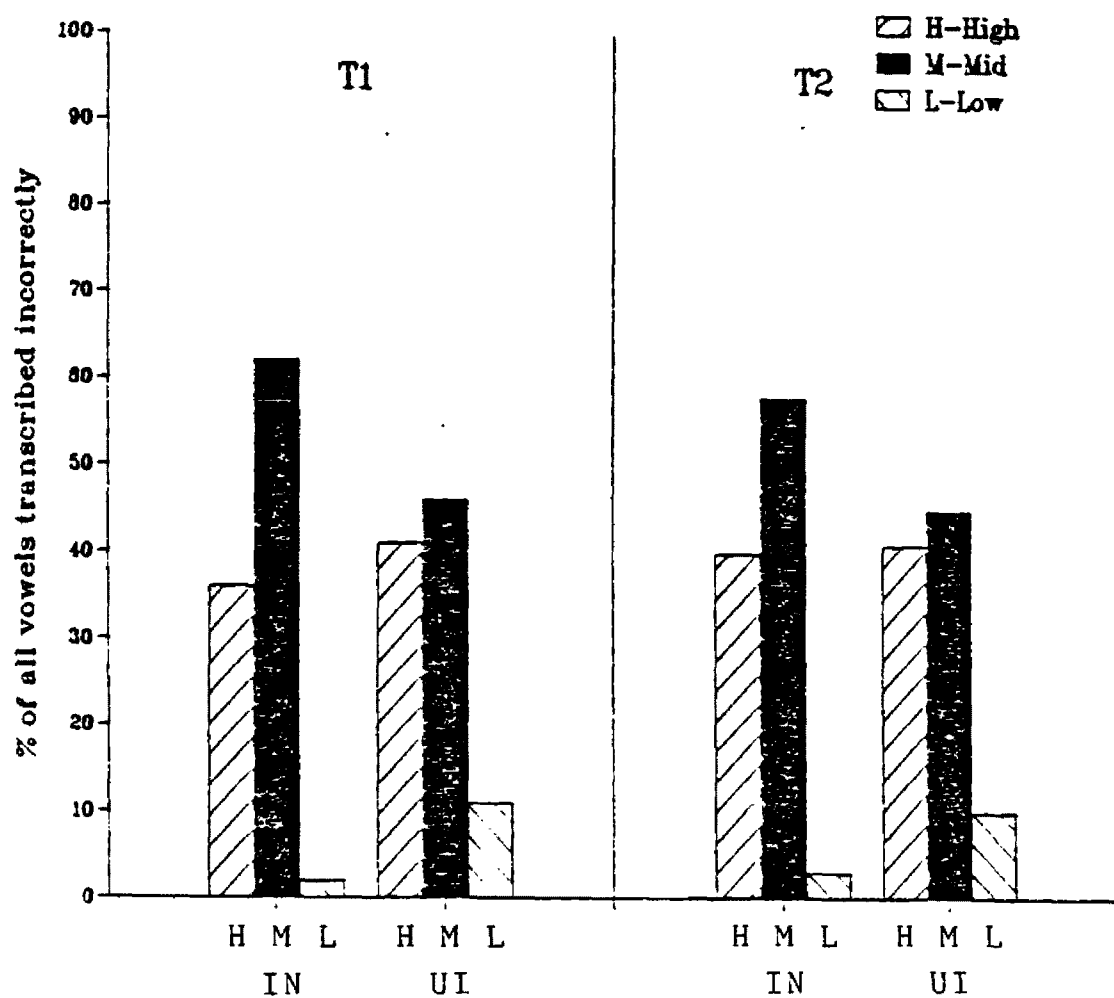


FIGURE 21. Percentages of all vowels transcribed incorrectly by degree of tongue height in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

conditions at both transcription times. Substantially more errors occurred on mid vowels in the intelligible condition than in the unintelligible condition during both transcription times. A slightly higher percentage of VEs occurred on low vowels in the unintelligible condition than in the intelligible condition at both transcription times. Results indicated slight differences between the two intelligibility conditions when high vowels were analyzed. Larger percentage differences existed between the intelligible and unintelligible conditions when mid and low vowels were analyzed. However, the relative frequency of occurrence of errors on high, mid, and low vowels was the same in both intelligibility conditions.

Total VEs were analyzed to determine the percentages of occurrence of substitution, omission, and addition errors in the intelligible and unintelligible conditions at Time 1 and Time 2. The percentages of VEs which were phoneme substitution errors were 72% at T1 and 75% at T2 in the IN condition and 75% at T1 and 74% at T2 in the UI condition. The percentages of VEs which were vowel omission errors were 10% at T1 and 8% at T2 in the IN condition and 13% at T1, T2 in the UI condition. The percentages of VEs which were vowel addition errors were 17% at T1, T2 in the IN condition and 12% at T1 and 13% at T2 in the UI condition. Results, Figure 22, indicated the majority of transcription errors on vowels in both intelligibility conditions at both transcription times were substitution errors. Slightly more addition errors were present in the IN condition than in the UI condition at both transcription times. However, the overall

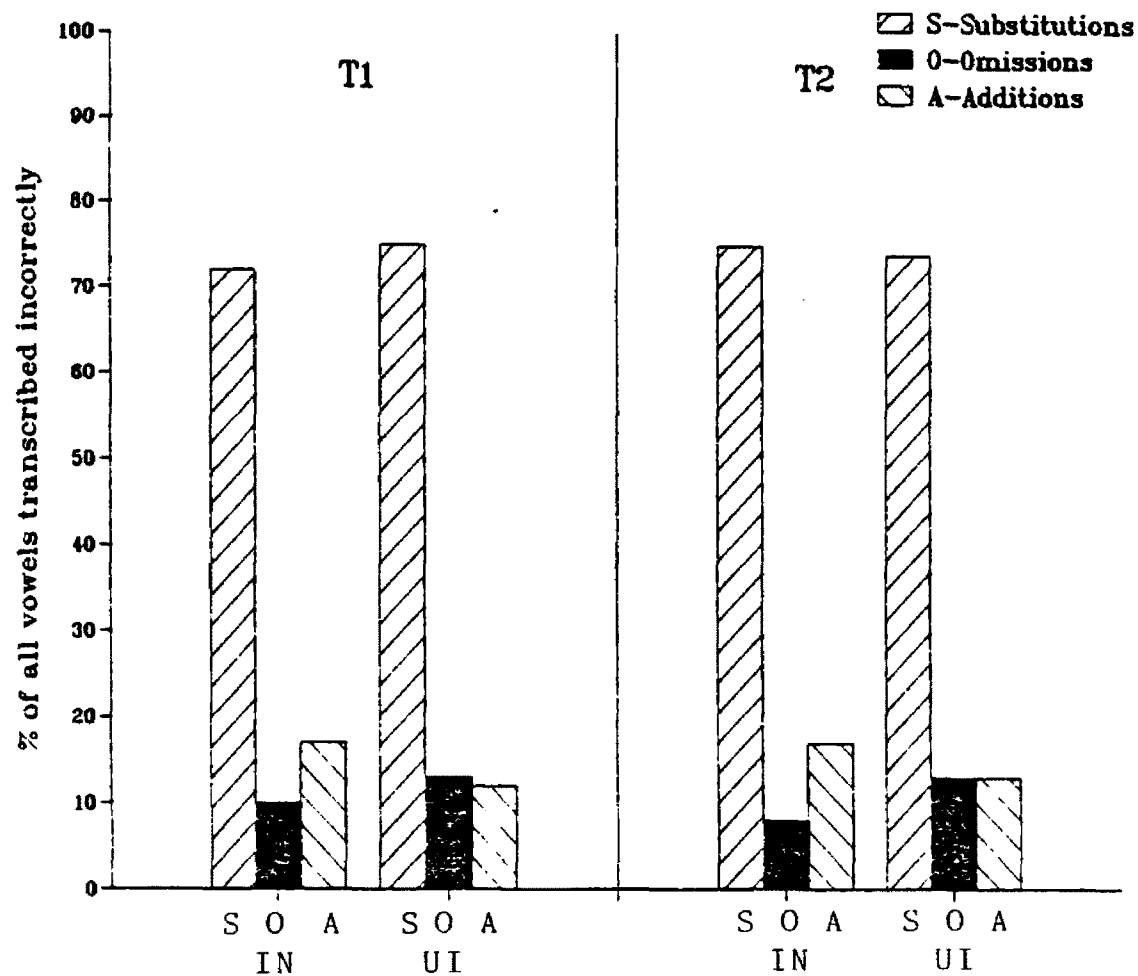


FIGURE 22. Percentages of all vowels transcribed incorrectly which were substitution, omission, or addition errors in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

percentages of occurrence of omission and addition errors were similar during both conditions at both transcription times. Results indicated essentially no differences between the intelligible and unintelligible conditions at both transcription times when vowel transcription errors were analyzed according to error type (substitutions, omissions, additions).

