# LIBRARY Michigan State University

### PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE	

1/98 c:/CIRC/DateDue.p85-p.14

# SPEECH NATURALNESS OF NORMAL SPEAKING CHILDREN AND ADOLESCENTS

By

Suzanne S. Coughlin

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

### DOCTOR OF PHILOSOPHY

Audiology and Speech Sciences

# ABSTRACT SPEECH NATURALNESS OF NORMAL SPEAKING CHILDREN AND ADOLESCENTS By Suzanne S. Coughlin

This study investigated the speech naturalness ratings of normal speaking children and adolescents between the ages of 8-16 years using the 1-9 point Likert scale used by Martin, Haroldson, and Triden (1984). In addition, listener-generated perceptual cues for speech rated as highly natural and highly unnatural were listed and weighted. Thirty two unsophisticated adult listeners rated 30 second audio-visual speech samples of children in conversation. Speech samples from 60 normal speaking children--6 males and 6 females at 8,10,12, 14, and 16 years of age--and 10 communicatively disordered (CDO) children were rated. Following naturalness rating tasks, listeners generated perceptual cues which influenced their ratings of natural and unnatural speech and weighted the influence of their cues.

Results revealed overall validity of the 1-9 point naturalness scale. Normal speaking children were rated more naturally than CDO speakers. Male and female speakers were rated comparably, whereas 8 year old speakers were

#### Suzanne S. Coughlin

rated significantly differently than 12, 14, and 16 year old counterparts. Considerable variability in listener agreement was seen in both normal and CDO speaker groups. Intra-rater reliability was consistent with findings in the literature.

Eight perceptual cue categories emerged when classifying listener generated cues. Listeners designated the same three categories as most influential and with similar weights in cueing both natural and unnatural speech characteristics: speech flow, articulation, and ability to be understood received 61% and 77% of the points distributed for natural and unnatural speech, respectively. Speech rate was not identified as a significant perceptual cue. This result was in contrast to results found in previous literature.

Suggestions for clinical applications and future research of speech naturalness of children were discussed.

Copyright by Suzanne S. Coughlin 1998

### **DEDICATION PAGE**

TO:

<u>ALMIGHTY GOD</u>: for my life, parents, and family; <u>MOM AND DAD</u>: Lenida and John Sobaski--who by their actions, unfailing support, and unconditional love, define *HUMANITARIAN*; <u>DGA DGA KARAS</u>: who inspired courage, laughter, and fun;

<u>GRAMMA KARAS</u>: Saint Julia: whose Christian faith allowed her to show us purpose in everything we do;

and

BILL AND ERIN COUGHLIN: who allowed my goal to become theirs.

#### ACKNOWLEDGMENTS

Writing this acknowledgment page is the culmination of years of effort and the generously offered help from many. I am deeply grateful to the administration and families at Sacred Heart Academy and to Connie Parkhurst for their help in recruiting 60 children to participate as speakers. Technological support was a significant part of this study. I am greatly indebted to the finest team of videographers anywhere: Steve Zlotolow, Dan Bracken, and Ben Bracken, whose creative genius, persistence, and hours of dedicated time elevated this study's technology to CD-ROM. Throughout my entire course of study and especially throughout completion of this dissertation, my colleagues in the Department of Communication Disorders at Central Michigan University have been a constant source of encouragement, ideas, and inspiration. Special thanks to Mark Lehman for his statistical genius in saving me hours of time and insuring data accuracy, and to Brad Swartz for his computer savvy which brought my dissertation defense presentation to life.

Dr. Peter LaPine, Dr. Renny Tatchell, Dr. Leo Deal, and Dr. Paul Cooke were my dissertation committee members. I wish to thank them for their expert direction, time, and encouragement. Special thanks to Renny Tatchell, my colleague for his hours of effort in document development and moral support. Dr. Leo Deal and Dr. Paul Cooke deserve exceptional appreciation beyond dissertation assistance. Dr. Deal served as my academic advisor. His support of my part-time doctoral status was the single most important variable that encouraged my continuance in the program. His kind words of direction and faith in my ability carried me through many times of adversity.

Dr. Paul Cooke served as my dissertation chair. His expertise in research development, suggestion of areas to be considered, practical advise, and excitement for my project encouraged forward momentum to completion. His qualitative guidance and respectful regard for students brings learning to life. In addition, his high clinical standards and earnest desire to keep "our clients' " interests paramount have inspired me to continue to develop relevant clinical questions and reach to be the best therapist possible.

Authorship of this document is not complete unless the names--Bill and Erin Coughlin--appear. Their sacrifices to make this project possible were immense. They unconditionally yielded the right-of-way to its completion. To them--I am forever grateful.

# TABLE OF CONTENTS

LIST OF TABLESx
LIST OF FIGURESxi
LIST OF ABBREVIATIONSxii
CHAPTER 1
CHAPTER 2

CHAPTER 3	47
Listener Reliability and Agreement Rating Range and Scale Use Inter-Listener Agreement Naturalness Characteristics of Speakers Range of Scale Values Received Mean of Scale Values Received Mode of Scale Values Received	
Age and Gender Considerations Analysis of Perceptual Cues	
CHAPTER 4 DISCUSSION	63
CHAPTER 5 SUMMARY AND CONCLUSIONS	77
APPENDICES	85
BIBLIOGRAPHY	101

.

# LIST OF TABLES

1	Comparison of mean naturalness scores of adult stutterers and nonstutterers in 5 studies on a 1-9 scale	12
2	Raters' degree of agreement on duplicated presentations	49
3	Age distribution of listeners included in study	50
4	Cumulative number and percent of inter-rater agreement pairs for the speech naturalness ratings of the 60 normal speaking children	52
5	Cumulative number and percent of inter-rater agreement pairs for the speech naturalness ratings of the 10 communicatively disordered speakers.	53
6	Differential range of rating intervals assigned to the 60 normal speaking samples	54
7	Differential range of rating intervals assigned to the 10 communicatively disordered speakers	54
8	Mean rating, range, range width, and standard deviation of ratings specific to age and gender of the 60 normal speaking children	55
9	Mean rating, range, and standard deviation ofratings of the 10 communicatively impaired children	56
10	Frequency of occurrence as mode in normative and communication disordered samples	57
11	ANOVA results for main effects of age and gender for the normal speaking group	57
12	Post-hoc multiple comparisons of group means	58
13	Perceptual cue categories and examples of verbatim cues listed in each category	60

14	Bar scale point distribution of perceptual cues to identify speech as natural	61
15	Bar scale point distribution of perceptual cues to identify speech as unnatural	62
16	Number of inter-rater pairs per age group of normal speaking children	67
17	Number of "1" ratings received by normal speaking groups by age	67
18	Bar scale point distribution of perceptual cues to identify speech as natural and unnatural	73

## LISTS OF FIGURES

- 1 Types of disfluencies noted during development......20

# LISTS OF ABBREVIATIONS

DAF	Delayed Auditory Feedback
EAI	Equal Appearing Intervals
DME	Direct Magnitude Estimation
СDО	Communicatively Disordered Speakers
ASHA	American Speech, Language, Hearing Association

### Chapter 1

### INTRODUCTION AND REVIEW OF THE LITERATURE

#### Introduction

### Speech: A Natural Communication Form

Speech has been recognized throughout recorded history as the most natural form of communication for many distinguishable reasons (Liberman, 1992). When contrasting the relationship of speech to graphic forms of communication, Liberman (1992) elaborated on the differences in naturalness between spoken and written forms of communication. Speech is a universal phenomenon, specific to the human species. Every human community has developed a form of spoken language. Under normal circumstances speech emerges in each individual as the primary form of communication, whereas graphic systems develop much later. The naturalness of speech is supported in the idea that, as a form of communication, speech is learned and does not need to be taught. Thus speech is a precognitive process requiring only that its users are human and exposed to the pattern of the ethnographic region. Liberman (1992) also indicated that, as opposed to graphic forms of communication, specific brain regions or areas are dedicated to the development and use of speech.

Finally, as a natural process, speech as a tool of communication "...is capable of expressing and conveying an indefinitely numerous variety of messages" (p. 168). In concluding the inherent nature of "speech" as a communication form, Liberman (1992) indicated that speech is a natural outgrowth of biological development of each human. Consequently, it is a natural phenomenon.

The term "speech naturalness" has appeared in professional literature for the past four decades as a concept that views speech output as a product that "...sounds normal or natural" (Finn & Ingham, 1994; Ingham, Gow, & Costello, 1985; Ingham, Ingham, Onslow, & Finn, 1989; Ingham & Onslow, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Martin, Haroldson, & Triden, 1984; Onslow, Hayes, Hutchins, & Newman, 1992; Parrish, 1951; Nichols, 1966). In addition to the concept of "speech naturalness," the idea of evaluating speech quality in terms of naturalness was also suggested more than forty years ago. In 1951 Parrish stated that "...naturalness in speech is a virtue" (p. 448) and has been commonly accepted as such since the time of Aristotle. Parrish proposed the "concept of naturalness" as central to the idea of desirable speech behavior and distinguished between a speaker's perception of natural speech production and a listener's perception of natural speech sound. In his attempts to define naturalness and depict its importance in communication, Parrish (1951) promoted the idea that interpretations of naturalness should be in the realm of what seems natural to the listener rather than what feels natural to the speaker. He supported the notion that any methods of teaching speech patterns that result in productions that are judged natural by listeners should be applauded and are critical when determining oral effectiveness. The essential component of Parrish's "concept of naturalness" that speech that is "natural" focuses the listener's attention on the meaning of the words spoken rather than the speech

pattern used in conveying the meaning. This idea of naturalness served as the catalyst for subsequent research.

Later, in the first published attempt to reliably measure naturalness, Nichols (1966) developed a tool for rating listeners' perceptions of speech naturalness using a 9-point interval scale (1= high naturalness and 9= low naturalness). He was primarily interested in two aspects of naturalness. First, Nichols sought to establish whether significant differences in naturalness would occur in written and spoken sentence presentations and in sentence lists. Second, he was concerned in the rating reliability of listeners. Nichols developed two sentence lists, one with frequently used words and one with low frequency used words. Two trained speakers read the lists to 20 listeners who were college sophomores in an oral reading class. Each sentence was rated using the 1-9 scale as indicated. In a conclusion supporting Parrish's (1951) earlier postulation, Nichols (1966) reported a listener's reliability range (r values) of .74 to .84 and concluded that "naturalness" of sentence readings was a concept that could be reliably rated by audiences. However, Nichols primary contribution to the naturalness literature was that "...the reliability of the rating of one listener is not sufficiently high to be useful, but the mean rating of an audience of twenty raters has apparently a high enough reliability to be useful in experimental studies and in classroom demonstration" (p. 159). Nichols also found that the vocabulary level of sentences influenced naturalness ratings and he postulated that sentences containing frequently used vocabularies were perceived as more natural than sentences containing relatively infrequently used vocabulary.

#### **Review of the literature**

#### **Speech Naturalness and Communication Disorders**

In the past two decades, ratings of speech naturalness have been primarily used to rate perceptual quality of speech that differed from normal production, specifically stuttered, synthesized, and dysarthric speech (Finn & Ingham, 1994; Ingham, Gow, & Costello, 1985; Ingham, Ingham, Onslow, & Finn, 1989; Ingham & Onslow, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Linebaugh & Wolfe, 1984; Martin, Haroldson, & Triden, 1984; Onslow, Hayes, Hutchins, & Newman, 1992; Sanders, Gramlich, & Levine, 1981). However, the definition of the term naturalness has been inconsistent, creating confusion. In some cases it has been equated with other perceptual terms such as intelligibility. As stated earlier, speech naturalness has been defined as speech output that sounds normal or natural (Finn & Ingham, 1994; Ingham, Gow, & Costello, 1985; Ingham, Ingham, Onslow, & Finn, 1989; Ingham & Onslow, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Martin, Haroldson, & Triden, 1984; Onslow, Hayes, Hutchins, & Newman, 1992; Parrish, 1951; Nichols, 1966). In contrast to this definition of naturalness, most studies have defined the term intelligibility as the ability of a listener to understand what a speaker is saying (Carney, 1994). Intelligibility studies have been widely conducted in areas of articulation, hearing impairment, synthesized speech, and motor speech disorders (Braverman, 1974; Beukelman & Yorkston, 1979; Carney, 1994; Fudala, 1970; Hudgins, 1949; Keeler, Clement, & Strong, 1976; Laddaga, Sanders & Suppes, 1981; Linebaugh & Wolfe, 1984; Shriberg & Kwiatkowski, 1982; Thomas, 1964). It is important to distinguish the term "naturalness" as a broad-based perceptual component from "intelligibility," the ability to be understood, in order to provide accurate definition of the perceptual

quality being studied and to eliminate confusion in the use of these terms in the professional literature.

#### Synthesized Speech: Naturalness

Although thought to be an important element of evaluating disordered speech, the term "naturalness" most often has been broadly defined. A study of naturalness done by Sanders, Gramlich, and Levine (1981) provided a very important distinction between naturalness and intelligibility, two perceptual arenas often confused. These authors studied the relationship between prosodic manipulations and naturalness of synthesized speech. They purposely defined naturalness as speech which "...sounds as if a normal native speaker produced it" (p. 488). They stated their definition of naturalness in order to distinguish it from other perceptual studies of synthesized speech dealing with intelligibility (Keeler, Clement, & Strong, 1976; Laddaga, Sanders & Suppes, 1981) . As Sanders, Gramlich, and Levine stated, "...intelligibility only indicates that the speech can be understood, not whether it is natural sounding and easy to listen to" (p. 487). Sanders and associates offered a model of speech quality containing three separate components: intelligibility, naturalness, and clarity. While maintaining interaction between these variables, Sanders and associates maintained that each of these components "...differentiate aspects of speech quality" (p. 488). In addition to reinforcing the distinction between "naturalness" and "intelligibility," the most significant finding of Sanders, Gramlich and Levine's study was that the listeners reliably used a 1-9 interval rating scale when evaluating different sentences and treatments. While other aspects of their work devoted much attention to isolating specific naturalness aspects of synthesized speech, the reliable performance of listeners reinforced the parameter of naturalness as a "scaleable" concept.

Many intelligibility studies of synthesized speech devices have been conducted to establish the intelligibility ratings of specific speech or augmentative communication output devices available for children and adults with disabilities (Hoover, Reichle, Van Tassell, & Cole, 1987; Logan, Pisoni, & Greene, 1985; Mirenda & Beukelman, 1987 and 1990). Rather than pursue <u>naturalness</u> of the synthesized output, the major aim of that body of literature provided ratings of <u>intelligibility</u>, or the ratings of the abilities of listeners to comprehend the messages produced. With recent incorporation of digitized speech output in augmentative devices, studies of speech intelligibility are not as necessary because of the more "natural" sound produced (Beukelman & Mirenda, 1992).

#### **Dysarthria Studies: Naturalness**

In contrast to the distinctions between naturalness and intelligibility provided by Sanders and colleagues (1981), studies of "naturalness" in the area of dysarthria have not necessarily defined the parameter to be observed. In 1984, Linebaugh and Wolfe studied relationships between articulation, rate, intelligibility, and naturalness in dysarthric speakers. Using a 7 point scale of naturalness, five certified speech and language pathologists rated audiorecorded portions of the "My Grandfather" passage for 14 spastic, 14 dysarthric, and 14 normal speaking individuals. However, in contrast to Sanders and associates (1981), these authors equated the terms "intelligibility" and "naturalness" by having them connected on the rating scale. Consequently, number "7" on the scale implied "...100% intelligibility and normal, respectively" (p.199). As noted in previous studies, results revealed strong inter-judge reliability for intelligibility/naturalness ratings (r=.93). Naturalness/intelligibility was negatively affected by mean syllable duration; when the mean syllable duration increased, naturalness ratings decreased. Although naturalness and

intelligibility have been differentiated in studies, Linebaugh and Wolfe (1984) integrated the terms naturalness and intelligibility and concluded that dysarthria rehabilitation "...might be enhanced by focusing on those elements of a particular patient's speech that contribute most to decreased intelligibility and naturalness" (p.204).

In another study, naturalness was differentiated from intelligibility but not distinctly defined. Bellaire, Yorkston, and Beukelman (1986) studied the effects of breathing pattern modification in increasing speech naturalness. In this case study, three speech and language pathologists made a global rating of naturalness following the completion of treatment goals aimed at varying breath group parameters and using pausing intermittently. Each clinician was asked "Did he sound more natural?" (p.278). However, beyond this question, no attempt at defining naturalness or delineating its distinction from intelligibility was explored. Nevertheless, these authors concluded that naturalness was a worthwhile pursuit of clinical treatment goals.

Speech intelligibility has been a primary goal of dysarthric speakers (Yorkston, Beukelman, & Bell, 1988). Although a preponderance of dysarthric research has dedicated effort to establishing parameters of intelligibility, in general a relationship between intelligibility and naturalness of speech was stated. Yorkston, Beukelman, and Bell (1988) maintained that achieving the goal of intelligible speech "...is a prerequisite for other aspects of speech performance, including naturalness" (p. 157).

#### Application of Speech Naturalness To Fluency

During the past two decades researchers studying fluency and fluency disorders have focused on "speech naturalness" to determine whether perceptual differences in naturalness ratings exist between stutterers (treated

and nontreated) and normally fluent speakers (Finn & Ingham, 1994; Ingham, Gow, & Costello, 1985; Ingham, Ingham, Onslow, & Finn, 1989; Ingham & Onslow, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Martin, Haroldson, & Triden, 1984; Onslow, Hayes, Hutchins, & Newman, 1992; Onslow, Costa, Andrews, Harrison, & Packman, 1996). Consequently, information about naturalness related to normally fluent speaking persons has been attained primarily throughout the research related to disorders of fluency.

Research interest in "speech naturalness" in the area of fluency disorders has intensified because some fluency-inducing procedures have been criticized for producing a speech quality that sounds unnatural or different from the norm (Ingham & Packman, 1978; Runyan & Adams, 1978, 1979; and Runyan, Bell, & Prosek, 1990.) The assumption that speech quality is an outcome measured in terms of rate and stuttering fluency, often relied upon in fluency shaping programs, may be naive (Onslow & Ingham, 1987). The study of speech quality with particular reference to fluency disorders is crucial, given the fact that many current therapies employ fluency-shaping procedures that utilize unnatural sounding speech patterns in the course of treatment (Onslow & Ingham, 1987). In addition, speculation regarding the role of post-treatment unusual speech quality as an indicator of potential relapse has been an interest of researchers (Martin, Haroldson, & Triden, 1984).

#### Naturalness Ratings of Adults With Fluency Disorders

To appreciate the specific knowledge of naturalness obtained to date, review of significant findings of studies with adults and the application of the "Naturalness Scale" (Haroldson, Martin, & Triden, 1984) is necessary. Without question, the preponderance of studies relating the concept of speech naturalness to fluency disorders have investigated the speech of adults. Even

prior to the recent popularity and use of the term "naturalness," studies have attempted to note whether the fluent speech of adult stutterers is perceptually different from speech regarded as normally fluent. Love and Jeffress (1971) found that stutterers' fluent speech was perceptually different from the speech of normally fluent cohorts, in that stutterers used a greater number of silent pauses compared to the speech of nonstutterers. Later, in 1978 and 1979, Runyan and Adams found that both unsophisticated and sophisticated listeners judged the fluent speech of stutterers who successfully completed therapy regimes, as perceptually different from speech of normally fluent counterparts. Recently, Finn (1997) found that the speech of persons who recovered from stuttering without a formal treatment program was rated more unnatural than the speech of normally fluent adults. Other research has examined naturalness ratings of speakers judged as having dialectal difference from the general dialect of an ethnographic region. Mackey, Finn and Ingham (1997) found that in a study of three speaker groups, stutterers' speech was judged as least natural, normally fluent adults' speech was judged as most natural, and the speech of speakers with a dialect was judged as more natural than the speech of stutterers but less natural than the speech of the nondialect fluent speakers.

