

11800

000000007

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 00627 7473



This is to certify that the

dissertation entitled

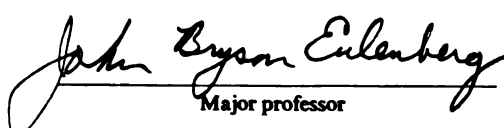
Children's Attitudes Toward Synthesized
Speech Varying in Quality

presented by

Sheila Bridges Freeman

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Audiology and
 Speech Sciences


Major professor

Date October 18, 1990

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		
SEP 13 2001	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU is An Affirmative Action/Equal Opportunity Institution

**CHILDREN'S ATTITUDES TOWARD
SYNTHESIZED SPEECH VARYING IN QUALITY**

By

Sheila Bridges Freeman

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

1990

list

spee

This

per

the

voi

exa

mea

the

for

and

ABSTRACT

CHILDREN'S ATTITUDES TOWARD SYNTHESIZED SPEECH VARYING IN QUALITY

By

Sheila Bridges Freeman

The goal of this research was to examine child listener's preferences and attitudes in response to natural speech and varying types of voice-output communication aids. This research addresses the following questions:

- 1) Does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used?
- 2) Do child listeners express a preference for different types of synthetic and natural speech?
- 3) Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

To investigate these questions two tasks were performed. In the first task, children's attitudes toward the hypothetical user of four types of speech (natural child voice, SmoothTalker 3.0, RealVoice, and DECTalk) were examined by the administration of a modified attitude measure, the CATCH (Rosenbaum, Armstrong, King, 1986). In the second task the subjects rated the quality of the same four voices along a 5-point Likert scale (Miranda, Eicher, and Beukelman, 1981).

S

assign

In Tas

monolo

questi

each c

C

and ov

signi

prefe

found

findi

aware

choic

This

shou

leve

atti

inte

Seventy-eight fourth grade children were randomly assigned to one of the four vocal conditions cited above. In Task I each subject heard an audio recording of a child's monologue in the assigned voice then completed the CATCH questionnaire. In Task II the subjects rated samples of each of the four voices on a 5-point Likert scale.

Children's overall attitudes toward the child VOCA user and overall voice quality preference were positive. A significant correlation was found to exist between voice preference and attitude. However, this correlation was found only to explain 13% of the variance. The major findings of this study indicate that children are clearly aware of differences in voice type and express a distinct choice for voices they like versus those they do not like. This suggests that a voice-output communication device should not be blindly prescribed. Its level of quality and level of listener's preference may contribute to the attitude of the listener toward its user and thus the social interaction between speaking and nonspeaking child peers.

I dedicate this research to my parents John and Anne Bridges and my sister, Annita M. Bridges for their eternal belief in me and for their unconditional love and to my dear husband Everette J. Freeman, whose words of wisdom, patience, and abiding love have steadied my course and helped me not to lose sight of my dreams.

wh

ac

sp

an

co

co

my

Dr

ou

an

St

f.

e.

My

f

st

my

in

fo

ACKNOWLEDGEMENTS

The completion of this dissertation reflects a journey which evolved through the course of my professional, academic, and personal development. This acknowledgement speaks to the many named and unnamed mentors, colleagues, and friends who inspired me along the way.

My sincere appreciation is extended to my dissertation committee for their mentoring and scholarly guidance in the completion of this research as well as their contribution to my professional and personal growth. My special thanks to Dr. John Eulenberg for teaching me we are only limited by our imaginations, to Dr. Gloriajean Wallace for her faith and trust "through Him all things are possible", to Dr. Ida Stockman for taking me under her wing and encouraging me to fly, and to Dr. Donald Burke for his constant support and encouragement.

I thank the Mason Public School District, particularly Mr. Kenneth McNeil, Dr. Korie Creaser, and Ms. Joyce Lincoln for making this research possible.

Many friends and loved ones have given their loving support. To them I express deep appreciation. I also extend my gratitude to Drs. Charlena and Harry Seymour for inspiring me to pursue my dreams and providing me a path to follow.

CHA

CHA

I

CHA

I

TABLE OF CONTENTS

CHAPTER

I.	INTRODUCTION	1
	Attitudes Toward Vocal Characteristics	6
	Attitudes Toward Disordered Vocal Characteristics	7
	Communicative Interaction Between Verbal and Nonverbal Communicators	9

CHAPTER

II.	REVIEW OF THE LITERATURE	12
	Speech Synthesizers	15
	Vocal Quality	19
	Intelligibility of Speech Synthesizers	21
	Attitudes Toward and Preference for Synthetic Speech	27

CHAPTER

III.	METHOD	36
	Overview of Goals and Focus of the Study ..	36
	Subjects	37
	Experimental Stimuli	39
	Behavioral Measures	46
	Data-gathering Tasks	50
	Data Analysis Procedures	54

CHAP

IV

CHA

AP

CHAPTER

IV. RESULTS	57
Attitude Measure	57
Does the Attitude of the Child Listener Toward the Child VOCA User Vary as a Consequence of the Different Types of Synthetic and Natural Speech Used? ...	60
Do Child Listeners Express a Preference for Different Types of Synthesized and Natural Speech?	68
Is There a Relationship Between the Child Listener's Preference for Natural and Synthetic Speech and His/Her Attitude Toward the Child VOCA User?	73

CHAPTER

V. DISCUSSION	75
Correlation Between Attitude and Voice Preference	77
Extentions of the Study	79
Implications for Future Research	82
Application and Clinical Implications	92

APPENDICES

Appendix A Parent Consent Form	95
Appendix B Lee's Monologue	98
Appendix C Mrs. Harris's Monologue	99
Appendix D Practice Form	100
Appendix E Modified CATCH	101
Appendix F Rating of Voice Quality	107
Appendix G Instructions for Test Administrators ...	108

APPENDICES (cont)

Appendix H Instructions for Students	110
Appendix I Factor Analysis of the Modified CATCH ..	112
LIST OF REFERENCES	113

LIST OF TABLES

Tables

1. Pertinent Subject Characteristic	38
2. Statistical Review of Attitude (CATCH) Scores According to Voice Type, Gender, and School	62
3. Analysis of Variance for Total CATCH and Subtests	65
Total CATCH attitude source table	65
Affective attitude source table	65
Behavioral attitude source table	66
Cognitive attitude source table	66
4. Statistical Review of Quality Rating According to Voice Type, Gender, and School	70
5. Analysis of Variance for Vocal Preference	72

CHAPTER I

INTRODUCTION

The impact of voice synthesis and computer technology has provided a new direction in the realm of augmentative and alternative communication. Many once speechless persons have been given the means by which to verbally communicate, converse, lecture, debate, orate, pray, and even sing for the very first time. Voice-output communication has become a reality for thousands of individuals (Dahmke, 1982). However, it is suspected that even more individuals could potentially benefit from such technology (Cohen & Palin, 1986) due to advancements in both dedicated augmentative communication systems (e.g. Alltalk, Touch Talker, Light Talker, Epson SpeechPAC, Vois 136) and adapted laptop microcomputers (e.g. RadioShack Models 100, 102, 200, Toshiba T1100 Plus, and Zenith 183). In the past decade, both approaches to communication systems design have been increasingly successful in meeting the needs of sensorily, physically, and communicatively handicapped individuals in their flexibility, accessibility, compatibility, expandability, and portability (Schwartz & Koenig, 1987). The social and cognitive development of handicapped children has been particularly enhanced by the current technology (hardware and software) which enables them to become active participants in their environment.

When compared to nonspeech or unaided augmentative modes of communication, voice-output communication aids (VOCAs) do offer a number of advantages. VOCAs provide the opportunity to experience speech production for the first time, to interact with individuals across the room or even at greater distances by telephone. VOCA users are no longer restricted to a visual mode of communication where initiation of interaction is dependent upon first establishing eye contact nor are they limited to interacting with sighted and literate individuals. The auditory feedback of speech synthesis provides the user immediate feedback upon the occurrence of an error in constructing a message, in order to self-correct and to facilitate the learning of new vocabulary. Consequently, the frequency of communication breakdowns is significantly decreased with less familiar interactants who may not be familiar with manual signing, picture symbol or other non-oral communication system. As most text-to-speech systems generate an unlimited and/or programmable vocabulary with an extensive storage capacity, most users are not limited by lack of access to an elaborate vocabulary as needed. Such systems afford the user multiple output options including a display and printer (Vanderheiden, 1983; Eulenberg & Rahimi, 1978; Goossen & Kraat, 1985; Harris, 1982; Kraat, 1985; Locke & Mirenda, 1988).

On the other hand, there are limitations to voice-output communication aids. It is still unrealistic to expect VOCA

use to
techn
vocal
conti
of re
(hum
tell
cons
cont
diff
dial
ind
ada
pro

imp
as
of
im
an
qu
hi
"u
wh
es

q

use to directly parallel spoken communication. While current technological advancements have significantly improved the vocal quality and intelligibility of speech synthesis, VOCAs continue to fall short in these areas, lacking the capability of replicating the intonation found in emotional expression (humor, sarcasm, anger, compassion, etc.) necessary for telling a joke, debating an issue, expressing sympathy, consoling, reprimanding, etc. Speech synthesis also continues to lack the capacity to express appropriate differences in gender, age, social, regional and ethnic dialect as well as personality characteristics of the individual user. Equally limiting is the lack of adaptability to less than optimal speaking conditions (i.e. projection over noise or whispering in a quiet environment).

To date, little research has examined the possible impact of these VOCA limitations on their functional adequacy as a communicative tool. The following are among the kinds of questions raised by this issue: How does intelligibility impact on social interaction, communicative effectiveness, and attitude of the listener? How does the "unnatural" quality of synthesized speech impact on the listener, and on his/her attitude toward the user? Initially, does this "unnatural" quality evoke negative attitudes from child peers which may be detrimental to social interaction and the establishment of social relationships?

This study was not designed to answer all these questions. Rather, it has focused generally on children's

preferences for and acceptance of synthesized speech in anticipation that the results would have practical implications for the design and selection of voice-output communication aids. Specifically, this investigation addressed the question of whether a relationship exists between children's voice quality preferences for varying types of voice-output communication aids and child peer social acceptance.

This question is considered relevant to the potential child VOCA user as research has indicated that the lack of social and communication experiences as a result of disability places the disabled child VOCA user at risk for experiencing reduced expressive communication skills (Kraat, 1985). Research suggests that children with developmental conditions resulting in communication impairment and physical disabilities may have reduced social, communicative, and cognitive experiences (Harris & Vanderheiden, 1980; Morris, 1981; Yoder & Kraat, 1983). Developmental delay has been identified as a consequence of limited social participation (Strain & Shores, 1977). While not all severely physically limited children evidence impoverished experience, they reflect a high risk population. Physical limitations restrict independence and the exploration of objects. Within the home environment the caregiver often fails to recognize the need to provide stimuli important to social and cognitive development which the child is unable to independently provide or request. Disabled children have few opportunities

to play and interact with individuals outside the environment of their immediate family where social relationships are mandatory (Richardson, 1969). These relationships are maintained for a long time with limited opportunity or exposure to voluntary social relationships with a variety of other people (Richardson, 1969).

With the 1975 passage of The Education for All Handicapped Children's Act (P.L. 94-142), there has been an integration of nonspeaking physically disabled children into the education system, carrying out a philosophy of improving educational quality and opportunity with the hope of changing the attitudes of nonhandicapped peers. Despite the growing phenomenon of heterogeneous groups of handicapped and nonhandicapped children in public schools, severely physically handicapped children tend to interact with adults (teachers and aides) more than with their peers.

Guralnick (1981) cites experimental evidence that in small play groups nonhandicapped children can play an important role in educational and therapeutic programs (Guralnick, 1976; Apolloni & Cook, 1978). While nonhandicapped children have been observed to adjust the complexity and characteristics of their communication to the level of their handicapped peers (Guralnick & Paul-Brown, 1980) failure at their initial attempts toward interaction reduces and eventually extinguishes further chances for repeated attempts (Goldstein & Terrell, 1987).

While the provision of an augmentative communication system is often seen as a way to increase peer social interactions thereby reducing this social isolation, this goal obviously could not be met if children have negative attitudes toward VOCA voice quality. Numerous studies provide evidence which suggests that speech characteristics can evoke negative attitudes. In the 1960's studies in sociolinguistics in particular revealed sensitivity to and awareness of the attitudes evoked by dialectal and language variations (Anisfeld & Lambert, 1964; Fishman 1969; Tucker, 1969; Ervin-Tripp, 1967). Similar findings have been reported as a result of examining disordered and normal vocal characteristics as described in the following section.

Attitudes toward vocal characteristics. Giles and Powesland (1975) attribute the tendency to make judgements and inferences about people on the basis of physical and acoustic characteristics as representing the application of our own "implicit" personality theories. In this way, we construct impressions about individuals based on the information which has been made available to us. Giles and Powesland (1975) cite the following studies which examine the contributions of voice quality to listener's attitudes and impressions.

Mehrabian and Weiner (1967) examined the relative contributions of vocal qualities and verbal content to the impressions people form regarding receptiveness (like/dislike) of the person they are addressing. They

qua
con
inf
and
cir
con

ev
fo
ma
ei
sp
dr
co
mo
th
th
ev

cl
t
s
p

d
c
f

quantified these contributions as; 7% being from speech content, 38% from vocal qualities, and 55% from facial information. This was consistent with the findings of Argyle and associates (1970) suggesting that under certain circumstances noncontent cues can carry more weight than the content itself.

Seligman, Tucker and Lambert (1972) lend supporting evidence regarding the importance of noncontent speech cues for evaluating personality. Student-teachers were asked to make subjective judgements using a 7-point scale regarding eight 3rd-grade boys based on photographs, tape-recorded speech and a sample of their work (composition and a drawing). Results showed that boys with "good" voices were consistently rated more favorably (i.e. more intelligent, more privileged, self-confident, gentle and enthusiastic) than the boys with "poor" voices. The authors thus concluded that speech style was an important clue to teachers in their evaluations of students.

Addington (1968) found that idiosyncratic characteristics of voice (thinness, flatness, nasality, tenseness, throatiness and orotundity) evoked personality stereotypes (i.e. breathy-feminine, pretty; thinness-social, physical and mental immaturity).

Attitudes toward disordered vocal characteristics. The distinction made between normal and disordered vocal characteristics is often a subjective judgement. Yet, we find a high rate of agreement among listeners with regard to

those vocal characteristics deemed "unusual". The social impact of disordered vocal quality can not be underestimated.

Beukelman (Kraat, 1985) cites a case study where a dysarthric woman seeking employment found certain characteristics of her voice quality evoked quite negative reactions. This case study is one of many which illustrate the need to systematically determine those characteristics which are most detrimental to social interaction and attitude.

Blood, Mahon and Hyman (1979) examined the effect of voice disorders on personality and appearance judgements. Results indicated a significant difference between the ratings of normal and disordered speakers. Voice disordered speakers elicited significantly more negative responses on judgements pertaining to speaker's personality and appearance.

Silverman (1976) conducted a study to determine whether a lateral lisp was perceived to be a speech defect by naive listeners. Two groups of students rated "The Person Speaking" (lateral lisp speaker or normal speaker) based on a 49-scale semantic differential. Results, indicated the lateral lisp was reported to call adverse attention to the speaker such that 37 of the 49-item scale were judged negatively.

These studies have examined perceptions evoked by disordered speech characteristics (lateral lisp) and vocal quality. Semantic differentials and categorical ratings

illustrate the negative perceptions which span dimensions of personality, appearance, acceptability, and handicapp. It is quite possible to generalize such findings to synthesized speech. The insight found in Sarah Blackstone's statement, "If these synthesizers were alive we would be treating them in our augmentative clinics!" (Blackstone, 1988) illustrates an analogous relationship between synthesized and disordered speech.

It is quite evident from these studies that vocal characteristics of the normal and disordered voice contribute greatly to listener's attitudes and impressions toward the speaker, hence contributing in some instances more than verbal content to receptiveness and social attractiveness, as well as evoking personality stereotypes such as femininity, attractiveness, and social and mental maturity.

Communicative interaction between verbal and nonverbal communicators. Unfortunately, little is known about the communicative interaction between speakers and users of electronic augmentative communication devices. However, the research suggests that the availability of an augmentative communication system does not necessarily promote or lead to effective communication or social interaction. The passive/respondent role of the augmentative system user has been cited by a number of studies (Calculator & Luchko, 1983; Harris, 1982; Calculator & Dollaghan, 1982; Ferrier & Shane, 1981). Among studies relevant to this issue is the research of Harris (1978) (as cited by Kraat, 1985). In this study

three children using AutoComs were observed interacting with their teachers in three contexts. The nonspeaking children were found to use aids minimally, rarely interact with peers, infrequently initiate exchanges and communicated primarily through one word responses and non-verbal behaviors. This behavior contrasts with the interaction among nondisabled children which occurs primarily with peers.