Total vowel transcription substitution errors, omission errors, and addition errors were each analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height. Results, Table 7, indicated the majority of substitution errors occurred on unrounded high-front and mid-front vowels in both intelligibility conditions at both transcription times. The percentage of substitution errors was equivalent between tense and lax vowels in both intelligibility conditions at both transcription times. The majority of omission errors occurred on unrounded lax mid-front and mid-central vowels in the intelligible condition and on unrounded lax high-front and mid-front vowels in the unintelligible condition. The majority of addition errors occurred on unrounded lax mid-central vowels in both intelligibility conditions at both transcription times. Thus, differences were present between the two intelligibility conditions on the vowel characteristic of tongue position when transcription omission errors were analyzed. Analysis of the types of transcription substitution and addition errors which occurred indicated essentially no differences between the two intelligibility conditions at both transcription times within each error type.

TABLE 7. Percentages of all vowel substitution errors, omission errors, and addition errors analyzed according to tenseness of production, lip configuration, tongue advancement, and tongue height in the intelligible (IN) and unintelligible (UI) conditions at Time 1 (T1) and Time 2 (T2) with the data collapsed across the media factor.

	TIME 1						TIME 2					
	IN			UI			IN			UI		
TENSENESS	S	O	A	S	O	A	S	O	A	S	O	A
Tense	49	4	11	47	26	29	49	10	5	43	26	21
Lax	50	96	89	50	72	69	51	72	95	53	74	77
LIP												
Rounded	29	4	16	33	3	39	27	8	9	30	2	31
Unrounded	71	96	83	63	98	59	73	92	91	66	98	67
ADVANCEMENT												
Front	61	46	20	52	96	23	65	36	23	53	96	22
Central	11	50	66	14	2	53	11	56	71	16	3	53
Back	28	4	14	30	1	23	24	8	6	27	1	22
HEIGHT												
High	41	38	14	40	58	25	46	38	11	42	60	20
Mid	56	60	83	43	42	68	51	59	85	52	38	71
Low	2	2	3	13	0	5	3	3	4	12	2	7

S=Substitution
O=Omission
A=Addition

Vowel Summary

Results indicated approximately half of all transcription errors on sounds (SEs) were errors on vowels for both intelligibility conditions at both transcription times. The overall percentage of vowels transcribed incorrectly (VEs) was slightly higher for the unintelligible condition than the intelligible condition at both transcription times. VEs were analyzed according to tenseness of production, lip rounding, and tongue position with data collapsed across the media factor. Descriptive analyses revealed unrounded lax high and mid-front vowels were most frequently transcribed incorrectly in both intelligibility conditions at both transcription times. Some variability related to percentage of occurrence tended to occur within each characteristic between the two intelligibility conditions. Higher percentages of transcription errors occurred on rounded and mid-central vowels in the intelligible condition than in the unintelligible condition. Higher percentages of transcription errors occurred on unrounded and low vowels in the unintelligible condition than in the intelligible condition. Despite percentage variability, the relative frequency of occurrence of errors on vowels was similar for both intelligibility conditions at both transcription times.

Summary of Speaker Intelligibility

Analysis-of-variance (2x2x2, time/intelligibility/media) results indicated a significant main effect for speaker intelligibility during whole word broad phonetic transcriptions. Analyses were performed with the data collapsed across the media factor. Results indicated higher interobserver percentages of agreement occurred for the intelligible condition than the unintelligible condition. Intraobserver agreement percentages indicated subjects tended to be consistent in reference to the scoring keys across time regardless of the degree of intelligibility or clinical experience of the subject. However, intraobserver agreement percentages tended to be lower in the unintelligible condition than in the intelligible condition. Analysis of errors which occurred on consonants and vowels indicated similar errors were made during both intelligibility conditions although there were some differences in the overall percentages of error. Thus, results indicated fewer errors tended to occur during whole word broad phonetic transcriptions of the more intelligible speaker, but the relative frequency of occurrence of sounds which were transcribed incorrectly were similar between the two intelligibility conditions at both transcription times when sound characteristics were analyzed.

CHAPTER 4

DISCUSSION

Sixteen subjects performed whole word broad phonetic transcriptions of an intelligible and an unintelligible child from audiotaped recordings and from audio-videotaped recordings. The purpose of this study was twofold: 1) To determine whether interobserver and intraobserver reliability calculated from the transcriptions was different between the media (audio-only versus audio-visual) from which the transcriptions were obtained; 2) To determine whether the degree of intelligibility of the speaker made a difference when interobserver and intraobserver reliability was calculated from audiotaped and audio-videotaped recordings.

Results indicated no significant difference between the effects of the audiotaped and the audio-videotaped recording condition on calculated interobserver and intraobserver reliability. A statistically significant difference was present between the intelligibility conditions on calculated interobserver reliability. Intraobserver percentages of agreement with reference to the scoring keys were consistent across time in both conditions but were poorer in the unintelligible than the intelligible condition. Overall, subjects performed whole word broad phonetic transcriptions better from the intelligible speech sample, regardless of recording medium.

Descriptive analyses of the data indicated the type and percentages of transcription errors which occurred on sounds were essentially the same in the audiotaped condition as in the audio-videotaped condition. In addition, while variability existed between the two intelligibility conditions on the percentages of transcription errors made on sounds, few differences existed between the two conditions on the types of errors which occurred. As previously noted, descriptive analyses were conducted with reference to the total number of transcription errors on the scoring keys rather than to the frequency of occurrence of each characteristic being analyzed. For example, results indicated 63% of all consonant transcription errors occurred on alveolar consonants in the audiovisual condition. However, results did not indicate what percentage of alveolar consonants presented in the speech samples were transcribed incorrectly. Therefore, results presented in this study should be interpreted accordingly.

The remainder of this chapter has been divided into two sections. The first section focuses on the clinical implications of this study for using audio-only and audio-visual recording conditions to phonetically transcribe the speech of intelligible and unintelligible speakers. The second section focuses on suggestions for future research as a means for overcoming some of the limitations of the present study by further defining and extending the results of this study.

Clinical Impressions

Audiotaped versus Audio-Videotaped Recordings

The results of this study indicated clinical reliability is essentially the same between the two recording conditions, audio-only and audio-visual. A review of the literature found no previous studies which had compared the effects of audio-only versus audio-visual recordings on calculated interobserver and intraobserver reliability of whole word broad phonetic transcriptions.

One might have expected differences to have occurred with higher reliability in the audio-visual condition based on previous research indicating the addition of visual cues to an audiotaped speech recording held constant between listening sessions increases correct responses by subjects (Irwin & Krafchick, 1965; Mencke, et al., 1983; Monsen, 1983; Stephens & Daniloff, 1977). In fact, the lack of differences in subject agreement between the two recording conditions in the present study was surprising in light of what was thought to be a more difficult task (i.e., whole word broad phonetic transcriptions) than correct/incorrect identification of consonants (Mencke, et al., 1983) and words (Monsen, 1983) or the identification of the occurrence/nonoccurrence of consonant errors (Irwin & Krafchick, 1965; Stephens & Daniloff, 1977). Perhaps the nature of the transcription task in this study was too difficult for the subjects to take

advantage of the cues visual information is thought to provide for correct phoneme identification. That is, since subjects were only allowed to listen to one production of each stimulus word in each recording condition (as also occurred in the studies cited above), visual contextual information may not have been utilized to augment the more complex whole word transcription task. The notion that visual information may be distracting appears to be supported by Wright's (1954) findings that visual distractions appeared to occur in live observation settings resulting in lower correlations and percentages of agreement on the analysis of sounds in single word production than when the same task was accomplished with auditory information alone. The results of the present study suggested even when one decreases the amount of apparent distractions of a live observation situation (e.g. body movements) by videotaping only a head and shoulders view of a speaker, the best agreement subjects can achieve on whole word phonetic transcriptions (keeping all other variables the same as those in the present study) is close agreement between the audio-only and audio-visual recording conditions. This may occur because visual information requires extra attention which does not occur in an ideal listening situation with only one stimulus repetition even when subjects are requested to pay close attention to the visual component. All subjects except one were observed paying close attention to the color monitor during the audio-visual tasks. One subject attended after being reminded after every stimulus production. Thus, even though subjects attended to visual stimuli, they evidently

did not make use of the visual information in addition to the audio information for performing whole word broad phonetic transcriptions.