A review of the literature has credited Martin, Haroldson, and Triden (1984) as publishing the first paper to establish the concept of naturalness of speech as it relates to differences between normal and abnormal speech patterns in stutterers and nonstutterers. In order to establish a tool and method to reliably rate the naturalness of speech of adult stutterers and nonstutterers, the authors adapted the use of the 9-point scale previously used by Nichols (1966). The scale required listeners to rate speech on a 9 point interval scale between 1 (highly natural) and 9 (highly unnatural). No definitions were provided to raters for the intervals between 1 and 9. Martin and colleagues demonstrated

that listeners' ratings using this scale could reliably discriminate between the speech of adult stutterers, non-stutterers, and the Delayed Auditory Feedback (DAF) induced speech of adult-stutterers. On this scale the mean ratings found for stutterers, DAF treated stutterers, and nonstutterers were 6.52, 5.84, and 2.12, respectively. In addition, on tasks of rerating the speech samples three weeks later, 88% of the ratings were within  $\pm$  1 of the original ratings. The authors concluded that scaling speech naturalness would be a useful concept in the management of fluency disorders. Given that many current therapy programs rely on some form of prolonged speech, rhythm, or DAF in the treatment of stutterers, the resultant speech shaped by these programs may produce a quality that is perceptually different from that of normal speech (Martin, Haroldson, Triden, 1984). These authors indicated that potential use of the "parameters of naturalness" would be a helpful aspect of management.

Since Martin, Haroldson, and Triden's (1984) application of the naturalness scale to stutterers' speech, others have attempted to explore its usefulness with adults in clinical speech treatment. Ingham, Martin, Haroldson, Onslow, and Leney (1985) found that throughout the course of therapy, adult stutterers--when given feedback (rating scores from the 1-9 scale) from listeners who rated naturalness--were able to improve their naturalness ratings and reduce stuttering frequency without specific instructions explaining what they might manipulate in their speech pattern to improve speech output. These authors also concluded that listeners are able to make reliable speech naturalness judgments that may be helpful during treatment. In 1989, Ingham, Onslow, and Finn also found that adult stutterers could modify their speech in attempts to improve naturalness and could judge improvements in naturalness similarly to listeners. A recent study by Onslow, Costa, Andrews, Harrison, and Packman (1996) systematically examined speech naturalness as a treatment

aspect as well as an outcome measure of a prolonged speech fluency shaping program using the 1-9 point scale. In their findings, Onslow et al. discovered that stutterers were able to maintain those naturalness ratings taken at discharge as long as twelve months following treatment. While slight variation in the stutterers' naturalness ratings were noted during post-treatment assessment periods, the stutterers' naturalness scores remained in the range associated with nonstutterers. Use of speech naturalness ratings as a treatment outcome/measurement was first noted in fluency treatment literature by Onslow et al. (1996).

#### Naturalness Ratings of Normal Speaking Adults

Using the 1-9 point naturalness scale, several studies incorporating adult nonstutterers have reported mean naturalness scale scores of normally fluent individuals that range from 2.12 to 3.55 (Ingham, Gow, & Costello, 1985; Martin, Haroldson, & Triden, 1984; Metz, Schiavetti and Sacco, 1990; Onslow, Hayes, Hutchins, & Newman 1992). Table 1 illustrates the naturalness rating mean scores of nonstutterers on the 1-9 point rating scale. The variability involved in the range of mean naturalness may be related to listener variability or to differences in the speakers themselves.

Lis Study Sophisti Unsophisti	steners: cated (S) cated (U)	Nonstutterers Mean Rating	Stutterers Mean Rating (Post Therapy)
Ingham, Gow, Costello, 1985	30 (S)	2.39 (n=15)	4.26 (n=15)
Martin, Haroldson, & Triden, 1984	30 (U)	2.12 (n=10)	
Metz, Schiavetti, & Sacco, 1990	30 (S)	3.55 (n=20)	5.92 (n=20)
Onslow, et al., 1992	29 (S)	3.25 (n=7)	5.49 (n=7)
Runyan, Bell, & Prosek, 1990	10 (S)	2.79 (n=140)	

TABLE 1. Comparison of mean naturalness scores of adult stutterers and non-<br/>stutterers in 5 studies on a 1-9 scale used by Haroldson, Martin, and Triden, 1984.1= highly natural and 9= highly unnatural (adapted from Onslow et al., 1992)

Note: Data for post-transfer stutterers following a prolonged stuttering treatment program was not available as a component of the Martin, Haroldson, & Triden (1984) and the Runyan, Bell, & Prosek, (1990) studies.

In contrast to the mean ratings of nonstutterers listed in Table 1, three of the aforementioned 5 studies (Ingham, Gow, & Costello, 1985; Metz, Schiavetti, & Sacco, 1990; Onslow et al., 1992) included the naturalness ratings of post-transfer stutterers who had been enrolled in prolonged stuttering treatment programs. In each of these studies, the mean naturalness ratings assigned to post-transfer stutterers were higher (more unnatural) than those assigned to the normally fluent speakers. Calculated differences in mean ratings of the post-transfer stutterers ranged from 1.87 to 2.37 scale values and were therefore judged to be "more unnatural" than those of the nonstutterers.

#### Naturalness Raters: The Listeners

Within these five studies, both inexperienced and experienced listeners were used. Onslow et al. (1992) and Curlee (1993) suggested that, ultimately, establishing normative data would be best served by using unsophisticated listeners who demonstrated a high degree of reliability. Clinicians could then be trained to listen using the normative data noted by the unsophisticated listeners. Earlier, Runyan and Adams (1979) stated that unsophisticated listeners would best represent the judgments of the general listening public and that their evaluations of naturalness should be the primary concern of investigating researchers.

Curlee (1993) also suggested that the most valuable assessment of an adult stutterer's disability and treatment efficacy was that of listeners' perceptual judgments of naturalness. Later, Kalinowski, Noble, Armson, and Stuart (1994) reinforced Curlee's premise. These authors used the 1-9 scale (Martin, Haroldson, & Triden, 1984) in a listening/rating task in which 64 naive listeners rated the pre- and post-treatment speech of 10 stutterers. Kalinowski et al. concluded that although post treatment speech was free from stuttering events, overall stutterers' post treatment speech was rated as more unnatural. Consequently, Kalinowski et al. and Onslow et al. (1996) added support to Curlee's (1993) idea that treatment goals in fluency shaping programs should aim to produce fluent and natural sounding speech. Curlee (1993) recommended that the most advantageous format for assessment of naturalness involves repeated assessments of speech from samples taken both within and outside the clinic. He also advocated the use of 3 to 5 raters to evaluate the speech samples for assignment of values on the 1-9 scale promoted by Martin, Haroldson, and Triden (1984).

Curlee (1993) also suggested that the need for establishment of normative naturalness values is great, and should take into account "...a much larger number of nonstuttering speakers so that the resulting data base will likely encompass the range of fluent and nonfluent speakers in the nonstuttering population" (p. 325). His suggestion arose from the fact that studies published to that date typically employed a 1-1 ratio (stutterer to nonstutterer) design. Following Curlee's recommendation, studies incorporating a large normative

group are warranted if they are to better represent the range of variation found in normally fluent speakers.

#### Naturalness Studies with Children

In contrast to the accumulating data available regarding adults, no information exists specifically concerned with "naturalness ratings" of speech patterns of normally fluent children under the age of 16. However, many studies have provided a considerable amount of information about speech pattern differences in early childhood. Disfluencies in preschoolers vary greatly from child to child (Haynes & Hood, 1977; Yairi, 1981; 1982). Disfluency patterns that are seen most frequently in 2 to 3 year olds are interjections, pauses, and revisions (Wexler & Mysak, 1982). However, these patterns generally decrease after age 3 (Wexler & Mysak, 1982). Yairi (1982) added that 2-year olds may demonstrate a considerable number of part-word repetitions.

While studies have aimed at identifying the characteristics of fluency differences in young children, naturalness ratings of normally fluent children have not been systematically researched. Little is known about possible perceptual differences between the fluent speech of stuttering and nonstuttering children beyond preschool age. The few studies that have attended to the perceptual differences of speech of children <u>have not</u> specifically targeted "speech naturalness." Rather, studies of perceptual differences have aimed to ascertain whether the speech of normally fluent children and fluent stuttering children is perceptually different. The two principal investigations which involved a comparison of the perceptual differences between normally fluent and CDO children are those of Krikorian and Runyan (1983) and Colcord and Gregory (1987).

In 1983 Krikorian and Runyan examined fluent speech samples of stuttering and nonstuttering children (using a 1 to 1 matched sample ratio), whose ages ranged from 4:2 to 6:11, to determine whether perceptual differences existed. Audio-taped stimuli were created from the story portions of the <u>Goldman-Fristoe Test of Articulation</u>, and listeners were asked to identify from each sample whether each speaker was a stutterer or a normally fluent speaker. Krikorian and Runyan concluded that sophisticated listeners <u>could not</u> distinguish speech of the stuttering children from that of the normally fluent children at statistically significant levels.

Later, in a similar task, Colcord and Gregory (1987) studied whether identical and fluent audio-taped speech samples of stuttering children could be identified as perceptually different from samples of normally fluent children. This study expanded the upper level of the age range of children used in the Krikorian and Runyan (1983) study by including children aged 4:1 to 9:6 years. Like Krikorian and Runyan, they also included the use of sophisticated judges. In their conclusions, Colcord and Gregory supported the premise that perceptual distinctions between the fluent speech of stuttering children and their normally fluent counterparts were difficult to make. Judges in their study were able to correctly identify the speech of stuttering children in up to 1/2 of the samples provided. In the remaining cases, no discrimination could be made and one normally fluent child was misidentified as a stutterer.

Like Krikorian and Runyan (1983), Colcord and Gregory also postulated that either 1) perhaps as stuttering children mature, additional perceptual cues are inserted into the speech pattern as a learned mechanism in the child's response to his/her environment that results in speech that is distinguishable from normally fluent adolescents; or 2) as a child receives speech therapy, perceptual differences become distinct as an artifact of therapy procedures.

Krikorian and Runyan (1983) also speculated that perhaps "...judges in perceptual investigations using child subjects may use a standard of fluency that tolerated deviations from and differences in speech production to a greater degree than they would when listening to more mature subjects" (p. 288). Neither study explored the naturalness of the speech samples involved for comparison between stuttering and normally fluent children; rather, both studies asked listeners to detect which speaker was a stutterer and which was a normally fluent speaker.

Review of the naturalness literature reveals that to date only one study has investigated the use of naturalness ratings with children. Ingham and Onslow (1985) conducted a study aimed at the measurement and modification of speech with 5 adolescent stutterers (one age 10, two age 13, and two age 14). The subjects were enrolled in a fluency shaping program reliant on rate control procedures. Corresponding to results obtained with adults, the results of the study revealed that natural sounding speech was not automatically a product of a therapy program designed to remove disfluencies, control rate, and utilize transfer skills. In addition, when given only numerical feedback of speech naturalness ratings (using the 1-9 scale [Martin, Haroldson, & Triden, 1984]), the subjects were able to improve naturalness rating scores. While Ingham and Onslow did not pursue perceptual differences in the speech of their adolescent subjects and age cohorts, the treatment naturalness target criterion imposed on their subjects was 2.40 (1-9 naturalness scale). This value was derived from a mean naturalness value assigned to the nonstuttering population determined by Ingham, Gow, and Costello (1985). Unfortunately, the naturalness targets used were calculated from a predominately adult population sample and not derived from cohorts more closely resembling the ages of the subjects involved.

Colcord and Gregory (1987) concluded their discussion by proposing the need for additional research of the perceptual, acoustic, and physiological

dimensions of

the fluent speech of older stuttering children as they approach adolescence and thus acquire increased neuromotor competencies for speech. With these data, investigators may be able to determine a specific period during development when stutterers' fluent speech productions start to differ perceptually, acoustically, or physiologically from nonstutterers (p.194).

#### **Communication Development: Maturational Considerations**

Technological advances have facilitated increased development in scientific knowledge of neurology and the neural basis of language. Primary neural development begun during gestation is completed in early childhood (Mateer, 1993). Cortical and subcortical structures are present by birth.

By three years of age, neuron and glial cell counts are fixed and myelinization is 90% that of adult levels. While the subcortex is myelinated by three years of age, myelinization of intracortical cells continues up to 60 years of age (Mateer, 1993, p. 4).

Myelinization of intracortical cells refers to those fibers connecting the cerebral hemispheres, the largest group of those being the corpus callosum. Neural development of these interconnecting fibers involves a process of elimination of fibers rather than the emergence of new structures. In referring to the corpus callosum, Mateer indicated that "Myelinization of the fibers begins at the end of the fetal period, increases in childhood and plateaus by seven to ten years" (p.5).

Neurophysiological evidence made possible through the use of electrophysiological measures indicates that the cerebral hemispheres of humans develop at differing ages and rates (Thatcher, Walker, & Giudice, 1987). Thatcher and associates found two growth patterns that are involved in hemispheric development: one being a continuous growth progression and the second being various growth spurts. Timing of the growth spurts appears to coincide with the developmental milestones of Piaget. In addition, these researchers have also suggested that hemispheric development is accomplished through elimination of connections throughout development. A similar pattern is seen in development through elimination of intracortical cells.

As noted by Mateer (1994) and Thatcher, Walker, and Guidice (1987), significant neural development of the cortex advances through mid-childhood. The patterns of neural development lend support to the suggestion by Colcord and Gregory (1987) that as children and adolescents develop, neuromotor competencies increase with age. With increased neuromotor competencies, integrated control of the articulatory, phonatory, and respiratory systems is maximized. This motoric developmental process reinforces the need for systematic study of naturalness patterns of children and adolescents to determine any emergent process of naturalness that may be attributed to neurological structures developing concurrently.

Research of systematic development of communication skills related to children's increasing age has been conducted by Dawson (1929) and Kowal, O'Connell and Sabin (1975) who studied the development of rate and disfluencies, respectively. Dawson studied the development of speech rate patterns in 200 children at twelve grade levels by measuring the number of phonemes produced in selected 15 second periods. He concluded that rate develops predominately in the grades one to three, followed by smaller, yet steady, increases as children approach grade twelve. Speech rate development also varied by gender throughout development. Dawson noted that girls spoke faster than boys until approximately the age of twelve (girls producing 80 phonemes and boys producing 70 phonemes in 15 second intervals). Then, between the ages of twelve and nineteen, minor inconsistent reductions and accelerations in speed were noted by both males and females. At age 20, males produced approximately 92 phonemes in a 15 second interval compared to the 80 reported for females.

Kowal, O'Connell, and Sabin (1975) were able to illustrate varying aspects of fluency development in their study of the disfluencies produced by 168 children at seven different age levels (twelve boys and twelve girls at each level). Frequencies of nonfluencies (normal), duration of unfilled pauses, and length of utterance were the specific attributes of fluency that changed. Results revealed that vocal disfluency development is a complicated process that fluctuates throughout childhood and adolescence, with increasing disfluencies evident between kindergarten and fourth grade. Results also showed a slight decrease in grade six, an increase by grade eight, and finally, a decrease at grade twelve, similar to the disfluency rate seen during kindergarten. Kowal and associates also noted not only that the frequency of disfluencies fluctuates but also that the types of vocal disfluencies altered during development. Figure 1 illustrates the number of nonfluencies indicated by type and age. Four types of nonfluencies were differentially noted: parenthetical ("You know" noted as PR); false starts (word, phrase, or utterance correction noted as FS); filled pauses ("um," "ah," "hm" noted as FS); and repeats (repetition of any portion of utterance noted as R). The average number of all types of disfluencies combined is represented as "M" (x) calculated across all ages studied.



**FIGURE 1**: Types of disfluencies noted during development. PR=Parenthetical; FS=false starts; FP=filled pauses; R=repeats; and M=mean of all types of disfluencies noted. (Kowal, O'Connell, & Sabin, 1975).

For example, at younger ages (kindergarten) false starts and repetitions of an element of a word or phrase were prevalent. Yet, by second grade these types of nonfluencies decrease, whereas filled pauses ("um", "ah") increase. By fourth grade, false starts and "parenthetical" utterances have greatly increased, repeats have slightly increased, and filled pauses have reduced. Sixth grade is marked by a decrease in all types of disfluencies, with filled pauses revealing the slightest reduction of all types. Eighth grade is characterized by a dramatic increase in parenthetical remarks and a slight increase in repeats, whereas filled pauses and false starts reduce somewhat. Grade ten is characterized by increases in all types of disfluencies except repeats, which show a decline. By grade twelve, a decline is seen in all types of disfluencies. Examination of the average number of disfluencies (the "M" on Figure 1 graph) reveals that two general periods of increase occur during these ages. The first occurs between

kindergarten and fourth grade and the second during sixth and tenth grades.

As with rate and fluency, use of linguistic stress develops as children

### grow. Starkweather (1980) maintained that

stressed syllables require more time and more effort and are consequently less fluently produced than unstressed syllables. In addition, the beats of lexical and linguistic stress create a characteristic rhythm by momentarily slowing down the speed of articulatory movement for stressed syllables in order to place extra time around them, and then speeding up again for unstressed syllables. The irregularity with which articulation of speech changes, along with the high speed of syllable production itself, taxes the speech mechanism to it limits, and when demand is made for even faster speech, further unstressing occurs leading to a reduction in stress contrast. At normal speeds, stress is likely to be found on words that carry more information (p. 167-168).

As children use speech, the patterns of stressed and unstressed syllables produce "speech-like" rhythm. With age, children gradually acquire more control over use of stress, timing and alternating rhythm.

Starkweather (1980), in summarizing his review of normative development literature, indicated that "...speech rate follows a steady, if step-wise, course of development in children and a somewhat different course in boys than in girls" (p. 193). Girls tend to speak at faster rates than boys until age twelve when a change occurs and boys' speech rate becomes more rapid than girls. Then, between the ages of "...twelve and nineteen, there is minor see-sawing back and forth between the sexes, until at age twenty boys seem to talk much faster" (Starkweather, 1980, p. 158). A correspondence in the development of rate of speech and types of nonfluencies is seen. As rate of syllable production increases, rate of nonfluencies also increases. Starkweather (1980) has also suggested that rhythm becomes irregular as children attempt to produce a continual, rapid flow of information.

The significance of this information, as it may relate to speech naturalness, is that speech development throughout childhood and adolescence. particularly demonstrated in the course of the parameters of fluency, is a developing process. Children's speech development is marked by considerable and noticeable change from age 5 through age 18, as physical, cognitive, and affective systems mature. As demonstrated in studies of fluency development, data related to speech nonfluencies, rate, and stress patterns depict variance among age groups that continues throughout childhood and adolescence. Studies of naturalness with stutterers have reinforced the relationship of fluency to naturalness ratings (Ingham, Gow, & Costello, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985: Ingham & Onslow, 1985: Martin, Haroldson, Triden, 1984). In general, findings support the concept that faster and more fluent speech was rated as more natural. Consequently, given the relationship of rate and fluency to naturalness seen in adult populations, consideration needs to be given to the guestion: how is naturalness perceived in the speech of children and adolescents during later childhood and earlier adolescence, a period marked by significant physical, cognitive, and affective growth?