Augmentative system users have a reduced number of interactants, communicating less often with peers, younger children, and less familiar persons (Ferrier & Shane, 1981; Yoder & Kraat, 1983) to the detriment of not only the handicapped child but also his/her nonhandicapped peers. The benefits of such interaction lie not only in providing a necessary language learning, socially and emotionally enriching experience to the handicapped child but also to his/her speaking counterpart.

Consequently, this study examined the relationship between the listener voice quality preference for varying types of VOCAs and the level of peer (child VOCA user) acceptance as measured by scores on an attitudinal measure. This study attempts to answer the following questions:

- 1) Does the attitude of the child listener toward the child VOCA user vary as a consequence of the varying types of natural and synthetic speech used?
- 2) Do child listeners express a preference for different types of synthesized and natural speech?

3) Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

These questions were addressed by eliciting from child subjects, (1) their attitudes toward the hypothetical users of four types of speech (three synthesized and one naturally produced) and, (2) their ratings of voice quality preference for the same four voices. This study hopefully can lead to a clearer understanding of the social impact of synthesized speech.

The literature review which follows examines these issues in two parts. Part one addresses the vocal characteristics and problems associated with synthesized speech. Part two reviews current studies examining intelligibility and listener's preference for natural and synthetic speech.

S
C
P
C
J
S
A
H
S
V
A
K
V
T
C
I
A
E
H
R
G
A

CHAPTER II

REVIEW OF THE LITERATURE

"It is the listener's subjective reactions to synthesized speech that are important to the ultimate utility of a particular speech synthesizer" (Nusbaum, Schwab, & Pisoni, 1984). The acceptance of a VOCA is a major consideration in its selection and for eventual success in its functional use. The use of a device must take place in a social interactive situation to be functional, thus its acceptance must be by both the user as well as those in his/her environment. In his discussion on "presentation of self" Goffman (1959) stated that communication is one way in which we present ourselves to other people in this world and attempt to influence what they see and think of us. This presentation is made in several ways, including our choice of words, our tone of voice, the attitudes we display and the topics we choose (cited in Kraat, 1985). An essential characteristic of speech or spoken communication, necessary for effective communication is its contribution to conveying attributes of the speaker, such as the speaker's personality, emotions, geographical, ethnic and socioeconomic background, his or her relative relationship to the listener, and a natural voice quality consistent with the speaker's age, gender, and developmental level (Newell, 1984; Eulenberg et al., 1985).

Unfortunately, many of these characteristics essential to natural speech are lacking in synthesized speech, thus evoking much criticism regarding the unnatural and questionable intelligibility of speech synthesizers (more specifically, VOCAs). Such criticism has been well documented. Some aid users have expressed very negative feelings about the unnatural quality and poor intelligibility of synthetic speech. Negative feelings are clearly conveyed in this statement by Holmquist, a VOCA user, "I must confess, I can't identify myself with the voice" (Kraat, 1985, p. 123).

VOCA users and listeners have expressed concern for intelligibility and need for more appropriate expressive intonation (Eulenberg et al. 1985). Bernstein (1985) pointed out that VOCA users wish to have the ability to express themselves with the full range of emotions and attitudes as most people, ranging from gentle/compassionate to impatient/angry. All people want voices which are appropriate to their ages and gender. Bernstein identified three ways in which VOCAs need to be improved upon to address these issues. First, increased phonetic accuracy is needed to increase the similarity between a synthesized and a real voice and to facilitate word identification; second, indexical flexibility, enabling the user to select speech attributes that identify the speaker's age, gender, size, regional accent, and dialect; and third, paralinguistic

control, giving the user control over those attributes of speech conveying the emotional attitude of the speaker (i.e. urgency, friendliness, impatience, etc.). In the absence of these features current systems have been described as unnatural.

In Klatt's, (1986) discussion of the evolution of text-to-speech synthesis, he identified error rate, unnatural timing, intonation, and voice quality (vowel quality change as a function of stress and phonetic environment) as parameters contributing to the listeners' impressions of unnaturalness.

According to Galyas, (1988) (cited by Crabtree et al., 1989) quality of synthetic speech is dependent upon two related but independent dimensions, intelligibility and naturalness. Because of the high correlation found between intelligibility and naturalness (Klatt, 1985), low intelligibility may be highly predictable of unnatural voice quality. Intelligibility studies have shown that many of the currently available speech synthesizers have low intelligibility, with scores below the 81% intelligibility level needed for speech to be understood (Blackstone, 1988).

In summary, concern has been expressed regarding the male robotic voice quality, lacking in appropriate gender (female) and age (child) which characterizes many speech synthesizers. The unnaturalness of the speech reflects the lack of proper paralinguistic features of inflection, pause,

rhythm, and rate, resulting in a lack of emotional tone and reduced intelligibility (Bernstein, 1986; Eulenberg et al., 1985; Klatt, 1986).

Speech Synthesizers

Improvement of vocal quality in speech synthesis is apparently no small task, as the development of human sounding speech synthesizers has been in the making for well over 30 years. The technology of speech synthesis has been in existence since the 1950's where its early use was found in military and industrial application (Blackstone, 1988). However, its first application to augmentative communication only came into being a little more than a decade ago.

In May 1978, the Phonic Mirror HandiVoice appeared on the market as the first commercially available VOCA (Eulenberg et al., 1985). The Votrax VSH voice synthesizer, a low-power single-board was manufactured for this device. The Phonic Ear Vois, an improved product of the same line was introduced in 1982 using the Votrax SC01 synthesis chip (Eulenberg et al., 1985). These early versions were described as being very "male and robotic" sounding. In response to the criticisms of VOCA consumers and concerned others, regarding the poor intelligibility, inappropriate intonation, and male robotic quality, other technologies were developed. Quality was somewhat improved with the 1982 introduction of the Vocaid (Texas Instrument) which was based on linear predictive coding (Eulenberg et al., 1985). In

spite

qualit

a vari

two mo

synthe

produ

voice

natur

throu

(Coh

repli

techn

throu

appro

exces

stor

for

appl

(Ada

(Pre

spec

the

comp

prin

spite of its improved quality it still possessed a male vocal quality and generated a limited vocabulary. Since that time a variety of technologies have been employed in VOCAs, the two more popular being digitization and text-to-speech synthesis.

Digitization represents a high-quality of speech production most closely simulating the quality of the human voice. Digitized speech recording is produced by collecting natural speech samples with a microphone and passing it through a series of filters and a digital-to-analog converter (Cohen & Palin, 1986). The output is a high fidelity replication of the original signal. The advantage of this technique is its ability to generate a variety of voices through the recording of gender-, age-, and family-appropriate speech. However, a major disadvantage is the excessive amount of computer memory required for coding and storing each individual sample, thus lacking the flexibility for generating a spontaneous vocabulary by text. The application of digitized speech is found in the Alltalk (Adaptive Communication Systems, Inc.) and the IntroTalker (Prentke Romich, Co.).

Manufacturers of speech synthesizers utilize a highly specialized patterned formula which has not been released to the public. Due to the competitiveness of the market, companies use proprietary technology. However, the general principle of text-to-speech synthesis employs a flexible

m
a
i
a
s
a
t
s

mathematical algorithm which represents rules for combining acoustic properties and rules for pronunciation. Its design is based on an explicit model or set of rules of how sounds are related to words, phrases, and punctuation. As a result, standard orthographic text input as generated by keyboard or alphanumeric membrane, is converted into phonetic text and then changed into a parametric description of speech which specifies the frequency and amplitude of various sound sources and resonances at each point in time. This parametric stream is then used to synthesize an acoustic signal (Mirenda & Beukelman, 1987; Blackstone, 1988).

The variation in quality found among speech synthesizers can in part be attributed to the complexity of the series of linguistic processes by which input is accepted and processed. DECTalk (Digital Equipment Corp.) generates signal parameters based on a constructive synthesis algorithm designed by Dennis Klatt (1980). Text input by computer keyboard is first converted to a pronunciation code using a dictionary and set of algorithmic rules (Cohen & Palin, 1986). By generating speech parameters from scratch (without using portions of prerecorded speech segments) the parameters specify the frequency and amplitude of various sound sources and filters (Bernstein, 1988). The code is then used to create speech via a digital-to-analog converter (Cohen & Palin, 1986). This method is employed to produce the

DECTalk, male, female and child voices (Mirenda and Beukelman, 1987).

The most important differences between the principle of constructive synthesis found in the DECTalk versus other synthesizers (contributing to its high level of intelligibility) is the "level of detail in which phonetic events are modeled and the number of context-sensitive rules used to imitate natural segment-to-segment transitions" (Bernstein, 1988).

Much of the naturalness of speech found in the higher quality of speech synthesis can be attributed to the use of concatenated diphones. Diphones are segments of speech that begin at the steady-state frequency midpoint of one phoneme and ends at the steady-state midpoint of the succeeding phoneme, thus preserving the natural transition found between phonemes (as opposed to allophones or triphones). The application of concatenated diphones is found in DECTalk, SmoothTalker 3.0 (male and female) as well as the RealVoice (Mirenda & Beukelman, 1990; Blackstone, 1988). SmoothTalker 3.0, developed by First Byte and found in Prentke Romich Company's Touch Talker and Light Talker, uses a software driven approach for generating speech. RealVoice is used by Adaptive Communication Systems, Inc., in the SpeechPac and ScanPac and most recently has been made available to be compatible with a variety of computers and other products (Blackstone, 1988).

Vocal Quality

Limited research has been generated examining the vocal quality of voice-output communication aids. The paucity of literature in this area may be partially attributed to the just recently improved higher quality of speech synthesis. However, past research has examined natural voice quality as ascribed to human speech and the attitudes thus evoked. Before examining the quality issue of synthesized speech it is important to examine a general framework of voice quality and those characteristics to which it has been ascribed.

Vocal characteristics. Voice quality is a complex acoustic phenomenon composed of many features such as volume, pitch, pause, duration, inflection, rhythm, and rate. Subject to the ear of the individual listener, as well as the listener's age, gender, culture, and context, perception of voice quality is quite subjective.

Personality traits such as "kind"- "cruel", "strong"- "weak", "pretty"- "homely", "fast"- "slow", are often ascribed to vocal characteristics. Speech characteristics affect the way in which we perceive and evaluate others. Initial impressions are often based on perceived vocal quality resulting in quick judgements about who we would or would not trust, or would not like as a friend, colleague or neighbor (Newcombe, 1986). Descriptive labels are often subjectively attached to voice quality reflecting personality traits (e.g. "pleasant"), esthetic quality (e.g. "beautiful"),

or acoustic characteristics (e.g. "hoarse, "nasal"). While pleasantness is subjective and may vary depending on the speaker's culture, age, gender, as well as context, guttural, harsh, throaty, hoarse, thin, nasal, denasal, and breathy are most commonly agreed upon as unpleasant or undesirable qualities (Newcombe, 1986). The following ascribed characteristics serve to illustrate the degree of negativity they evoke. Nasal quality suggests laziness. A breathy quality suggests fragility and helplessness. Thin vocal quality is interpreted as weak or immature and strident quality as shrewish (Newcombe, 1986).

Voice quality can not be completely isolated from other distinguishing vocal features such as volume, pitch, pause, duration, inflection, rate, and rhythm. Pitch patterns can convey meaning, clarifying stressed versus unstressed syllables (e.g. im'port vs. im port') (Newcombe, 1986). Variation of pitch or vocal inflection serve to identify statements as declarative, interrogative, imperative, or explanatory (Newcombe, 1986). Pitch also indicates emotions such as anger, sarcasm, humor, and connotative meaning serving to stimulate and sustain the listener's level of interest. The absence of pitch variation (monotone inflection) suggests coldness, boredom, ignorance, and mechanical or robotic characteristics (Newcombe, 1986).

Pauses mark the end of a thought sequence, establish mood, or indicate change in time, place, and characters. The

number, length, and location affect the rhythm, perceived fluency, rate and meaning of oral communication. An excessive number of pauses disrupts fluency, such that the listener perceives the rate as excessively slow and labored. Number of words or syllables pronounced within a specific amount of time also determine rate.

Rate is a major factor in determining intelligibility of a message. Like the other measures, pause and pitch, rate is subjective, dependent upon the listener's level of interest in the topic, emotional involvement, and time restrictions. A rate which is considered slower than normal is perceived as lethargic, monotonous, or lacking in self-confidence. A rate considered as faster than normal is perceived to be pushy, aggressive, or eager (Newcombe, 1986). Fluency of diction and consequently intelligibility are sacrificed in instances of both excessively fast or excessively slow rates.

Intelligibility of Speech Synthesizers

Within the past decade much of the research concerning speech synthesis has been in the area of synthetic versus natural speech intelligibility and listener perception. Much of this research has been generated by Pisoni and colleagues at the Speech Research Laboratory (Department of Psychology) at Indiana University. Various methodologies have been used to examine intelligibility and listener perception. For instance, Greene, Logan and Pisoni (1986), and Greene, Manous and Pisoni (1984) used the Modified Rhyme Test (MRT) (House,

Williams, Hecker, and Kryter, 1965), a closed-response test in which the subjects were presented stimulus items in stem form on an answer sheet and were then required to supply the missing letter (based on his/her perception of the auditory stimulus) given six response alternatives.

In another study examining constraints on the perception of synthetic speech generated by rule, Nusbaum and Pisoni (1985) utilized both a closed- and open-response format version of the MRT, the Harvard psychoacoustic sentence (i.e. sentences which are both syntactically correct and meaningful), and the Haskins syntactic sentences (i.e. sentences which are syntactically correct but not meaningful). These tasks examined word recognition in sentences.

Greene and Pisoni (1988) examined listening comprehension in connected speech in which 15 narrative passages were presented with a set of multiple-choice questions taken from standardized adult reading comprehension tests.

In the research of Mirenda and Beukelman (1987, 1990) the intelligibility of speech synthesizers was examined using the Computerized Assessment of Intelligibility of Dysarthric Speech Software (CAIDS: Yorkston, Beukelman, & Traynor, 1984). This software was used to construct the stimulus sentence and word intelligibility tasks. Utilizing an Apple IIe computer the authors were able to generate stimulus

sentences from a master pool of 100 sentences. Word lists were generated in a similar way.

These studies represent just some of the procedures used to examine intelligibility and listener perception. Due to variability in methodology as well as variability in issues related to intelligibility, it is not possible to make direct comparisons among these studies. However, general findings suggest that, (1) listeners can comprehend speech from the text-to-speech synthesizers in question (i.e. MITalk-79, Prose 2000, and DECTalk) at fairly high levels of performance compared to other speech synthesizers, (2) context is a powerful aid to understanding, and (3) only after a few minutes of exposure to synthesized speech output, subjects' comprehension level improves substantially (Greene & Pisoni, 1988). At this point the intelligibility studies of Mirenda and Beukelman (1987, 1990) will be discussed.

Mirenda and Beukelman (1987) compared the intelligibility of four different voice types (EchoII+, Votrax Personal Speech System, DECTalk and a natural voice) by listeners from three different age groups (6-8 year olds, 10-12 year olds, and adults) in single word and sentence intelligibility tasks. The results indicated that the word intelligibility scores were lower than the sentence intelligibility scores for all speech synthesizers but not for natural speech. In the word intelligibility task significant differences in intelligibility were found among

the four speech types ranging from natural voice, DECTalk (male, female, and child), Votrax, and Echo II. No difference between age groups was found.

In the sentence intelligibility tasks differences were found between the voice types and age groups. Intelligibility scores were found to decrease with decrease in age, however, no significant difference in intelligibility was found between intelligibility scores for the natural voice and the three DECTalk voices (male, female, and child).

Examining the performance of the 6-8 year olds, word intelligibility scores were reported to range from 93.6% (natural voice) to 30.8% (Echo+ with standard english spelling). The DECTalk child voice (Kit the Kid) received an intelligibility score of 59.6%.

Sentence intelligibility scores ranged from 94.2% (natural voice) to 35.8% (Echo+ with standard English spelling). The DECTalk child voice received a sentence intelligibility score of 80.9%. The authors expressed some concern regarding the interpretation of these result. Beukelman and Yorkston (1979) (as cited by Mirenda & Beukelman, 1987) found that subjects' ability to answer questions about passages read by a dysarthric speaker deteriorated rapidly when the speaker's intelligibility decreased below approximately 81%. This evidence suggests that while DECTalk voices were not found to be significantly different in intelligibility from natural speech, it may be

significantly inferior to natural speech in terms of information transfer (i.e. level of comprehension). Because the level of information transfer begins to decrease rapidly at 81% in adults, it may be suspected that children would experience even greater difficulty, as in the sentence intelligibility task, intelligibility was found to decrease with decreasing listener age.