The phonetic context of the speech samples was not controlled in the present study. That is, the speech samples were not loaded equally with "visible" and "non-visible" phonemes (Binnie, et al., 1974; Erber, 1971; Fisher, 1968; Woodward & Barber, 1960). This was not done in order to keep the task representative of tasks typically performed by practicing speech/language pathologists when evaluating children's speech as well as to keep the speech sample representative of the frequency of sound occurrences in the English language (Mines, Hanson, & Shoup, 1978). The majority of consonant transcription errors in both recording conditions occurred on alveolar sounds, which are not highly visible. Alveolar consonants comprised approximately half of the sounds presented to the subjects (see Appendix G). In addition, the majority of vowel transcription errors occurred on unrounded vowels, which are also not highly visible. Furthermore, unrounded vowels comprised approximately three-fourths of the sounds presented to subjects (see Appendix G). Possibly, the calculated reliability of the transcriptions for the audio-visual condition would have been higher than for the audio-only condition had the speech sample contained more visible sound contexts.

Research which cites the contribution of visual cues to speech perception has indicated a listener makes greater use of visual cues for perception as the signal to noise (S/N) ratio in the listening

environment decreases (Binnie, et al., 1974; Erber, 1969; Miller & Nicely, 1955; Neely, 1956; O'Neill, 1954). For example, O'Neill (1954) found under favorable listening conditions with a S/N ratio of +10dB, vision contributed 5% to vowel scores and 17% to consonant scores on a word recognition task as compared to 29% to vowel scores and 57% to consonant scores at a S/N ratio of -29dB. The present study was conducted in a quiet room under fairly ideal listening conditions. Perhaps agreement on whole word phonetic transcriptions would have been higher in the audio-visual condition than the audio-only condition if the listening environment had been less than ideal. In other words, if clinicians transcribe speech samples in a quiet setting utilizing a good quality audio signal, the addition of visual information may not contribute much to the reliability of the transcription.

The influence of the number of presentations of each stimulus the subjects were allowed to hear will be discussed next. Subjects were allowed to hear only one presentation of each stimulus since time to transcribe was provided between stimuli, the volume setting for stimuli presentation was held constant among all subjects, and optimal visual and auditory clarity were provided. Suggestions have been made in the literature to replay a stimulus two or three times prior to transcribing and then to replay the stimulus a maximum of two times if a confident transcription is not made within a few seconds of hearing (Shriberg & Kent, 1982). This investigator frequently had to pause the playback system longer than the time period allowed between word

presentations on the tapes suggesting a need existed to replay some stimuli for some subjects. Amorosa et al. (1985) found interobserver reliability calculated on initial consonant transcriptions could be increased up to 15% by allowing subjects to replay stimuli as often as they wished. These suggestions indicated the percentages of agreement calculated could possibly have been increased by allowing stimulus replays. In addition, earlier discussion of the role of visual cues indicated opportunities for stimulus replays may have allowed subjects additional opportunities to take advantage of visual information.

The complexity of the task, whole word broad phonetic transcription, was discussed earlier. One might have expected subjects to have performed more reliably in both recording conditions had they been required to transcribe only consonants, or identify correct versus incorrect phonemes. Fristoe and Goldman (1968) found subjects could listen for two or three sounds per word as well as for one sound when identifying occurrence and type of errors in an audio-only condition. Percentages of agreement ranged from 87% to 91% for the more complex task of rating error types. Henderson (1938) found judges had higher agreement on correct/incorrect identifications (90%) versus exact consonant transcription (85%) in an audio-only condition. Results of the present study indicated interobserver agreement of approximately 84% occurred in both recording conditions. This percentage of agreement is only slightly poorer than either of the reported percentages in the studies cited above indicating the whole word transcription task was not more difficult than transcription of single

sounds. Whether either recording condition would have been augmented substantially by a "simpler" task is only a matter of conjecture although one might suspect subjects may have been able to concentrate more on visual cues in the audio-visual condition possibly resulting in higher reliability than the audio-only condition.

Thus, discussion regarding the results of this study indicated speech/language clinicians might obtain more reliable results on transcriptions obtained from audio-videotaped recordings than from audio taped recordings if they typically transcribed in noisy conditions or if they replayed stimuli several times. However, no major differences appeared between audiotaped versus audio-videotaped recordings for whole word broad phonetic transcriptions of single words in ideal listening conditions. A consideration of the issue of intelligibility may provide further conclusions regarding the use of audiotaped versus audio-videotaped recordings.

Speaker Intelligibility

The results of the present study indicated a statistically significant difference in interobserver reliability between the two intelligibility conditions regardless of recording medium. Calculated interobserver percentages of agreement differed between the two intelligibility conditions such that they were lower in the unintelligible condition regardless of recording medium. A review of the literature found no previous studies which addressed the effects of

speaker intelligibility on the calculated interobserver and intraobserver reliability of whole word broad phonetic transcriptions obtained from audiotaped and audio-videotaped recordings. The results of the present study were similar to previous findings by Stephens and Daniloff (1977) on acceptability of /s/ production and by Moller and Starr (1984) on speech ratings, indicating higher calculated interobserver reliability for more intelligible speakers regardless of listening condition.

The results of the present study were not surprising with regard to the finding of differences in calculated reliability from whole word broad phonetic transcriptions of an intelligible (PCC severity rating=84%) versus an unintelligible (PCC severity rating=50%) speaker. Frequent references were found in the literature with regard to the effect of the degree of speaker intelligibility on various phoneme judgment tasks indicating unintelligible speech is typically more difficult to analyze (e.g., Amorosa, et al., 1985; Fristoe & Goldman, 1968; Mencke, et al., 1983; Stephens & Daniloff, 1977; Yorkston & Beukelman, 1978).

The speech samples presented by each speaker in the present study were similar to each other with regard to percentages of occurrence of various consonant and vowel characteristics (Appendix G). These characteristics were also typical of those found in normal speaking adults and children (Mines, et al., 1978; Shriberg & Kent, 1982). The percent consonants correct (PCC) severity adjective (Shriberg &

Kwiatkowski, 1982) assigned each speaker was obtained on single word utterances rather than on a connected speech sample as those authors suggested because speakers tend to make fewer errors in single word productions (Dubois & Bernthal, 1978; Faircloth & Faircloth, 1971; Irwin & Krafchick, 1965; Johnson, Winney, & Pederson, 1980). Therefore, the PCC adjective assigned each speaker was representative of the child's intelligibility at the level in which the whole word broad phonetic transcriptions occurred. In addition, neither speaker demonstrated gross phoneme distortions, compensatory articulation patterns, or voice quality differences such as those known to characterize the speech of some individuals who have cleft palate (Van Demark, 1964), or dysarthria (Darley, Aronson, & Brown, 1969; Platt, Andrews, Young, & Quinn, 1980), or who are deaf (Hudgins & Numbers, 1942). Thus, differences between the two speakers in the present study appeared to have been based on degree of intelligibility differences rather than qualitative differences. This was supported by the fact that similar percentages of transcription errors occurred on consonants and vowels, and the types of characteristics of the sounds transcribed incorrectly were similar in both intelligibility conditions. Since differences in reliability occurred based on degree of intelligibility in the present study, one would expect to find greater differences in calculated reliability and in the descriptive analyses of transcription errors on sounds if there had been qualitative differences between speakers, i.e., if the unintelligible speaker had demonstrated gross phoneme distortions or prevalent phonological process errors.