#### Naturalness Ratings: Audio vs. Audiovisual Data

With the exception of a study done by Martin and Haroldson (1992), all perceptual studies of speech naturalness incorporated the use of speech samples that were audio-taped. Martin and Haroldson attempted to determine whether ratings of speech naturalness would be affected by the medium of presentation: audio-only versus audiovisual. Ten adult stutterers and ten nonstutterers, matched for gender and age (within 2 years), offered 1 minute samples of spontaneous speech that were both audio and videotaped. No effort was made to remove disfluencies from the stutterer's samples, but all reference
to the topic or word "stuttering" was edited from the samples. All samples were rated by 24 unsophisticated listeners. Results revealed that the naturalness of the speech of nonstutterers was rated similarly on both audio taped and video taped presentations. As a group, the naturalness ratings of the stutterers were systematically higher (more unnatural) than those of the nonstuttering subjects. However, a comparison between the audio-only and the audiovisual presentations of the range across individual stutterers' samples revealed that the audiovisual presentations of the individual stutterers rated more unnatural than the audio-only sample (between .21 and 1.51 scale points on a Likert 1-9 point scale). As Martin and Haroldson reiterated, studies of perceptual differences between the stutter-free speech of treated stutterers and normally fluent cohorts have employed audio-only procedures. "It would be informative to determine whether the addition of the visible aspects to the audio speech samples of successfully treated, stutter-free stutterers would affect perceived naturalness judgments of the speech samples" (Martin & Haroldson, 1992, p. 526). The possibility exists that visible perceptual differences may add cues to listeners that would affect naturalness ratings.

#### Perceptual Correlates of Naturalness

Speech naturalness is a multidimensional property. "Unlike a simple attribute, such as height, there is no single measurable property of speech which determines the perception of speech naturalness" (Johnson, 1987, p.13). Studies of naturalness with stutterers have reinforced the concept that rate and fluency are perceptual parameters of naturalness ratings by demonstrating that faster, more fluent speech was rated as being more natural (Ingham, Gow, & Costello, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Ingham & Onslow, 1985; Martin, Haroldson, & Triden, 1984). Additional perceptual cues of

naturalness exist, because studies controlling rate and fluency continued to find differences in ratings of speech naturalness in the fluent speech of adult stutterers and age-matched counterparts (Ingham, Gow, & Costello, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985). Although rate and fluency are contributing factors to perceptions of naturalness, other dimensions must play a role.

Hotchkiss (1973) attempted to identify perceptual cues used by listeners when differentiating the fluent speech of stutterers from their normally fluent counterparts. Listeners were asked to list cues in the speech samples that allowed them to detect speech of stutterers. Listeners who accurately distinguished stutterers from nonstutterers could list various cues that assisted in their choices. Cues identified as helpful in distinguishing stutterers from nonstutterers are listed in Figure 2. Raters who were unreliable were trained using information from the reliable raters. With the assistance of perceptual cues, the reliability of listeners improved significantly. Hotchkiss's involvement of listener identification of cues was valuable in training listeners who could not distinguish the normally fluent speech from the stutter-free speech of stutterers.

Perceptual Cues Differentiating Stutterers Fluent Speech from Normally Fluent Speech							
Laryngeal Behaviors	Rate/Pause Behaviors	Articulatory Behaviors					
Laryngeal Tension Vocal Tremor Monotone Low Intensity	Slow rate Abnormal Pauses Hesitations	Exaggerated word initiation Longer Syllable Duration Imprecise Articulation					

FIGURE 2. Perceptual Cues used by highly reliable listeners when differentiating fluent speech of stutterers from normally fluent cohorts. From Hotchkiss (1973)

Later, in an application of Hotchkiss's work related to "naturalness," Johnson (1987) stated that "...identifying the measurable correlates of speech naturalness would enhance the value of perceptual ratings of speech naturalness as a measure of speech" (p. 13). Implications for future study could then incorporate the relationships of each perceived attribute of naturalness and its correlation to a speaker's naturalness performance. In addition, therapeutic programs could include identified correlates of naturalness to enhance speech production, specifically fluency programs. Although somewhat varied, some stutterers have been able to improve their naturalness ratings after being given feedback (Ingham, Martin, Haroldson, Onslow, & Leney; 1985; Ingham & Onslow, 1985). However, with knowledge of specific naturalness correlates, treatment programs could become directional, focusing on the modification of the naturalness correlates in order to develop the ultimate target: natural sounding speech.

Johnson (1987) attempted to identify some of the dimensions of speech naturalness in the speech of stutterers (fluent and disfluent) and normally fluent speakers. One hundred twenty listeners (unsophisticated) who were adult native English speakers rated 30 second speech samples from 30 different speakers; 10 from normally fluent speakers, 10 from stutterers displaying disfluency, and 10 from stutterers' fluent speech. Naturalness ratings were obtained on each of the 30 second samples of speech. Later, each speech sample was rated with a 1-5 equal interval scale along the following parameters: rate, fluency, and expressiveness. These parameters were chosen because of their previous use in studies to evaluate speech quality following stuttering treatment (Goldsmith & Anderson, 1984; Ingham & Packman, 1978; Perkins, 1973). Johnson concluded that rate and fluency were important dimensions of naturalness. She also stated that "...listeners may rely on different cues when judging the naturalness of

various types of disordered speech" (p. 78). Although Johnson (1987) stated that identification of naturalness correlates was imperative, listener identification or generation of cues (as in Hotchkiss' s design) was not employed. Rather, attributes of rate, effort, and expressiveness--retrieved from previous fluency studies--were correlated to listeners' judgments of naturalness. Analyses of Johnson's results present the question: what more specific cues do listeners use when judging the naturalness of children's speech? Furthermore, Johnson indicated that

if ratings of speech naturalness are to have clinical value, studies of the speech naturalness of disordered populations must address the question of what are the "normal" and "abnormal" ranges of speech naturalness...Until that range is identified, studies which attempt to improve unidimensional ratings of speech naturalness are limited to demonstrating that change has taken place without being able to demonstrate that a speaker has moved from "abnormal" to "normal speech naturalness (p. 78).

#### The Naturalness Scale: Construct Validity

Onslow and Ingham (1987) published a paper reviewing measurements of speech quality and management of stuttering. In this article, the authors reinforced the need for a measure of speech quality, citing the criticism of unnatural sounding speech patterns as a result of fluency shaping programs. They stated that "...the need for such a measure has become increasingly urgent because of the growing use of therapies that employ unnatural sounding patterns" (p. 2). Onslow and Ingham concluded that perceptual judgments would serve an important function in therapy management but would only flourish following the establishment of refined validity and reliability measures.

Investigations of the construct validity of scaling methods have determined that use of an equal appearing interval scale (EAI) such as the 1-9 naturalness scale (Martin, Haroldson, and Triden, 1984) is appropriate for scaling the complex dimension of speech naturalness (Metz, Schiavetti, & Sacco, 1990; Schiavetti, Martin, Haroldson, & Metz, 1994). Perceptual judgments of speech naturalness are complex and multidimensional (Johnson, 1987; Metz, Schiavetti, & Sacco, 1990). Stevens (1975) has indicated that perceptual dimensions can be defined as one of two perceptual continua: prothetic or metathetic. Stevens (1975) defined these continua by indicating that

The prototypes of the two kinds of perceptual continua are exemplified by loudness and pitch. Loudness is an aspect of sound that has what can be best described as degrees of magnitude of quantity. Pitch does not. Pitch varies from high to low; it has a kind of position, and in a sense it is a qualitative continuum. Loudness may be called a prothetic continuum, and pitch a metathetic one. The criteria that define those two classes of continua reside wholly in how they behave in psychophysical experiments (p. 13).

It is important to determine to which continua, prothetic or metathetic, an attribute belongs because measurement recommendations for each continuum differ. A prothetic continuum "... is an additive, quantitative continuum that is best scaled with direct magnitude estimation (DME) because observers cannot subdivide a prothetic continuum into equal intervals" (Metz, Schiavetti, & Sacco, 1990, p. 516). In contrast, perceptual judgments which are defined as metathetic are a substitute, or qualitative dimension which may be scaled by using either DME or EAI methods (Metz, Schiavetti & Sacco, 1990; Schiavetti, Martin, Haroldson, & Metz, 1994). Two studies have determined that the continuum of speech naturalness is metathetic. Metz, Schiavetti, and Sacco (1990)--using audio-taped speech samples--and Schiavetti, Martin, Haroldson, and Metz (1994)--using videotaped speech samples and a psychophysical comparison of scaling data--concluded that scaling speech naturalness, whether audio or videotaped, was a metathetic continuum. As a metathetic continuum ...either interval scaling or direct magnitude estimation is an appropriate procedure for the measurement of this dimension" (Metz, Schiavetti, & Sacco,

1990, p. 523). In addition, Metz and associates promoted continued use of the 1-9 naturalness scale (Martin, Haroldson, & Triden, 1984) to assist in similar and comparable research findings.

#### **Research Directions from Current Literature**

Systematic study of naturalness ratings of normally fluent children has considerable merit. Given the scant, yet inquisitive nature of the studies comparing normally fluent children and their stuttering counterparts, specific information relating to naturalness ratings of normally speaking children is nonexistent. While Colcord and Gregory (1987) conclude that all children, normally fluent and otherwise, have speech neuromotor competencies which are maturing to adult potentials, systematic study of naturalness ratings of normally fluent children would assist researchers in comparing naturalness ratings of age cohorts with communication disabilities. Particularly important are the establishment of normative naturalness data of normal speaking children as they approach adolescence and acquire more mature neuromotor systems. In addition, another advantage of a broad judgment of speech naturalness, such as the judgments required by the 1-9 scale, is that it can accommodate the variance in interactions that speakers have.

Since the initial application of the "naturalness scale" in the measurement of the speech quality of stutterers (Martin, Haroldson, & Triden, 1984), many studies have validated its usefulness as an assessment and therapy tool. Predominately, research has explored the naturalness of adult stutterers' speech, in both stuttering and stutter-free contexts. Using the 1-9 point interval scale published by Martin, Haroldson, and Triden (1984), mean naturalness values have been published for adult nonstutterers that range from 2.12 to 3.55 and for the fluent speech of post therapy stutterers that range from 4.26 to 5.92. Studies rating naturalness of stutterers' speech that included disfluency have reported mean values of 6.52 (Martin, Haroldson, & Triden, 1984) and 6.4 and 6.81 (Martin & Haroldson, 1992). Research results in the literature generally agree that the application of the concept of naturalness and the "Naturalness Scale" is a reliable tool for scaling perceptual differences of speakers in singular settings as well as over time. Most researchers agree that the fluent or stutter-free speech of non-fluent adults remains perceptually distinguishable from the speech of normally fluent counterparts.

With the exception of the Martin and Haroldson (1992) study, all investigations have relied on audio-only speech samples in measuring naturalness. Martin and Haroldson examined the role of audio only versus audiovisual data in rating the naturalness of speech of stutterers and nonstutterers. In order to gain a full perspective of communication naturalness in "natural speaking" environments which acknowledges the role of visual communication characteristics, additional audiovisual study of speech samples is imperative, particularly in investigations of perceptual cues of speech naturalness.

Few studies have attended to perceptual differences in the fluent or stutter-free speech of stuttering children in comparison to their normally fluent cohorts. Those studies that have addressed this issue (Colcord & Gregory, 1987; Krikorian & Runyan, 1983) have examined a wide range of children (ages 4;1 to 9;0) as a singular group. In addition, both studies pursued the format of identifying a speaker as a stutterer or a normally fluent speaker. Neither study pursued the concept of naturalness of the speakers using the "Naturalness Scale". Both studies utilized a forced choice model with regard to the identification of each of the subjects used as a stutterer or normally fluent speaker. Conclusions reached in the two studies available with children suggest

that studies of perceptual differences between the fluent speech of stuttering and nonstuttering children need to take into account children's neuromotor maturation, the possibility of therapy artifacts developing in children's speech, and the possibility of a variable listener standard in judgments of children's fluency as opposed to that of adults. However, to date no research has aimed at a systematic study of naturalness characteristics of normally fluent children during specified chronological age periods in order to later compare naturalness ratings of children with communication disorders.

The only published application of the naturalness scale to children's speech (5 stuttering adolescents) was recorded in a study by Ingham and Onslow (1985). This study was primarily concerned with the effects of feedback of naturalness scores on treatment outcomes and did not explore the question of whether or not perceptual differences in the fluent speech of stuttering children and normally fluent children exist. Thus, to be able to compare the speech naturalness of CDO with their normal-speaking peers, naturalness ratings of normally fluent children need to be established.

In summary, there is much to be learned about naturalness. Specifically, many research gaps exist with regard to the application of "naturalness" ratings with children. Recent perceptual studies aimed at comparing normal speaking and CDO children have not incorporated the use of a naturalness scale. Perhaps most needed, yet still not examined, is the systematic comparison of naturalness ratings of normally fluent children that may account for age or developmental maturation. While information which has been obtained in studies of perceptual speech differences of normal speaking and disordered populations, most research designs have incorporated a one-to-one subject match (1 normal speaking child to 1 communicative handicapped child). Given the speculation regarding the potential variation in neuromotor speech

development in childhood and adolescence, the one-to-one age match model provides limited information relating to these naturally occurring variations in developing children. In fact, results of the two major studies aimed at perceptual differences in the speech patterns of normally fluent and stuttering children both revealed that distinctions in the speech patterns of these two groups were difficult to detect. The degree to which a larger normative representation may result in greater differentiation between the speech of normally fluent and fluent stuttering children would be a beneficial research effort.

Other research limitations have involved the exclusive use of audio-only speech samples, with the exception of one study with adults (Martin & Haroldson, 1992). Given that researchers continue to wrestle with identification of cues to the attribute "naturalness," it seems restrictive to study speech samples without the benefit of visual cues. As communication technology continues to develop, applications using visual displays will be included. For example, distance communication of the future via phone and computer interface will have a visual component. Such trends are now operative in computer interfaced, interactive, educational technology linking classrooms throughout the United States and Europe.

It has long been accepted in the study of communication disorders that visual components are crucial to the understanding and treatment of some disorders, most critically stuttering, articulation and voice disorders (Van Riper & Emerick, 1990). By establishing verbal and nonverbal correlates of "speech naturalness" as suggested by Johnson (1987), therapeutic programs (particularly fluency enhancing programs) could then include directly targeted naturalness parameters to enhance speech production. Therefore, if the ultimate clinical application of naturalness is to affect treatment programs, noting the visual components of naturalness is as pertinent as noting verbal attributes.

Studies have not adequately identified perceptual cues of naturalness. While studying the perceptual difference and recognition of speech of normally fluent speakers and fluent speech of adult stutterer cohorts, Hotchkiss (1973) asked listeners to list cues in the speech samples which allowed them to detect the speech of stutterers. Results indicated that listeners, who accurately distinguished stutterers from nonstutterers, could list various cues that assisted in their choices. By training raters who previously had not performed reliably in the rating task, the reliability of listeners improved significantly. Hotchkiss's involvement of listener identification of cues was valuable in training listeners who could not distinguish normally fluent speech from stutter-free speech of stutterers. Listener nomination of cues to naturalness would be extremely valuable in identifying the correlates of naturalness. Although Johnson (1987) stated that such identification of cues has not been employed in studies subsequent to Hotchkiss's (1973).

# Purpose of the Study

As previously stated, application of Johnson's (1987) results present the

question of what cues listeners use when judging the naturalness of children's

speech. Furthermore, Johnson indicated that

if ratings of speech naturalness are to have clinical value, studies of the speech naturalness of disordered populations must address the question of what are the "normal" and "abnormal" ranges of speech naturalness....Until that range is identified, studies which attempt to improve unidimensional ratings of speech naturalness are limited to demonstrating that change has taken place without being able to demonstrate that a speaker has moved from "abnormal" to "normal speech naturalness (p. 78).

Consequently, the purpose of this study is to identify and compare

speech naturalness ratings of normally speaking adolescents ages 8-16.

Specifically, the interest of this study is to

1. establish normative naturalness data for normal speaking children between the ages of 8-16,

2. establish any differences in naturalness ratings for normal speaking children that are attributed to age,

3. establish any differences in naturalness ratings for normal speaking children that are attributed to gender,

4. establish the degree of speech naturalness variability, leading to the range of normal naturalness at age demarcations,

5. establish the degree of listener intra- and inter-reliability on judgments of naturalness made,

6. identify perceptual cues of listeners that account for naturalness ratings interpreted as highly natural and highly unnatural,

7. determine the influence of the weighted perceptual cues on the group naturalness ratings.

#### Chapter 2

#### METHODOLOGY

In order to establish naturalness normative data of 8 -16 year olds and to compare effects related to gender, age, and naturalness range, the following methods were employed.

#### **Speakers**

Two speaker groups were used for this study: normal speaking children and adolescents and age-matched communicatively impaired counterparts.

<u>Normal Speakers</u>: Sixty children--30 males and 30 females equally distributed at the age levels of 8, 10, 12, 14, and 16 years--were chosen from area elementary and secondary schools. Participants were recruited via advertisements, flyers, and through the assistance of school administrators. Each participant passed a speech screening prior to selection administered by the researcher. Speaking participants ages 10 and above read the <u>My</u> <u>Grandfather</u> passage, and 8 year old participants were administered the <u>Goldman Fristoe Test of Articulation</u>. In addition, a case history, as provided by the parent or guardian (see Appendix A), was reviewed to ascertain the presence of any concomitant medical conditions that may confound results. Six males and six females were selected at each age level. In order to control for age variability, speakers at each age level were restricted to  $\pm$ 4 months from the mid year of birth anniversary. For example, for inclusion at level 8 years, a participant's age must be between 8 years 2 months and 8 years 10 months.

Communicatively impaired cohorts: Approximating the national statistic that as many as one in every ten persons has a communication impairment (Van Riper & Emerick, 1990), a group of children representing this portion of the population at large was also included as speakers. Ten children (approximately 14% of total speaker sample) between the ages of 8 and 16 served as the communicatively impaired sample. Individual case history information, as provided by the parent or guardian (see Appendix A), was reviewed to ascertain the history of specific communication disorders present. Two children, representing each of the following types of communication disorders, were included: articulation, fluency, voice, language, and hearing. The communication impaired cohorts were judged by ASHA certified speech and language pathologists not involved in this study to have moderate-to-severe manifestations of their specific disorders. This degree of severity was chosen in order to allow the potential for the full use of the naturalness scale range. National statistics do not provide age and disorder specific data because of the difficulty of accurate reporting and the occurrence of concomitant disorders are not accounted for adequately (Van Riper & Emerick, 1990). In addition, Hegde (1995) has stated that contributing to the inaccuracy of incidence reporting is the fact that data collected have been dependent on the entry criteria of the agencies reporting; thus, the data may represent only the number of individuals eligible to receive treatment in that facility. Consequently, the reported national statistics may under-represent the actual number of individuals having communication disorders. Given this possibility, exceeding the national average by including 10 communicatively disordered children (14% of total speaker sample) allowed for a more adequate representation of communication disabilities seen in childhood. In addition, to reflect the prevalence values related to gender, the communicatively impaired cohort group included 7 males and 3 females. Case

history information relating each communicatively impaired child's history was reviewed by an ASHA certified speech and language pathologist, independent of other aspects of this study, to verify eligibility for this sample.

#### **Speech Samples**

Each speaker was video-taped in color using an RCA Video HI-8 camcorder, placed six to eight feet from the participant. Speakers were taped while sitting in a chair at a table, so that a front view of the upper body was clearly visible. After being seated, each child was asked some introductory questions to assist habituation to the setting and maximum comfort (see Appendix B). Following these questions, each speaker was instructed that his/her task was to tell the examiners what "...kids like to spend their time doing when they are not in school." In addition, the examiner presented each speaker with 6 possible suggestions related to this topic if a conversation topic was not readily accessed by the child. Each of the 6 suggestions was presented to subjects on five by eight inch index cards and placed within view on the table. As each suggestion was placed on the table before the subject, the examiner read the card outloud to preclude interference by any reading disability a child may have had. A minimum of three minutes of monologue or conversation was elicited from each speaker. In addition, a speaker was allowed to select more than one topic if needed to facilitate three minutes of recorded speech. Appendix B illustrates the instructions and suggestions provided for each speaker.