More recently, a second study by Mirenda and Beukelman (1990) compared word and sentence intelligibility among natural speech and seven speech synthesizers (SmoothTalker 3.0 male, SmoothTalker 2.0 male, SmoothTalker 2.0 female, RealVoice female, Artic R65B, Votrax SC02, and Lightwriter voice) by listeners from three age groups (7-8 year olds, 11-12 year olds, and adults). Different patterns of intelligibility were found across the three age groups and eight vocal conditions. In spite of differences found among the three age groups an overall trend was noted. Natural speech was found to be more intelligible in all of the data sets (with the exception of the adult sentence task where the SmoothTalker 3.0 male was statistically equivalent in intelligibility to the natural speech). Four groupings of voice intelligibility emerged, a) natural speech; b) SmoothTalker 3.0 male and RealVoice; c) Votrax SC02, Artic R65B, and Lightwriter; and d) SmoothTalker 2.0 female and male.

Examining the performance of the younger child group (7-8 year olds) intelligibility scores for the word intelligibility tasks ranged from 96.29% (natural speech) to 27.71% (SmoothTalker 2.0 male). Word intelligibility for the RealVoice and the SmoothTalker 3.0 male was 48.57% and 45.43% respectively. Comparison of these scores with that obtained for the 6-8 year olds for the DECTalk child (59.6%) as reported by Mirenda and Beukelman (1987) (in the previously cited study) indicates that word intelligibility is poorer for these two synthesizers.

Scores for the sentence intelligibility tasks for 7-8 year olds ranged from 94.76% (natural speech) to 41.11% (SmoothTalker 2.0 male and female). Sentence intelligibility for the RealVoice female and SmoothTalker 3.0 was 50.63% and 67.46% respectively. Natural speech, SmoothTalker 3.0, and RealVoice intelligibility scores were found to be significantly different. Comparison of these scores with that obtained for the 6-8 year olds for the DECTalk child (80.9%) as reported by Mirenda and Beukelman (1987) indicates that sentence intelligibility is poorer for these two synthesizers.

Naturalness of voice quality may or may not include intelligibility, as speech may be quite intelligible but lack naturalness due to what the listener perceives to be a "male robotic" quality. Yet, naturalness may be sacrificed as a consequence of poor intelligibility. Naturalness thus

reflects the degree to which speech resembles that of human speech. In order to examine the issue of natural quality it is necessary to investigate the subjective perception and level of acceptance expressed by listeners.

Attitudes toward and preference for synthetic speech.

Only within the past several years have studies begun examining listener's preference for and acceptability of various types of synthesized speech. Speech synthesizers examined have included DECTalk, SmooothTalker, RealVoice, Echo II+, Votrax Type-N'-Talk, MITalk, Prose 2000, and the Artic R65B. Preference and acceptability were found to be largely determined by many of the same qualities evaluated in natural speech, including naturalness, intelligibility, and age- and gender- appropriateness. As a consequence, in most instances natural voice was preferred over synthetic speech (Crabtree et al., 1989; Quist & Lloyd, 1988; Nusbaum, Schwab & Pisoni, 1984).

Listener preference for natural and synthetic speech.

A variety of methods have been used as a way to "get at the subjective measure of speech quality". Many of these methods have evaluated speech quality in terms of subjective preference (Nusbaum et al., 1984). Nusbaum and associates (1984) identified the following methods of preference ratings.

The relative preference method utilized reference samples of speech spanning a range of speech quality.

Listeners then make pairwise preference comparisons between the test and reference samples. Other methods have used rating scales and multidimensional scaling of judgements along with Osgood's semantic differential technique. This technique elicits judgements of speech quality along different rating scales as defined by opposing adjectives (e.g. "annoying" vs. "pleasant"). Voiers (1977) (as cited by Nusbaum et al., 1984) generalized the rating scale approach found in the Diagnostic Acceptability Measure. In utilizing this approach listeners rated speech samples on several different rating scales which measured a different perceptual measure of quality (i.e. raspiness).

In the absence of a systematic investigation of perceived quality, studies have elicited measures of intelligibility, preference, and naturalness using one or a combination of the previously described methods. In Voiers' (1980) (as cited by Nusbaum et al., 1984) comparison of several vocoders in different noise conditions he concluded that acceptability is strongly related to intelligibility of speech.

Nusbaum and associates (1984) examined human speech, the Votrax type-'n-talk, and the MITalk-79 under 4 conditions. Subjects were required to: 1) complete a test of comprehension based on passages read by each of the 3 vocal conditions, 2) rate the speech on each of 17 scales defined by opposing adjective pairs, 3) provide an estimate of how

much they would trust different kinds of information provided in the speech they had heard, and 4) answer several questions on their reactions to various aspects of the experiment. These activities served to lead to the development of an evaluation test that indicated subjective differences between natural and synthetic speech and provided information about the acceptability of different types of text-to-speech synthesizers.

This study resulted in identifying suprasegmental qualities of speech that served to target differences between natural and synthetic speech (i.e. to reflect the listener's perception of the naturalness of speech). Qualities ascribed by listeners to synthetic speech were that it was more choppy, coarse, old, harsh, rough, and foreign than natural speech. These qualities were related to general prosodic characteristics, intonation and timing. It was further concluded that acceptability was closely related to the intelligibility of the speech. Adjective pairs describing the acceptability of speech relative to intelligibility were comfortable/frustrating, and annoying/ pleasant. The degree of effort involved in understanding the speech samples (i.e. easy/hard, clear/confusing, distracting/ improves concentration) suggested the listeners' perceived level of intelligibility.

Mirenda, Eicher, and Beukelman (1989) examined the preferences of 4 age groups (6-9 year olds, 10-12 year olds,

adolescents and adults) in rating 11 different voices (4 natural and 7 synthetic) along a 5-point Likert Scale given 6 different contexts. Varying contexts were presented given the following hypothetical users: adult male, adult female, child male, child female, computer, and self.

The eleven voices used in the study included: a) DEctalk male; b) DEctalk female; c) DEctalk child voice; d) Echo II+ (standard English spelling entered); e) Echo II+ (phonetically entered); f) Votrax (standard English spelling entered); g) Votrax (phonetically entered); h) natural adult female voice; i) natural adult male voice; j) natural child female voice and; k) natural child male voice. The subjects were required to rate each recording of a two minute story read by the eleven voices, along a 5-point Likert scale, ranging from "I would like it alot" to "I wouldn't like it at all". In the rating of each voice the subjects were asked to imagine that the speaker was related to them in some way (i.e. father, mother, brother, sister, self, computer).

The results indicated that the female listeners (across all ages) found only the natural female voice acceptable for their own voice while rejecting all other alternatives (i.e. male natural and all synthetic voices). The male listeners were more receptive to voices which were gender-appropriate, while accepting the female voices for the hypothetical female and child users. Age groups differed in preference for an acceptable computer voice: children preferred synthetic

speech, and adults preferred a more natural sounding voice.

One interesting observation of this study was the lack of a strong correlation between intelligibility and vocal preference. Past research findings suggest DECTalk voices (male, female, and child) were rated comparable in intelligibility to that of a natural voice (Mirenda & Beukelman, 1987). However, in this study, subjects consistently rated the DECTalk voices below neutral (i.e. not acceptable) and third place or lower when compared to the other voices.

In summary, the findings reported by Mirenda et al. (1989) suggest that a preference for voice type(s) appears to be strongly influenced by gender appropriateness, with subjects preferring women to sound like women and men like men. While children indicated a preference for computers to sound like computers (synthetic speech) adults preferred computers to sound more like people (natural voice).

Crabtree, Mirenda and Beukelman (1989) presented a follow-up study in the examination of age and gender preference for synthetic and natural speech. In this study the preferences of younger and older male and female listeners for natural and synthetic speech given six different contexts were rated along a 5-point Likert scale. The twelve voices represented included four natural voices (male and female adult and child voices), and eight synthetic (SmoothTalker 2.0 male and female; SmoothTalker 3.0;

Lightwriter; Artic R65B; RealVoice female; Votrax SC02; Echo II+). The 40 subjects representing ages 6-8 year olds, 10-12 year olds, 15-17 year olds, and 25-45 year olds rated the voices in the context of: a) an adult male user; b) adult female user; c) child male user; d) child female user; e) computer user; and f) self as user.

Results indicated that few of the synthesizers were rated as high or comparable to the natural voices across hypothetical contexts (with the exception of the computer as user context). Male and female subjects (across age groups) consistently agreed in ratings in four of the six contexts (adult male, adult female, child male, and computer questions).

The child female question indicated significant variation due to gender x age x voice interaction (i.e. variation by gender, age and voice). Gender appropriateness and naturalness were the determining factors for acceptability for older female listeners (i.e. child female, adult female). Naturalness was the determining factor for acceptability for the younger female listeners (i.e. accepting child female only). For the male listeners in both age groups naturalness (preference for child female, adult female, child male, and adult male) was the primary criterion for acceptability followed by gender appropriateness (RealVoice female).

A different pattern of preference was indicated for male versus female listeners in the self as user context.

Subjects preferred voices which were most natural and gender appropriate, (i.e. in the context of "self" as user). Thus male listener's preferred those voices which reflected a "more natural male" voice type (adult and child male voices) and females preferred those which reflected a "more natural female" voice type (child and adult female voices).

While this study is quite similar to that of Mirenda et al., (1988), different results were noted in several instances. Unlike the findings reported by Mirenda and associates (1989), where an age group difference was noted in the female child question, both an age and gender difference was observed by Crabtree et al., (1989). In addition, results were also found to vary for the computer question such that males and females young and old agreed that most voices were acceptable for a computer.

A study of listener acceptability judgements of human and synthesized speech was performed by Lloyd and Quist (1988). In this study a paired comparison design using all possible combinations of 9 speech samples was used (i.e. normal speech, monotonic speech, dysarthric speech, electrolaryngeal speech (Western Electric), artificial laryngeal speech (Tokyo Reed), esophageal speech, and synthesized speech (DECTalk, EchoII+, and Votrax type-n-talk).

Thirty-two adult female subjects listened to the 40 paired stimuli under one of four conditions. In the first task subjects were to make judgements of rank among the speech samples. In the second task they again listened to the speech samples (individually) and completed a questionnaire indicating comments listing likes and dislikes. Rank ordering of preference judgements indicated the following results: 1) Normal speaker, 2) Monotone speaker, 3) DECTalk speech synthesizer, 4) Tokyo Reed, 5) Western Electric, 6) Echo II+, 7) esophageal speaker, 8) Votrax Type-n-talk, and 9) dysarthric speaker. In summary, the human speech (with the exception of dysarthric speech) was preferred over synthesized. As cited in previous studies DECTalk was preferred above all other speech synthesis, followed by the EchoII+ and Votrax. The pneumatic device (Tokyo reed) was preferred over the alaryngeal and synthesized speech.

The collection of studies here cited suggest that judgements of preference are subject to the complex interaction of speech features and the multidimensionality of vocal quality (i.e. intelligibility, age- and gender-appropriateness, and naturalness). The absence of or incongruous presentation of speech features serve to decrease the "humanness" of vocal quality, subsequently rendering it "less acceptable". The purpose of this study is to determine, (1) whether children respond differentially (i.e.

as indicated by attitude and preference) to four vocal conditions (three synthetic speech and one natural speech), and (2) whether a relationship exists between the child listener's vocal quality preference and his/her attitude toward a peer child VOCA user (using the four previously sampled voices).

CHAPTER III

METHOD

Overview of Goal and Focus of the Study

The overall goal of this study is to determine whether children respond differentially (as measured by preference and attitude) to different vocal conditions varying in physical characteristics (i.e. synthetic and natural speech). The methodology of this study was designed to answer the following questions:

- 1) Does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used?
- 2) Do child listeners express a preference for different types of synthesized and natural speech?
- 3) Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

A two-task approach was used in order to answer these questions. Task I elicited children's attitudes toward a hypothetical child user of three types of commercially available voice-output communication aids and natural speech. Task II elicited a scaled rating of the child listeners' (cited in Task I) preferential judgement of the

same four voices (three commercial synthetic voices and naturally produced speech).

The speech synthesizers chosen for this study represent three of the more "high-quality" speech synthesizers currently available on the commercial market. Due to the higher quality of speech generated utilizing a principle of concatenated diphones, the DEctalk, RealVoice, and SmoothTalker 3.0 were the speech synthesizers of choice for this research.

Subjects

Criteria for Subject Selection

The subjects were seventy-eight fourth-grade children enrolled in a local public school system. There were a number of reasons for choosing this age range. Previous attitudinal studies have typically examined populations ranging from grades 5 through 7, (Armstrong, Rosenbaum and King, 1986, 1987; King, Rosenbaum, Armstrong, 1988; and Westervelt and Turnbull, 1980). Fourth-grade children were selected for the present study because they represent a group for which there is currently little information. In addition, it was felt that fourth graders possess the skills necessary to communicate their attitudes with regard to affect, cognition, and behavior.

As homogeneity of subject population must be controlled for in research design, all subjects were required to meet the following conditions: fourth grade level, Mason public

school system, and Mason residential region. Additional requirements included normal hearing and vision (with or without visual aids) and base reading at a 4th grade level, for purposes of homogeneity as well as for meeting the prerequisites established for test administration (i.e. ability to see and read the questionnaire and adequately hear the audio-recording). Vision and hearing screenings had been passed within the past two years based on biannual health records. Health (i.e. no debilitating chronic health condition) and academic ability (i.e. reading aptitude scores within grade level as indicated by annual academic reading aptitude tests) were screened based on each child's school record. Table 1 displays subject selection criteria.

Table 1. Pertinent subject characteristics.

1.	Fourth grade academic level
2.	Mason public school system
3.	Mason residential region
4.	Normal hearing and vision (with or without visual aids)
5.	Base reading at a 4th grade level
6.	No debilitating chronic health conditions

Procedure for Subject Selection

Four elementary schools within the Mason school system participated in the study. Fourth grade teachers were given

Talking Computer Project Parent Consent Forms (see Appendix A) to distribute among their 4th grade students. The dissemination of Parent Consent Forms was followed by the examiner addressing each of the classrooms to provide the students with information about the Talking Computer Project.

"Hi. I'm Sheila Bridges from Michigan State University. I'm doing a project about children who use talking computers. I would like to see what you kids and other kids your age know and think about children who use computers to talk. I think this is a pretty neat project, so I would like you to ask your parents to sign this form so that you can participate. Are there any of you who need a form for your parents to sign?"

Parent Consent forms were collected and records were reviewed for each student whose parent had given consent. The school record review included a review of the student's name, age, vision, hearing, health, and reading screening scores. Only those students who had met the aforementioned criteria were included in the subject population resulting in the selection of 80 children, 20 children per school.

Experimental Stimuli

Description of Stimuli

The four vocal conditions examined in this study included three synthetic voices and one natural voice; (1) natural child's voice; (2) the DECTalk, child voice setting ("Kit the Kid") manufactured by Digital Equipment Corporation; (3) the SmoothTalker 3.0 voice, female voice setting as implemented in the Touch Talker and Light Talker

(TM), manufactured by Prentke Romich Co.; and (4) the RealVoice female voice setting as implemented in the Epson SpeechPac, manufactured by Adaptive Communication Systems, Inc.

Preparation of Stimuli

All four voice samples were based on a language sample elicited during an interview of a 9-year-old female child. This procedure was used for the purpose of maintaining a naturalness and age-appropriateness typical of a 9-year-old. The language sample was transcribed verbatim (i.e. preserving the child's idiosyncratic choice of grammar and semantics) for this purpose. This transcription was edited (i.e. abbreviated) by the examiner in order to derive the actual wording to be used in the monologue presented to the subjects in Task I. The exact same words were used in all four voice samples in Task I (see Appendix B).

The editing eliminated all references to the physical condition or gender of the speaker. The speaker's name was given as "Lee". In this edited form, the "Lee monologue" was used verbatim in each of the four voice samples, retaining its orthography and punctuation. Orthographic text-to-speech entry was used in creating each of the synthesized speech samples. For the natural voice sample, the speaker who provided the initial language sample read the monologue verbatim while maintaining natural vocal inflection as used in the original language sample.

Recording Procedures

In Task I the Lee Monologue (as described above) was entered and stored in each of the three speech synthesizers, to be retrieved later for audio-recording purposes. For each speech synthesizer, the monologue was retrieved then recorded on a high fidelity TDK SA 90 audio cassette tape using a Nakamichi BX 300 Cassette Deck. Amplitude levels were normalized across voice samples. Levels were monitored during recording using a Realistic SA-102 Integrated Stereo Amplifier and the Nakamichi BX 300 dB peak level meter which provides for visual display of amplitude level. Amplitude was maintained at levels 7 through 10 as indicated by the Nakamichi BX 300 Peak Level Meter. Similar procedures were used for recording the natural speaking voice.