Consideration must be extended to the lack of effect intelligibility had on whole word phonetic transcriptions obtained from audiotaped versus audio-videotaped recordings. Possibly, the inherent similarities between the two speech samples with regard to the phoneme characteristics presented in each sample resulted in the lack of difference in reliability between the two recording conditions regardless of intelligibility. As mentioned in the previous section, increasing the number of stimulus replays might have increased calculated reliability in both recording conditions. Extra replays might be a critical factor for reliable transcriptions of an unintelligible child's speech. Indeed, replays might be most effective in the audio-visual condition when visual information might provide the cues needed for reliable transcription in an "adverse" listening condition. Possibly, stimulus replays would have increased the calculated interobserver and intraobserver reliability in both recording conditions more for the unintelligible speaker than for the intelligible speaker resulting in closer agreement between the two intelligibility conditions in the present study. The notion that replays would be most helpful for the more unintelligible speaker is supported by the finding of Amorosa et al. (1985) regarding a substantial percentage of agreement increase for the most unintelligible speaker in the study in an audio-only condition in which subjects replayed tapes as often as needed versus hearing stimuli once in a live observation condition. In addition, Monsen (1983) found sentence repetitions contributed

significantly to the perceived intelligibility of unintelligible speakers.

Thus, discussion of the effects of intelligibility on interobserver and intraobserver reliability calculations indicated degree of intelligibility affected reliability but did not tend to have any effect on the recording medium from which the transcription was obtained, at least, under the conditions of the present study. Visual inspection of the percentages of occurrence of all sound characteristics in both speech samples (see Appendix G) and descriptive analyses results indicated no real differences in the quality of the intelligibility of the two speakers. That is, both speakers made the same types of correct and incorrect sound productions and the speakers differed only in the overall frequency of occurrence of the errors. The unintelligible speaker did not demonstrate any distortions or unusual phonological patterns. Perhaps the lack of "unusualness" did not create a need for subjects to make more use of visual cues during the transcriptions in the audio-visual recording medium. Certainly, the results of the present study indicated recording medium makes no difference in calculated interobserver and intraobserver reliability of whole word broad phonetic transcriptions if the two speech samples differ only in the degree of intelligibility.

Future Research

A number of limitations in this study warrant further research to help resolve the issue of which recording condition, audio-only or audio-visual, provides the best medium for obtaining high interobserver and intraobserver reliability calculated from whole word phonetic transcriptions. These limitations and research suggestions are discussed below.

Although the two speakers were selected for this study based on intelligibility, results indicated the quality of intelligibility did not differ between them. Therefore, further work needs to be done exploring the differences in intelligibility between speakers with articulation disorders. That is, the calculation of the degree of intelligibility must go beyond a percent consonants correct score to include the presence/absence of phonological processes, compensatory articulations, and/or sound distortions. This study could be replicated utilizing a child with a mild articulation problem similar to the intelligible child in the present study, as well as a child whose level of intelligibility was at about the 50% level due to multiple articulation deficits. This study could also be replicated comparing two unintelligible children at about the 50% level who differ on the reasons for the 50% rating. Only when differences in intelligibility are really differences in articulatory ability, can the issue of the

recording medium best suited to analyze whole word phonetic transcriptions begin to be resolved.

Related to the degree of intelligibility and transcription medium is the issue of the length and context of the speech sample. Possibly whole word phonetic transcription of single words is not sufficient for identifying which recording medium yields higher reliability. This study should be replicated utilizing phrases and/or connected speech to determine whether differences exist between the recording media and whether the same intelligibility conditions as presented in this study exert any effect on calculated reliability between the recording media. In addition, the context of the sounds presented in the speech samples could also be controlled and compared between single word and connected speech presentations in the audio-only versus audio-visual recording conditions in order to further assess the issue of sound visibility effects on correct sound identification (Binnie, et al., 1974; Montgomery, Walden, & Prosek, 1977).

Subjects in this study were only allowed to hear stimulus words one time during each transcription. Possibly, a need exists to hear a repetition of a stimulus two or three times in order to best perceive the stimulus. In addition, stimulus replays might allow subjects to attend better to visual cues. This study should be replicated so differences which the number of stimulus repetitions have on calculated reliability between recording media may be explored. In addi-

tion, the need for visual cues and additional replays as related to the quality of intelligibility should also be studied further.

The subjects in this study were allowed to see a written gloss of the child's intended production. Possibly, more transcription errors occurred due to transcribing according to the expected production for the more unintelligible child as Oller and Eiler's (1975) results indicated. In addition, the use of an articulation test as familiar as the Goldman-Fristoe Test of Articulation from which to obtain the speech samples may be so familiar to most clinicians that phonetic expectation may occur regardless of whether a gloss is provided. A replication of this study could be done to explore utilizing an unfamiliar word list and controlling the presence of a gloss for word presentations. In addition to studying the effects of gloss/no gloss, the order of word presentation should be randomized to reduce listener expectations about phonetic productions between transcription times.

Although the use of whole word broad phonetic transcriptions did not appear to be a particular limitation of the present study, the need to accurately describe a child's articulatory deficits in order to provide services suggests a need to consider the use of narrow phonetic transcriptions. Since reliability would seem more difficult to achieve due to the more difficult nature of the task (Amorosa, et al., 1985), a clinician would need to use audiotaped or audio-video-taped recordings for reliability purposes. Thus, differences in calculated interobserver and intraobserver reliability between the

two recording conditions would provide useful information for determining the most reliable source of narrow phonetic transcriptions. In addition, one could compare whole word broad versus whole word narrow phonetic transcriptions in an audio-visual medium to determine whether subjects make better use of visual cues for narrow transcriptions due to the increased "adversity" (i.e., task difficulty) of the listening situation.

The task of reliably identifying speech problems raises issues of validly assessing the speech behaviors one thinks one is assessing. Thus, while studying the use of broad versus narrow phonetic transcriptions and audiotaped versus audio-videotaped recordings approaches this goal, one should consider recent suggestions for the need to utilize acoustic analyses to supplement perceptual analyses of speech samples (Amorosa, et al., 1985; Riley, Hoffman, & Damico, 1986; Shriberg, et al., 1984). Many clinicians do not have access to such equipment or the financial resources for purchasing the equipment. Thus, there is a need to determine what equipment and perceptual diagnostic measurements a clinician can make the best use of to achieve the most reliable diagnostic and assessment results within the confines of the equipment available or affordable. In addition, there is evidently a need to study differences in diagnostic recommendations which occur based on perceptual versus perceptual plus physiological-acoustic analyses.

Finally, as the quality of audiotape and audio-videotape recording equipment continues to change, studies should be designed to continually update information with regard to the most practical types of equipment for practicing speech/language pathologists to purchase. One important point of focus for such studies would be to determine how much high quality audio output is needed to suit the purpose of phonetically transcribing speech with high reliability. This would seem particularly important in light of the cost of "high quality" audio and video equipment.

When speech/language pathologists analyze a child's speech for the purpose of identifying an articulation problem and the need for therapy, planning therapy, and determining the time for dismissal from therapy, they should utilize a medium in addition to or instead of live observations to perform whole word phonetic transcriptions. This is important in order to provide the opportunity to obtain the most reliable analyses possible. Given that many of the children seen for speech/language diagnosis and therapy demonstrate "typical" articulation delays, there appears to be no difference between the reliability calculated on transcriptions obtained from audiotaped versus audio-videotaped recordings. However, for children who demonstrate more severe types of articulation problems which would appear to exacerbate the quality of intelligibility, the results are far from conclusive as to the best recording medium from which to reliably obtain whole word broad phonetic transcriptions to be used for diagnosis, assessment, and treatment decisions.

ENDNOTES

1. The use of Pearson product-moment correlation coefficients was found to be inappropriate for accurately describing interobserver and intraobserver reliability on whole word phonetic transcriptions. The coefficients obtained were weak to moderate and were statistically significant at $<.01$ in most instances. However, the data from which the coefficients were obtained were skewed with extreme margins such that most subjects were correct most of the time on their transcriptions of each phoneme. The number of comparisons of disagreements (+/- as compared to ++ or --) were minimal. Therefore, there was little opportunity for subjects to show patterns of disagreements. Apparently, the small number of (+/-) occurrences were sufficient to result in low correlation coefficients within and between subjects. Thus, the data were best reflected by percentages of agreement ($[\# \text{ agreements} / \# \text{ agreements} + \text{ disagreements}] \times 100$).