#### Speech Sample Validity, Reliability, and Randomization

The researcher reviewed all speech samples and selected a 30 second continuous speech segment. The 30 second sample of speech chosen was the

first continuous sample of each subject's speech (without extraneous pauses) from the middle portion of the sample. Once all seventy segments were selected, an ASHA certified speech and language pathologist, without knowledge of the project, reviewed each of the seventy 30 second segments to verify 1) the normative speech pattern produced by each child in the normal speaking group and 2) the presence of a communication disorder for each child in the group of communicatively impaired children.

Following verification of speech sample eligibility for each of the 70 speech samples, 14 speech samples (20% of all speakers) were chosen to serve as duplicate presentations for purposes of intra-rater reliability analysis. Twelve of the 14 were from the normal speaking group and 2 samples were from the communicatively impaired group (CDO). Consequently, eighty-four 30 second speech samples (70 original + 14 duplicate presentations) were rated by each listener.

Prior to development of the CD-ROM, all speech samples were ordered using a quasi-random distribution procedure. To accomplish the quasi-random ordering of the speech samples, the 60 normative samples were randomly selected into five groups, each containing 12 samples. Then of the 10 speech samples from the CDO speakers, 2 samples were randomly selected and included in each of the 5 subgroups previously established. As a result, each of the 5 subgroups contained a total of 14 speech samples. The 14 speech samples in each of the five subgroups were randomly ordered. Finally the order of the 5 subgroups was randomized to complete the ordering of stimuli to be rated by listeners. For purposes of intra-rater reliability, the 14 speech samples chosen to serve as duplicate presentations were randomly ordered in a subgroup presented following the 70 randomly ordered speech stimuli in a manner similar to Martin and Haroldson's (1992) procedure. Consequently, eighty-four speech

samples were presented with the first 70 representing samples from the 60 normal speaking and 10 communicatively impaired children, and the final 14 segments representing the samples chosen as duplicate presentations.

#### **CD**-ROM Construction

A CD-ROM of all speech samples was made as the audio-visual medium for raters to view. Technical construction of the CD-ROM was conducted by professional videographers in a media resource center at a mid-western university. The 30 second speech sample from each of the seventy children included in this study was selected from the middle portion of the original 3 minute video-taped sample. Each selected 30 second video tape speech sample was transferred to digital video using *MEDIA 100* software. Each digitized segment was edited into a *QUICKTIME* file. Once stored as a computerized file via *QUICKTIME*, all files were transferred to a *MACINTOSH 8100* audio visual authoring station, using the software program *DIRECTOR*. The software program *DIRECTOR* governed the scripting and navigation system of the CD-ROM playback, so all recorded speech samples were viewed in the order specified. Once the navigation system for all speech samples was completed, a CD-ROM was created using an audio-visual program called *Toast*.

#### Listeners

The listeners chosen for rating tasks were 39 adult native English speaking persons without specialized knowledge of communication disorders. The listeners were chosen from a small mid-Michigan community having a regional university. Listeners were recruited through newspaper ads and flyers placed in local businesses and paid \$10.00 for participation in this study. The listener group was composed of 19 males and 20 females between the ages of 18 and 56. Gender distribution in the groups of listeners conformed to national gender statistics (95.1 males to 100 females nationally), according to the U.S. Bureau of the Census (1994). Each listener completed a demographic survey giving information specific to his/her background in order to eliminate any persons with specialized training or coursework in speech and language disorders. Additional information obtained about the listeners was used to gauge the amount of interaction with children within the age ranges studied (see Appendix C). Each listener also passed a bilateral puretone hearing screening at 20 dB for the frequencies of 500, 1000, 2000, 4000, and 8000 Hertz.

#### Listening Tasks

One listening task containing three activities was incorporated into the methodology of this study. Each listener was seated individually in a quiet room, approximately 10' by 10' in dimension, during the entire rating/response period. The listening task incorporated three separate activities: 1) rating each speech sample provided using a 1-9 Likert scale (Martin, Haroldson, & Triden, 1984); 2) listing perceptual cues relaying why (s)he rated a speaker as "at or near highly natural" or "at or near highly unnatural;" and 3) weighting each of the perceptual cues listed on a provided bar scale. A researcher remained in the room with each listener for monitoring and assistance purposes.

To accomplish the listening task, the CD-ROM containing eighty-four 30 second speech samples was inserted into a *POWER MACINTOSH* Computer (Model 6100/66), which was connected to an *APPLE COLOR PLUS* 14 inch monitor. The listeners were instructed how to use the *APPLE DESKTOP MOUSE II* to advance the CD-ROM program to play each speech sample in the sequence to be rated. Each listener wore a *SONY HEADSET* (MDR-009) to maximize audio reception of the speech sample and reduce ambient noise in the

listening area. Following directions related to equipment use, each listener was given the following instructions:

# Activity 1: Rating Naturalness

To instruct the raters, adapted language employed by Martin and

Haroldson (1992) was read and given in written form.

We are studying what makes speech "natural" or "unnatural." You will be played eighty-four 30 second video-taped speech samples. Each sample will be introduced by the sample number. Your task is to rate the speech naturalness of each sample. If the speech is highly natural to you <u>circle</u> the number 1 ("highly natural") on that sample's scale in the packet in front of you. If the speech is highly unnatural to you, <u>circle</u> the number 9 ("highly unnatural") on that sample's scale. If the speech is somewhere between highly natural and highly unnatural, <u>circle</u> the appropriate number on the scale. Do not hesitate to use the ends of the scale (1 or 9) when appropriate. Be sure to rate each sample. An example of the rating scale is located on the bottom of this page.

At the end of each sample, the computer monitor will fade black and pause to signal you that the sample has ended. When you see the black screen, make your naturalness rating then click on the red arrow to go on to the next sample.

Naturalness will not be defined for you. Make your rating based on how natural or unnatural the speech is to you.

You may view each sample only once. This <u>is not</u> a timed task, so proceed at the speed comfortable to you. Breaks are permitted if needed. Remember, however that it is important that you rate each sample provided.

Any Questions?

### Sample of rating scale sheet: (Please CIRCLE the number of your rating)

SAMPLE #:\_\_\_\_

HIGHLY						H			Y
NATURAL						UNN			I <b>AL</b>
1	2	3	4	5	6	7	8	9	

Following these instructions, three practice ratings were provided to listeners. Stimuli used for these samples were three digitized, thirty second samples of normal adult speech, one female and two males. Practice samples preceded all of the samples to be rated in this study. After each listener rated the practice stimuli, the following instructions were provided:

Do you have any questions about the task now that you have practiced?

If no questions were asked, the researcher continued with the following

instructions:

At the end of this listening task I'll have further instructions for you. When you are ready, you may begin by double clicking the cursor on the red arrow below the practice sample.

# Activity 2: Listing Cues for Naturalness

After all samples were rated, the researcher instructed the listener by

reading and presenting the following written instructions on a worksheet:

# **Instructions to listeners:**

You have completed the first part of this job. For the second part of this job, I will play you the samples you rated as highly natural followed by the samples that you rated as highly unnatural. After each playback group, I want you to think about your recent use of the 1-9 point scale and using this piece of paper, list why you rated samples as either "at or near highly natural" or "at or near highly unnatural." You may list as many reasons as you wish for each column. Five spots are provided for you to list those items. If you do not need all five spaces that is fine. If you need more spaces, please turn the sheet over and use as much space as you may need. If you cannot think of a single word to describe a reason for rating, please describe what you mean, using a phrase or sentence. Take as much time as needed.

Any Questions?

Here are the samples you rated as "highly natural."

#### Sample of Listener's response sheet:

#### **ACTIVITY 2:**

#### <u>Please List and describe as many factors you heard or saw that helped you to</u> <u>decide speech samples were:</u>

At or near <u>"Highly Natural"</u>	At or near " <u>Highly Unnatural"</u>				
1	1				
2	2				
3	3				
4	4				
5	5				

The process for activity 2 began immediately after completion of naturalness ratings. Replays of all samples rated highly natural were conducted before replays of those rated highly unnatural.

The researcher reviewed the respondent's rating sheet from activity 1, noting the first speech sample that was rated as 1 (Highly <u>Natural</u>). In the event the listener's most natural rating was a 2, samples rated as 2 were then used in the replay activity. After identifying the first sample rated highly natural and while the listener reviewed its replay, the researcher simultaneously noted all other remaining samples for replay. The listener was required to listen to all replays prior to beginning activity 2. When the replays were completed, the researcher instructed the listener as follows:

Now, please proceed with your listing. Let me know when you are finished. Take as much time as needed.

The rater then completed listing factors (s)he identified as contributing to highly natural speech. The researcher then informed the rater by saying:

Here are the samples you rated as "Highly unnatural."

While the listener was writing factors related to natural speech, the researcher reviewed the respondent's rating sheet from activity 1, noting the speech samples that were rated as 9 (Highly <u>Unnatural</u>). In the event the listener's most unnatural rating was an 8, samples rated as 8 were then used in the replay activity. The listener was required to listen to all replays prior to beginning activity 2. When the replays were completed the researcher instructed the listener as follows:

Now, please proceed with your listing. Let me know when you are finished. Take as much time as needed. When you have completed this task, please give your list to the researcher.

# Activity 3: Prioritizing/Weighting factors listed

Once the listener completed activity 2, the researcher read and presented

the following instructions and worksheet:

Now that you have listed or described as many factors as possible, please distribute your reasons for rating along the provided bar scales. Please assign each reason you listed to occupy a portion of the bar, based on the amount you feel that item influenced your decision. Therefore, the reason most influential to you receives the largest portion of the bar, the next most influential receives the next largest portion, and so on, continuing to the least influential factor, which receives the smallest portion of the bar.

The following <u>example</u> demonstrates the use of the bar scale for a listing of:

Factors listed for what I like to do after work:

	1. Relax.	
2	Homework.	
3.	Go shopping.	

Using the scale bars provided, please indicate the amount of importance or influence each of the factors you listed played in your rating of naturalness.

Scaling Factors for:

"At or Near Highly Natural"

Scaling Factors for:

"At or Near Highly Unnatural"

#### **Review of Perceptual Cues Listed**

In activity 2, each listener was asked to list or describe the cues that enabled him/her to rate a speaker as either at or near "highly natural" and at or near "highly unnatural". In activity 3, each cue listed was assigned a bar scale value in proportion to its importance in determining natural or unnatural quality. In a procedure used by Hotchkiss (1973), the examiner--in concert with another certified speech and language pathologist--reviewed all cue descriptions in order to analyze and cluster cues listed into any classification of similar meaning. Cues that were listed by highly reliable listeners were analyzed related to assigned weight of influence in determining speech as highly natural and highly unnatural.

#### **Statistical Analyses**

The intent of this study was to observe and measure the perceptual listener phenomena of "naturalness." Consequently, the principal intention of this study was to describe the "naturalness" ratings of the normal adolescent population, with specific interest in the variables of gender and age. This purpose is what Doehring (1988) described as a study of "Group Description." Information is obtained under controlled conditions, but there is no interjection of an independent variable as in the creation of an experiment. Rather, the focus of this descriptive research was to observe the relationship between "attribute variables" of the population samples (Ventry and Schiavetti, 1986). The primary focus of statistical results was to describe the range of variation and differences in the gender and ages of the normal population sample examined.

Various statistical analyses were performed in order to evaluate the data obtained in the research questions listed. The data obtained from observations of listeners on the 1-9 Likert scale were ordinal. Ordinal scales imply that the item being evaluated can be "...arranged in ranks or levels such as "greatest to least or most severe to least severe " (Ventry & Schiavetti, 1986 p. 146). The descriptive statistics of group mean, mode, and standard deviation were used to outline the characteristics of the group of normal speaking children at each age group studied and for each gender. These computations allowed for the determination of variability or range of "naturalness" for each age group and gender studied.

Inferential statistics were used to establish relationships of variables and differences in group means. Differences related to the attribute variables of speaker age and gender were analyzed using a two-way analysis of variance (ANOVA) to determine whether any statistically significant differences in group factors of age and gender existed.

Both inter- and intra-listener reliability measures were a critical portion of this data analysis to examine the precision of listener judgments. Intra-rater reliability was included in this methodology through rate-rerate design procedures. In listening activity one, 14 speech samples were duplicated following presentation of the 70 tasks to be rated. Values assigned by each rater were analyzed in two ways. Initially, each listener's original ratings on the 14 stimuli were compared to the 14 duplicate presentations using two-tailed, paired sample t-tests. Secondly, the reliability formula which defined listeners as reliable if 75% of their duplicate speech naturalness ratings were within  $\pm 1$  interval from the original rating was employed (Ingham, Gow, & Costello 1985).

Since 75% of the fourteen duplicated reliability samples equaled 10.5, the criterion for intra-rater consistency was increased to 79% (11 of 14 samples). Consequently, all listeners' initial ratings were compared to their ratings of naturalness on each of the 14 duplicated samples and those which were within  $\pm 1$  interval on 11 samples or greater were considered reliable.

Inter-rater (listener) agreement (the extent to which listeners agreed in the rankings of each sample) was analyzed using the identical procedure used by (Tinsley & Weiss, 1975). Computing inter-rater agreement was accomplished by systematically comparing each sample rating given by each listener to the ratings of each of the remaining listeners' ratings. In doing so, a total of 34,720 paired rating comparisons were analyzed utilizing a computer application written for the *MACINTOSH* computer using the *MICROSOFT QUICK BASIC* SOFTWARE PROGRAM. From these data, percentage of rater agreement was computed for pairs rated identically,  $\pm 1$ ,  $\pm 2$  and so on to  $\pm 8$  scale values.

Each listener listed and weighted perceptual cues that influenced him/her in using the naturalness scale. Descriptive statistical analysis, consisting of the relative assigned weight of influence, determined how each of the cues rated contributed to the overall group rating of naturalness.

#### Chapter 3

#### RESULTS

The results of this study will be reported in four sections: 1) listener reliability and agreement; 2) overall considerations in the use of the naturalness scale; 3) naturalness data for children related to age, gender, and variability findings; and 4) identification of perceptual cue themes and their influence on listener judgments.

#### **Listener Reliability and Agreement**

Thirty-nine listeners (20 female and 19 male) who ranged in ages from 18 to 56 participated in this study. The average age of listeners was 22 years 4 months for females and 27 years 4 months for males.

Intra-listener reliability on replicated samples was analyzed prior to main effect analysis of this study to verify the responses of listeners and to make appropriate interpretations. Fourteen speech samples, 12 from the normal speaking group and 2 from the communicatively impaired group, were chosen to serve as duplicate presentations for purposes of intra-rater reliability analysis. The presentation of these samples followed the presentation of the original 70 quasi-random ordered samples. Intra-listener reliability was analyzed in two ways. Two tailed t-Tests were conducted, comparing the naturalness ratings of the 14 duplicate presentations to the original ratings these samples received (within the first 70 ratings). Results of the 39 paired sample t-Tests revealed that all of the listeners were reliable at the .05 level of confidence (See Appendix D for table of t - Tests).

However, closer examination of the listeners' responses was conducted in a similar manner used by Ingham, Gow, and Costello (1985), which defined listeners as reliable if 75% of their speech naturalness ratings on the same 1-9 Likert rating scale were within plus or minus 1 interval from the original rating. For example, if a listener rated an original presentation as a 4, then a re-rating of 3, 4, or 5, would be considered reliable. Given 75% of the fourteen duplicated reliability samples equaled 10.5, the criterion for intra-rater consistency was increased to 79% (11 of 14 samples). Consequently, all listeners' initial ratings were compared to their ratings of naturalness on each of the 14 duplicated samples and those which were within plus or minus 1 interval on 11 samples or greater were considered reliable. For this investigation, 32 of the 39 listeners met this criterion. Thus, because of their unreliable judgments, the remaining 7 listeners' ratings (Female 11, and Males 1, 3, 5, 6, 13, and 16) were eliminated from further data analysis. Table 2 illustrates each listener's reliability percentages on duplicate rating tasks.

Rater	# Samples within +1	# samples rated	total ratings of	Percentage	
	interval	identically	reliable	Reliability	
Female 1	5	6	11/14	79	
Female 2	4	10	14/14	100	
Female 3	5	9	14/14	100	
Female 4	1	13	14/14	100	
Female 5	4	9	13/14	<u>93</u>	
Female 6	4	9	13/14	93	
Female 7	7	4	11/14	<b>79</b>	
Female 8	4	8	12/14	86	
Female 9	6	8	14/14	100	
Female 10	8	6	14/14	100	
Female 11	4	4	8/14	57 *	
Female 12	3	11	14/14	100	
Female 13	2	11	13/14	93	
Female 14	1	12	13/14	93	
Female 15	3	11	14/14	100	
Female 16	2	12	14/14	100	
Female 17	4	10	14/14	100	
Female 18	4	9	13/14	93	
Female 19	0	14	14/14	100	
Female 20	9	4	13/14	93	
Maio 1	5	5	10/14	71*	
Male 2	4	10	14/14	100	
Male 2 Male 3	6	4	10/14	71*	
Male 4	4	8	12/14	86	
Male 5	4	5	Q/14	64*	
Malo 6	6	2	10/14	<u> </u>	
Male 0 Male 7	6	8	14/14	100	
Male 8	6	6	12/14	86	
Male 0	1	13	14/14	100	
Male 3	3	8	11/14	79	
Malo 11	1	13	14/14	100	
Male 12	5	15	14/14	100	
Male 12	3	5	10/14	71*	
Male 13	3	11	14/14	100	
Malo 15	3	10	12/17	03	
Male 16	<u>5</u>	<u></u>	<u>10/14</u>	<del></del> 57*	
Male 17	5 F	3	0/14 14/44	57 100	
Male 17	5	9	14/14	100	
Male 10	3	11	14/14	100	
Male 19	5		11/14	/8	

Table 2. Rater's degree of agreement on duplicated presentations

\* Raters not meeting reliability criteria

Using the 75% reliability rule, seven of the original thirty-nine listeners were eliminated from further data analysis. Therefore, the data from the thirtytwo listeners (19 female and 13 male), who ranged in age from 18 to 56, was used in analysis of the main effects in this study. Average age of the included listeners was 22 years 7 months for females and 29 years 4 months for males. Table 3 illustrates the age distribution of the 32 remaining listeners.

Remaining Listeners	Number	Mean Age	Median Age	Standard Deviation	
Females	19	23	19	9.6	
Males	13	29	27	10.6	

Table 3. Age distribution of 32 listeners included in study data.

#### **Rating range and Scale Use**

Cursory examination of listeners' responses revealed noteworthy observations that give credence to more specific findings. First, listeners did perform as instructed in the use of the scale. Eighty-four percent (27/32) of the listeners used the complete 9 point Likert scale. Of the five remaining listeners 4 used 8 intervals and 1 used a 7 point scale. Analysis of the use of the scale ends revealed that ninety-one percent of all listeners (29/32) used the scale value of "1" (highly natural). The three listeners who did not use the rating of "1" were female listeners. Ninety-four percent of all listeners (30/32) used the scale value of "9" (highly unnatural). The two listeners who did not use the rating of "9" were male listeners.

#### Inter-Listener Agreement

Listener (rater) agreement was defined as the degree to which raters assigned similar naturalness scale values to a given speech sample (Tinsley &

Weiss, 1975). Determination of inter-rater agreement was computed by comparing the rating assigned to a given sample by one rater to the naturalness ratings assigned by each of the other raters. Computing inter-rater agreement was accomplished by systematically comparing each sample rating given by each listener to the ratings of each of the remaining listeners' ratings. For example, the rating assigned to sample 1 by listener 1 was compared with the ratings for sample 1 given by listener 2 through 32. Next, the ratings assigned to sample 1 by listener 2 was compared to the ratings given by listeners 3 through 32. This process continued, comparing all possible listener rating pairs (496 comparisons) for each speaker sample. With 60 normal speaking samples, the total number of possible paired comparison observations was 29,760. For each of the 5 age levels in the normative sample group (8, 10, 12, 14, and 16), 5,952 rater comparison pairs were made. Table 4 provides the number of rater pair comparisons along each possible scale value and the cumulative percentage of inter-rater agreement for the normative group, in a manner initiated by (Martin, Haroldson, & Triden, 1984).