SmoothTalker recording. Lee's monologue was entered utilizing the text-to-speech capability of the Touch Talker. The Touch Talker has a 128-key flat membrane keyboard. The keys are assigned values by the Touch Talker operating system according to numerous levels, which are graphically represented by plastic overlay sheets provided by the manufacturer. On the so-called "custom overlay" level, the keyboard is configured according to the QWERTY arrangement. This overlay level was used in this study for orthographic input of the Lee monologue, as well as for access to the text once it had been stored. One limitation of the Touch Talker is the delay which occurs in accessing stored input due to its inability to read messages which extend beyond

its 40 character LCD display. This delay was reduced by a rapid succession of keystrokes utilizing a predetermined pattern of encoding (i.e. alphabetically based).

As previously stated, input followed rules of standard orthography (i.e. not adjusting for mispronunciation through phonetic entries or misspellings). In this way, intelligibility was determined by the device's ability to accomodate (i.e. based on the use of extensive phonological rules and/or an exception dictionary) those rules characteristic of English pronunciation.

Vocal parameters which were not restricted to default values were set by the examiner. Volume was set at level 8 (on a scale of 1 to 9) and pitch was set at level 7 (on a scale of 1 to 9). These settings were the ones recommended by the manufacturer's Michigan representative for optimal female voice. Gender was programmed at ^FB (i.e. female, base). Gender is programmable given the options of ^M (male), ^F (female), B (base) and T (treble).

Instrumentation used for recording purposes included the Nakamichi BX 300, the Realistic SA-102 amplifier, and the Touch Talker VOCA. The Touch Talker was accessed through the external speaker port and connected to the Realistic SA-102 in order to control for quality and amplitude level of the voice output (which tends to be distorted by the poor quality characteristic of the internal speakers found in many VOCAs). Amplitude was monitored by

maintaining recording levels 7 through 10 (Nakamichi BX 300 Peak Level Meter reading).

RealVoice recording. In preparing the RealVoice sample, Lee's monologue was entered orthographically via the Epson SpeechPac keyboard and spoken out using the device's text-to-speech function. As in the case of the Touch Talker, neither phonetic spelling nor misspelling were used to accomodate anticipated mispronunciation of orthographic entries. The monologue was stored using the Epson SpeechPac's memo capability which provides the storage of messages for delayed written or spoken recall. This permitted the storage of the messages for quick access and repeated recall. Vocal parameters accessible to the user are quite limited, restricting the user to the use of default values for female gender, pitch, and rate. However, volume was programmed at 3 (\V3) as suggested in the owner's guide for the use of amplification.

Recording instruments and procedures followed those methods previously described for the SmoothTalker.

DECTalk Recording. The portable DECTalk, "Kit the Kid" child voice, was interfaced with the Radio Shack TRS-80 Model 100 portable laptop computer, in order to utilize its text-to-speech capabilities for orthographic input. The monologue was entered orthographically without phonetic or misspelled entries (as previously described in the recording procedures for the SmoothTalker and RealVoice). The monologue was stored for immediate access by a single

keystroke, utilizing MinDEC, a customized software program by Dr. John Eulenberg (Artificial Language Laboratory, Michigan State University). Recording instruments and procedures followed those methods previously described for the SmoothTalker and RealVoice.

Natural Speech Recording. The natural human voice sample was based on the voice of a 9-year-old female chosen by the investigator. This 9-year-old speaker ("the speaker") was given a copy of her monologue (the "Lee monologue", described previously) to rehearse prior to the final audio-recording. Upon successfully achieving a quality reading characterized as "natural", the subject was recorded reading the monologue in a sound treated audiological testing booth. Recording instruments included the Nakamichi BX 300, the Realistic SA-102 amplifier, the JVC KD-15 Dolby System Stereo Cassette Deck and a Panasonic Cardioid Dynamic (WM-1151) microphone. The JVC KD-15 was used to monitor the recording level. Because the recording of human voice samples involved a microphone rather than a direct audio connection, as was the case with the voice synthesizers, the JVC KD-15 replaced the Nakamichi BX 300 in this instance.

The microphone was mounted 12 inches (30 cm) from the speaker, and recording levels were maintained at a peak level reading ranging from 7 to 10.

Recording of practice sample. A voice sample presented for practice purposes, consisted of a short monologue by an

adult female named "Mrs. Harris". The speaker (an adult female chosen by the investigator) was recorded reading a monologue in a sound treated audiological testing booth. Recording instruments included the Nakamichi BX 300, the Realistic SA-102, the JVC KD-15 Dolby System Stereo Cassette Deck and a Panasonic Cardioid Dynamic (WM-1151) microphone.

The microphone was mounted 12 inches (30 cm) from the subject and recording levels were maintained at a peak level reading ranging from 7 to 10. The text of the "Mrs. Harris" monologue is given in Appendix C and D.

Recording of Voice Preference Samples. For each of the four subject groups, a set of four voice samples were audio-recorded, one sample for each of the four voice sources under consideration. The fourth presentation was in each group a sample of the voice presented in Task I for that group. The order of the first three selections of each group was randomized among the remaining three voice sources. The order was based on random numbers generated by a TI-36 calculator.

Each sample in a given set was a dubbed audio recording of the first few lines of the full passage prepared in Task I. Each sample was dubbed from the original cassette recording using a Realistic MPS-5 Model 32-2031 Dubbing Cassette Recorder.

The text of each sample in Task II was a spoken (human voice or text-to-speech) interpretation of the following passage:

"Hi. My name is Lee and I'm 9 years old. I've lived in Lansing Michigan for two years. I used to live in Tyler Texas."

Determination of loudness levels for stimulus presentation. The Bruel and Kjoer sound level meter was used to establish uniform volume levels of the sound source (Sony and Marantz cassette players) at 70 dB. To determine the appropriate settings for each cassette player, it was positioned with its speaker elevated one meter from the floor. The sound level meter was held at arm's length 1 meter from the sound source. The measurements were taken at 0 degree azimuth.

Behavioral Measures

In conducting the two tasks of this study it was necessary to use behavioral measures of the two concepts, acceptance of the child VOCA user by his or her peers (i.e. "listener attitude") and listener preference of voice type (i.e. "voice quality preference").

The CATCH Attitudinal Measure: Peer Acceptance

In quantifying the concept of peer acceptance, this study drew upon an already existing measure of attitudes, the CATCH (Chedoke-McMaster Attitudes Toward Children with Handicaps) developed by Rosenbaum, Armstrong, and King (1986). The CATCH is a 36-item self-report attitude scale designed to measure children's attitudes toward handicapped peers. Developed as a measure of children's attitudes toward disabled peers, it also serves as a tool for the

evaluation of interventions designed to improve attitudes. The design is based on a 5-point Likert scale ranging from 0 - 4. Each of the 36 items reflects the alternate presentation of positive and negative statements ("I feel sorry for handicapped children." versus "I would stick up for a handicapped child who was being teased."). Response to each item designates the respondent's level of agreement ranging from "strongly disagree" to "strongly agree".

The measure, based on an attitude model proposed by Triandis (1971), identifies three dimensions of attitudes: (a) an affective component with statements of feelings; (b) a behavioral intent component, suggesting what the child might do, and (c) a cognitive component, expressed by statements of belief (Rosenbaum et al., 1986). These three subcategories of attitude, affective, behavioral, and cognitive, are each represented by 12 questions (e.g. a total of 36 questions).

Reliability analysis of the CATCH was conducted by Rosenbaum et al. (1986) examining the reliability of the total CATCH and each of its components. The test developers reported reliability for the total CATCH and its components by the calculation of a coefficient (Cronbach) alpha. The coefficient alpha for Factor 1 was .91; for Factor 2, .74; for Factor 3, .65; and .90 for the total CATCH indicating that the components and total CATCH were strongly associated.

The test-retest reliability as reported by the test developers, indicated a reliability coefficient of .70, .63, .44 and .73 for Factor 1, Factor 2, Factor 3, and total CATCH scores, respectively (Rosenbaum et al., 1986) indicating that the CATCH was a reliable and internally consistent measure.

Factor 1 and Factor 3 each contained a mixture of affective and behavioral intent items, accounting for 24.4% and 4.4% of the variance, respectively. Factor 3 consisted of cognitive items, accounting for 8.9% of the variance. Behavioral intent items were difficult to tease out as they were so closely intertwined with affective items. Thus the subcomponents affect and cognition may be effective measures of attitudes in addition to the total CATCH.

The 36 statements of the CATCH were based on common feelings and experiences of children 9 to 13 years of age. In the development of this measure a pool of statements was developed according to the affective, behavioral intent, and cognitive dimensions of Triandis' attitude model. The recommendations of school teachers and principals were solicited based on their evaluation of the appropriate grammar and reading level (for the target population) of each item. Statements were reworded or deleted as recommended by these evaluators.

The CATCH was selected for this study because it provides a validated quantitative measure of children's attitudes toward handicapped peers. It therefore matched

the current study in both the subject population and the population toward which attitudes were to be elicited. The current study differs in focus, however, from that of the study for which the CATCH was designed. In the original study, the attitude-eliciting characteristic was "handicapped", whereas in this research it was voice quality. This necessitated a modification of the stimulus questions. The term "handicapped child" in each of the 36 statements of the original version was in this study replaced by the gender-neutral name, "Lee" (see Appendix Measure of Listener Preference: Vocal Quality

As previously discussed in Chapter II, "voice quality" may have many interpretations, each based on the context in which voice is used. For this study, voice quality was viewed conceptually as a multidimensional subjective judgement of the features of intelligibility, naturalness, and age- and gender- appropriateness.

It is implicitly assumed that the children's preference ratings of synthetic and natural voice reflect differences in quality. In this study, the focus was on children, both as users and perceivers of voice-output systems of varying quality. Therefore, a measure of voice quality simple enough for a child to understand and express was selected. This measure reflected both the aesthetic features of voice quality (i.e. intelligibility and naturalness) and the appropriateness features (i.e. apparent age and gender).

The instrument used in this study for capturing these judgements contributing to listener preference was a 5-point Likert scale, presenting the subject with a choice of five statements ranging from a rating of 4, "I like it very much" to a rating of 0, "I don't like it at all". These statements do not contain an explicit reference to the hypothetical person using the voice under consideration. However, the spoken and written instructions to the subjects referred to the user of this voice. To aid the subject in making the choice, each statement was accompanied by a simple line drawing indicating a corresponding facial expression. This arrangement was used by Mirenda, Eicher, and Beukelman (1989) in their examination of listener preference for synthetic and natural speech. The rating form for Task II of this study is given in Appendix F.

Data-gathering Tasks

Setting and Assignment of Subjects to Task Activities

Eighty 4th graders meeting the previously described selection criteria were selected to participate in the study. Twenty children were selected from each school. These subjects were assigned numbers which were randomized by stratified randomization (i.e. according to school) and assigned to one of the following four vocal conditions (1) natural voice; (2) DECTalk; (3) SmoothTalker; and (4) RealVoice. In each school five children were assigned to each of the four vocal conditions in order to establish balanced groups. While eighty children were selected to

participate in the study, only a total of 78 children were in attendance the days the study took place. The anticipated four balanced treatment groups of twenty students were therefore reduced to three groups of twenty and one group of eighteen.

Three trained graduate students in speech-language pathology were instructed individually regarding the procedures for administering the two tasks. Two of the four groups were observed simultaneously, one group by the examiner and the other by one of the three trained graduate students. By completing two groups simultaneously it was possible to complete each testing site within a two day period of time, thus controlling for the element of time.

The study took place during the noon hour in classrooms in a generally quiet location. Each speech recording was presented free field on a Marantz Superscope 104 Professional Cassette Recorder or a Sony TC-205 Cassette-Corder. The cassette recorder was centrally located in each group presentation.

Practice Sample

Prior to administering the Task I stimulus to each group, the examiner presented a practice stimulus in order to familiarize the subjects with the procedures and the terminology, ("agree" versus "disagree") for completing the CATCH questionnaire. "Place an "X" across the statement that best describes how you feel". The practice stimulus was a brief audio-recorded monologue of the voice of an adult

woman referred to as "Mrs. Harris" (see Appendices C and D). The students were instructed to listen to the recording and then mark their responses to three questions about Mrs. Harris. The subjects discussed each item as a group under the direction of the examiner or a trained graduate student assistant. This served to insure that the procedures were clearly understood by each child. Appendices C and D display the "Mrs. Harris" monologue and the response sheet, respectively.

Task I: Presentation of the CATCH

Task I response elicitation began immediately following the presentation of this practice set. Each subject listened to one of the following four vocal conditions, (1) natural child voice, (2) female SmoothTalker voice, (3) female RealVoice, and (4) DECTalk child voice ("Kit the kid"). In each group the corresponding audio-recorded voice samples of "Lee's Monologue", were presented as described earlier (see Appendix B). The speech samples ranged in duration from about one minute (natural voice) to about three minutes (SmoothTalker voice). Following the presentation of the randomly assigned voice, each child then completed the 36-item CATCH (see Appendix E). Each child was presented written and verbal instructions regarding the purpose of the CATCH questionnaire and procedures for its completion (see Appendices G and H).

Task II: Presentation of Stimulus Voice Quality Preference
Rating

The second task required the same subjects to express a personal preference for each of the previously described voices (i.e. 3 speech synthesizers and 1 natural speech). Each of the four vocal conditions were rated along a five-point Likert Scale utilizing descriptors ranging from "I like it very much" to "I don't like this at all", in expressing preference.

Procedures for administering preference ratings. The subject group was presented Task II immediately following its completion of Task I. The instructions for this task, were given orally by the examiner and also presented at the top of the rating form, as follows:

"You have just listened to one of Lee's voices. However, Lee has 4 different voices. I would like to know how much you like each of Lee's 4 voices. You will hear all 4 voices twice. The first time all 4 voices will be presented at once. You will just listen. The second time you will listen to each voice. After each voice place an "X" across the face under the statement that best describes how much you like the voice you have just heard".

In this task each group of children was presented an audio recording of a set of four voice samples. For each group, the recording of the entire set was played twice through. In the first playing of the set (1:53 minutes in length), subjects were instructed to listen only. During the second playing the playback was paused after each voice sample to allow the subjects time to rate it on a rating form (see Appendix F) on a scale of 0 to 4 ("I like it very much"

to "I don't like it at all"). Rating consisted of placing an "X" over one of five drawings paired with a corresponding statement of degree of liking.

Data Analysis Procedures

Scoring the CATCH

The CATCH contains 36 items, 12 items in each of the following components: affective (ASUM), behavioral (BSUM), and cognitive (CSUM). The components are arranged in random order with an equal number of positively and negatively worded statements presented in alternating order. Each of the 36 items are scored on a 5-point Likert scale with values ranging from 0 (strongly disagree) to 4 (strongly agree), with negative statements being inversely scored (e.g. 4 - 0).

Each child's responses were entered item by item as recorded on his/her CATCH response form (refer to Appendix E). The raw data were stored to disk in this fashion and calculated using PC-CALC, (CALC, 1985) a software program with the capability of manipulating numbers and performing a variety of mathematical computations. Data were manipulated in this way to derive subscores for affective (ASUM), behavioral (BSUM), and cognitive (CSUM) responses per child. Total scores (TSUM) were derived by the summation of the subscores of the three components (i.e. $ASUM + BSUM + CSUM = TSUM$).

The total CATCH score (TSUM) per child had a possible range of 0 to 144. A score of 72 was a neutral response

(predominantly, "I can't decide" responses). For the purpose of this study, scores below 72 were interpreted as negative attitudes and those above were positive.

Scoring Voice Quality Preference Rating

Judgements of voice quality preference was rated along a 5-point Likert scale, presenting a choice of five statements ranging from a score of 4, "I like it very much" to a score of 0, "I don't like it at all". A score of 2 indicated "I don't care either way" (refer to Appendix F). Each statement corresponded to one of the four vocal conditions (i.e. DECTalk, SmoothTalker, RealVoice and natural voice).

Raw data were entered and calculated using PC-Calc in much the same manner as described above. Individual and group scores were obtained for each of the four vocal conditions, thus facilitating transfer for further simple statistical and global analysis using SPSS/PC+ V2.0.

Analysis of the Data

Simple statistical analysis (mean, range and standard deviation) as well as a three-way analysis of variance (ANOVA) were performed using SPSS/PC+ V2.0. An ANOVA was computed to determine the relationship between the independent variables school, voice type and gender and the dependent variable, the CATCH attitude measure, in order to address the question "Does the attitude of the child listener vary toward the "child VOCA user" as a consequence of the natural and varying types of synthetic speech used?"