2. The purpose of this study was to determine the effects of recording media and speaker intelligibility on whole word broad phonetic transcription reliability. Therefore, the transcription scoring keys were utilized as the reference for both interobserver and intraobserver reliability calculations in order to determine the accuracy and consistency with which the subjects each performed whole word broad phonetic transcriptions. Intraobserver reliability is typically calculated by determining an examiner's agreement between two or more sets of obtained data without an outside reference (Shriberg & Kent, 1982). In order to calculate such agreement for each subject in the present study, the transcription obtained at Time 1 should have been compared on a point-by-point basis to the transcription obtained at Time 2 for each speaker without reference to a standard key. When intraobserver agreement was calculated for each subject without utilizing the scoring keys as a reference, the resulting percentages of agreement ranged approximately 2% to 10% higher than the percentages calculated utilizing the scoring keys. Since this investigator was concerned with accuracy and consistency rather than agreement in the present study, intraobserver reliability was calculated utilizing the transcription scoring keys and percentages of difference between the Time 1 and Time 2 transcriptions were reported.

APPENDIX A

Griffith's (1967) modified word recognition list.
All subjects and non-subject observers scored 100%.

1. done	2. pitch	3. team	4. lodge	5. mark
dug	pig	teethe	lob	park
dung	pick	teel	long	bark
dub	pip	tease	log	dark
dug	pit	teeth	laws	lark
6. sin	7. pup	8. red	9. fizz	10. game
pin	puff	fed	fin	shame
win	pub	led	fib	same
tin	puck	shed	fill	tame
fin	putt	wed	fig	came
11. thigh	12. hold	13. ten	14. tong	15. have
lie	gold	pen	tog	has
thy	sold	den	tall	half
high	told	hen	toss	hash
fie	cold	then	talks	hag
16. pig	17. tin	18. gay	19. zip	20. sick
rig	thin	nay	lip	thick
big	shin	they	nip	kick
wig	sin	way	ship	chick
dig	gin	may	dip	pick
21. bass	22. wean	23. shin	24. leave	25. see
batch	weal	thin	liege	zee
bat	weave	kin	leach	knee
bash	weed	fin	leash	lee
badge	we're	tin	leap	thee

APPENDIX B

Percentage of consonants transcribed correctly by two non-subject observers and sixteen subjects on the Grand Quiz of Phonetics Tape 4A (Shriberg & Kent, 1982) as part of entrance criterion (pass=75%+).

<u>Observer</u>	<u>% correct</u>
1	84
2	84
<u>Subject</u>	
1	82
2	81
3	81
4	83
5	77
6	85
7	83
8	80
9	78
10	78
11	84
12	85
13	76
14	85
15	85
16	83

APPENDIX C (a)

Name: _____

Date: _____

Subject #: _____

Subject Survey

1. Are you a student? yes _____ no _____
 Year in school? Junior _____
 Senior _____
 1st Year Graduate _____
 2nd Year Graduate _____
2. Are you a practicing speech/language pathologist?
 yes _____ no _____
 Number of years of experience: _____
 Age of primary caseload: _____
 Speech/language disorder of primary caseload: _____
3. Approximate number of clinical clock hours earned: _____
4. Have you had a class in phonetic transcriptin methods?
 yes _____ no _____
 Date of class enrollment: _____
 Grade received: _____
5. How frequently do you perform phonetic transcriptions?

6. Do you typically perform phonetic transcriptions: live _____;
 from audiotape _____; from videotape _____; other (specify)
 _____.
7. Do you typically transcribe: whole-words _____
 target sounds _____
8. Do you have normal or corrected vision? yes _____ no _____

APPENDIX C (b)

Subject Survey Results Summary

Subject	Phonetics Grade	Frequency Transcription	Medium	Targets	Vision
1+	A	rare	live	words	+
2*	A	weekly	audio	sounds	+
3+	B	rare	audio	words	+
4+	B	rare	audio	words	+
5*	A	rare	live	words	+
6+	A	rare	audio	words	+
7+	A	rare	live	words	+
8+	A	rare	live	words	+
9+	A	rare	live	words	+
10+	A	rare	audio	words	+
11+	A	rare	live	words	+
12*	A	rare	live	words	+
13*	B	rare	live	sounds	+
14*	A	rare	audio	words	+
15*	A	rare	audio	sounds	+
16*	B	weekly	live	words	+

*=speech/language pathologist
+=student

APPENDIX D

Percent Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982) Modified Sampling and Scoring Rules.

Sampling Rules

Single word responses included on the transcription key of the Goldman-Fristoe Test of Articulation were used for the PCC; i.e. other responses not targeted for transcription in this study were not used.

Scoring Rules

Consonant sound changes were counted as incorrect if they were:

- a. deletions of the target consonants;
- b. substitutions of another phoneme for the target consonant;
- c. or additions of a consonant not targeted.

PCC Calculation Formula

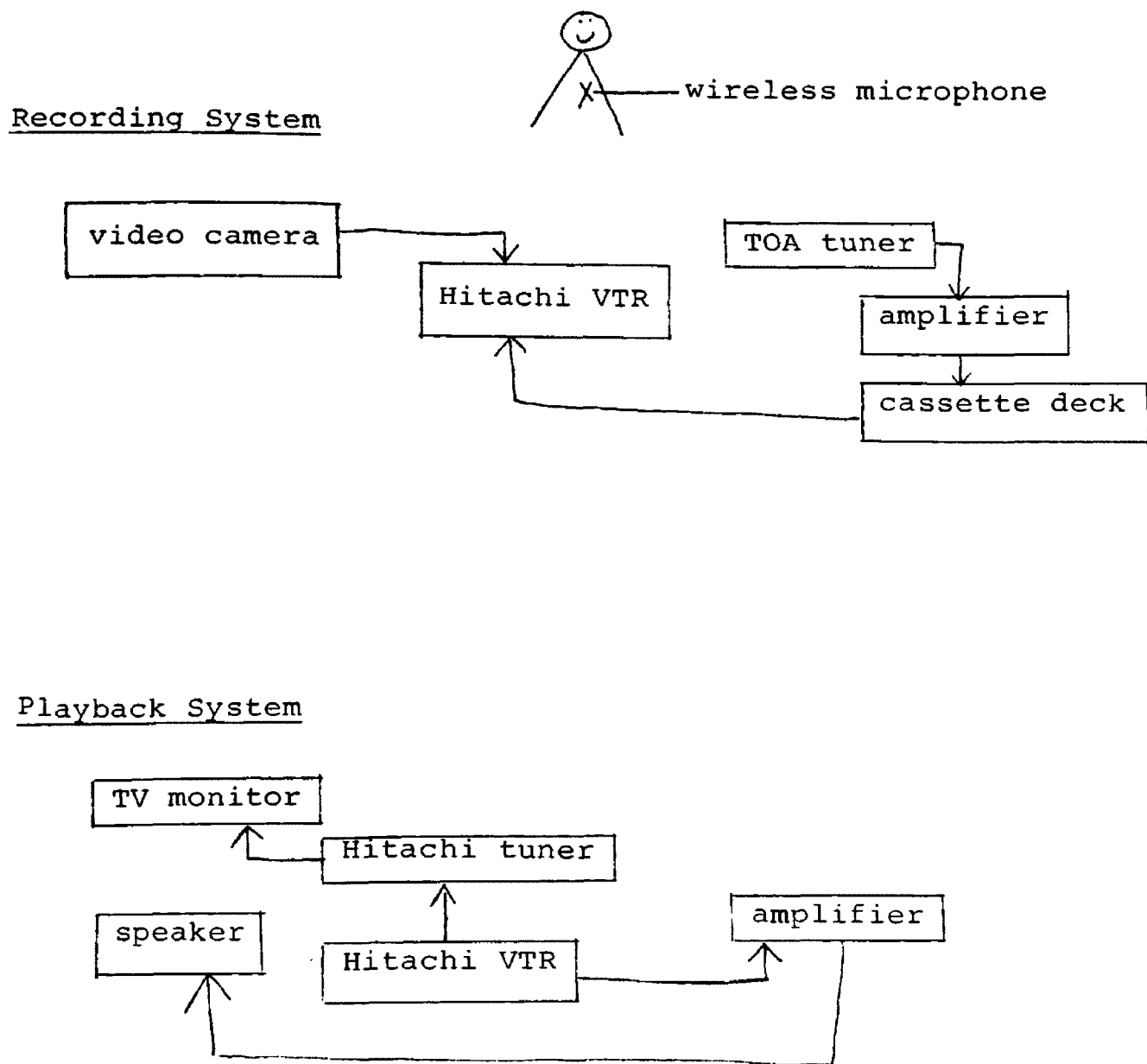
$$\text{PCC} = \frac{\text{number of correct consonants}}{\text{number of correct + incorrect consonants}} \times 100$$