	Agreement and Difference in Scale Values								
Speakers	Identical	<u>+</u> 1	±2	±3	±4	±5	±6	±7	±8_
8 yr. olds [% of Total] [Cum points	1333 [22] 5] 1333	2051 [35] 3384	1202 [20] 4586	692 [12] 5278	366 [6] 5644	201 [3] 5875	88 [1] 5933	19 [.003] 5952	0
[Cum %]	[22] =========	[57]	[//]	[87]	[95] =======	[98]	=======	[100] =======	
10 yr. olds [ <u>% of Total]</u> [cum points]	1593 [27] ] 1593	2305 [38] 3898	1205 [21] 5103	517 [8] 5620	254 [5] 5874	67 [.8] 5941	11 [.2] 5952	0	0
[Cum. %]	[27] 	[65]	[86]	[94] =======	[99]	[99.8] 	[100]		
12 yr. olds [ <u>% of Total]</u> [cum points [Cum. %]	1787 [ <u>30]</u> ] 1787 [30]	2205 [37] 3992 [67]	1106 [19] 5098 [86]	506 [8] 5604 [94]	223 [4] 5827 [98]	94 [1.4] 5921 [99.4]	27 [.5] 5948 [99.9]	4 [.1] 5952 [100]	
14 yr. olds [ <u>% of Total</u> [cum points [Cum. %]	1846 ][31] 6] 1846 [31]	2306 [39] 4152 [70]	1041 [17] 5193 [87]	541 [9] 5734 [96]	153 [2.9] 5887 [98.9]	52 [.8] 5939 [99.7]	13 [.3] 5952 [100]	-	<u>.</u>
16 yr. olds [ <u>% of Total</u> [cum points [Cum. %]	2230 ][37] s]_2230 [37]	2204 [37] 4434 [74]	798 [14] 5232 [88]	337 [6] 5569 [94]	193 [2.8] 5762 [96.8]	100 [1.6] 5862 [98.4]	75 [1.3] 5937 [99.7]	10 [.2] 5947 [99.9]	5 [.1] 5952 [100]
Total Norm. Speak. [ <u>% of Total</u> [cum points [Cum %]]	8789 ] [30] s 8789 [30]	11071 [37] 19860 [67]	5352 [18] 25212 [85]	2593 [8] 27805 [93]	1189 [4.4] 28994 [97.4]	514 [1.7] 29508 [99.1]	214 [.7] 29722 [99.8]	33 [.1] 29755 [99.9]	5 [.1] 29760 [100]

<b>Table 4</b> .	Cumulative number and percent of inter-rater agreement pairs for the speech	
naturalne	ss ratings of the 60 normal speaking children.	

With the 10 communication disordered speech samples, a total of 4,960 paired comparison observations was made. Table 5 provides the number of rater pair comparisons along each possible scale value and the cumulative percentage of inter-rater agreement as calculated for the normative group, in a manner initiated by (Martin, Haroldson, & Triden, 1984).

Agreement and Difference in Scale Values									
Speakers	Identical	_±1	±2	±3	±4	±5	±6	±7	±8_
10 CDO speakers [# of pairs] [% of Total]	2264 [46]	1413 [28]	664 [14]	273 [5]	151 [3]	88 [1.8]	69 [1.4]	21 [.4]	17 [.4]
[cum points] [Cum %]	2264 [46]	3677 [74]	4341 [88]	4614 [93]	4765 [96]	4853 [97.8]	4922 [99.2]	4943 [99.6]	4960 [100]

Table 5. Cumulative number and percent of inter-rater agreement pairs for the speech naturalness ratings of the 10 communicatively disordered speakers.

### **Naturalness Characteristics of Speakers**

### **Range of Scale Values Received**

Considerable variability of naturalness ratings was seen for the normal speaking samples. The number of different scale values assigned to any one individual sample illustrates the range of natural ratings received. For the sixty normative speakers, the range of intervals any one sample received was a minimum of 3 and a maximum of 9. Perhaps the best illustration of the wide range of ratings the normative samples received by the listeners is noted in the fact that seventy-seven percent (46/60) samples received between six to nine intervals in naturalness ratings. Of the 60 samples of normal speakers, 27% (16/60 samples) were rated using an eight interval differential. Table 6 illustrates the range of intervals used for rating the normal speaker samples. For

example, 2 speakers received ratings that used all 9 possible scale values, 16

speakers received a range of 8 scale values, and so on.

Range of Scale value	Number of the 60 Normative Samples Receiving Interval range				
9 scale intervals	2				
8 scale intervals	16				
7 scale intervals	14				
6 scale intervals	14				
5 scale intervals	7				
4 scale intervals	5				
3 scale intervals	2				
2 scale intervals	0				
1 scale interval	0				

 Table 6. Differential range of rating intervals assigned to the 60 normal speaking samples.

A comparable distribution of interval rating ranges was found in the ratings of the ten speakers having communication disorders. Similar to the variation in ratings seen for the normative speaking samples, 70 percent (7/10) of the CDO samples received six to nine intervals in naturalness ratings. Table 7 illustrates the ranges of values received by the ten CDO samples.

 Table 7. Differential range of rating intervals assigned to the 10 communicatively disordered speakers.

Range of Scale value	Number of the 10 CDO Samples Receiving Interval range
9 scale intervals	2
7 scale intervals	4
6 scale intervals	1
4 scale intervals	2
3 scale intervals	1

## **Mean of Scale Values Received**

The naturalness rating tasks of this study deliberately included a population of communicatively disordered children. As expected, speech samples of normal speaking children received more natural ratings than the speech samples of those speakers with communication disorders. The mean naturalness ratings assigned by reliable listeners for the 60 normal speakers ranged from 1.22 to 5.66 while the mean naturalness ratings for the 10 communicatively disordered speakers ranged from 6.31 to 8.81. Table 8 lists the mean, range of mean values, and standard deviation specific to age levels and gender of the 60 normal speaking children.

<b>Table</b> 8	<b>3.</b>	Mean	rating,	range,	range v	vidth,	and standard	l deviation	of ratings	specific to
age and	d ge	ender	of the	60 norm	al speal	king c	hildren.			

AGE	Gender	Mean (X)*	Range of Mean Values *	Range Width	Standard Deviation
8	male	2.65	2.00 to 3.10	1.10	1.59
	female	3.45	1.94 to 5.66	3.72	2.13
	combined	3.05	1.94 to 5.66	3.72	1.92
10	male	2.63	1.94 to 3.78	1.84	1.43
	female	2.10	1.66 to 2.66	1.00	1.17
	combined	2.36	1.66 to 3.78	2.12	1.33
12	male	2.33	1.59 to 3.28	1.69	1.44
	female	1.96	1.44 to 2.31	.87	1.18
	combined	2.14	1.44 to 3.28	1.84	1.32
14	male	2.20	1.79 to 2.65	.86	1.29
	female	1.87	1.22 to 2.31	1.09	1.03
	combined	2.03	1.22 to 2.65	1.22	1.18
16	male	1.95	1.31 to 3.09	1.78	1.31
	female	2.03	1.22 to 3.66	2.44	1.47
	combined	1.99	1.22 to 3.66	2.44	1.39

# DISTRIBUTION of NORMAL SPEAKER RATINGS

\* 1 = Highly natural to 9 = Highly unnatural

Table 9 lists the mean, range of mean values, and standard deviation

specific to the group of children with communication disorders.

 Table 9. Mean rating, range, and standard deviation of ratings of the 10 communicatively impaired children.

SPEAKERS	Mean (X)*	Range of Mean Values *	Standard Deviation
CDO	8.00	6.31 to 8.81	1.49
* 1 = Highly natural to	o 9 = Highly unnatura	1	

# **DISTRIBUTION of CDO SPEAKER RATINGS**

# Mode of Scale Values Received

The mode ratings for the 8-16 year old normal speaking group in contrast to the CDO group showed the difference in naturalness values assigned. Ninety-seven percent of the normative samples had mode scores of 1, 2, 3, or 4 whereas, 100% of the CDO samples had mode scores of 7, 8, or 9. Table 10 displays the frequency of occurrence of each of the nine scale ratings as the mode score for the 60 normative samples and the 10 communicatively disordered speakers.

Naturainess Scale Value:	Frequency of scale value as mode:	Frequency of scale value as mode:	
	(60 normal speakers)	(10 CDO Speakers)	
1	30	0	
2	19	0	
3		0	
4	1	0	
5	0	0	
6	0	0	
7	2	1	
8		1	
9	0	8	
	60 ratings	10 ratings	

# Table 10. Frequency of occurrence as mode in Normative and communication disordered Samples

# Age and Gender Considerations

Group means of the 60 normal speaking samples were compared to determine whether ratings related to gender or age were significantly different. A two-way ANOVA revealed no significant differences in the naturalness ratings of males and females (p. < .715). However, statistical significance (p. < .005) between naturalness ratings of age groups was seen. Table 11 illustrates ANOVA results.

		ANO	VA	
Main Effect	Sum of Squares	Df	F	Sig.
GENDER	7.30	1	.135	.715
AGE	9.094	4	4.201	.005*
GENDER x A	GE 3.485	4	1.610	.186

 Table 11. Analysis of variance results for main effects of age and gender for the normal speaking group.

\* mean difference is significant at the .05 level

Post-hoc analysis was used to specify which of the possible comparisons accounted for the statistical significance noted in speaker ages. Significant

differences were noted between the mean ratings for the following multiple comparisons: 8 year olds and 12 year olds, 8 year olds and 14 year olds, and 8 year olds and 16 year olds. No significant differences were seen in the ratings between 8 and 10 year olds. In addition, no other significant differences were noted in any remaining age group comparisons. Table 12 illustrates the Tukey results.

Age	Group	Comparisons	Mean Difference (I-J)	Std Error.	Sig.
Α	ge (l)	Age (J)		-	
8	years	10 years	.6875	.305	.175
8	years	12 years	.9063*	.305	.034
8	years	14 years	1.0156*	.305	.013
8	years	16 years	1.0625*	.305	<u> </u>
10	years	8 years	6875*	.305	.175
10	years	12 years	.2188	.305	.951
10	years	14 years	.3281	.305	.817
10	vears	16 years	.3750	.305	.733
12	years	8 years	9063*	.305	.034
12	years	10 years	2188	.305	.951
12	years	14 years	.1092	.305	.996
12	vears	16 years	.1562	.305	.986
14	years	8 years	-1.0156*	.305	.013
14	years	10 years	3281	.305	.817
14	years	12 years	1094	.305	.996
14	vears	16 years	4.687	.305	1.00
16	years	8 years	-1.0625*	.305	.008
16	years	10 years	3750	.305	.733
16	years	12 years	1562	.305	.986
16	vears	14 years	-4.690	.305	1.00

Table 12. Post-hoc multiple comparisons of group means.

#### 

#### **TUKEY TEST**

---------

\* The mean difference is significant at the .05 level.
#### Analysis of Perceptual Cues

Following each listener's naturalness ratings of the 84 speech samples, the researcher replayed the samples rated as highly natural or highly unnatural that were within the first 70 unduplicated samples (60 normal speech; 10 impaired speech). A range of 4-7 samples rated as highly natural (1) and 5-7 samples weighted as highly unnatural (9) were replayed for each listener. Samples rated as highly natural were replayed as a group and perceptual cues were listed, followed by the same procedure for the samples rated as highly unnatural. Each listener then listed and weighted perceptual cues that influenced him/her in rating speech as either "highly natural" or "highly unnatural". Raters listed 134 perceptual cues for speech rated as natural and 143 perceptual cues for speech rated as unnatural. The researcher with another certified speech and language pathologist reviewed all cues listed or described in order to analyze and cluster them into categories of similar meaning or recurrent themes suggested by Guba (1981). Convergence of descriptive cues involved establishing categories that internally are similar in meaning yet differ from each other. Once established, each of the categories was listed with its assigned percentage of weighted importance.

As a result, the following 8 content categories or clusters emerged in classifying listed cues for both rating naturalness as well as rating unnaturalness: speech flow, articulation, understanding or clarity of speech, style/ease of speaking, voice, body language, rate of speech, and knowledge of subject/language ability. Table 13 displays these categories with examples of verbatim cues listed by raters.

<u>C</u> (	ontent Category	Natural/Unnatural	Verbatim Examples listed by raters
1.	SPEECH FLOW	Natural	speaks smoothly, no breaks in speech, speech was fluid, not a lot of pauses, dead time is little, no unusual pauses.
		Unnatural	stutters, broken unclear words, halting pauses, many 'ums' & 'ers' unusual tempo, not flowing.
2.	ARTICULATION	Natural	pronounces words well, no accent, articulates well, good pronunciation, diction good for age.
		Unnatural	lisps, a lot of slurring, poor pronunciation, doesn't say sounds correctly, slushy sound.
3.	UNDERSTANDING or	Natural	clear speech, easy to understand, speaks clear, words clear, clearly conveyed message.
	CLARITY	Unnatural	can't make out what saying, unintelligible, not talking clearly, words not recognizable, mumbled.
4.	SPEECH RATE	Natural	good speed, good rate, talks at a regular pace, tempo.
	·····	Unnatural	Slower speech.
5.	STYLE & EASE of	Natural	calm, confident sounding, ease, confidence, not nervous, comfortable, not bored.
	PRESENTATION	Unnatural	too much effort, insecurity in speaking, words struggled out, person seemed frustrated, confidence level low.
6.	KNOWLEDGE	Natural	uses descriptive words, proper grammar, wide vocabulary, knowledge of subject, intelligent, expresses well, tells story in order.
_		Unnatural	no complete sentences, grammar use poor, sentences/thoughts unfinished.
7.	BODY LANGUAGE	Natural	eye contact, facial expressions, body language, physical presentation.
		Unnatural	facial expressions, not very good eye contact, physical presentation, body movement, mouth not moving properly.
8.	VOICE	Natural	oud clear voice, alters voice inflections, good voice level, natural volume,
		Unnatural	pitch too high, nasal voice, quiet voice, low volume, variation in tone, unusual pitch, spoke too quietly.

 Table 13. Perceptual cue categories and examples of verbatim cues listed in each.

\_\_\_\_

=

L.

-----

=

Each of the 32 listeners (19 female and 13 male) distributed 100 points of a bar scale on those items (s)he listed as cueing natural speech and then followed the same procedure in distributing the items listed as cueing unnatural speech. Consequently 3200 points were distributed for cues related to "natural" speech and an equal number was distributed for cues related to "unnatural" speech. Overall, the 3200 points attributed for cues related to "unnatural" speech. Overall, the 3200 points attributed to rating naturalness were assigned to the categories of cues as follows: understanding (31%), flow (16%), articulation (14%), style (11%), knowledge/use of language (10%), rate (7%), voice (7%), and body language (4%). Table 14 displays the distribution of the bar scale points for the perceptual cues identified in rating speech as natural.

WEIGHTED CUE CATEGORIES FOR NATURAL RATINGS											
	F	A	u/c	R	S/E	K/L	BL	v			
19 FEMALE LISTENERS	350	194	648	86	180	191	84	161			
% of 1900 points	18%	10%	34%	5%	9%	10%	5%	8%			
13 MALE LISTENERS	175	240	338	142	156	122	41	64			
% of 1300 points	13%	19%	26%	11%	11%	10%	3%	5%			
TOTAL LISTENER GROUP	525	_ 434	986	228	336	313	125	238			
% of 3200 points	16%	14%	31%	7%	11%	10%	4%	7%			

Table 14. Bar Scale point distribution of perceptual cues to identify speech as NATURAL.

(F-flow; A-articulation; U/C-understanding/clarity; R-rate; S/E-style, ease; K/L-knowledge of subject, language; BL-body language and V-voice).

In weighting the cues attributed to rating unnaturalness, the 3200 points were assigned to the following categories of cues: understanding (44%), flow (22%), articulation (11%), voice (8%), style/ease (7%), body language (4%), knowledge/use of language (3%), and rate (<.009%). Table 15 displays the distribution of the bar scale points for the perceptual cues identified in rating speech as unnatural.

Tab	<b>e</b> 15.	Bar	Scale point	distribution o	f perceptua	i cues to i	denti	fy speec	h as
-----	--------------	-----	-------------	----------------	-------------	-------------	-------	----------	------

#### UNNATURAL.

	F	A	U/C	R	S/T	K/L	BL	V
19 FEMALE LISTENERS	414	238	930	3	121	66	35	107
% of 1900 points	22%	13%	49%	<.02%	6%	3%	2%	6%
13 MALE LISTENERS	312	125	483	0	121	25	85	149
% of 1300 points	24%	10%	37%	0%	9%	2%	7%	11%
TOTAL LISTENER GROUP	714	363	1413	3	242	91	120	256
% of 3200 points	22%	11%	44%	<.009	7%	3%	4%	8%

of subject, language; BL-body language and V-voice).

In listing perceptual cues for determining both natural and unnatural features of speech, listeners as a group designated the same three classifications as having the most influence with similar weight: understanding, flow, and articulation. First, listeners identified their ability to understand the speaker as the most heavily rated perceptual cue of natural speech (31%) and unnatural speech (44%). Second, listeners identified cues related to the flow or smoothness of speech as influential in determining both natural speech (16%) and unnatural speech (22%). Finally, articulation or pronunciation of speech was weighted as the next most influential factor in rating natural speech (14%) and unnatural speech (11%). Of the 5 remaining categories weighted as characteristics of natural speech, the category related to "style and ease of delivery" received 11% weight, and all others received weights between 4% and 10%. Similarly, of the 5 remaining categories weighted as showing characteristics of unnatural speech, all received weights between <.009% and 8%.

## Chapter 4

## DISCUSSION

This study primarily focused on the speech naturalness of children and adolescents. Listeners were asked to rate the naturalness of speech on a 1-9 scale; they were also asked to identify perceptual cues of naturalness. Chapter 3 presented the findings; this chapter will discuss the results relative to the following: the naturalness ratings, the reliability of listeners, and the perceptual cues of naturalness.

## **Normative Naturalness Data of Children**

While rating "naturalness of speech" has received increasing attention in the literature particularly with adults, systematic research of the naturalness ratings of children's speech, normal or disordered, is scarce. Development of children's communication skills--specifically rate and types of disfluencies--have been noted to fluctuate, with increasing disfluencies evident between kindergarten and fourth grade (Kowal, O'Connell, & Sabin, 1975). Do listeners perceive the naturalness of children's speech in a manner that fluctuates during this same period of development? Rating naturalness of adults' speech has been demonstrated to be reliable and valid (Ingham, Gow, & Costello, 1985; Martin, Haroldson, & Triden, 1984; Metz, Schiavetti, & Sacco, 1990) . Can similar functions be noted when rating naturalness of children who, by fluency data, demonstrate variation among age groups that continues through childhood

and adolescence? In a manner similar to data presented in the literature regarding the naturalness of adults' speech and listeners' reactions, the data accumulated and analyzed in this study address these questions and focus on naturalness measures relating to children.

Overall, the validity of the 1-9 point Likert scale was demonstrated. Group mean values as well as group mode values revealed the ratings of the normal speaking children as more natural than the ratings for the communicatively disordered children. Similar to the naturalness ratings noted in studies of adult speech naturalness (Martin Haroldson, & Triden, 1984; Ingham, Gow, & Costello, 1985; Metz, Schiavetti, & Sacco, 1990; Runyan, Bell, & Prosek, 1990; Onslow, Hayes, Hutchins, & Newman, 1992), normal speaking children in this study received more natural ratings (mean of 2.32) than the speakers with communication disorders (mean of 8.0). This group mean for the normal speaking children is comparable to the range of mean ratings associated in various studies of adults' speech naturalness (2.12 to 3.55-- Martin, Haroldson, & Triden, 1984; Ingham, Gow, & Costello, 1985; Metz, Schiavetti, & Sacco, 1990; Runyan, Bell, & Prosek 1990; Onslow, Hayes, Hutchins, & Newman, 1992). The group mean assigned to the communicatively impaired children's speech in this study was higher (more unnatural) than the mean rating typically associated with adults who stutter (naturalness mean of 6.52 on the same 1-9 scale Martin et al., 1984). These results are consistent with naturalness findings reported for adult speakers in the professional literature. This result demonstrated that listeners can validly distinguish a difference in the naturalness ratings of normal speaking and CDO <u>children</u>.