To identify the source of significant differences between the schools a Tukey-b post hoc follow-up test for one-way analysis was conducted.

The question "Do children's responses to synthesized and natural speech vary in terms of listener preference" was addressed by computing a three-way ANOVA to determine the relationship between the independent variables school, voice type, and gender and the dependent variable, a scale of voice preferences. To identify the source of significant differences among the voice types a Tukey-b post hoc follow-up test for one-way analysis was conducted.

The Kendall Rank Correlation Coefficient (a nonparametric measure examining relationships for ordinal-level variables) was computer to address the question of whether a relationship exists between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user.

The coefficient alpha was computed to determine the reliability of the CATCH scores. To establish that the three dimensional structure of the CATCH measure had been maintained following its modification, factor analysis was computed (i.e. for comparative purposes) to identify the three factors representing the components of the CATCH and to determine the degree to which they accounted for the percentage of variance (i.e. item loading to determine percentage of variance attributed to any of the three components or a combination thereof).

CHAPTER IV

RESULTS

The overall goal of this study is to determine whether the vocal quality of a voice-output communication aid is related to child peer social acceptance. This research addresses the following questions:

- 1) Does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used?
- 2) Do child listeners express a preference for different types of synthetic and natural speech?
- 3) Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

To investigate these questions two tasks were performed. In the first task, children's attitudes toward the hypothetical user of four types of speech were elicited and examined. In the second task the subjects rated the quality of the same four voices.

Attitude Measure

Modification of the CATCH

The CATCH contains 36 items, 12 items in each of the following components: affective (ASUM), behavioral (BSUM),

and cognitive (CSUM). The components are arranged in random order with an equal number of positively and negatively worded statements presented in alternating order. Each of the 36 items are scored on a 5-point Likert scale with values ranging from 0 (strongly disagree) to 4 (strongly agree), with negative statements being inversely scored. Total scores (TSUM) were derived from the summation of the subscores of the three components (i.e. ASUM + BSUM + CSUM = TSUM).

The CATCH was selected for this study because it provides a validated quantitative measure of children's attitudes toward handicapped peers, a population consistent with the study at hand. However, the current study does differ in focus ("attitude toward child peer VOCA user" rather than "attitude toward a handicapped child") from that of the study for which the CATCH was designed, thus necessitating the modification of the stimulus questions (i.e. substituting the term "Lee" for "handicapped"). Due to this modification the reliability of the modified CATCH was examined.

Reliability of the Modified CATCH

Factor analysis: Modified CATCH. For the purpose of this study, the Cronbach alpha (Norusis, 1988) was computed using SPSS/PC+ V2.0 to determine the reliability of the modified CATCH. The modified CATCH was found to be a reliable measure of attitude with an alpha of .92.

Reliability analysis of the CATCH was further determined by factor analysis, (i.e. principle component analysis) computing the Kaiser Normalization and varimax rotation (Norusis, 1988). Factor loading served to determine whether an underlying pattern of relationships existed, consistent with the theoretical construct of the CATCH as reported by Rosenbaum et al., 1986. Factor 1 consisted of a mixture of affective, behavioral and cognitive intent items, while Factor 2 contained mostly cognitive intent items. Twenty-six questions were loaded on Factor 1, while six were loaded on Factor 2 and four were loaded on Factor 3. Typical items loaded on Factor 1 included "I would like having Lee live next door to me" and "I would be happy to have Lee for a special friend" (see Appendix I). These findings were consistent with those reported by Rosenbaum et al. (1986) identifying affect and cognition as the two major components of attitude.

For the purpose of the current study the Pearson correlation coefficient was computed to examine the correlation of each of the component variables to the (total) measure of attitude. The two factors found to be most strongly correlated with attitude (CATCH) were affect (ASUM) and behavior (BSUM) ($r = .94$, and $.92$, $p < .001$, respectively). Correlation for cognition (CSUM) was also found to be significant ($r = .73$, $p < .001$).

Does the Attitude of the Child Listener Toward
the Child VOCA User Vary as a Consequence of
Different Types of Synthetic and Natural Speech Used?
A Measure of Attitude: Statistical Analysis of the CATCH

A split-plot stratified randomized (mixed) block design was used for the examination of the independent factors, school, gender, and voice type. School and gender were introduced as independent factors as past studies have suggested that these variables have differentially influenced overall CATCH attitude scores (Rosenbaum et al., 1986; King, Rosenbaum, Armstrong, & Milner, 1988). Thus this design allows for examination of these variables as possible main effects contributing to the variation of the measure in question, attitude.

Simple statistical analysis of these variables is shown in Table 2. The mean, sum score, and standard deviation for total CATCH score (TSUM) are illustrated for each of the independent factors. CATCH scores have a possible range of 0 to 144. A score of 72 is a neutral response (predominantly, "I can't decide" responses). In the case of this study, scores below 72 were interpreted as negative attitudes and those above were positive.

Examination of the data reveals that for overall mean attitudinal scores, SmoothTalker received the highest score of 101.55 while both RealVoice and natural voice received the lowest (96.60 and 96.50 respectively). A comparison of mean attitudinal scores for the RealVoice and natural voice

suggest negligible differences. It should be noted that lower sum attitudinal scores for DECTalk may be misleading as a smaller group of subjects participated in the DECTalk voice condition (i.e. 18 rather than 20 subjects).

The mean scores were all positive with a limited range of 12.80 ranging from 90.50 (male rating of the RealVoice) to 103.30 (female rating of the DECTalk). Attitudes toward all the speech synthesizer were positive and not significantly different from those observed in response to natural speech.

A comparison of the mean attitude (CATCH) scores indicate variation between schools. School A assigned the highest CATCH scores to the RealVoice (103.60) and the DECTalk voice (102.33). School B assigned the highest scores to the SmoothTalker (116.80) and the natural voice (104.60). School C assigned the highest scores to natural voice (106.80) and DECTalk voice (105.40), while School D assigned highest scores to the SmoothTalker (100.20) and the RealVoice (97.00). No overall trend is evident from these findings. Differences in scores were found to be quite small, suggesting that such differences may be insignificant and no true pattern for total attitude score (TSUM) was evident for the independent variable.

A comparison of male and female mean attitude scores suggests female attitudes were most positive toward the DECTalk (mean = 103.30) and the SmoothTalker (mean 102.18) while CATCH scores for males were most positive toward the

Table 2. Statistical review (mean, sum, and rank) of attitude(CATCH) scores according to voice type, gender, and school.

ATTITUDE SCORES				
	DECTalk	Smoohtalker	RealVoice	Natural Voice
Overall				
Mean	98.33	101.55	96.50	96.60
Sum	1770.00	2031.00	1930.00	1932.00
SD	17.95	15.02	21.23	17.16
Female				
Mean	103.30	102.18	100.50	96.18
Sum	1033.00	1124.00	1206.00	1058.00
SD	21.30	13.30	24.74	14.37
Male				
Mean	92.13	100.78	90.50	97.11
Sum	737.00	907.00	724.00	874.00
SD	10.95	17.70	13.92	20.99
School A				
Mean	102.33	88.80	103.60	87.00
Sum	307.00	444.00	518.00	435.00
SD	15.28	5.20	15.65	11.81
School B				
Mean	98.40	116.80	98.00	104.60
Sum	492.00	584.00	490.00	523.00
SD	8.26	9.42	16.76	17.13
School C				
Mean	105.40	100.40	87.40	106.80
Sum	527.00	502.00	437.00	534.00
SD	20.90	9.76	28.66	21.09
School D				
Mean	88.80	100.20	97.00	88.00
Sum	444.00	501.00	485.00	440.00
SD	22.22	19.04	25.00	10.17

SmoothTalker (mean = 100.78) and the natural voice (mean = 97.11). Females assigned lowest scores to the natural voice, while males assigned the lowest score to the RealVoice. It is interesting to note that while the natural voice received the second highest scores from the male listeners, and the lowest score for the female listeners the attitude scores (mean = 97.11, male; mean = 96.18, female) are quite similar. Rosenbaum et al. (1986) report gender differences in comparing male and female attitudes toward handicapped children, where the female scores being the higher scores were reported to "shift slightly to the right of the male". In view of this slight deviation reported by past studies and evident in the current study, further investigation of gender and school was performed to determine their significance.

Analysis of Variance

A 4x4x2 analysis of variance was conducted to answer the question, Does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used? The independent variables were sex, school, and voice type. These variables were examined to determine their significance and possible main and interaction effects. One three-way analysis of variance (ANOVA), school (4) x voice type (4) x sex (2) was performed. The dependent variable was the CATCH measure of attitude including the scores for the the total CATCH

(TSUM), the affective component (ASUM), behavioral component (BSUM), and cognitive component (CSUM).

Tables 3a - 3d illustrate the ANOVA source table for each of the four attitude measures: total attitude, affective, behavioral, and cognitive scores respectively. Table 3a shows that neither the independent variables voice type, gender, and school nor their two-way or three-way interactions were significant main effects contributing to total attitude (TSUM).

Table 3b shows that neither the independent variables voice type, gender, and school nor their two-way or three-way interactions were significant main effects contributing to affective attitude (ASUM). Similar findings were found upon examination of the behavioral attitude measure (BSUM) shown in Table 3c. Neither the independent variables voice type, and school nor their two-way or three-way interactions were significant main effects.

Table 3d displays the source table for cognitive attitude (CSUM). School was the only independent variable found to be a significant main effect, while neither voice type or gender nor their two-way and three-way interactions (i.e. voice type x gender x school) were significant main effects.

In summary, school was the only independent variable found to be a significant main effect, evident in the cognitive (CSUM) component of the CATCH. This suggests that in the case of the four schools attended by the subjects,

Table 3. Analysis of variance for Total CATCH and subtests

3a. Source table for total attitude measure.

TOTAL CATCH ATTITUDE SOURCE TABLE					
	df	SS	MS	F	p
Main Effect	7	2346.449	335.207	1.071	.397
Sex	1	499.571	499.571	1.597	.213
School	3	1490.308	496.769	1.588	.205
Voice type	3	352.538	117.513	.376	.771
Two-way Interactions	15	4929.761	328.651	1.050	.425
Sex x School	3	491.156	163.719	.523	.668
Sex x Voice type	3	837.477	279.159	.892	.452
School x Voice type	9	3754.305	417.145	1.333	.247
Three-way Interactions					
Sex x School x Voice type	9	2590.079	287.787	.920	.517

3b. Source table for affective attitude.

AFFECTIVE ATTITUDE SOURCE TABLE					
	df	SS	MS	F	p
Main Effect	7	233.552	3.365	.578	.770
Sex	1	55.201	55.201	.956	.333
School	3	136.838	45.613	.790	.506
Voice type	3	39.705	13.235	.229	.876
Two-way Interactions	15	673.853	44.924	.778	.694
Sex x School	3	38.614	12.871	.223	.880
Sex x Voice type	3	104.151	34.717	.601	.617
School x Voice type	9	503.656	55.962	.969	.477
Three-way Interactions	9	542.011	60.223	1.043	.422
Sex x School x					

* p < .05 Significant F

Table 3. (cont)

3c. Source table for behavioral attitude.

BEHAVIORAL ATTITUDE SOURCE TABLE					
	df	SS	MS	F	p
Main Effect	7	321.567	45.938	.877	.532
Sex	1	100.410	100.410	1.917	.173
School	3	128.515	42.838	.818	.491
Voice type	3	89.132	29.711	.567	.639
Two-way Interactions	15	1072.779	71.519	1.366	.205
Sex x School	3	111.259	37.086	.708	.552
Sex x Voice type	3	129.576	43.192	.825	.487
School x Voice type	3	855.065	95.007	1.814	.091
Three-way Interactions					
Sex x School x Voice type	9	661.487	73.499	1.403	.215

3d. Source table for cognitive attitude.

COGNITIVE ATTITUDE SOURCE TABLE					
	df	SS	MS	F	p
Main Effect	7	338.021	48.289	1.674	.139
Sex	1	24.019	24.019	.833	.366
School	3	269.984	89.995	3.120	.035*
Voice type	3	47.172	15.724	.545	.654
Two-way Interactions	15	414.167	27.611	.957	.512
Sex x School	3	59.648	19.883	.689	.563
Sex x Voice type	3	180.474	60.158	2.085	.115
School x Voice type	9	194.610	21.623	.750	.662
Three-way Interactions					
Sex x School x Voice type	9	89.850	9.983	.346	.954

* p < .05 Significant F

children in attendance of one school responded differently from children attending another (regardless of their gender or the type of voice heard) with regard to questions of belief, such as "Lee wants lots of attention from adults" and "Lee feels sorry for herself/himself". In all other instances (ASUM, BSUM, and TSUM) neither the variables school, sex, and voice type, nor their two-way and three-way interactions were found to be significant.

Thus, in answer to the question, does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used, it would appear that voice type (natural and synthetic) does not contribute to this variation. Voice type was not a significant main effect in determining attitude. This may be simply stated as "I like you regardless of what voice, synthetic or natural, you choose to use." Neither voice type, the treatment in question nor the independent variables, sex or school are significant factors with regard to overall attitude.

Post Hoc: Examination of Trends

The schools attended by the children were found to be a significant factor in determining the listener's beliefs (cognitive attitude) regarding "Lee". Yet, all factors (i.e. voice, sex, as well as school) were found not to be significant in determining overall listener attitude. The Tukey-b post hoc follow-up test for one-way analysis was conducted to identify where the significant differences (.05

significance) occurred among the schools with a Tukey critical q - value of 3.58. The significant difference occurred between School B with a mean CSUM (cognitive component) score of 33.45 and School D with a mean CSUM score of 28.55.

Do Child Listeners Express a Preference for
Different Types of Synthetic and Natural Speech?
Vocal Preference Measure

The second task required the subjects' preferential rating of each of the four voice types (suggestive of perceived voice quality). The instrument used for listeners' preference was based on a 5-point Likert scale presenting subjects with a choice of statements ranging from "I like it very much" (4.0) to "I don't like it at all" (0.0). A score of 2.0 indicates a neutral response "I don't care either way".

A split-plot stratified randomized (mixed) block design was used for the examination of the independent factors, school, gender, and voice type. School and gender were introduced as independent factors, thus this design allows for examination of these variable as possible main effects contributing to the variation in preferential ratings of voice types.

Simple statistical analysis of these variables is shown in Table 4. The mean, sum score, and standard deviation are displayed for each of the independent factors under question, school, gender, and the four voices, DECTalk,

SmoothTalker, RealVoice, and natural voice. Examination of the data reveals that for overall mean preference ratings the natural voice received the highest rating with a mean of 3.69 followed by DECTalk with a mean of 2.31. The remaining two speech synthesizers both received ratings below 2.0. The SmoothTalker and the RealVoice received the lowest ratings of 1.9 and 1.8 respectively. Differences between the mean and sum SmoothTalker and RealVoice preference ratings appear to be insignificant. Again, it should be noted that lower sum preference ratings for DECTalk may be misleading as a smaller group of subjects participated in the DECTalk voice condition (i.e. 18 rather than 20 subjects).

A comparison of the mean preference ratings indicates some variation between the schools. School C assigned the highest preference rating to the natural voice (3.80) and the DECTalk voice (2.80). School D assigned the highest ratings to the natural voice (3.40) and the SmoothTalker (2.10). In all instances (across gender and school) natural voice received the highest rating. the natural voice Differences in scores were found to be quite small, suggesting that such differences may be insignificant and no true pattern for preference rating is evident for the independent variable school.

In comparing male and female preference ratings for SmoothTalker and RealVoice, some divergence was noted. While both groups rated RealVoice with an approximate mean of 1.8, it received the second lowest rating by females and the

Table 4. Statistical review (mean, sum, and rank) of quality rating according to voice type, gender, and school.

QUALITY RATINGS				
	DECTalk	SmoothTalker	RealVoice	Natural Voice
Overall				
Mean	2.31	1.90	1.80	3.69
Sum	180.00	149.00	142.00	288.00
SD	1.11	1.43	1.22	.76
Female				
Mean	2.38	1.73	1.84	3.70
Sum	105.00	76.00	81.00	163.00
SD	1.06	1.40	1.38	.76
Male				
Mean	2.20	2.10	1.79	3.68
Sum	75.00	73.00	61.00	125.00
SD	1.17	1.46	1.01	.77
School A				
Mean	2.28	1.94	1.78	3.67
Sum	41.00	35.00	32.00	66.00
SD	1.02	1.43	1.31	.49
School B				
Mean	2.15	1.45	1.65	3.90
Sum	43.00	29.00	33.00	78.00
SD	.99	1.19	1.11	.31
School C				
Mean	2.80	2.15	2.15	3.80
Sum	56.00	43.00	42.00	76.00
SD	1.06	1.66	1.35	.52
School D				
Mean	2.00	2.10	1.75	3.40
Sum	40.00	42.00	35.00	68.00
SD	1.26	1.41	1.15	1.27

lowest score by males. Thus, males responded positively to three of the four voices, rating them neutral (2.0) and above. Male listeners expressed equal liking for the DECTalk and SmoothTalker as indicated by preference ratings of 2.2 and 2.1, respectively. Results thus suggest males express a somewhat greater preference for synthesized speech than females. The significance of observed gender and school differences and their contribution to quality were further investigated.