Severity Adjective Assignment to PCC Scores

>85%	=	mild
65%-85%	=	mild-moderate
50%-65%	=	moderate-severe
<50%	=	severe

APPENDIX E

Recording System Schematic and Playback System Schematic



APPENDIX F

Counterbalanced Order of Presentation of Experimental Conditions to Subjects and Non-Subject Observers

Key: AI = audio-intelligible
 AU = audio-unintelligible
 VI = video-intelligible
 VU = video-unintelligible

	Day 1	Day 2	Day 3	Day 4
<u>Observer</u>				
1	AI-VU	VI-AU		
2	AU-VI	VU-AI		
<u>Subject</u>				
1	AI-AU	VI-VU	AU-AI	VU-VI
2	AU-VI	VU-AI	VI-AU	AI-VU
3	VI-VU	AI-AU	VU-VI	AU-AI
4	VU-AI	AU-VI	AI-VU	VI-AI
5	VI-VU	AU-AI	VI-VU	AI-AU
6	AU-AI	VI-VU	AI-AU	VU-VI
7	AI-VU	VI-AU	VU-AI	AU-VI
8	VU-VI	AU-AI	VI-VU	AI-AU
9	VI-AU	AI-VU	AU-VI	VU-AI
10	AU-AI	VU-VI	AI-AU	VI-VU
11	AI-AU	VU-VI	AU-AI	VI-VU
12	VU-VI	AI-AU	VI-VU	AU-AI
13	VI-AU	VU-AI	AU-VI	AI-VU
14	VU-AI	VI-AU	AI-VU	AU-VI
15	AI-VU	AU-VI	VU-AI	VI-AU
16	AU-VI	AI-VU	VI-AU	VU-AI

APPENDIX G (a)

The number and percentages of occurrence of all consonants in the speech samples of the intelligible and unintelligible speakers by place, manner, and voicing.

IN=intelligible

UI=unintelligible

PLACE	IN		UI	
	#	%	#	%
Stop-Plosives	40	31	37	37
Fricatives	36	28	29	29
Affricates	5	4	3	3
Nasals	24	19	23	23
Liquids	15	12	1	1
Glides	7	6	6	6
total	127		99	
MANNER	#	%	#	%
Bilabial	21	17	23	23
Labiodental	10	8	8	8
Linguadental	1	1	0	0
Alveolar	57	45	47	47
Palatal	7	6	5	5
Rhotic/Palatal	8	6	1	1
Labial/Velar	5	4	6	6
Velar	17	13	9	9
Glottal	1	1	0	0
total	127		99	
VOICING	#	%	#	%
Voiced	79	62	56	57
Voiceless	48	38	43	43
total	127		99	

APPENDIX G (b)

The number and percentages of all vowels in the speech samples of the intelligible and unintelligible speakers by diphthongs, tongue height, tongue advancement, tenseness, and lip configuration.

IN=intelligible

UI=unintelligible

	IN		UI	
	#	%	#	%
DIPHTHONGS	2	3	3	4
TONGUE HEIGHT	#	%	#	%
High	20	28	25	37
Mid	38	55	30	45
Low	<u>11</u>	16	<u>12</u>	18
total	<u>69</u>		<u>67</u>	
TONGUE ADVANCEMENT	#	%	#	%
Front	36	51	39	56
Central	21	30	13	19
Back	<u>12</u>	17	<u>15</u>	22
total	<u>69</u>		<u>67</u>	
TENSENESS	#	%	#	%
Tense	22	32	24	36
Lax	<u>47</u>	68	<u>43</u>	64
total	<u>69</u>		<u>67</u>	
LIP CONFIGURATION	#	%	#	%
Rounded	19	28	16	24
Unrounded	<u>50</u>	72	<u>51</u>	76
total	<u>69</u>		<u>67</u>	

APPENDIX H (a)

Transcription Scoring Key and Transcription Format
for the Intelligible Child

KEY: omission - * (e.g. /hə*/ for hot)
addition - + (e.g. /hətə/ for hot)

A/V

Subject _____
Condition _____
Session _____

- | | | | |
|-------------------|----------|--------------------|--------------|
| 1. house | raʊs | 26. orange | ɔrɪndʒ |
| 2. phone | fən | 27. bath | bæθ |
| 3. cup | kʌp | 28. bathtub | bæθtʌb |
| 4. gun | ɡʌn | 29. thumb | θʌm |
| 5. knife | naɪf | 30. finger | fɪŋɡəz |
| 6. window | wɪndə | 31. ring | rɪŋk |
| 7. wagon | wegən | 32. jumping | dʒʌmpɪŋ |
| 8. wheel | wiəl | 33. pajamas (2nd) | dʒəməz |
| 9. chicken | ˈtʃɪkən | 34. airplane | ˈeɪplən |
| 10. zipper | dʒɪpəz | 35. blue | blu |
| 11. scissors | ˈsɪzəz | 36. brush | bɹʌʃ |
| 12. duck | dʌk | 37. drum (2nd) | dɹʌm |
| 13. yellow | ˈjeloʊ | 38. flag | fleg |
| 14. vacuum | vækjʊm | 39. Santa Claus | sænə klɔz |
| 15. matches | mætʃɪz | 40. Christmas tree | kɹɪsməs ˈtri |
| 16. lamp | læmp | 41. squirrel | ˈskwɪrəl |
| 17. shovel | ˈʃʌvəl | 42. sleeping | ˈsli:pɪŋ |
| 18. car | kɑr | 43. bed (2nd) | bed |
| 19. rabbit | ˈræbɪt | 44. stove (2nd) | stev |
| 20. fishing | fɪʃɪŋ | | |
| 21. church | tʃɜ:ʃ | | |
| 22. feather (2nd) | ˈfe:zə | | |
| 23. pencils | ˈpensəlz | | |
| 24. that | dæt | | |
| 25. carrot | ˈkerət | | |

NOTE: Unless otherwise indicated,
please transcribe the 1st
production of a target word

APPENDIX H (b)

Transcription Scoring Key and Transcription Format
for the Unintelligible Child

A/V

KEY: omission - * (e.g. /hə*/ for hot)
addition - + (e.g. /hətə/ for hot)

Subject _____
Condition _____
Session _____

- | | | | |
|--------------------|-------------------------------------|--------------------------|-------------------------------|
| 1. house | <u> fʌɪs </u> | 26. orange | <u> wɔːndz </u> |
| 2. phone | <u> fəʊ </u> | 27. bath | <u> bæf </u> |
| 3. cup | <u> kʌp </u> | 28. bathtub | <u> bæskʌb </u> |
| 4. gun | <u> dʌn </u> | 29. thumb | <u> sʌm </u> |
| 5. knife (3rd) | <u> naɪf³ </u> | 30. finger | <u> fɪŋgə </u> |
| 6. window | <u> wɪnə </u> | 31. ring | <u> ɪŋ </u> |
| 7. wagon | <u> weɪdʒən </u> | 32. jumping | <u> dʌmpɪn </u> |
| 8. wheel | <u> wiːl </u> | 33. pajamas (2nd) | <u> dʌməs </u> |
| 9. chicken | <u> tʃɪkɪnz </u> | 34. airplane | <u> eɪpeɪn </u> |
| 10. zipper | <u> sɪp </u> | 35. blue (2nd) | <u> blu </u> |
| 11. scissors (2nd) | <u> sɪəz </u> | 36. brush | <u> brʌʃ </u> |
| 12. duck (2nd) | <u> dʌk </u> | 37. drum | <u> drʌm </u> |
| 13. yellow (3rd) | <u> əwɔ </u> | 38. flag (3rd) | <u> fæɪd </u> |
| 14. vacuum | <u> vækuəm </u> | 39. Santa Claus | <u> sæn təz </u> |
| 15. matches | <u> mætɪz </u> | 40. Christmas tree (2nd) | <u> krɪsməs ti </u> |
| 16. lamp | <u> læmp </u> | 41. squirrel | <u> skwɪ </u> |
| 17. shovel | <u> ʃəvəl </u> | 42. sleeping (2nd) | <u> sliːpɪŋ </u> |
| 18. car (2nd) | <u> kɑː </u> | 43. bed | <u> bed </u> |
| 19. rabbit | <u> ræbɪt </u> | 44. stove | <u> stəʊv </u> |
| 20. fishing | <u> fɪʃɪŋ </u> | | |
| 21. church (2nd) | <u> tʃɜːtʃ </u> | | |
| 22. feather (2nd) | <u> feðə </u> | | |
| 23. pencils (4th) | <u> pensɪz </u> | | |
| 24. this | <u> dɪs </u> | | |
| 25. carrot (3rd) | <u> kæɪ </u> | | |