Comparison of the ratings received by normal speaking children in this study revealed that ratings of "1 and 2" ("at or near highly natural") were the mode ratings, 50% and 32% respectively. Inversely, the ratings of "8" and "9"

("at or near highly unnatural") were the mode ratings for the communicatively disordered children 10% and 80% respectively. These results reinforce that listeners can validly distinguish a difference in the naturalness ratings of normal speaking and CDO <u>children</u> and supports the use of the 1-9 point scale as a valid means to rate naturalness.

Considerable variability was seen in both the normal and communicatively impaired speaker groups. The widest range of scale values assigned to any one speaker sample occurred in the normal speaking group where the range of intervals any one sample received was between 3 and 9 scale values. The fact that 77% of all normative samples received between six to nine intervals reinforces the idea that children--close in age, speaking on a self-selected topic and in a conversational setting--display variation in what is perceived as natural speech characteristics. A wide variability of ratings received by any one speaker was also seen in the communicatively impaired speaker group. Seventy percent of the communicatively impaired speakers received ratings that varied between 6 to 9 scale values and, as noted in the normal speaking group, all samples were rated within a minimum range of three scale values. These data suggest that the process of rating naturalness within normal and disordered speaking children must recognize and expect variability. While current literature does not elaborate on the variability of ratings received by adults, speech naturalness research has concluded that rating naturalness can only be satisfactorily completed with a group of raters and that use of a single rater should be avoided. The variability in ratings (5 or more scale values) received by 88% of the normal speaking children and 70% of the communicatively impaired children strongly supports the use of multiple raters when rating children's naturalness. The existence of variability in naturalness ratings is underscored by the fact that the ratings analyzed in this study were those assigned by

listeners found to be reliable. In considering that 18% of the original listeners were excluded, the breadth of ratings seen was large.

Two major questions of this study were to determine whether speech naturalness ratings of children were impacted by gender or age. Data in this study found that the speech naturalness ratings of male and female children were statistically comparable. The differences in speech development, specifically related to speech rate variation by gender (Kowal, O'Connell, & Sabin, 1975; Starkweather, 1980), had no effect on naturalness ratings of children's speech. Generally, gender-related studies have concluded that faster speech rates are correlated with higher naturalness ratings. One would postulate, given the documented gender differences in speech rate development between kindergarten through fourth grade (Kowal, O'Connell, & Sabin, 1975), that naturalness ratings during these ages would depict gender related differences. Consequently, this study's finding that male and female children were rated comparably was unexpected given the previously cited differences in speech development of male and female children during these ages.

Within the normal speaking groups, the results of this study revealed that listeners perceived the naturalness of eight year olds' speech as significantly different from the speech of 12, 14, and 16 year olds. Eight year old speakers had the largest range of mean values (1.94 to 5.66) and the largest standard deviation (1.92) of all normal speaker groups. Researchers (Kowal, O'Connell, & Sabin, 1975; Starkweather, 1980) have noted that disfluencies, rate, and stress patterns vary among age groups. Although there continues to be a variation throughout childhood, the greatest variance occurs between ages 5 to 9. The data of this study suggest that speech naturalness of 8 year olds is significantly different during this developmental period and is characterized by developing speech rate, elevated numbers of disfluencies, and stress pattern variations. As

a speaking group, 8 year olds' speech samples received the least number of identical ratings (1333). Table 16 illustrates that as age increased, speaker group samples received progressively increasing numbers of identical ratings from listeners.

AGE GROUP	8	10	12	14	16
PAIRS RATED					
IDENTICALLY	<u>   1333    </u>	<u>    1593                                </u>	<u> </u>	<u>    1846                                </u>	<u> </u>
% OF TOTAL	22	27	30	31	37

Table 16. Number of Inter-rater pairs per age group of normal speaking children.

Moreover, in contrast to age groups whose overall inter-rater agreement was 86% or greater considering pair differences up to  $\pm$  2 scale values, interrater agreement for the 8 year old group did not reach a comparable percentage until  $\pm$  3 scale value differences were taken into account. These data suggest that, as a group, the 8 year olds not only were more variable as speakers but were also more difficult to rate.

Three hundred eighty four ratings were made for each of the normal speaker age groups (12 speech samples x 32 listeners). When observing the ratings for each group, it was evident that 8 year olds received the least number of "1" (highly natural) ratings, whereas the 16 year olds received the most ratings of "1." It is interesting that with increasing age, a progressive, stepwise number of "1" ratings were seen, as displayed in Table 17.

Table 17.	Number of	"1" ra	atings	received	by	normal	speaking	groups
by age.			-		-			

AGE GROUP	8	10	12	14	16	
Number of "1"						
Ratings received	85	<u>    117    </u>	156	161	187	
% of total possible	22	30	41	42	49	

No significant differences were noted in the naturalness ratings of 10, 12, 14, and 16 year old speakers. However, as age increased, group mean naturalness ratings showed a trend toward more natural scale values. Although not statistically significant, it is interesting that listeners rated children's speech as more toward highly natural as age increased.

## Naturalness Ratings: The Listeners/Raters

#### **Overall Demographics**

The listeners involved in this study were untrained and represented various ages, occupations, and interaction frequency with children. Overall age range of reliable listeners was 18 to 56, with the average age of female and male listeners being 22 years 7 months and 29 years 4 months, respectively. Listeners also represented a wide array of occupations. Thirty-one percent of the listeners were employed in full time positions such as office staff, computer analyst, counselor, accountant, construction worker, administrator, and teacher. Sixty-nine percent of listeners were students enrolled at a mid-western university in programs related to a variety of curricula, such as, sociology, child development, broadcasting, special education, French, health fitness, teaching, and mechanical engineering. No rater had completed coursework in speechlanguage pathology. For purposes of post hoc data trend analysis, all listeners were asked whether or not they had children and how frequently they interacted with children between the ages of 8-16. Thirty-percent of the listeners had children of their own (age range 2-25 years), and 66% indicated frequent interactions with children between the ages of 8-16.

In contrast to listeners used in other studies of adult naturalness, the listeners included in this investigation represented a wider range of age.

Previously cited research employed university undergraduates as listener groups without reference to listener ages (Martin, Haroldson, & Triden, 1984; Martin & Haroldson, 1976; Mackey, Finn, & Ingham; 1997). Listeners in this study were untrained by design, a criterion seen in few studies of adult speech naturalness (Martin, Haroldson, & Triden, 1984; Mackey, Finn, & Ingham; 1997). Others have routinely employed more sophisticated judges such as speech-language pathology students (undergraduate/undergraduate), certified speech-language pathologists (the researchers), or practicing clinicians as judges (Ingham, Gow, & Costello, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Ingham, 1997).

The earliest definition of speech naturalness (Nichols, 1966) placed the listener as the focal point by stating that naturalness is a phenomenon defined by the listener. Emphasized in this perspective is the view that the listener's perception of what sounds natural is speech that focuses his or her attention on the meaning of the words spoken rather than the speech pattern used in conveying the message. Later, researchers suggested that establishing data from unsophisticated listeners should be the primary concern of researchers whereas the opinion of untrained listeners would best represent the judgments of the general listening public, a consideration which should be the primary concern of treatment programs (Runyan & Adams, 1979; Onslow et al., 1992; Curlee, 1993). Listeners used in this investigation met these criteria and made judgments that were exposed to rigorous standards.

#### Intra-judge Reliability

Use of a rigorous method to determine intra-judge reliability was warranted, given the subjective nature of Likert rating scales. The procedure

used to determine intra-judge reliability in this investigation was the implementation of the intra-rater agreement method used by Ingham, Gow, and Costello (1985) (duplicate presentations rated as identical or within  $\pm 1$  scale 75% of the time). In contrast to the results of two tailed t-Tests, where all listeners were considered reliable at the .05 level of confidence, use of the "75% rule" resulted in the elimination of 7 listeners, one female and 6 males (18% of the original listener pool). Of these 7 listeners, 2 were self-employed, 5 were university students, 1 had children, and all indicated either occasional or frequent interaction with children between the ages of 8-16. Other than gender, no significant trend or pattern in the backgrounds of the eliminated listeners was noted in post hoc analysis. Of the listeners remaining, 62% of male judges were 100% reliable on rate-rerate tasks and 50% of female judges were 100% reliable.

The number of listeners forfeited in this study (18%) was somewhat greater than the number of listeners eliminated in Ingham, Gow, and Costello's (1985) study. Ingham, Gow, and Costello employed the 75%  $\pm$  1 agreement standard, using the same naturalness scale, and eliminated 13% of listeners. While the number of listeners removed in both studies was considerable, the increased number of listeners eliminated in this study suggests that rating naturalness in normal speaking children is a more difficult task. Difficulty in rating speakers comparably was also noted in the percentage of duplicate ratings which were identical. On duplicate presentations, listeners re-rated normal speaking samples with identical values only 65% of the time while communicative disordered samples were re-rated identically in 80% of occurrences.

#### Inter-listener agreement

The paired comparison procedure used in this study to establish interlistener agreement was the identical procedure used in naturalness studies with

adults (Mackey, Finn, & Ingham, 1997; Martin, Haroldson, & Triden, 1984; Martin & Haroldson, 1992). In this investigation many results differed from those reported in adult speech naturalness research. Inter-rater agreement in rating children's speech naturalness in this study was more difficult to achieve than that noted in adult research, and raters demonstrated considerable variability in rater comparisons. In adult studies inter-rater agreement reached levels of 75% to 90% with all pairs within  $\pm 1$  scale values. However, in this investigation, 77% inter-rater agreement (the lower range reported in adult studies) was not achieved until all pairs within  $\pm 2$  scale values were considered. Additionally, inter-rater agreement of naturalness ratings (normal and communicatively impaired) did not reach 90% (the upper range of agreement reported in adult studies) until pairs rated within  $\pm 3$  scale values were considered.

In studies rating speech naturalness of adult stutterers and normal speakers (Martin & Haroldson, 1992; Martin, Haroldson, & Triden, 1984), rater differences of up to  $\pm 4$  scale values encompassed all possible rater pair comparisons (Refer to tables 4 and 5). However in this study 100% of rater paired comparisons was not achieved until pairs within  $\pm 8$  scale differences were considered. In this study, rather than 100% of all rater comparisons being encompassed by  $\pm 4$  scale values, 97% of pairs for the normal speaking group and 96% for the CDO group were within  $\pm 4$  scale values. These data suggest that judging the speech naturalness of children-- normal speaking or otherwise--is a more difficult, variable process than that seen with adults. Perhaps when rating children's naturalness, judges do not know whether to employ: 1) speech as produced by adult models, or 2) the speech (s)he presumes is appropriate at that particular child's age as the standard of communication. Judges may accept a wider range of speech styles with children because they are developing in communication skill, and do not yet depict the adult model. Because the

listeners in this study were adult, as the children became older, perhaps their speech more closely approximated the adult model and was regarded as more natural. The increased frequency of listener agreement as age increased suggests this perspective. Consequently, when considering younger children, perhaps those variations from the adult speech model may then pose greater problems of classification, as listeners-- particularly untrained--do not know what to expect. Related to this speculation is the possibility that rating children's speech with less naturalness variability is dependent on the listener's experience or familiarity with children's speech and language styles. Perhaps then, naturalness ratings of children provided by sophisticated or trained listeners should be studied for comparisons of degree of variability as a group.

In summary, inter-rater agreement of 80% or greater was achieved within  $\pm 2$  scale values with all age groups of normal speaking children except eight year olds. Inter-rater agreement of pairs rated identically or within  $\pm 1$  scale values increased with age. Overall, rating the speech naturalness of all children included in this study resulted in more variable scale values used and greater inter-rater agreement difficulty.

#### **Perceptual Cues of Naturalness**

Qualitative data in this study support the idea that speech naturalness is a multi-dimensional concept, influencing listeners with many variables. Perceptual studies of speech fluency have attempted to identify cues which distinguish the fluent speech of stutterers from speech of normally fluent adults (Hotchkiss, 1973; Johnson, 1987). These studies, using methods of scaling with researcher provided speech parameters, have identified that speech rate, fluency, and articulatory behavior have served as underlying cues for listeners in distinguishing normally fluent from stutterers' fluent speech. In contrast to

researcher-provided parameters, this study employed listener generated definitions written following review of samples rated as highly natural and after samples rated as highly unnatural. Data in this study revealed perceptual cues somewhat different from those previously noted in these studies which were concerned with fluency and speech naturalness.

Thirty-two listeners included in this study listed perceptual cues which signaled speech as natural or unnatural. The total number of cues listed was 277: 134 cues suggested natural speech, and 143 cues suggested unnatural speech. Content qualitative analysis procedures resulted in eight categories of cues: speech flow, articulation, understanding/clarity of speech, style/ease of speech, voice, body language, rate of speech, and knowledge of subject/language ability. Table 18 provides a summary of the number of points received and the percentage of the 3200 total points assigned on bar scales to each category of cues generated by the listeners of this study.

WEIGHTED CUE CATEGORIES										
Natural Speech	U/C	F	A	S/E	K/L	v	R	BL		
3200 Total Points	986	525	434	336	313	238	228	125		
Percentage of total	31%	16%	14%	11%	10%	7%	7%	4%		
	U/C	F	A	v	S/E	BL	K/L	R		
Unnatural Speech										
3200 Total Points	1413	714	363	256	242	120	91	3		
% of 3200 points	44%	22%	11%	8%	7%	4%	3%	<.009		

<b>Table 18</b> .	<b>Bar Scale</b>	point dist	ribution of	perceptual	cues to	<b>identify</b>	speech as	NATURAL
and UNN/	TURAL.					-	-	

(F-flow; A-articulation; U/C-understanding/clarity; R-rate; S/E-style, ease; K/L-knowledge of subject, language; BL-body language and V-voice).

Cues related to understanding of speech message, flow of speech, and articulation were weighted as the primary listener-generated cues when listeners rated both natural and unnatural speech. When defining natural speech cues, these primary categories received 61% of scale points possible. Likewise, when defining unnatural speech, these primary categories received 77% of possible points.

A listener's ability to understand speech was the most weighted perceptual category of both natural speech (31%) and unnatural speech (44%). This suggests that the standard of speech accepted by the community is expected by listeners; and when the model does or does not correspond to the standard, listener's perception of naturalness is affected. Understanding or intelligibility of speech as an antecedent to determining natural or unnatural speech was not identified in Johnson's (1987) or Hotchkiss's (1973) research. However, these results support the earlier notion of Sanders, Gramlich, and Levine (1981) who maintained that intelligibility is a function of understanding and is a separate quality of speech, not related to whether speech sounds natural. Speech naturalness should only be considered once intelligibility is established (Sanders, Gramlich, & Levine, 1981). The untrained listeners in this study, when generating cues of naturalness, prioritized the importance of intelligibility as critical to speech naturalness. This observation supports Yorkson, Beukelman, and Bell's (1988) notion that intelligibility of speech is a necessary prerequisite to considerations of naturalness. Consequently, without instructions to consider intelligibility or understanding of speech separately from naturalness, listeners related it as the primary variable of naturalness.

Similar perceptual cues rated by listeners in Hotchkiss's (1973) and Johnson's (1987) work were generated by the listeners in this study. However, some difference in the roles of the perceptual correlates was noted. This study verified previous reporting of speech flow (fluency) as related to listeners' perception of naturalness (Finn & Ingham, 1994; Martin & Haroldson, 1992; Martin et al., 1984; Onslow, Hays et al., 1992). As noted in the literature,

speech flow was identified as a significant perceptual cue which allowed differentiation between the speech of normal speakers and fluent stutterers. Likewise, listeners considered the role of articulatory behavior in cueing naturalness as primary, receiving 14% of points indicated for natural and 11% of unnatural speech.

Perhaps the most notable finding of the qualitative data generated in this study is its contrast to the findings of previous naturalness studies that suggested a strong relationship between naturalness and speech rate (Finn & Ingham, 1994; Johnson, 1988; Martin & Haroldson, 1992; Martin et al, 1984; Onslow et al., 1992). Rate of speech as a dimension of naturalness comprised only 7% of listener generated cues of natural speech and a negligible .009% of unnatural speech cues listed. Listeners' scant acknowledgment of rate as a variable of naturalness may be a function of attention to more salient co-occurring variables. In other words, perhaps untrained listeners when attending to speech with coexisting conditions (reduced intelligibility, interrupted speech flow, and imprecise articulation) are incapable of determining that rate may be an underlying variable affecting naturalness. Rather, untrained listeners may not isolate factors of rate but consider rate as part of the "Gestalt" of the speech being heard. Trained listeners, such as speech and language clinicians, may be more sensitive to parameters of rate because of their understanding of speech parameters.

Body language, identified by visually observed characteristics, was noted as a dimension of both natural and unnatural speech; but it received only 4% of the bar scale values assigned. Although speech is a communication form that has visual and auditory components, visual cues related by listeners to speech naturalness perceptions were insignificant. Martin and Haroldson's study (1992) concluded that speech naturalness ratings of adult nonstutterers speech were

similar between audio and audio video speech; but ratings of stutterers' speech was rated more unnatural using audiovisual media. Perhaps visual components were not given more emphasis by listeners in this study because the normal and communicatively disordered speakers involved, with the exception of two stutterers, did not manifest visual accessory behaviors typically seen in studies focusing on naturalness and disfluency.

The variety of unprompted perceptual cues elicited from speakers in this study reinforces the concept of speech naturalness as a multidimensional quality. Although naturalness of speech has received considerable attention as related to fluency, this study suggests that the boundaries of speech naturalness exist beyond their applications to fluency disorders. In fact, the variables which listeners attend to when rating naturalness of speech are many and are shared by normal speaking children as well as a wide array of communicatively disordered speakers.

#### Chapter 5

## **Summary and Future Research**

## Summary

While speech naturalness ratings have been studied since 1966, the major rekindling of interest in naturalness research occurred in 1984 as an attempt to investigate perceptual differences in the fluent speech of stutterers. Initially in research, the terms *naturalness* and *fluency* were equated, and it was theorized that objective definitions of fluency and naturalness would be similar. What has evolved from naturalness research with adults is the idea that fluency is only one aspect of naturalness, a concept that embraces other perceptual variables. In the past 9 years, the naturalness literature has suggested the clinical merit of determining naturalness data related to children and of defining the perceptual correlates of naturalness.

Results of this study investigated both of these major considerations. Systematic review of data from this study validated that speech naturalness is a scaleable parameter of the speech of normal speaking children and adolescents, as had been previously reported in the studies with adults. Without exception, the speech of normal speaking children was rated as more natural than the speech of the communicatively impaired speakers. Normative data from 8, 10, 12, 14, and 16 year old speakers were compiled.

Considerable variability of ratings were given to all speakers involved in this study. This suggests two major elements. First, variability of naturalness is

inherent in normal speaking processes. In reference to fluency, normal speech is found on a continuum, with variation in the range of disfluencies a speaker experiences which are related to the specific demands of the communication situation. These data suggest that ratings of naturalness may vary with a continuum of communication situations in the same manner as fluency. Naturalness data acquired in this study illustrate a range of rating variability at each age group involved. In a comparison of the ratings of children between the ages of 8-16, significant differences between the naturalness ratings of 8 year olds and 10-16 year olds were seen.

Second, results of this study strongly reinforce the idea that valid ratings of naturalness are only accomplished with a group of raters. Inter-rater agreement data revealed that with all children, particularly 8 year olds, a wider degree of rated scale differences was needed to achieve the inter-rater agreement levels seen with adults. Therefore, the task of rating children's speech, particularly the younger groups, is a more variable, difficult task.