Analysis of Variance

An analysis of variance was conducted to examine the significance of voice type with regard to preference rating and to address the question "Do child listeners express a preference for different types of synthetic and natural speech?" Again, independent variables were sex, school, and voice type. These variables were examined to determine possible main and interaction effects. A three-way analysis of variance (ANOVA) of school (4) x voice type (4) x sex (2) was performed. The dependent variable was the 5-point Likert Scale (0 - 4 for the purpose of this study) developed by Mirenda et al. (1989) as a measure of voice preference based on a subjective rating of quality.

Table 5 displays the ANOVA source table for voice preference examining the independent variables, school, sex, and voice. Voice type was found to be a significant main effect. However, neither sex, school nor their two-way and three-way interactions were significant. These results

suggest that the children's responses to synthesized and natural speech do vary in terms of listener preference. Children do express varying degrees of preference in response to synthetic and natural speech.

Table 5. Analysis of variance for vocal preference.

Vocal Preference Source Table					
	df	SS	MS	F	p
Main Effects	7	185.743	26.535	19.625	.000*
Sex	1	.303	.303	.224	.636
School	3	10.117	3.372	42.494	.060
Voice	3	175.485	58.495	43.262	.000*
Two-way Interactions	15	17.469	1.165	.861	.608
Sex x School	3	4.588	1.529	1.131	.337
Sex x Voice	3	3.562	1.187	.878	.453
School x Voice	9	8.945	.994	.735	.677
Three-way Interaction	9	6.783	.754	.557	.831
Sex x School x Voice	9	6.783	.754	.557	.831

* $p < .05$ Significant F

Post Hoc Test

The Tukey-b post hoc follow-up test for one-way analysis was conducted to identify where the significant differences (.05 significance) occurred among the voice types with a Tukey critical q - value of .819. The natural voice, which received the highest preference rating, with a mean of 3.69, was found to be most significantly different

when compared to all other voices. DECTalk voice which was rated closest in preference relative to the natural voice with a mean of 2.31, was also found to be significantly different when compared to SmoothTalker. The natural voice and DECTalk were the only two voices which received positive ratings above 2.0, while mean scores for SmoothTalker and RealVoice were below 2.0 with mean scores of 1.9 and 1.8 respectively.

Is There a Relationship Between the Child Listener's
Preference for Natural and Synthetic Speech and
His/Her Attitude Toward the Child VOCA User?
Correlation of Attitude and Quality

Upon completion of the two tasks examining attitude toward and preference for four different voice types, it is necessary to again turn our attention to the initial investigative question; Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

To investigate the correlation between voice quality and attitude, the Kendall Rank Correlation Coefficient was computed. Results indicated a correlation between attitude and quality ($r = .359$). A one-tail z test of significance indicated a z score of 4.648 ($p < .001$), suggesting that the correlation between the two variables was significantly different from zero.

Thus in answer to the question "Is there a relationship between the child listener's preference for natural and

synthetic speech and his/her attitude toward the child VOCA user? results indicate that yes, there is a relationship. Based on these results, it can thus be concluded, children's attitudes toward a (hypothetical child) user of a voice-output communication aid are related to the listener's preference for the voice type.

CHAPTER V

DISCUSSION

The development of computer technology and voice synthesis has provided many nonspeaking individuals the opportunity to utilize spoken communication. For the nonspeaking individual it has brought hope of overcoming the restrictions of social isolation and of playing a more active/initiator role in social interactions. Despite the benefits, these devices pose a number of problems and unanswered questions regarding the limitations such technologies may impose on the lives of nonspeaking children. The use of voice-output communication aides does not parallel spoken communication. It often lacks appropriate intelligibility, intonation, emotional expression as well as the gender, age, social, regional and personality characteristics of the individual child user. Consequently, vocal quality of a voice-output communication aid may affect child peer social acceptance.

This study has addressed the questions:

- 1) Does the attitude of the child listener toward the child VOCA user vary as a consequence of different types of synthetic and natural speech used?
- 2) Do child listeners express a preference for different types of synthetic and natural speech?

3) Is there a relationship between the child listener's preference for natural and synthetic speech and his/her attitude toward the child VOCA user?

This study is unique in that it has placed under direct examination the attitudes of an often overlooked segment of our augmentative communication population, the non-communicatively impaired child interactant. In addition it has examined three of the more popular commercially available speech synthesizers, the DECTalk, the SmoothTalker 3.0, and the RealVoice, as well as a natural voice.

The major findings of this study suggest that child listeners of synthetic (DECTalk, SmoothTalker 3.0 and RealVoice) and natural speech are aware of differences in voice type and express a preference for voices they do versus do not like. Further more, findings suggest that a child listener who likes a given voice very much is more likely to express a positive attitude toward the child VOCA user. It is important to examine the implications of these findings.

Many attitudinal studies have reported listener's attitudes and personality characteristics ascribed to speakers based on the listener's perception of vocal quality. The results of this research suggest that such findings must be carefully examined prior to applying these generalizations to children. Children's attitudes towards "child VOCA users" of varying types of speech (synthesized

and natural) does correlate significantly with listener vocal quality preference. Variance in attitude is not attributable to gender, school, or voice type. Nor has a cause effect relationship been established between attitude and voice preference.

Correlation Between Attitude and Voice Preference

Acceptance and Selection of a VOCA

The results of this research confirm that children's attitudes toward child VOCA users do correlate significantly with voice preferences. The significant implications of these findings can not be underestimated as acceptance (i.e. positive attitude) is essential to the eventual success and functional use of a communication aid. Vanderheiden and Lloyd, (1986) substantiate the significance of this evidence as they point out the importance of VOCA acceptance by the user and the interactant.

"Aids must be both acceptable to and motivating for the individual. Lack of use or the unwillingness of the individual to utilize the components in certain environments, may otherwise result. Similarly, they must be acceptable to the family, to peers and friends and to those in the individual's educational or work environments, or the offending components may be ignored or not made available to the individual."

p. 58

In the present study attitudes are in general positive. Overall voice preferences are also positive, ranging from 3.69 to 1.8 (i.e. no ratings of 1.0, "I don't like it," or less). Consequently, both being positive, attitude and voice quality preference are significantly correlated. The

large sample population size further contributed to the significance of this correlation. Although correlation is significant it is also low, accounting for only 13% of the variance. Alternatively, 87% of the variance is unaccountable for. Ideally a strong correlation should account for 50% or more of the variance, thus suggesting some degree of predictability.

The voices sampled in this study represent the state-of-the-art technology to date. These voices not only represent improved intelligibility, gender- and in the case of the DEctalk, age-appropriate voices but also the more expensive devices on the commercial market. As a result much of the general nonspeaking community find these devices to be unaffordable with prices beginning at \$2,000. While selection of such devices may be desirable based on voice quality (listener and user acceptability), a number of design factors, may serve to make it less than an optimal selection. Design factors include: lack of user accessibility for turning devices on and off and programming new vocabulary, lack of lightweight portability for the ambulatory user, lack of flexibility and user accessibility for shifting modes of input and output and computer interfacing, inability to be integrated with other communication and environmental control systems, slow rate of speech output, lack of durability for use in a variety of settings, and other design limitations (Blackstone, 1990).

Research is currently underway to further identify and overcome some of these design limitations.

Extention of the Study

This study has sought to reveal child listener's attitudes toward the "child VOCA user" as a consequence of varying types of speech (natural and synthetic). The findings suggest that child listener's overall attitudes are positive and are significantly correlated with voice quality preference. However, positive attitude was not attributed to voice type.

Based on previous studies DECTalk, SmoothTalker 3.0, and RealVoice have been reported to be more gender appropriate (portraying a female voice) and more intelligible when compared to other speech synthesizers. In addition, these synthesizers all utilize a similar principle of synthesis. The current market represents a vast array of VOCAs varying in technological design, quality (rate, intonation, pronunciation, gender- and age-appropriateness) and intelligibility. It is quite possible that children may express a broader range of attitudes (i.e. negative bias) when exposed to a broader, more diverse range of VOCAs than presented in this study. The absence of such variety may be a limitation of this study. This factor as well as other study limitations will be discussed.

Selection of Vocal Conditions

The devices selected for the purpose of this study utilize "state-of-the-art" speech synthesis based on a

principle of concatenated diphones. Children's overall attitudes were found to be overwhelmingly positive across all four vocal conditions. However, given a wider selection of devices or a broader range in voice quality (i.e. greater variation in design, intelligibility, and age- and gender-appropriateness) these measures may have elicited more variable responses and negative attitudinal bias. These findings may simply reflect attitudes and preferences for devices utilizing similar technology.

Attitude Formation in Young Children

Adult attitude studies have reported well established and often pejorative attitudes relative to voice differences and speech disorders. In view of the young age of the subjects in this study it is quite possible that negative attitudes, at this point, may not have formulated with regard to vocal differences and voice quality preference. Young children may lack exposure to experiences and attitudinal bias resulting in stereotypic and prejudicial responses. Examination of an older age group where attitudes may be more firmly established may thus elicit more pejorative responses.

Attitudinal and Voice Preference Measures

In order to reduce the level of task difficulty for the child participants it was necessary to use behavioral measures of the two concepts, attitude and voice preference, simple enough for a child to understand and express. In the selection of these measures, the attitudinal (CATCH) and

voice preference measures may have lacked sufficient sensitivity to "tease out" attitudinal and preference differences. In addition, the voice scale may have been too limited in range (i.e. ranging from 0 - 4) to be sensitive to the degree of perceived differences in voice quality.

Further investigation examining a wider variety of VOCAs or speech synthesizers more divergent in quality, intelligibility, and age- and gender- appropriateness, an older subject population or broader more sensitive measurements of attitudinal and voice preference ratings are warranted.

An Examination of Contrived Conditions

The context in which these voices were presented were contrived in order to establish controlled conditions for specifically examining voice quality in isolation. Attitudes were based on a self-reported scale rather than observable behavior, and were elicited based on a hypothetical user as opposed to an actual interactant. Factors such as vocal quality appropriate to age, gender, personality, social, regional, ethnic or family dialect could not be judged in the absence of a "real live Lee". In addition, voices were rated in the context of having heard four other voices presented in a monologue, within a quite limited time span. How would such findings generalize to actual speaking-nonspeaking child interaction? How might attitudes vary in the context of such an interaction given prolonged exposure and opportunity for social interaction?

si

po

10

a

F

(

9

h

9

F

9

h

1

t

e

S

1

1

1

1

1

1

1

1

These questions are not presented to dismiss the significance of the research findings cited here, but to point out that this is but the first of what will hopefully lead to many other studies in the examination of attitudes and the impact of speech synthesis on child interaction.

Implications for Future Research

Further Examination of Variables Contributing to Attitude

Gender and attitudinal differences. Past studies (Rosenbaum, et al., 1986; King et al., 1988) have reported a gender difference in the expression of attitudes toward handicapped child peers. King et al., (1988) report that girls scored significantly higher on the CATCH than boys. Rosenbaum et al. (1986), report preliminary data suggesting girls interact positively significantly more often than boys, who engage in more nonverbal independent play. In a forced contact situation girls provide physical assistance to their gender-matched disabled peers significantly more often than boys.

The current study shows that while gender was not a main effect for attitude and attitudes expressed by both males and females were generally positive, female attitudes were somewhat more positive than the attitudes expressed by males. The most dramatic gender difference was evident in the rating of DECTalk. DECTalk was rated highest by females but received the next to the lowest rating by the males. The natural voice received the next to the highest rating by males but the lowest rating by females. While such gender

diff
slig
atti
reve

Rich
stin
disa
info
othe
asse
ima
comp
inf
of
"fu
han

fut
inf
Sch

the
der
par
pos
cor

differences may not be significant, there appears to be some slight yet perceivable difference in male versus female attitudes which this design is not sensitive enough to reveal.

Rosenbaum et al. (1988) cite findings reported by Richardson (1970) indicating that girls tend to prefer stimulus children with "functional" rather than "cosmetic" disabilities, while the opposite was true for boys. As no information was provided about the hypothetical child Lee, other than "Lee uses a computer to talk", variables associated with disability are left up to the child's imagination or his/her concept of what a child who "uses a computer to talk" would be like. The absence of descriptive information about Lee may serve to decrease the significance of male/female difference previously attributed to "functional" vs "cosmetic" variables associated with handicap.

Examination of gender differences may be targeted in future research by providing the listener more descriptive information regarding the hypothetical user.

School and Attitudinal Differences

School was found to be a significant main effect for the cognitive component of attitude. In the absence of demographic data (i.e. economic, social, occupational, parent's native language, educational level, etc.) it is not possible to identify the specific variables which contributed to the difference found between schools.

Ho

ep

ed

sl

ac

pr

p

c

m

s

R

C

3

C

However, King and associates (1988) report in their epidemiological study of children's attitudes that educational and occupational levels of parents were not significant factors contributing to students' attitudes. In addition, they reported no significant difference in the performance of those children whose parents consented to participate in the study versus the performance of those children whose parents refused.

The cognitive (CSUM) component of the CATCH is a measure of knowledge or belief. School was only found to be significant in the context of this component. It is quite possible that belief, unlike emotional and behavioral components, may be amenable to educational programs. That is, the school program may influence the child's beliefs or cognitive responses without largely impacting on the other components.

Rosenbaum et al. (1986) found school programs to be a significant factor upon examining the attitudes of children in traditionally segregated programs for disabled students versus children in schools with a long standing policy of integrating disabled children into regular classrooms. The attitude scores of children in traditionally segregated programs were found to be higher. However, there was no relationship between percentage of visibly disabled students in the school and the mean CATCH scores.

Generalizing these findings to the study at hand, school difference may not be attributed to parent consent,

pa

of

sc

fu

se

f

i

d

c

a

h

y

parent's educational or occupational levels, nor the number of visibly disabled/communicatively impaired students in the school. Having ruled out these factors suggests the need for further investigation of contributing school program (e.g. segregated versus integrated programming) factors.

Programs which offer instructional intervention (i.e. formal lessons, movies, and simulated interaction) to improve attitudes, in the absence of actual contact with the disabled or nonspeaking child may simply serve to change one's belief system without resulting in an overall attitudinal or behavioral change. Intervention which provides direct contact over a period of time has been purported to be the most effective method for improving attitudes (Voeltz, 1980, 1982; Armstrong et al. 1987). Armstrong and associates (1987) found that randomly assigning children to buddy groups (pairing of gender matched disabled and non-disabled children in social activities for 3 months) resulted in improved overall CATCH attitude scores.

There is an urgent need to encourage speaking and nonspeaking child interaction. Attitude change is the first step. Some evidence suggests that school integration of handicapped and nonspeaking children alone may not result in improved attitude. Gottlieb et al., (1974) (cited by Rosenbaum et al. 1988) reported attitudes to be less favorable in schools with exposure to children requiring special education. Also, successful integration requires a

str

al.

inv

th

pa

su

on

st

gu

i

E

T

a

i

M

3

3

structured program such as the buddy program (Armstrong et al. 1987) and the support of an interdisciplinary team.

A team approach, given the full support and active involvement of administrators, educators, physical therapists, habilitative specialists, speech-language pathologists as well as family members is necessary for a successfully integrated educational and communicatively oriented program. Neiswander (1978) provides a descriptive study of an administrative plan designed to serve as a guideline for administrators with the goal of identifying, initiating, and implementing a computer-based Communication Enhancement program for communicatively disordered students. The major findings of this study were that educational administrators have the responsibility of effectively integrating the varied philosophies and professional knowledge of various disciplines in order to:

1. Introduce new technology.
2. Introduce a new level of interdisciplinary cooperation.
3. Use computer technology as a focal point for communicative and educational purposes.
4. Acknowledge and overcome resentment toward new technology.

Further research is needed in this area to determine the true benefits of such a program and its long-term effects on the interaction of nonspeaking and speaking child.

Further

Preferen

Voi

results

be a sig

consequ

Signifi

natural

synthes

signifi

Thus, t

indicat

closely

voice.