NOTE: Unless otherwise indicated,
please transcribe the 1st
production of a target word

REFERENCES

- American National Standards Institute. (1978). American National Standard Methods for Pure-Tone Threshold Audiometry. S3.21.
- American Speech-Language-Hearing Association. (1984). "Proposed Revisions of Guidelines for Identification Audiometry". ASHA, 47-50.
- Amorosa, H., von Benda, U., Wagner, E., & Keck, A. (1985). "Transcribing Phonetic Detail in the Speech of Unintelligible Children: A Comparison of Procedures". British Journal of Disorders of Communication, 20, 281-287.
- Andrews, G., Platt, L.G., & Young, M. (1977) "Factors Affecting the Intelligibility of Cerebral Palsied Speech to the Average Listener". Folia Phoniatica, 29, 292-301.
- Baer, D.M. (1977) "Reviewer's Comment: Just Because It's Reliable Doesn't Mean That You Can Use It". Journal of Applied Behavior Analysis, 10, 117-119.
- Bernstein, D.K. & Tiegerman, E. (1985). Language and Communication Disorders in Children. Columbus: Charles E. Merrill Publishing Company.
- Beukelman, D.R. & Yorkston, K.M. (1980). "Influence of Passage Familiarity on Intelligibility Estimates of Dysarthric Speech". Journal of Communication Disorders, 13, 33-41.
- Binnie, C.A., Montgomery, A.A., & Jackson, P.L. (1974). "Auditory and Visual Contributions to the Perception of Consonants". Journal of Speech and Hearing Research, 17, 619-630.
- Chapman, R.S. (1981). "Computing Mean Length of Utterance in Morphemes" in J.F. Miller (Ed.), Assessing Language Production in Children. Baltimore: University Park Press, 22-25.
- Cherry Wilkinson, L., Clevenger, M., & Dollaghan, C. (1981). "Communication in Small Instructional Groups: A Sociolinguistic Design". In W.P. Dickson (Ed.), Children's Oral Communication Skills. New York: Academic Press.
- Code, C. & Ball, M. (1984). Instrumentation in Speech-Language Pathology. San Diego: College Hill Press.

- Dale, P.S. (1978) "What Does Observing Mean?" In G.P. Sackett (Ed.), Observing Behavior, Volume I. Baltimore: University Park Press, 219-237.
- Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). "Clusters of Deviant Speech Dimensions in the Dysarthrias". Journal of Speech and Hearing Research, 12, 462-496.
- Dinoff, M., Newmark, C., Barnhart, S., Holm, A., Stern, S., & Saunders, T.R. (1970). "Reliability of Video-Tape Interviewing". Psychological Reports, 27, 275-278.
- Dinoff, M., Stenmark, D.E., & Smith, R.E. (1970). "Comparison of Video-Tape and Face-to-Face Interviewing". Psychological Reports, 27, 53-54.
- Dollaghan, C. & Miller, J. (1986). "Observational Methods in the Study of Communicative Competence" in R.L. Schiefelbusch (Ed.), Language Competence: Assessment and Intervention. San Diego: College Hill Press, 99-129.
- DuBois, E.M. & Bernthal, J.E. (1978). "A Comparison of Three Methods for Obtaining Articulatory Responses", Journal of Speech and Hearing Disorders, 43, 282-294.
- Eisler, R.M., Hersen, M., & Agras, W.S. (1973). "Videotape: A Method for the Controlled Observation of Nonverbal Interpersonal Behavior". Behavior Therapy, 4, 420-425.
- Elphick, R. (1984). "Comparison of Live and Video Presentation of a Speechreading Test with Children". British Journal of Audiology, 18, 109-115.
- Erber, N.P. (1969). "Interaction of Audition and Vision in the Recognition of Oral Speech Stimuli". Journal of Speech and Hearing Research, 12, 423-425.
- Erber, N.P. (1971). "Effects of Distance on the Visual Reception of Speech". Journal of Speech and Hearing Research, 14, 848-857.
- Erber, N.P. (1979). "Auditory-Visual Perception of Speech with Reduced Optical Clarity". Journal of Speech and Hearing Research, 22, 212-223.
- Faircloth, M. A. & Faircloth, S.R. (1970). "An Analysis of the Articulatory Behavior of a Speech-Defective Child in Connected Speech and Isolated Word Responses". Journal of Speech and Hearing Disorders, 25, 51-61.

- Fisher, C.G. (1968). "Confusions Among Visually Perceived Consonants". Journal of Speech and Hearing Research, 11, 796-804.
- Foster, S.L. & Cone, J.D. (1980). "Current Issues in Direct Observation". Behavioral Assessment, 2, 313-338.
- Fristoe, M. & Goldman, R. (1968). "Comparison of Traditional and Condensed Articulation Tests Examining the Same Number of Sounds". Journal of Speech and Hearing Research, 11, 583-589.
- Goldman, R. & Fristoe, M. (1969). Goldman-Fristoe Test of Articulation. Minnesota: Circle Press, American Guidance Service.
- Griffiths, J.D. (1967). "Rhyming Minimal Contrasts: A Simplified Diagnostic Articulation Test". Journal of the Acoustical Society of America, 112, 236-241.
- Harris, F.C. & Lahey, B.B. (1982). "Recording System Bias in Direct Observational Methodology: A Review and Critical Analysis of Factors Causing Inaccurate Coding Behavior". Clinical Psychology Review, 2, 539-556.
- Henderson, F.M. (1938). "Accuracy in Testing the Articulation of Speech Sounds". Journal of Educational Research, 31, 348-356.
- Hodson, B.W. & Paden, E.P. (1983). Targeting Intelligible Speech: A Phonological Approach to Remediation. San Diego: College Hill Press.
- Holm, R.A. (1978). "Techniques of Recording Observational Data" in G.P. Sackett (Ed.), Observing Behavior, Volume II. Baltimore: University Park Press, 99-108.
- Hoops, H.R. & Curtis, J.F. (1971). "Intelligibility of the Esophageal Speaker". Archives of Otolaryngology, 93, 300-303.
- Hudgins, C.V. & Numbers, F.C. (1942). "An Investigation of the Intelligibility of the Speech of the Deaf". Genetic Psychology Monographs, 25, 289-392.
- Ingram, D. (1976). Phonological Disability in Children. New York: Elsevier North Holland, Inc.
- Irwin, R.B. (1970). "Consistency of Judgements of Articulatory Productions". Journal of Speech and Hearing Research, 13, 548-555.

- Irwin, R.B. & Krafchick, I.P. (1965). "Audio-Visual Test for Evaluating the Ability to Recognize Phonetic Errors". Journal of Speech and Hearing Research, 8, 281-290.
- Johnson, J.P., Winney, B.L., & Pederson, O.T. (1980). "Single Word Versus Connected Speech Articulation Testing". Language, Speech, and Hearing Services in the Schools, 11, 169-174.
- Jordan, E.P. (1960). "Articulation Test Measures and Listener Ratings of Articulation Defectiveness". Journal of Speech and Hearing Research, 10, 141-150.
- Kazdin, A.E. (1977). "Artifact, Bias, and Complexity of Assessment: The ABC's of Reliability". Journal of Applied Behavior Analysis, 10, 141-150.
- Kent, R.N. & Foster, S.L. (1977). "Direct Observational Procedures: Methodological Issues in Naturalistic Settings" in A.R. Ciminero, K.S. Calhoun, & H.E. Adams (Eds.). Handbook of Clinical Assessment. New York: John Wiley & Sons.
- Kent, R.N., Kanowitz, J., O'Leary, D., & Cheiken, M. (1977). "Observer Reliability as a Function of Circumstances of Assessment". Journal of Applied Behavior Analysis, 10, 317-324.
- Kent, R.N., O'Leary, K.D., Dietz, A., & Diament, C. (1979). "Comparison of Observational Recordings In Vivo, Via Mirror, and Via Television". Journal of Applied Behavior Analysis, 12, 517-522.
- Kipfmuller, L.J. & Prins, T.D. (1971). "Consonant Intelligibility: Reliability and Validity of a Speech Assessment Procedure". Journal of Speech and Hearing Research, 14, 559-564.
- Lipinski, D. & Nelson, R. (1974). "Problems in the Use of Naturalistic Observation as a Means of Behavioral Assessment". Behavior Therapy, 5, 341-351.
- Lovaas, O.I., Koegel, R., Simmons, J.Q., & Long, J.S. (1973). "Some Generalization and Follow-Up Measures on Autistic Children in Behavior Therapy". Journal of Applied Behavior Analysis, 6, 131-166.
- Mash, E.J. & McElwee, J.D. (1974). "Situational Effects on Observer Accuracy: Behavioral Predictability, Prior Experience, and Complexity of Coding Behaviors". Child Development, 45, 367-377.