Differences in speech naturalness ratings attributed to gender were not seen. This result was unexpected since studies of speech maturation have delineated significant gender differences in rate and disfluency development as well as the significant physical, cognitive, and affective growth of children during these ages.

In focusing on analyzing listener-generated cues of verbal and non-verbal correlates of speech naturalness, eight content categories of perceptual cues emerged in the tasks of this study. Three of the eight categories incorporated items correlated to perceptual studies of fluency, lending support to the notion that naturalness may be an all-encompassing quality of speech perception which includes the correlate of fluency. Ability to understand speech, the flow of speech, and precise articulation patterns emerged in that order of importance as

the primary definitions of both speech naturalness and unnaturalness. Other categories of less consideration were, style/ease of speech, knowledge of subject/language use, rate of speech, voice, and body language. These were weighted differently for influences on natural and unnatural speech.

Rate as an important perceptual cue of naturalness was not found. This result was surprising given the positive correlation made between faster speech rates, fluency, and more natural speech ratings. In contrast to the findings of previous studies, reference to rate as a perceptual cue was meager. This result may be an artifact of more salient cues perceived by the untrained listeners used.

## **Clinical Application**

Results of this study added valuable information which can be useful in clinical management with children. Professional literature has indicated a great need to establish normative naturalness data that encompass an increased number of normal speakers, rather than the limited numbers available in published studies to date. The naturalness data compiled in this study were vast, incorporating 384 ratings at each age level studied. In doing so, the data of this study have demonstrated the variability of speaker naturalness for children ages 8 through 16. Data from twelve normal speaking subjects, at each age level, offer clinicians a speech naturalness rating range which could be useful as a targeted treatment outcome which allows comparison of a normal speaking age cohort with CDO children in clinical management programs.

Implication for use of the naturalness scale extends beyond its value in the fluency applications noted to this date. Data from this study suggest that fluency is a component of speech naturalness. Instances where CDO clients have been given speech naturalness ratings have been documented in the fluency literature as a useful way to improve speech naturalness. The listener-generated

naturalness correlates noted in this study encompassed parameters beyond those related to fluency. This wider scope of naturalness cues suggests that providing CDO clientele in various types of therapy programs with feedback of naturalness ratings and other attributes may be useful in improving speech naturalness.

Speech naturalness of normal speaking as well as communicatively impaired children was reliably rated by inexperienced listeners in this study. Perhaps naturalness ratings from a number of inexperienced raters would serve as the most valuable assessment of treatment efficacy both within the clinic and outside clinical situations. Naturalness ratings of clientele by multiple listeners would be valuable feedback to both clients and clinicians in decisions related to therapy transfer tasks, consideration of treatment dismissal, and post-treatment maintenance.

The goal of natural speech is as desirable in treatment programs for children with articulation, voice, and hearing impairments as in treatment programs for fluency disorders. While some fluency treatment programs now incorporate speech naturalness measures, other therapy strategies have not systematically included these. Recent emphasis in therapy treatment procedures has been devoted to documenting specific measurable components relative to the disorder being treated. Perhaps in the profession's quest for quantified treatment data, more emphasis has been directed toward development of objective measures of specific parameters of speech while subjective measures of the integrated speech act have been overlooked. It would appear that in addition to specific object measurements used in treatment programs, clientele would benefit from the employment of speech naturalness ratings as a general indicator of progress toward the overall goal of therapy: the attainment of natural speech.

## **Future Research**

From initial studies of speech naturalness, researchers have realized its potential as a tool for clinical practice. Future research suggested as a result of this study will continue to pursue the use of speech naturalness measures in various therapy programs.

From the information found in this investigation, many directions for future research appear. The age-related difference in naturalness ratings of 8 year olds in comparison to 10-16 year olds needs further study. What speech differences in addition to the perceptual impressions of listeners can be attributed to these differences in ratings between speakers age 8 and others? Specific analysis related to acoustic parameters would allow exploration of the relationships between the definitions of naturalness in physical terms and in perceptual terms.

The fact that some speakers in both the normal speaking and communicatively impaired groups received ratings that were separated by between 4 and 8 scale intervals is interesting. Future research with data from this study should address detailed analysis and comparison of the speech of those children in each age range who received the most natural and unnatural ratings. Study of these samples by sophisticated listeners with the incorporation of in-depth acoustic measurements may explain the variability in ratings seen in this study. In addition to re-examination of those samples rated as most and least natural per age level, it would be valuable to provide raters the perceptual cues identified in this study so that they may be simultaneously rated and correlated to naturalness ratings. Finally, specific to the methods involved in this study, it would be beneficial to duplicate this task with another group of untrained listeners from an area without access to a university. As a group of listeners, those involved in this study were older, more educationally and professionally

diverse; in addition, they had different amounts of interactions with children. However, even with the increased range of experiences and ages, 69% of the listeners were students enrolled at a mid-western university. Duplicating this study with untrained listeners--having similar demographics, but who are not affiliated with a university--would provide insight as to how representative and effective the study methods incorporated in this project were.

In addition, results explaining the most naturally and unnaturally rated speakers would have merit in training listeners to be more reliable in rating naturalness with children's speech. Eighteen percent of the original group of listeners did not achieve 75% intra-rater reliability and were eliminated from this study. What can be done to improve intra-rater reliability? To fully interpret the perceptual data compiled from the untrained listeners, further investigation using the tasks of this study with trained listeners would be beneficial. Would experienced and untrained listeners utilize similar criteria in referencing internal variables of naturalness? Although arguments have been made for the use of untrained listeners as judges to better reflect the opinions of the general public, clinical definition of naturalness by trained listeners may generate responses that more readily translate into speech parameters that can be used for therapeutic programs. In other words, although the definitions provided in this study were broad, specific interpretation from trained listeners may lead to more quantifiable parameters using physiological and acoustic measures.

This concept gives rise to the various sources of feedback communicators receive. All published studies to date have used adult listeners, either unsophisticated or sophisticated. However, it seems reasonable to suggest that future studies should explore naturalness ratings of age cohorts, especially with children. It would be interesting to obtain naturalness ratings of speakers from this study by age cohorts. The data received from children listeners rate would

allow comparison between the ratings of the adult listeners obtained in this study. Should peer raters demonstrate less variability in rating these samples, the idea of adult listeners using expectations of adult models when rating children's naturalness, may be substantiated. In addition, it would be interesting to pursue naturalness ratings of the children included in this study by other communicatively impaired children.

The relationship between rate and pause behaviors to ratings of naturalness needs to be explored in view of previous significance attributed to rate and naturalness in the literature and the insignificant reference to rate in this study. Replication of this study with speech and language clinicians as raters would permit comparison of the naturalness ratings in scale value, the listeners variability seen, and the definition of naturalness/unnaturalness cues.

Results of this study suggest that the boundaries of speech naturalness and their clinical application exist beyond application to fluency disorders. Clinical use of speech naturalness ratings, as a self-monitoring devise, has been shown to assist clients in fluency shaping programs. Although the focus of this study was not on naturalness ratings of the disorders represented in the communicatively impaired group, the fact that all speakers in this group rated comparably with stutterers suggests that the use of this scale should be applied to other disorder groups. This dimension of research is exciting not only for its clinical application but also as a means of establishing a more complete definition of what naturalness entails.

Finally, previous studies have verified that the naturalness scale is sensitive to situations in which speech flow changes. Future study of perceptual correlates and scale validity should investigate the changes in perceived speech naturalness when speakers are systematically instructed to vary the listener-

related cues identified in this study. In doing so, significant information may result and may in turn translate into tangible therapeutic approaches.

APPENDICES

# APPENDIX A

# SPEAKER PARTICIPANTS INFORMATION SURVEY

1.	Name:			Phone Numb	er <u>:</u>		
2.	Address:	<u></u>					
3.	Age:	Years	Months	_ Date Of Birth:			
4.	Person fi	lling out	Questionna	ire:			
5.	Relations	hip to Pa	articipant: _				
Pl	ease Circl	<del>9:</del>					
6.	Does par •norm •norm •Any o	<b>ticipant I</b> nal or com nal hearin current m speech o If Yes, P	have: rected to nor ng? nedical condi or hearing? lease explai	mal vision? tion that affects n:	Yes Yes Yes 	or or or	No No No
7.	Has partic or hea If Yes	ipant eve aring ther , please	ər received s rapy? explain:	peech, language,	Yes	or	No
8.	Is Particip Speed If Yes	ant curre ch & Lang , Please	ntly under th guage Patho Explain:	e treatment of a logist or Audiologist?	Yes	or	No

# **APPENDIX B**

Questions asked of speakers at onset of taping:

1. What's your name?\_\_\_\_\_

2. Do you go to school? \_\_\_\_\_\_Where?\_\_\_\_\_

3. What's your favorite part of school?\_\_\_\_\_

# Instructions:

We really appreciate your helping out with this job. I think you will find helping us very easy. We are interested in how kids like to spend their time when they are not in school. So we need you to think about what you really like to do when you are not in school. We want you to tell us as much as possible about what it is that you like to do. If you would like some ideas, I have some for you. For example: you could talk about your:

- 1.) Favorite Sport.
- 2.) Favorite Hobby.
- 3.) The funniest thing that ever happened to you.
- 4.) Your most or least favorite family pet.
- 5.) Something you love to watch on T.V. and Why
- 6.) The best or worst vacation you ever had.

What would you like to talk about?

Whenever you are ready, you can begin.

# **APPENDIX C**

# LISTENER INFORMATION SURVEY

1.	Name:Phone Number:	
2.	Address:	
3.	Age:	
4.	Are you a College Student?: (Please Circle) Yes or No If yes, what is your major?	
5.	Have you <u>completed</u> any coursework in Communication Disorders (CDO)Yes or No If yes, please list CDO courses:	
6.	Are you Currently enrolled in Coursework in CDO?Yes or No If yes, please list CDO courses:	
7.	<b>Do you have children?</b> If So, Please list each child's gender and age; (for example: Female, age 5.)	_
		-
8.	Please circle the one most representative of your frequency of experiences INTERACTING with children between the ages 8-16.	of of

FREQUENTLY------OCCASIONALLY-----SELDOM-----NEVER

9. <b>Do</b>	<ul> <li>You have:</li> <li>•normal or corrected to normal vision?</li> <li>•normal hearing?</li> <li>•Any current medical condition that affects speech or hearing?</li></ul>	Yes Yes Yes	or or or	No No No
10. Ha	ave you ever received speech, language, or hearing therapy? If Yes, please explain:	. <b>Yes</b>	or	No
11. Ar	e you currently under the treatment of a Speech & Language Pathologist or Audiologist? If Yes, Please Explain:	Yes	or	No

# **APPENDIX D**

# TABLE of t-values

# Paired Sample t -Tests\*

Listener	t-Value	Listener	t - Value
Female 1	.818	Male 1	.205
Female 2	1.000	Male 2	.336
Female 3	.671	Male 3	.407
Female 4	.366	Male 4	.583
Female 5	.028**	Male 5	.861
Female 6	.165	Male 6	.373
Female 7	.221	Male 7	1.000
Female 8	.263	Male 8	.302
Female 9	.104	Male 9	.366
Female 10	1.000	Male 10	.071
Female 11	.218	Male 11	.336
Female 12	.583	Male 12	.189
Female 13	.136	Male 13	.273
Female 14	.431	Male 14	.082
Female 15	.583	Male 15	.082
Female 16	.671	Male 16	.216
Female 17	.336	Male 17	.671
Female 18	1.000	Male 18	.583
Female 19	.263	Male 19	.686
Female 20	.793		

\* df=13

\*\* p <.05 significant

## **APPENDIX E**

## 1. Analysis of descriptions/definition of naturalness:

Total Listeners: 19 Female (those considered reliable) 13 Male (those considered reliable)

Each listener distributed 100 points of a bar scale on those items (s)he listed as natural. Below are the content categories, with the descriptions given by the listeners. Immediately right of the description is the point value assigned by the listener on the bar scale.

## **General Demographics:**

#### Eight content categories emerged:

- 1. Flow
- 2. Articulation
- 3. Understanding/clarity
- 4. style/ease

- 5. Voice
- 6. Body language
- 7. rate 8. Knowledge of
  - Subject/language

19 female raters distributed=1900 possible points13 maleraters distributed=1300 possible pointstotal points rated=3200 point values

I. CONTENT CATEGORY:

FLOW (6% of total points)

FEMALE	(18% of female points)		MALE (13% of ma	ale points)
Speaks smooth	nly	21	fluid speech	30
no ums	•	20	not repetitive	10
didn't stumble		12	dead time is little	20
no breaks in sp	eech	24	smooth flow	9
smooth		9	nat. shorten of words	10
not a lot of pau	Ses	16	did not stutter	18
spoke at good	pace	32	good flow	19
speech flowing		31	no unusual pauses	25
sentences flow	ed well	25	not hesitating-speech	21
language flows		39	easy flow	<u>13</u>
uses "um" less	frequently	14	-	175
speech was flu	id	16		
few pauses		20		
doesn't repeat		23		
flowing uninter	rupted	34		
good flow		15		
no stuttering		8		
not alot of ums	5	<u>23</u>		
		350		

# II. CONTENT CATEGORY: ARTICULATION (14% of total)

FEMALE	(10% of female points)	MALE	(18% of male points)	
Articulates W	/ell	30	no lisps or accents	15
no accent		9	clear pronunciation	40
pronounced	words well	13	good articulation	31
pronounces	well	35	good articulation	24
uses good so	ounds	13	diction good for age	37
says words o	orrectly	12	good articulation	12
clear pronun	ciation	42	good pronunciation	45
normal speed	ch	<u>40</u>	good articulation	<u>36</u>
		194	-	240

# **III. CONTENT CATEGORY:**

# **CLARITY/EASY TO UNDERSTAND** (31% of Total Points)

FEMALE (34% of female points)		MALE (26% of male points)	
Clarity	39	clearly conveyed message	23
clear speech	31	understood words well	21
able to understand	5	clear/understandable	60
clear tone	35	clear speech	45
easy to understand	10	general clarity	27
spoke clearly	20	easy to understand	36
able to understand	11	clear voice	48
able to understand	31	understandable	32
clear voice	20	clear tone	31
words clear	34	clear tone	<u>15</u>
easy to listen to	23		338
clear speaking	40		
speaks clear/understandable	62		
word/sounds understandable.	35		
understandable pronuncia.	67		
clear tone	27		
understandable/clear	60		
clarity	40		
easy to understand	26		
talks from diaphragm/clear	<u>32</u>		
	648		

## IV. CONTENT CATEGORY:

# RATE (7% of total points)

FEMALE	(5% of female points)		MALE	(11% of male points)
speed		10	good speed	25
speed		32	good rate	34
talks at a reg	jular pace	12	pace of speech	31
spoke at a g	ood pace	<u>54</u>	fast speech	20
	·	86	tempo	<u>32</u>
			•	142

## V. CONTENT CATEGORY: DELIVERY STYLE, RELAXED, CONFIDENCE, ENTHUSIASM, EASE OF PRESENTATION (11% of total possible points)

FEMALE	(9% of female points)		MALE	(12% of male points)
calm	· · · ·	7	comfort level	35
comfortable		6	confidence	19
not nervous		49	ease	32
not bored		7	confidence	30
didn't fidget		4	enthusiasm	<u>40</u>
no extra soul	nds/mouth move.	26		156
confident sou	Inding	52		
more express	sion	16		
voice tone ex	citement	<u>13</u>		
		180		

# VI. CONTENT CATEGORY:

## KNOWLEDGE/LANGUAGE ABILITY (10% of Total Points)

FEMALE	(10% of female points)		MALE	(9% of male	points)
knowledge o	of subject	30	knew subj o	discussed	21
great gramm	nar	20	wide vocab	ulary	12
explains info	in detail	26	logical orde	or of speech	9
proper gram	mar	21	intelligent/e	xpress well	15
sentences w	rere complete	15	vocabulary/	choice wds	13
uses descrip	otive words	8	talkative/lot	s of words	52
tells a storv	in order	10			122
good vocabi	ulary	7			
speaks in co	mplete sentences	16			
use correct v	wording	13			
normal Engli	ish	25			
0		191			

# VII. CONTENT CATEGORY:

# **BODY LANGUAGE (4% of Total Points)**

FEMALE	(4% of female points)		MALE (S	3% of male points)
eye contact		11	physical presenta	tion 6
facial expres	ssions	9	eye contact	<u>35</u>
eve contact		13	-	41
body langua	ge	17		
eye contact	•	<u>34</u>		
-		84		

## VIII. CONTENT CATEGORY: VOICE-TONE-VOLUME (7% of Total Points)

FEMALE (8% of female points)		MALE (5%	of male points)
deep voice	12	loud speech	7
loud	16	natural volume	23
good voice level	18	normal tone of voice	ə <u>34</u>
loud, clear voice	14		64
alters voice infection	11		
loud voice	29		
loud enough	20		
lower tone	10		
good tone	12		
soft calm voice	<u>19</u>		
	161		

## **Composite of Categories:**

	flow	artic	under	rate	style	know	Ы	volce	
FEMALE	350	194	648	86	180	191	84	161	
% of points	18	10	34	5	9	10	5		
MALE	175	240	338	142	156	122	41	64	
% of points	13	19	26	11	11	10	3	5	
TOTAL GROUP	525	434	986	228	336	313	125	238	
% of points	16	14	31	7	11	10	4	7	

(Flow=Flow; Artic=Articulation; Under=Understanding; Rate=Rate; Style=Style or ease of presentation; Know=Knowledge of subject/language ability; BL=Body Language; Voice=Voice)
# MOST TO LEAST WEIGHTED CATEGORIES INFLUENCING PERCEPTIONS OF <u>NATURAL</u>

FEMALE LISTENERS	MALE LISTENERS	COMBINED LISTENERS
Understanding-34%	Understanding-26%	Understanding31%
Flow18%	Articulation19%	Flow16%
Articulation10%	Flow13%	Articulation14%
Knowledge/lang10%	Rate11%	Style11%
Style9%	Style12%	Knowledge/lang-10%
Voic <del>e8</del> %	Knowledge/lang-10%	Voice7%
Body language5%	Voic <del>e5</del> %	Rate7%
Rate5%	Body language3%	Body language4%

# **APPENDIX F**

# 2. Analysis of descriptions/definition

of unnatural:

Total Listeners: 19 Female (those considered reliable) 13 Male (those considered reliable)

Each listener distributed 100 points of a bar scale on those items (s)he listed as natural. Below are the content categories, with the descriptions given by the listeners. Immediately right of the description is the point value assigned by the listener on the bar scale.