T

consis

Mirenc

(1989)

femal

prefe

prefe

only

Other

appr

been

list

Further Examination of Variables Contributing to Voice Preference

Voice preference and perception of quality. The results of this study indicate that voice type was found to be a significant factor in determining voice preference and consequently preference for type of speech synthesizer. Significantly greater preference was indicated for the natural voice when compared to the three speech synthesizers. In addition, DECTalk was found to be significantly different when compared to SmoothTalker. Thus, these findings were consistent with past studies indicating a greater preference for synthetic voices most closely approximating the natural quality of the human voice.

The preference for natural voice was found to be consistent with the findings of a number of other studies. Mirenda, Eicher and Beukelman (1989) and Crabtree et al. (1989) report that adults as well as children, males, and females expressed a preference for the natural voice. This preference has been attributed to judges' tendency to base preference on "naturalness". Naturalness however, may be only one of several criteria used in determining preference. Other criteria include intelligibility, age- and gender-appropriateness of the voice, and context. However, it has been found that these criteria may vary as a factor of the listener's age and gender.

Naturalness, age- and gender-appropriateness. In a study examining listener acceptability judgements of human and synthesized speech, Quist and Lloyd (1988) reported that adult female subjects expressed a preference for unassisted human speech (i.e. normal, monotone, and esophageal) to synthesized speech (DECTalk, Echo+, and Votrax). The DECTalk was preferred over the other speech synthesizers as well as artificial laryngeal devices. It was believed by the researchers that the more speech approximates the "norm" or "standard", the more it will be preferred. Alternatively, the more divergent it is from the norm the more likely it is to be rejected.

There appears to be a small "margin of safety" between a divergence from the norm that can be described as bizarre enough to be interesting versus excessively divergent to be foreign and unpleasant. Such a distinction may be unique to the young and curious, who have grown up with "Star Wars", Pac Man, an assortment of video games and talking toys. This variability was evident in the study at hand by some of the comments made by subjects following exposure to the voice treatment. Comments ranged from "If I had a talking computer I wouldn't go to school" to "That sounds neat!" What appeals to adults obviously can not be indiscriminately generalized to today's youth.

A similar age difference was reported by Mirenda et al. (1989), when subjects were asked "If you were using the computer in the slide to learn something new, how would you

fe

a

su

lo

in

s

s

e

e

feel if it had this voice?" The younger subjects expressed a preference for the four synthesized voices. The older subjects preferred the four natural voices, expressing lowest preference for those voices which were least natural in quality.

It is important to note, in repeating this study using seven of the most current commercially available synthesizers, these findings were not replicated (Crabtree et al., 1989). Most of the voices preferred by subjects, crossing all ages and gender, were natural in quality but also included the SmoothTalker and the Votrax.

As technology continues to improve, we will find speech synthesis more closely approximating natural speech. Increased exposure in an increasingly technologically oriented society may serve to desensitize listeners of all ages. Differences previously witnessed among and between ages and gender may gradually become less divergent.

Today, children are experiencing greater exposure to high quality synthetic speech in the form of DECTalk or digitized speech through such avenues as automated devices, airport vehicles, movies, cars, toys, etc. As a consequence, the yardstick by which they measure the naturalness of other voice types may be based on an increasingly stringent criterion.

Naturalness does play a critical role in determining preference. However, it is important to recognize that preference is influenced by a variety of factors pertaining

to quality as well as the complex interaction of various features.

Nusbaum, Schwab and Pisoni (1984) identify some of these qualities in the form of adjective pairs used to indicate differences between natural and synthetic speech. Qualities ascribed to natural speech included; interesting, easy, gentle, clear, pleasant, smooth, friendly, improves concentration, etc. Many of the characteristics ascribed to synthetic speech; hard, frustrating, confusing, annoying, halting, distracting, etc. were based on or reflected degree of intelligibility. As a consequence, rank ordering of natural and synthetic speech were based on both the perception of naturalness and intelligibility.

A strong relationship exists between preference and intelligibility. Many studies examining preference judgements comparing different synthetic voices have found that the preferred voice was always the more intelligible voice (Logan & Pisoni, 1986; Crabtree et al., 1989; Quist & Lloyd, 1988).

Mirenda and Beukelman (1987, 1990) have reported that DECTalk most closely resembles natural speech in intelligibility followed by SmoothTalker 3.0 and RealVoice. Crabtree et al., (1989) report intelligibility was found to be the deciding factor in expressing preferences among eight different speech synthesizers as the female RealVoice and the SmoothTalker were preferred by more subjects in more contexts.

o

c

n

R

e

cl

v

g

a

l

"p

s

j

t

c

t

f

c

e

h

y

w

a

These findings were found to be consistent with those of the study at hand. Children's preference ratings correlated with the degree of intelligibility, ranking natural voice first, followed by DECTalk, SmoothTalker and RealVoice.

In addition to naturalness and intelligibility it is essential that speech synthesis be able to communicate those characteristics of the speaker normally ascribed to natural voice. These include (but are not limited to) the age and gender of the speaker. It may be that gender- and age-appropriateness represent some of the "natural" qualities looked for in speech synthesis, features that add to its "humanness".

Mirenda, Eicher, and Beukelman (1989) report the significance of this criteria in subjects' preference judgements concerning six hypothetical users. It was found that synthetic speech was preferred when the natural voice conflicted with the age- and/or gender-appropriateness of the hypothetical user. In other words, synthetic speech was found to "win by default" when natural speech was not consistent with the gender and/or age of the user.

This may lend further insight to the preferences expressed by subjects in the present study. Lee, the hypothetical child, was presented as a gender-neutral 9-year-old. The natural voice being that of a 9-year-old girl was preferred in every respect (natural, intelligible, age- and possibly gender-appropriate). DECTalk's "Kit the Kid",

the second choice and the only other voice to receive a rating above neutral, was comparable to the natural voice in intelligibility, and was both age and neutral gender appropriate. SmoothTalker and RealVoice (both with ratings below 2.0) exhibited reduced intelligibility and were obviously female in gender (which may have confirmed the preconceived judgement by the listener that Lee was a girl). However, both voices represented adults, suggesting inappropriate age, further lending to their low ratings.

It would appear that, while certain qualities (naturalness, intelligibility) are quite critical as criteria for judging preference, the number/frequency of inconsistencies is detrimental to acceptance.

While it is beyond the realm of this study, evidence suggests the need to examine other variables of synthetic voice quality such as appropriateness relative to ascribed personality characteristics as well as family, social, regional, and ethnic dialect. These serve as essential speech characteristics which should be incorporated in the next generation of speech synthesizers.

Application and Clinical Implications

The results of this study suggest a need for greater sensitivity to children's expressed preferences for perceived voice quality (and other related variables) in the prescription of voice-output communication aids. Past research has presented evidence which suggests esthetic characteristics of voice (quality, intelligibility, etc.)

contribute to the approachability and social attractiveness of the speaker. Peer interaction is essential to the social, psychological, cognitive, and communicative development of children. Communication devices should serve as vehicles to facilitate interaction. Given a measure serving as a prognostic indicator of social acceptability among child peers, the clinician would be greatly assisted in prescribing a more individualized/personalized match that is both pleasing to the child user and his/her peers.

In conclusion, a significant correlation has been found to exist between voice preference and attitude. However, this correlation was found to be low such that vocal preference accounts for a small percentage of the attitudinal variance. Children do perceive a difference in voice types and express a preference for natural and synthesized speech which may reflect perceived vocal quality. This suggests that a voice-output communication device should not be blindly prescribed, as its level of quality and level of listener's preference may contribute to the attitude of the listener toward its user and thus the social interaction among child peers.

APPENDICES

Sub

De

I a
a
ou
us
re
li
qu
wo
BE

In
re
in
ac
he
Ne
fr
co

If
re
sc

It
th
we
th
be

To
de
Fo
ch

Th

S

S
D

M

APPENDIX A

Subject: Parent Consent Form for
The Talking Computer Project

Dear Parent,

I am requesting permission for your child to participate in a research project. The purpose of this project is to find out what children know and think about other children who use a computer to talk. Your child will be presented a tape recording of a child using a computer that talks. After listening to this recording he/she will be given questionnaire which asks questions regarding what your child would do if he/she were to meet this child. THIS WILL NOT BE PRESENTED DURING YOUR CHILD'S CLASS INSTRUCTION TIME.

In order for your child to participate in this project, a review process is necessary to determine his/her eligibility in this study. I am therefore asking your permission to have access to your child's school records to determine his/her hearing and reading levels within the past two years. Neither school records nor copies of records will be removed from the school premises. Information will be kept strictly confidential.

If this information is not available, I will provide a reading and hearing screening. The results of your child's screening will be made available to you upon your request.

It is my strong belief that your child's participation in this project will be a learning experience for him/her as well as provide a major contribution to our understanding of those factors which may facilitate child interaction and how best to use talking computers for communication.

To indicate consent to your child's participation in the described project, please sign the attached Parent Consent Form and the Consent Checklist and return them to your child's teacher by April 18th.

Thank you for your consideration.

Sincerely,

Sheila J. Bridges, M.A. CCC-Sp
Department of Audiology
and Speech Sciences
Michigan State University

For more information, please
contact Sheila Bridges
Phone: (517) 353-5399
(517) 355-2825

I .

of

re

re

pr

re

pe

re

al

so

Talking Computer Research Project

**Michigan State University
Department of Audiology
and Speech Sciences**

PARENT CONSENT FORM

I _____ freely consent to the participation
(parent's name)

of _____ in the previously described
(child's name)

research project. It is my understanding that my child will remain completely anonymous. In the event that a review process requires information pertaining to my child's reading and hearing performance, Ms. Bridges has been given permission to have access to _____ academic
(child's name)

records and, should such information not be available, she also has permission to administer a reading and/or hearing screening.

(signature)

(date)

Talking Computer Research Project

**Michigan State University
Department of Audiology
and Speech Sciences**

CONSENT CHECKLIST

- _____ I am consenting to the participation of my child
_____ (child's name) in the Talking
Computer Research Project which has been fully and
clearly explained and summarized.
- _____ I understand that the test administration will take
an estimated 45 - 60 minutes and that it will not be
presented during class instruction time (i.e. lunch
period and recess). Additional time may be required
in the event a hearing and/or reading screening is
necessary.
- _____ I understand that the study does not involve any
inherent risk or discomfort to my child.
- _____ I freely and voluntarily consent to my child's
participation in this study with the understanding
that he/she may discontinue participation at any
point in time during the project.
- _____ I understand that my child's participation in this
study will remain in the strictest confidence.
- _____ While maintaining strict confidentiality (as
promised above) it is my understanding that the
project results may be presented at a professional
conference and submitted to a professional
journal/publication.

parent name (print)

(parent signature)

(date)

APPENDIX B

LEE'S MONOLOGUE

Hi. My name is Lee and I'm 9 years old. I've lived in Lansing Michigan for 2 years. I use to live in Tyler Texas. In Tyler its hot all the time. Almost all the time. And everybody I know lives down there. Well, at least all my family does that I know, live down there.

Lansing's all right. I like the stores in Lansing. They're better than Tyler. Meijer's is my favorite. It has clothes and books.

I use to go to Gary when I was in the first grade in Tyler. Now I'm in third grade and I go to Maple Grove. We do spelling, math, social studies, english, and science. I hate science. I like social studies and math a lot. In social studies we're talking about Alexander Graham Bell. He's the person that invented the telephone. We do art every Thursday. And music, usually we have music every day. Not lunch time, but at recess sometimes we have music. I know I can play two songs. One is called the "Bunny Hop" and the other is "I Want a Piece of Pie", which sounds country. I don't really like country.

APPENDIX C

Mrs. Harris' Monologue

Hi. My name is Mrs. Harris. I live in Lansing, MI. I have lived in East Lansing for two years. I have two daughters. Their names are Tanja and Nichole. Tanja had her birthday two days ago and she is 7 now. Nichole my younger daughter is five years old.

APPENDIX D

EXAMPLES OF HOW TO FILL OUT THE FORM:

Read each statement carefully. Decide how you feel about the statement. Put an "X" through the choice (strongly disagree.... strongly agree) that best describes how you feel.

Sample

1. Mrs. Harris is a man.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

2. I would really like talking to Mrs. Harris.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

3. I would be embarrassed if my mom invited Mrs. Harris to my birthday party.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

APPENDIX E

- THERE ARE NO RIGHT OR WRONG ANSWERS.
- I JUST WANT TO KNOW YOUR IDEAS.
- PLEASE DO NOT READ AHEAD.
- THINK ABOUT EACH SENTENCE CAREFULLY.
- WHEN FINISHED, REVIEW ITEMS TO MAKE SURE NONE WERE SKIPPED

1. I wouldn't worry if Lee sat next to me in class.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

2. I would not introduce Lee to my friends.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

3. Lee can do lots of things for herself/himself.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

4. I wouldn't know what to say to Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

5. Lee likes to play.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

6. I feel sorry for Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

7. I would stick up for Lee if she/he was being teased.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

8. Lee wants lots of attention from adults.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

9. I would invite Lee to my birthday party.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

10. I would be afraid of Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

11. I would talk to Lee even though I don't know her/him.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

12. Lee doesn't like to make friends.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

13. I would like having Lee live next door to me.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

14. Lee feels sorry for herself/himself.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

15. I would be happy to have Lee for a special friend.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

16. I would try to stay away from Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

17. Lee is as happy as I am.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

18. I would not like Lee as much as my other friends.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

19. Lee knows how to behave properly.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

20. In class I wouldn't sit next to Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

21. I would be pleased if Lee invited me to her/his house.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

22. I would try not to look at Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

23. I would feel good doing a school project with Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

24. Lee doesn't have much fun.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

25. I would invite Lee to sleep over at my house.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

26. Being near Lee would scare me.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

27. Lee is interested in lots of things.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

28. I would be embarrassed if Lee invited me to her/his birthday party.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

29. I would tell my secrets to Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

30. Lee is often sad.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

31. I would enjoy being with Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

32. I would not go to Lee's house to play.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

33. Lee can make new friends.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

34. I would feel upset if I saw Lee.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

35. I would miss recess to keep Lee company.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

36. Lee needs lots of help to do things.

Strongly Disagree	Disagree	Can't Decide	Agree	Strongly Agree
----------------------	----------	--------------	-------	-------------------

APPENDIX F

Rating of Voice Quality

You have just listened to one of Lee's voices. However, Lee has 4 different voices. I would like to know how much you like each of Lee's voices. You will hear all 4 voices twice. The first time all 4 voices will be presented at once. You will just listen. The second time you will listen to each voice. After each voice place an "X" across the face with the statement that best describes how much you like the voice you have just heard.

Voice 1

I like it
very much



I like it



I don't care
either way



I don't
like it



I don't like
it at all



Voice 2

I like it
very much



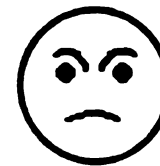
I like it



I don't care
either way



I don't
like it



I don't like
it at all



Voice 3

I like it
very much



I like it



I don't care
either way



I don't
like it



I don't like
it at all



Voice 4

I like it
very much



I like it



I don't care
either way



I don't
like it



I don't like
it at all



APPENDIX G

INSTRUCTIONS FOR THE TALKING COMPUTER PROJECT

INSTRUCTIONS FOR TEST ADMINISTRATOR

INTRODUCTION

I'm _____ from Michigan State University. We are doing a project. We would like to find out what you know and think about children who use a computer to talk. When your teacher calls your name please come up to get your questionnaire. You will then leave as a group to go to another room to listen to a tape recording of a child talking using a computer.

(CHILDREN WILL BE GIVEN THEIR QUESTIONNAIRE AND A PENCIL AND GO TO THE ASSIGNED ROOM).

EXPLANATION OF THE CATCH

You will be listening to a tape recording of a child named Lee. This is not a test, so there is no right or wrong answer. These are questions asking what you think about Lee after you have listened to the recording. It will include questions about things that you would or would not like to do if you were to meet Lee.

I will explain this form step-by-step.

(PRESENT POSTER ILLUSTRATING EXAMPLES)

On the next page, there are some examples of how to fill out this form. Let's practice these examples by listening to a recording of Mrs. Harris. This is ONLY FOR PRACTICE.

(PLAY RECORDING OF MRS. HARRIS)

First, read the statement to yourself and then decide how you feel about the statement. You have 5 choices to choose from (POINT TO THEM AND READ EACH ONE ALOUD). The sample shows that you must put an "X" across the choice you think best states your feelings.

When you agree with the sentence it means "this is exactly the way you feel. This is just what you think, so you AGREE". If you agree with Sample A then you think Mrs. Harris is a man.

To disagree means this is not the way you feel. You do not think this at all. So to DISAGREE with Sample A, means that you DO NOT AGREE, you do not think Mrs. Harris is a man. If you Disagree very much, then you place the "X" across STRONGLY DISAGREE (POINT TO THE SAMPLE).