- Mencke, E.O., Ochsner, G.J., & Testut, E.W. (1983). "Listener Judges and the Speech Intelligibility of Deaf Children". Journal of Communication Disorders, 16, 175-180.
- Metz, D.E., Samar, V.J., Schiavetti, N., Sitler, R.W., & Whitehead, R.L. (1985). "Acoustic Dimensions of Hearing-Impaired Speakers' Intelligibility". Journal of Speech and Hearing Research, 28, 345-355.
- Miller, G.A. & Nicely, P.E. (1955). "An Analysis of Perceptual Confusions Among Some English Consonants". Journal of the Acoustical Society of America, 27, 338-352.
- Mines, M.A., Hanson, B.F., & Shoup, J.E. (1978). "Frequency of Occurrence of Phonemes in Conversational English". Language and Speech, 21, 221-235.
- Moller, K.T. & Starr, C.D. (1984). "The Effects of Listening Conditions on Speech Ratings Obtained in a Clinical Setting". Cleft Palate Journal, 21, 65-69.
- Monsen, R.B. (1978). "Toward Measuring How Well Hearing-Impaired Children Speak". Journal of Speech and Hearing Research, 21, 197-219.
- Monsen, R.B. (1983). "The Oral Speech Intelligibility of Hearing-Impaired Talkers". Journal of Speech and Hearing Disorders, 48, 286-296.
- Montgomery, A.A., Walden, B.E., & Prosek, R.A. (1987). "Effects of Consonantal Context on Vowel Lipreading". Journal of Speech and Hearing Research, 30, 50-59.
- Neely, K.K. (1956). "Effect of Visual Factors on the Intelligibility of Speech". Journal of the Acoustical Society of America, 28, 1275-1277.
- Norris, M., Harden, J.R., & Bell, D.M. (1980). "Listener Agreement on Articulation Errors of Four- and Five-Year-Old Children". Journal of Speech and Hearing Disorders, 45, 378-389.
- Oller, D.K. & Eilers, R.E. (1975). "Phonetic Expectation and Transcription Validity". Phonetica, 31, 288-304.
- O'Neill, J.J. (1954). "Contributions of the Visual Components of Oral Symbols to Speech Comprehension". Journal of Speech and hearing Disorders, 19, 429-439.

- Platt, L.J., Andrews, G., & Howie, P.M. (1980). "Dysarthria of Adult Cerebral Palsy: II. Phonemic Analysis of Articulation Errors". Journal of Speech and Hearing Research, 23, 41-55.
- Platt, L.J., Andrews, G., Young, M., & Quinn, P.T. (1980). "Dysarthria of Adult Cerebral Palsy: I. Intelligibility and Articulation Impairment". Journal of Speech and Hearing Research, 23, 28-40.
- Prins, D. & Bloomer, H.H. (1965). "A Word Intelligibility Approach to the Study of Speech Change in Oral Cleft Patients". Cleft Palate Journal, 2, 357-368.
- Prins, D. & Bloomer, H.H. (1968). "Consonant Intelligibility: A Procedure for Evaluating Speech in Oral Cleft Patients". Journal of Speech and Hearing Research, 11, 128-137.
- Redfield, J. & Paul, G.L. (1976). "Bias in Behavioral Observation as a Function of Observer Familiarity with Subjects and Typicality of Behavior". Journal of Consulting and Clinical Psychology, 44, 156.
- Reid, J.B. (1970). "Reliability Assessment of Observation Data: A Possible Methodological Problem". Child Development, 41, 1143-11.
- Riley, K., Hoffman, P.R., & Damico, S.K. (1986). "The Effects of Conflicting Cues on the Perception of Misarticulations". Journal of Phonetics, 13, 481-487.
- Romanczyk, R.G., Kent, R.N., Diament, C., & O'Leary, K.D. (1973). "Measuring the Reliability of Observational Data: A Reactive Process". Journal of Applied Behavior Analysis, 6, 175-184.
- Rosenblum, L.A. (1978). "The Creation of a Behavioral Taxonomy" in G.P. Sackett (Ed.), Observing Behavior, Volume II. Baltimore: University Park Press, 15-24.
- Sackett, G.P. (1978). "Measurement in Observational Research" in G.P. Sackett (Ed.), Observing Behavior, Volume II. Baltimore: University Park Press, 25-43.
- Sanders, D.A. & Goodrich, S.J. (1971). "The Relative Contribution of Visual and Auditory Components of Speech to Speech Intelligibility as a Function of Three Conditions of Frequency Distortion". Journal of Speech and Hearing Research, 14, 154-159.

- Shriberg, L.D. & Kent, R.D. (1982). Clinical Phonetics. New York: John Wiley & Sons.
- Shriberg, L.D. & Kwiatkowski, J. (1982). "Phonological Disorders III: A Procedure for Assessing Severity of Involvement". Journal of Speech and Hearing Disorders, 47, 256-270.
- Shriberg, L.D., Kwiatkowski, J., & Hoffman, K. (1984). "A Procedure for Phonetic Transcription by Consensus". Journal of Speech and Hearing Research, 27, 456-465.
- Steele, J.A., Binnie, C.A., & Cooper, W.A. (1978). "Combining Auditory and Visual Stimuli in the Adaptive Testing of Speech Discrimination". Journal of Speech and Hearing Disorders, 43, 115-122.
- Stephens, M.I. & Daniloff, R. (1977). "A Methodological Study of Factors Affecting the Judgement of Misarticulated /s/". Journal of Communication Disorders, 11, 583-589.
- Subtelny, J.D. (1977). "Assessment of Speech with Implications for Training" in F. Bess (Ed.), Childhood Deafness. New York: Grune and Stratton, 183-194.
- Taplin, P.S. & Reid, J.B. (1973). "Effects of Instructional Set and Experimenter Influence on Observer Reliability". Child Development, 44, 547-554.
- Tikofsky, R.S. & Tikofsky, R.P. (1964). "Intelligibility Measures of Dysarthric Speech". Journal of Speech and Hearing Research, 7, 325-333.
- Van Demark, D.R. (1964). "Misarticulations and Listener Judgements of the Speech of Individuals with Cleft Palates". Cleft Palate Journal, 1, 232-245.
- Walden, B.E., Prosek, R.A., & Worthington, D.W. (1974). "Predicting Audiovisual Consonant Recognition Performance of Hearing-Impaired Adults". Journal of Speech and Hearing Research, 17, 270-278.
- Woodward, M.F. & Barber, C.G. (1960). "Phoneme Perception in Lipreading". Journal of Speech and Hearing Research, 3, 212-222.
- Wright, H.N. (1954). "Reliability of Evaluations During Basic Articulation and Stimulation Testing". Journal of Speech and Hearing Disorders Monograph Supplement 4, 19-27.

Yorkston, K.M. & Beukelman, D.R. (1978). "A Comparison of Techniques for Measuring Intelligibility of Dysarthric Speech". Journal of Communication Disorders, 11, 499-512.

Yorkston, K.M. & Beukelman, D.R. (1981). "Communication Efficiency of Dysarthric Speakers as Measured by Sentence Intelligibility and Speaking Rate". Journal of Speech and Hearing Disorders, 46, 296-301.