#### **General Demographics:**

The identical categories identified in the "natural" definition section were used to determine if listeners used similar considerations in defining unnatural. Listener definitions were found in one of these categories:

1. Flow

- 5. Voice
- 2. Articulation
- 3. Understanding/clarity
- 4. Style/ease

- 6. Body Language 7. Rate
- 8. Knowledge of

Subject/language

19 female	e raters	distributed=	1900 possible points
13 male	<b>raters</b>	distributed=	1300 possible points
	total	points rated=	3200 point values

## I. CONTENT CATEGORY:

# FLOW (22% of total points)

FEMALE	(21% of female points)		MALE (24% of male points)	
stutter		20	repeated self	28
broken unclea	r words	13	stuttering	20
stuttering		20	repetitions in speech	10
stuttering		17	stuttering	25
a lot of pauses	i	8	lots of dead time	5
too many gaps	bet. words	14	unable to get flowing	11
stuttering		25	repeat, but not stutter	25
speech blocks		17	severe stuttering	30
pauses		18	stuttering	43
stuttered alot		9	halting pace/pauses	12
stuttered		4	too many pauses	7
long breaks be	ot. words	6	many "ums" & "ers"	9
stuttering		26	stuttering	12
stuttering		15	prolongations wds.	31
takes a long tir	ne to process	22	stutter	17
stutters freque	ntly	17	unusual tempo/stops	
			interruptions	17
drags words or	ut	16	flow/fluency	<u>10</u>
stutters		8		312
stuttering		13		
takes awhile		16		
excessive stut	tering	23		
stuttering	-	20		
words a little ju	impy	10		
not flowing/lots	s interrupt.	24		
stutters	•	<u>11</u>		
		402		

# II. CONTENT CATEGORY:

# **ARTICULATION (11% of Total Points)**

FEMALE	(13% of female points)		MALE (10% of male points)	
lisps		16	slurred speech	28
speech slurred	1	6	slurring	12
poor pronuncia	ation	30	off pron of consonants	14
trouble pronou	Incing words	15	slurring of speech	38
noises not in w	vords	5	inability to pronounce	15
lisp		9	lisp	4
doesn't say so	unds correctly	20	slushy sound	<u>14</u>
a lot of slurring	1	15	-	125
trouble pronou	incing wds.	14		
unclear pronui	nciation	50		
slurred speech	1	30		
just noises cor	ning out	15		
slurred speech	1	<u>13</u>		
·		238		

# **III. CONTENT CATEGORY:**

# CLARITY/EASY TO UNDERSTAND (44% of Total Points)

FEMALE (49% of female point	nts)	MALE (37% of male points)	
can't make out what saying	27	difficult to understand	30
hard to understand	20	could not understand	44
trouble. under what's said	18	unintelligible	21
mumbled words	16	can't understand	69
unable to understand	25	words not always intell.	52
hard to understand	21	can't understand	30
word not clear	33	words unclear	<b>56</b>
unintelligible	40	can't understand	55
poor communication	55	can't understand	31
could not understand	50	unintelligible	40
a hard time understand	72	cannot understand	55
couldn't understand	45		483
confused/hard to underst.	13		
not talking clearly	20		
speech hard to understand.	50		
speech not clear/slurred	31		
not understand	<b>5</b> 5		
no real words recognized	44		
can't understand	41		
difficulty understanding	16		
muffled tone	11		
hard to comprehend wds	40		
can't understand what said	40		
cannot understand	41		
difficult to understand	45		
words not recognizable	<u>64</u>		
-	930		

IV. CONTENT CATEGORY:

## RATE (0% of Total Points)

FEMALE (.002% of female points)		MALE	(0% of male points)	
slower speech	3			<u>-0-</u>
	3			0

## V. CONTENT CATEGORY: DELIVERY STYLE, RELAXED, CONFIDENCE, ENTHUSIASM, EASE OF PRESENTATION (9% of Total Points)

FEMALE	<u>(6% of female points)</u>		MALE (9% of male points)	
amount of e	ffort	30	too much effort	40
difficulty get	ting out words	15	confidence level low	15
person seer	ned frustrated	15	insecurity in speaking	22
must really	concentrate	17	high frustration level	21
too much ef	fort	32	effort getting wds out	<u>23</u>
words strug	gled out	<u>12</u>		121
		121		

# VI. CONTENT CATEGORY: KNOWLEDGE/LANGUAGE ABILITY (3% of Total Points)

FEMALE	(3% of female points)		MALE	(2% of m	ale points)
no complete :	sentences	9	grammar us	se poor	25
senten/thoug	hts unfinished	14	-	•	25
can't speak ir	n complete sent.	15			
can't express	compl. thought	18			
don't use con	rect word	<u>10</u>			
		66			

## VII. CONTENT CATEGORY: BODY LANGUAGE (4% of Total Points)

FEMALE	(2% of female points)		MALE	(7% of male points)	
facial expres	sions	20		eye contact	30
not very goo	d eye contact	<u>15</u>		mouth not move prop.	26
	-	35		physical presentation	14
				body movement	<u>15</u>
				-	85

## VIII. CONTENT CATEGORY: VOICE-TONE-VOLUME (8% of Total Points)

FEMALE (6% of female points)		MALE (11% of male points)	
pitch too high	20	spoke too quietly	19
soft voice	10	variation in tone	23
high intonation in voice	3	unnatural pitch	10
nasal voice	3	unusual pitch	10
soft, high pitch	14	too nasal a quality	1
higher pitch	3	pitch: too nasal/low/high	17
quiet voice	15	soft & high pitch	20
very high squeak force voice	27	low volume	20
poor nasal quality	8	unusual tone	15
high pitch	4	too soft	_5
	107		149

# **Composite Categories:**

	flow	artic	under	rate	style	know	Ы	voice	
FEMALE	414	238	930	3	121	66	35	107	
% of points	22%	13%	49%	<.002%	6%	3%	2%	6%	
MALE	312	125	483	0	121	25	85	149	
% of points	24%	10%	37%	0%	9%	2%	7%	11%	
TOTAL GROUP	714	363	1413	3	242	91	120	256	
% of points	22%	11%	44%	<.0009	7%	3%	4%	8%	

[Flow=Flow; Artic=Articulation; Under=Understanding; Rate=Rate; Style=Style or ease of presentation; Know=Knowledge of subject/language ability; BL=Body Language; Voice=Voice)

## MOST TO LEAST WEIGHTED CATEGORIES INFLUENCING PERCEPTIONS OF UNNATURAL

FEMALE LISTENERS Understanding-49% Flow----22% Artic-----13% Style------6% Voice------6% Knowledge-----3% Body Language-2% Rate------<.002%

#### MALE LISTENERS Understanding-37% Flow----24% Voice----11% Artic----10% Style----9% Body Language---7% Knowledge-----2% Rate-----0%

#### COMBINED LISTENERS

## MOST TO LEAST WEIGHTED CATEGORIES INFLUENCING PERCEPTIONS OF:

## NATURAL

#### FEMALE LISTENERS

#### MALE LISTENERS

Understandi	ng26%
Articulation-	19%
Flow	13%
Rate	11%
Style	11%
Knowledge/l	ang10%
Voice	5%
Body Langu	age3%

#### COMBINED GROUP

Understanding	
Flow1	6%
Articulation	-14%
Style1	1%
Knowledge/Lang	10%
Voice	7%
Rate	7%
Body language	4%

#### UNNATURAL

- FEMALE LISTENERS

#### MALE LISTENERS

Understanding	37%
Flow	-24%
Voic <del>e</del>	-11%
Articulation	
Style	9%
Body Language	7%
Knowledge/lang-	2%
Rata	0%
	- /-

#### COMBINED GROUP

Understanding44%
Flow22%
Articulation11%
Voic <del>e8</del> %
Style7%
Body Language4%
Knowledge/lang3%
Rate<.0009%

# **BIBLIOGRAPHY**

## BIBLIOGRAPHY

Bellaire, K., Yorkston, K. M., & Beukelman, D. R. (1986). Modification of breath patterning to increase naturalness of a mildly dysarthric speaker. <u>Journal of Communication Disorders</u>, 19, 271-280.

Beukelman, D. R., & Mirenda, P. (1992). <u>Augmentative and alternative</u> <u>communication: management of severe communication disorders in children and</u> <u>adults</u>. Baltimore: Paul Brookes.

Beukelman, D., & Yorkston, K. (1979). The relationship between information transfer and speech intelligibility of dysarthric speakers. <u>Journal of Communication Disorders</u>, 12, 189-196.

Braverman, J. H. (1974). <u>A comparison of oral form discrimination in deaf</u> and hearing children and the relationship of orosensory skill and other factors to <u>speech articulation in deaf talkers</u>. Unpublished Ph.D. dissertation, Columbia university.

Carney, A. E. (1994). Understanding speech intelligibility in the hearing impaired. In K. G. Butler (Ed). <u>Hearing Impairment and Language Disorders.</u> <u>Topic in Language Disorders Series.</u> Gaithersburg, Maryland: Aspen.

Colcord, R. D., & Gregory, H. H. (1987). Perceptual analyses of stuttering and nonstuttering child's fluent speech productions. <u>Journal of Fluency</u> <u>Disorders, 12</u>, 185-195.

Curlee, R. F. (1993). Evaluating treatment efficacy for adults: assessment of stuttering disability. <u>Journal of Fluency Disorders, 18,</u> 319-331.

Dawson, L. O.. (1929). A study of the development of the rate of articulation. <u>Elementary School Journal</u>. 29, 610-615.

Doehring, D. G. (1988). <u>Research strategies in human communication</u> <u>disorders.</u> Boston: College-Hill.

Finn, P., & Ingham, R. J. (1989). The selection of "fluent" samples in research on stuttering: conceptual and methodological considerations. <u>Journal of Speech and Hearing Research</u>. <u>32</u>, 401-418.

Finn, P., & Ingham, R. J. (1994). Stutterers' self-ratings of how natural speech sounds and feels. <u>Journal of Speech and Hearing Research. 37.</u> 326-340.

Finn, P. (1997). Adults recovered from stuttering without formal treatment: perceptual assessment of speech normalcy. <u>Journal of Speech</u>. <u>Language. and Hearing Research. 40.</u> 821-831.

Fudala, J., (1970). <u>Arizona Articulation Proficiency Scale: Revised</u>. Los Angeles: Western Psychological Services.

Goldsmith, T., & Anderson, D. (1984). The enigma of fluency: A single case study. <u>The South African Journal of Communication Disorders</u>. <u>31</u>, 47-52.

Guba, E., & Lincoln, Y. (1981). <u>Effective evaluation: improving the</u> <u>usefulness of evaluation results through responsive and naturalistic approaches</u>. San Francisco: Jossey-Bass.

Haynes, W., & Hood, S. (1977). Disfluency changes in children as a function of the systematic modification of linguistic complexity. <u>Journal of Communicative Disorders</u>, <u>11</u>, 79-93.

Healey E. C., & Ramig, P. R. (1986). Acoustic measures of stutterers' and nonstutterers' fluency in two speech contexts. <u>Journal of Speech and Hearing</u> <u>Research, 29</u>, 325-331.

Hegde, M. N. (1978). Fluency and fluency disorders: their definition, measurement, and modification. <u>Journal of Fluency Disorders. 3.</u> 51-71.

Hegde, M. N. (1995). <u>An introduction to communicative disorders.</u> 2nd edition. Austin, Texas: Pro-Ed.

Hoover, J., Reichle, J., Van Tassell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. <u>Journal of Speech</u> and Hearing Research. <u>30</u>, 425-431.

Hotchkiss, J. C., (1973). <u>Perceptual differences in the fluent speech of</u> <u>stutterers and non-stutterers</u>. Unpublished doctoral dissertation. Layfayette, Indiana: Purdue University.

Hudgins, C. V., (1949). A method of appraising the speech of the deaf. <u>Volta Review, 48,</u> 642-644.

Ingham, R. J., Gow, M., & Costello, J. M. (1985). Stuttering and speech naturalness: some additional data. <u>Journal of Speech and Hearing Disorders</u>, <u>50</u>, 217-219.

Ingham, R. J., Ingham, J. C., Onslow, M., & Finn, P. (1989). Stutterers' self-ratings of speech naturalness: assessing effects and reliability. <u>Journal of Speech and Hearing Research</u>, <u>32</u>, 419-431.

Ingham, R. J., & Onslow, M. (1985). Measurement and modification of speech naturalness during stuttering therapy. <u>Journal of Speech and Hearing Disorders, 50</u>, 261-281.

Ingham, R. J., Martin, R. R., Haroldson, S. K., Onslow, M., & Leney, M. (1985). Modification of listener-judged naturalness in the speech of stutterers. Journal of Speech and Hearing Research. 28, 495-504.

Ingham R. J., & Packman, A. C. (1978). Perceptual assessment of normalcy of speech following stuttering therapy. <u>Journal of Speech and Hearing</u> <u>Research. 21</u>, 63-73.

Johnson, L. (1987). <u>Multidimensional scaling of speech naturalness in</u> <u>stutterers</u>. Unpublished doctoral dissertation. University of Minnesota.

Kalinowski, J., Noble, S., Armson, J., & Stuart, A. (1994). Pretreatment and posttreatment speech naturalness ratings of adults with mild and severe stuttering. <u>American Journal of Speech and Language Pathology</u>, <u>3</u>, <u>2</u>, 61-66.

Keeler, K. D., Clement, G. L., Strong, W. J., & Palmer, E. P. (1976). Two preliminary studies of the intelligibility of predictor-coefficient and formant-coded speech. <u>Institute of Electrical and Electronics Engineers Transactions on</u> <u>Acoustics. Speech. and Signal Processing. 24.</u> 429.

Kowal, S., O'Connell, D. C., & Sabin, E. F. (1975). Development of temporal patterning and vocal hesitations in spontaneous narratives. <u>Journal of Psycholinguistic Research, 4</u>, 195-207.

Krikorian, C. M. & Runyan, C. M. (1983). A perceptual comparison: stuttering and nonstuttering children's nonstuttered speech. <u>Journal of Fluency</u> <u>Disorders, 8</u>, 283-290.

Laddaga, R., Sanders, W. R., & Suppes, P. (1981). Testing intelligibility of computer generated speech with elementary-school children. In P. Suppes (Ed.), <u>University-level computer-assisted instruction at Stanford: 1968-1980</u>. Stanford, California: Stanford University.

Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost, & L. Katz (Eds.), <u>Orthography. phonology. morphology. and meaning.</u> New York: Elsevier Science Publishers.

Linebaugh, C. W., & Wolfe, V. W. (1984). Relationships between articulation, intelligibility, and naturalness in spastic and ataxic speakers. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), <u>The dysarthrias:</u> <u>Physiology, acoustics, perception, management.</u> San Diego: College-Hill Press. Logan, J., Pisoni, D., & Greene, B. (1985). Measuring the segmental intelligibility of synthetic speech: Results from eight text-to-speech systems. <u>Research on Speech Perception (Progress Report No. 11).</u> Bloomington: Indiana University.

Love, L. R., & Jeffress, L. A. (1971). Identification of brief pauses in the fluent speech of stutterers and nonstutterers. <u>Journal of Speech and Hearing</u> <u>Research, 14</u>, 229-240.

Mackey, L., Finn, P., & Ingham, R. (1997). Effect of Speech Dialect on speech naturalness ratings: A systematic replication of Martin, Haroldson, and Triden (1984). Journal of Speech, Language, and Hearing Research. 40, 349-360.

Martin, R. R., & Haroldson, S. K. (1992). Stuttering and speech naturalness: audio and audiovisual judgments. <u>Journal of Speech and Hearing</u> <u>Research</u>, <u>35</u>,521-528.

Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. Journal of Speech and Hearing Disorders, 49, 53-58.

Mateer, C. (1993). Neural Bases of Language. In E. Boberg, (Ed.) <u>Neuropsychology of Stuttering.</u> Edmonton, Alberta: The University of Alberta Press.

Metz, D. E., Samar, V. J., & Sacco, P. R. (1983). Acoustic analysis of stutterers' fluent speech before and after therapy. <u>Journal of Speech and Hearing Research</u>, <u>26</u>, 531-536.

Metz, D. C., Schiavetti, N, & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. Journal of Speech and Hearing Disorders, 55, 516-525.

Mirenda, P., & Beukelman, D. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. <u>Augmentative and Alternative Communication. 3.</u> 120-128.

Mirenda, P., & Beukelman, D. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. <u>Augmentative and Alternative Communication, 6</u>, 61-68.

Nichols, A. C. (1966). Audience ratings of the "naturalness" of spoken and written sentences. <u>Speech Monographs. 33.</u> 156-159.

Onslow, M., Hayes, B., Hutchins, L., & Newman, D. (1992). Speech naturalness and prolonged-speech treatments for stuttering: further variables and data. Journal of Speech and Hearing Research, 35, 274-282.

Onslow, M., Costa, L., Andrews, C., Harrison, E., & Packman, A. (1996). Speech outcomes of a prolonged-speech treatment for stuttering. <u>Journal of</u> <u>Speech and Hearing Research. 39.</u> 734-749.

Packman, A., Onslow, M., & van Doorn, J. (1994). Prolonged speech and modification of stuttering: perceptual, acoustic, and electroglottographic data. Journal of Speech and Hearing Research. 37, 724-737.

Parrish, W. M. (1951). The concept of naturalness. <u>Quarterly Journal of</u> <u>Speech. 37</u>, 448-450.

Perkins, W. (1973). <u>Behavioral management of stuttering: Modification</u> of stuttering by behavioral manipulation of conversational speech. (SRS Final Report No 14-P-55281). Washington, DC: Social and Rehabilitation Service, Department of Health, Education and Welfare.

Prosek, R. A. & Runyan, C. M. (1982). Temporal characteristics related to the discrimination of stutterers' and nonstutterers speech samples. <u>Journal of Speech and Hearing Research, 25,</u> 29-33.

Runyan, C. M. & Adams, M. R. (1978). Perceptual study of the speech of "successfully therapeutized" stutterers. <u>Journal of Fluency Disorders, 3</u>, 25-39.

Runyan, C. M. & Adams, M. R. (1979). Unsophisticated judges' perceptual evaluations of the speech of "successfully treated" stutterers. <u>Journal of Fluency Disorders, 4</u>, 29-38.

Runyan, C. M., Bell, J. N., & Prosek, R. A. (1990). Speech naturalness ratings of treated stutterers. <u>Journal of Speech and Hearing Disorders. 55</u>, 434-438.

Sanders, W. C. Gramlich, C., & Levine, A. (1981). Naturalness of synthesized speech. In P. Suppes (Ed.), <u>University-level computer-assisted</u> instructions at Stanford: 1968-80 (pp. 487-501). Stanford, CA: Stanford University.

Shriberg, L. D., and Kwiatkowski, J. (1982). Phonological disorders III: a procedure for assessing severity of involvement, <u>Journal of Speech and Hearing</u> <u>Disorders. 47.</u> 256-270.

Starkweather, C. W. (1980). Speech fluency and its development in normal children. In N. J. Lass (Ed.), <u>Speech and language: advances in basic research and practice.</u> New York: Academic Press.

Stevens, S. S. (1975). <u>Psychophysics.</u> New York: John Wiley & Sons.

St. Louis, K. (1995). Investigations of speech naturalness ratings. In W. Starkweather and H. Peters (Eds.), <u>1st world congress on fluency disorders-proceedings. volume II.</u> Munich, Germany: International Fluency Association.

Tinsley, H. E. A., & Weiss, D. J. (1975). Inter-rater reliability and agreement of subjective judgments. <u>Journal of Counseling Psychology</u>, 22, 353-376.

Thatcher, R. W., Walker, R. A., & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. <u>Science. 230.</u> 110-113.

Thomas, W. G. (1964). Intelligibility of the speech of deaf children. <u>Proc.</u> <u>Int. Congr. on Educ. of the Deaf</u>. Washington, D. C.: U.S. Govt. Printing Office, 245-261.

U. S. Bureau of the Census. (1994). <u>County and City Data Book: 1994.</u> Washington, DC: U. S. Government Printing Office.

Van Riper, C., & Emerick, L. (1990). <u>Speech correction: an introduction</u> to speech pathology and audiology. New Jersey: Prentice Hall.

Ventry I. M., & Schiavetti, N. (1986). <u>Evaluating research in speech</u> <u>pathology and audiology</u>. second edition. New York: Macmillan Publishing Company.

Wexler, K., & Mysak, E. (1982). Disfluency characteristics of 2-, 4-, and 6-year old males. Journal of Fluency Disorders, 7, 37-46.

Winer, B. J. (1971). <u>Statistical principles in experimental design</u>. New York: McGraw Hill.

Yairi, E. (1981). Disfluencies of normally speaking two-year-old children. Journal of Speech and Hearing Research, 24, 490-495.

Yairi, E. (1982). Longitudinal studies of disfluencies in two-year-old children. Journal of Speech and Hearing Research, 25, 155-160.

Yorkston, K. M., Beukelman, D., R., & Bell, K. R. (1988). <u>Clinical</u> <u>management of dysarthric speakers</u>. Austin Texas: Pro-Ed.