The first example says: (READ IT ALOUD). If you would really hate talking to Mrs. Harris, then maybe you'd pick "Strongly Disagree", because you do not agree with the statement at all;

or maybe you would just dislike talking to Mrs. Harris, so you might pick "Disagree";

or maybe you just don't really know how you feel about the statement so you might pick "Can't Decide";

or maybe you might enjoy talking to Mrs. Harris, so you might pick "Agree";

or maybe you really would like talking to Mrs. Harris, so you might pick "Strongly Agree".

Decide how you feel about the statement and then mark one of the 5 boxes by putting an "X" through it.

(ASK THEM IF THERE ARE ANY QUESTIONS ABOUT HOW TO COMPLETE IT AND ASK THEM TO TRY THE NEXT EXAMPLE. EXAMINE EACH CHILD'S RESPONSE. REPEAT THE ABOVE STATEMENTS ABOUT EACH RESPONSE IF NECESSARY.)

(BEGIN LEE'S MONOLOGUE)

You have now finished the practice section of the project. You will now hear a cassette recording of Lee. At the end of the recording you will complete the questions asking you what you think about Lee. You will not begin writing until the recording has finished playing and I have told you to begin.

(READ THE INSTRUCTIONS AT THE TOP OF THE QUESTIONNAIRE).

(PLAY THE APPROPRIATE CASSETTE TAPE)

Now go ahead and complete the next five pages.

APPENDIX H

INSTRUCTIONS FOR THE TALKING COMPUTER PROJECT

INSTRUCTIONS FOR STUDENTS

You will be listening to a tape recording of a child named Lee. This is not a test, so there is no right or wrong answer. These are questions asking what you think about Lee after you have listened to the recording. It will include questions about things that you would or would not like to do if you were to meet Lee.

On the next page, there are some examples of how to fill out this form. Let's practice these examples by listening to a recording of Mrs. Harris. This is ONLY FOR PRACTICE.

First, read the statement to yourself and then decide how you feel about the statement. You have 5 choices from which to choose.

The sample shows that you must put an "X" across the choice you think best states your feelings.

When you agree with the sentence it means "this is exactly the way you feel. This is just what you think, so you AGREE."

If you agree with Sample A then you think Mrs. Harris is a man.

To disagree means this is NOT the way you feel. You DO NOT think this at all. So to DISAGREE with Sample A, means that you DO NOT AGREE, you do not think Mrs. Harris is a man. If you Disagree very much, then you place the "X" across STRONGLY DISAGREE.

Read example 1. If you would really hate talking to Mrs. Harris, then maybe you'd pick "Strongly Disagree" because you do not agree with the statement at all; or maybe you would just dislike talking to Mrs. Harris, so you might pick "Disagree"; or maybe you just don't really know how you feel about the statement so you might pick "Can't Decide"; or maybe you might enjoy talking to Mrs. Harris, so you might pick "Agree"; or maybe you really would enjoy talking to Mrs. Harris, so you might pick "Strongly Agree".

Decide how you feel about the statement and then mark one of the 5 boxes by putting an "X" through it.

TAPE RECORDING OF LEE

You have now finished the practice section of the project. You will now hear a cassette recording of Lee. At the end of the recording you will complete the questions asking you what you think about Lee. You will not begin writing until the recording has finished playing and I have told you to begin.

APPENDIX I

RELIABILITY ANALYSIS - SCALE (ALPHA)

Reliability Coefficients 36 items
Alpha = .9227

Factor Analysis of the Modified CATCH

FACTOR ANALYSIS

Variable	Communality	Factor	Eigenvalue	Pct Var	CumPct
Q01	.51882	1	11.30984	31.4	31.4
Q02	.29237	2	2.78662	7.7	39.2
Q03	.09745	3	2.01971	5.6	44.8
Q04	.09654				

Format = Sort Blank (.5)

Rotation = Varimax

LIST OF REFERENCES

REFERENCES

- Addington, D. (1968). The relationship of selected vocal characteristics to personality perception. Speech Monographs, 25, 492-503.
- Anisfeld, E., & Lambert, W. (1964). Evaluational reactions of bilingual and monolingual children to spoken language, Journal of Abnormal and Social Psychology, 69, 89-97.
- Apolloni, R. & Cooke, T. (1978). Integrated programming at the infant, toddler and preschool levels. In M. Guralnick (Ed.) Early intervention and the integration of handicapped and nonhandicapped children. Baltimore, MD: University Park Press.
- Argyle, M., Salter, V., Nicholson, H., Williams, M. Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and non-verbal signals. British Journal of Social and Clinical Psychology, 9, 222-231.
- Armstrong, R., Rosenbaum, P. & King, S. (1987). A randomized controlled trial of a "Buddy" programme to improve children's attitudes toward the disabled. Developmental Medicine and Child Neurology, 29, 327-336.
- Barnes, D. (1971). Preschool play norms: A replication. Developmental Psychology, 5, 99-103.
- Bennett, S. & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. Journal of Speech and Hearing Research, 16, 608-615.
- Bernstein, J. (1988). Linguistics in voice output communication aids. In L. Bernstein (Ed.) The vocally impaired: Clinical practice and research (pp. 186-205). Philadelphia: Grune & Stratton.
- Bernstein, J. (1986, April). Voice identity and attitude. The Official Proceedings of Speech Tech '86 (pp. 213-215). New York: Media Dimensions.
- Blackstone, S. (Ed.) 1988 - A good year for speech output? Augmentative Communication News, 1(1), 1-3.
- Blood, B., Mahon, B., & Hyman, M. (1979). Judging personality and appearance from voice disorders. Journal of Communication Disorders, 12, 63-68.

- Brown, B., Strong, W., Rencher, A., & Smith, B. (1974). Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. Journal of the Acoustical Society of America, 55.
- Brown, B., Strong, W., Rencher, A., & Smith, B. (1973). Perceptions of personality from speech: effects of manipulations of acoustical parameters. Journal of the Acoustical Society of America, 54, 29-35.
- Calculator, S. & Luchko, C. (1983). Evaluating the effectiveness of a communication board training program. Journal of Speech and Hearing Disorders, 48, 185-191.
- Cohen, C. & Palin, M. (1986). Speech syntheses and speech recognition devices. In M. Grossfeld & C. Grossfeld (Eds.) Microcomputer application in rehabilitation of communication disorders (pp. 183-211). Rockville, MD: Aspen.
- Coxon, L. & Laikko, P. (1983). Listener reactions to three nonvocal communication outputs, Unpublished master's thesis, Washington State University, WA.
- Crabtree, M., Mirenda, P., & Beukelman, D. (1989, Nov.) Age and gender preference for synthetic and natural speech. Paper presented at the annual convention of the American Speech-Language-Hearing Association, St. Louis, MO.
- Dahmke, M. (1982). Let there be talking people too. Byte, 7(9), 7-8.
- Eulenberg, J., Wood, C., & Finkelstein, S. (1985). Age- and gender-appropriate familial voice output: the Shanon Singer system. The Official Proceedings of Speech Tech'85. New York: Media Dimensions, Inc.
- Eulenberg, J. & Rahimi, M. (1978). Toward a semantically accessible communication aid. Proceedings of the National Electronics Conference, 32.
- Ervin-Tripp, S. (1967). An Issei Learns English. Journal of Social Issues, 23, 78-90.
- Ferrier, L. & Shane, H. (1981). Communication Skills. In J. Umbreit and Carullias (Eds.) Educating the severely physically handicapped: Curriculum adaptations (pp. 17-34). Columbus, OH: Special Press.

Fishman, J. (1969). Bilingual attitudes and behaviors. Language Science, 5, 5-11.

Galyas, K. (1988). Microcomputer technology and speech synthesis in communication aids. Stockholm, Sweden: Department of Speech Communication, Royal Institute of Technology.

Giles, H. & Powesland, P. (1975). Speech style and social evaluation. New York: Academic Press.

Goffman, I. (1959). The presentation of self in everyday life, New York: Doubleday.

Goldstein, H. & Terrell, D. (1987). Augmenting communicative interaction between handicapped and nonhandicapped preschool children. Journal of Speech and Hearing Disorders, 52, 200-211.

Goossens, C., & Kraat, A. (1985). Technology as a tool for conversation and language learning for the physically disabled. Topics in Language Disorders, 6(1), 56-70.

Greene, B., Logan, J. & Pisoni, D. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. Behavior Research Methods, Instruments, and Computers. 18 (2) 100-107.

Greene, B., Manous, L., & Pisoni, D. (1984). Perceptual evaluation of DECTalk: Final Report on Version 1.8. Research in Speech Perception Progress Report No 10. Bloomington, IN Speech Research Laboratory, Indiana University.

Greene, B., and Pisoni, D. (1988). Perception of synthetic speech by adults and children: Research on processing voice output from text-to-speech systems. In L. Bernstein (Ed.) The vocally impaired: Clinical practice and research (pp. 206-248). Philadelphia: Grune & Stratton.

Guralnick, M. (1981). The social behavior of preschool children at different developmental levels: effects of group composition. Journal of Experimental Child Psychology, 31, 115-130.

Guralnick, M. (1976). The value of integrating handicapped and nonhandicapped preschool children. American Journal of Orthopsychiatry, 46, 236-245.

- Guralnick, M. & Paul-Brown, D. (1980). Functional and discourse analyses of nonhandicapped preschool children's speech to handicapped children. American Journal of Mental Deficiency, 84(5), 444-454.
- Harris, D. (1982). Communication interaction processes involving nonvocal physically handicapped children, Topics in Language Disorders, 2(2), 21-37.
- Harris, D. (1978). Descriptive analysis of communicative interaction processes involving non-vocal severely physically handicapped children. Doctoral dissertation, University of Wisconsin-Madison, WI.
- Harris, D. & Vanderheiden, G. (1980). Enhancing the development of communicative interaction, In R. Schiefelbusch (Ed.) Nonspeech language and communication, Baltimore, MD: University Park Press.
- Hart, R. J. & Brown, B.L. (1974). Interpersonal information contained in the vocal qualities and in content aspects of speech. Speech Monographs.
- Holmquist, E. (1984). I am my own person. In E. Eulenberg (Ed.), Conversations with Non-Speaking People, Toronto: Canadian Rehabilitation Council for the Disabled.
- House, A., Williams, C., Hecker, M., Kryter, K. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America. 37:158-166.
- Kelly, M. & Chapanis, A. (1977). Limited vocabulary natural language dialogue. International Journal of Man-Machine Studies, 9, 479-501.
- King, S., Rosenbaum, P., Armstrong, R., & Milner, R. (1988). An epidemiological study of children's attitudes toward disability. Unpublished manuscript.
- Klatt, D. (1986). Text-to-Speech: Present and future. The Official Proceedings of Speech Tech'86 (pp. 221-226). New York: Media Dimensions, Inc.
- Kraat, A. (1985). Communication interaction between aided and natural speakers. Toronto, Canada: Canadian Rehabilitation of Council for the Disabled.
- Kraat, A. (1986). Developing Intervention Goals. In S. Blackstone (Ed.) Augmentative communication: An introduction (pp.267-290). Rockville, MD: ASHA

- Locke, P. & Mirenda, P. (1988). A computer-supported communication approach for a child with severe communication, visual, and cognitive impairments: A case study. Augmentative and Alternative Communication, 4(1), 15-22.
- Logan, J. & Pisoni, D. (1986). Preference judgements comparing different synthetic voices. Presented at the 111th meeting of the Acoustical Society of America, Cleveland, OH.
- Mehrabian, A. & Wiener, M. (1967). Decoding of inconsistent communications. Journal of Personality and Social Psychology, 6, 109-114.
- Mirenda, P. & Beukelman, D. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. Augmentative and Alternative Communication, 3(3), 120-128.
- Mirenda, P. & Beukelman, D. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. Augmentative and Alternative Communication, 6(1), 61-68.
- Mirenda, P., Eicher, D., & Beukelman, D. (1989). Synthetic and natural speech preferences of male and female listeners in four age groups. Journal of Speech and Hearing Research, 32 (1), 175-183.
- Morris, S. (1981). Communication interaction development at mealtimes for the multiply handicapped child: implications for the use of augmentative communication systems. Language Speech Hearing Services in the Schools, 12, 216-232.
- Neiswander, L. (1978). The initiation and administration of a computer-based communication enhancement pilot program. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.
- Newcombe, P. (1986). Voice and diction. Raleigh, NC: Contemporary Publishing
- Newell, A. (1984, June). Do we know how to design communication aids? Second International Conference on Rehabilitation Engineering, Ottawa, Canada.
- Norusis, M. (1988). SPSS/PC+ advanced statistics V2.0 for the IBM PC/XT/AT and PS/2. Chicago, IL: SPSS Inc.

- Nusbaum, H. & Pisoni, D. (1985). Constraints on the perception of synthetic speech generated by rule. Behavior Research Methods, Instruments, and Computers. 17: 235-242.
- Nusbaum, H., Schwab, E., & Pisoni, D. (1983). Perceptual evaluation of synthetic speech: Results from eight text-to-speech systems. Research on Speech Perception Progress Report No. 11. Bloomington, IN: Speech Research Laboratory Indiana University.
- Nusbaum, H., Schwab, E., & Pisoni, D. (1984). Subjective evaluation of synthetic speech: Measuring preference, naturalness, and acceptability. Research on Speech Perception Progress Report No. 10. Bloomington, IN: Speech Research Laboratory Indiana University.
- Quist, R. & Lloyd, L. (1988). Listener acceptability judgments of human and synthesized speech. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Boston, MA.
- Richardson, S. (1969). The effect of physical disability on the socialization of a child. In D. Goslin (Ed.) Handbook of socialization theory and research. New York: Rand McNally.
- Rosenbaum, P., Armstrong, R., & King, S. (1988). Determinants of children's attitudes toward disability: A review of evidence. Unpublished manuscript.
- Rosenbaum, P., Armstrong, R., & King, S. (1986). Children's attitudes toward disabled peers: A self-report measure. Journal of Pediatric Psychology. 11(4), 517-530.
- Schwartz, A. & Koenig, M. (1987). A comparison of microcomputer-based and dedicated augmentative communication systems. Seminars in Speech and Language, 8(2) 143-152.
- Seligman, C., Tucker, G., & Lambert, W. (1972). The effects of speech style and other attributes on teachers' toward pupils. Language in Society, 1, 132-142.
- Shrewsbury, R., Lass, N., & Joseph, L. (1985). A survey of special educators' awareness of, experiences with, in the schools. Language, Speech, and Hearing Services in the Schools, 16, 293-298.
- Silverman, E. (1976). Listeners' impressions of speakers with lateral lisps. Journal of Speech and Hearing Disorders, 41, 547-552.

Strain, P. & Shores, R. (1977). Social reciprocity: A clinical teaching perspective. Exceptional Children, 43, 526-530.

Triandis, H.C. (1971). Attitude and attitude change. New York: Wiley.

Tucker, G.R. (1969). Judging personality from language usage: A Filipino example. (mimeo).

Vanderheiden, G. (1983). Non-conversational communication technology needs of individuals with handicaps. Rehabilitation World, 7, 8-12.

Vanderheiden, G. & Lloyd, L. (1986). Communication systems and their components. In S. Blackstone (Ed.) Augmentative communication: An introduction (pp. 49-161). Rockville, MD: ASHA.

Voeltz, L. M. (1982). Effects of structured interactions with severely handicapped peers on children's attitudes. American Journal of Mental Deficiency, 86, 380-390.

Voiers, W. (1977). Diagnostic acceptability measure for speech communication systems. IEEE International Conference on Acoustics, Speech, and Signal Processing, Hartford, CT.

Voiers, W. (1980). Interdependencies among measures of speech intelligibility and speech "quality". IEEE International Conference on Acoustics, Speech, and Signal Processing, Hartford, CT.

Weeks, C., Kelly, M. & Capanis, A. (1974). Studies in interactive communication: Cooperative problem solving by skilled and unskilled typists in a Teletypewriter mode. Journal of Applied Psychology, 59, 665-674.

Westervelt, V. & Turnbull, A. (1980). Children's attitudes toward physically handicapped peers and intervention approaches for attitude change. Physical Therapy, 60, 890-901.

Wieder, S. & Kornet, R. (1983). Assessing pragmatic abilities in a multi-handicapped child, Unpublished manuscript, Queens College, CUNY, Flushing, NY.

Yoder, D. & Kraat, A. (1983). Intervention issues in non-speech communication. In J. Miller, D. Yoder, and R. Schiefelbusch (Eds.) Contemporary issues in language Intervention, ASHA Report #12. Rockville, MD: ASHA.

Yorkston, K., Beukelman, D., & Trayner, C. (1984).
Computerized assessment of intelligibility of
dysarthric speech. Austin, TX: Pro-Ed.

MICHIGAN STATE UNIV. LIBRARIES



31293006277473