

7-24-1975

A Comparative Study of the Relative Incidence of Various Vocal Parameters Among Black and White College Females

Barbara Kuhl Jacobs
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Gender, Race, Sexuality, and Ethnicity in Communication Commons](#), and the [Other Communication Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Jacobs, Barbara Kuhl, "A Comparative Study of the Relative Incidence of Various Vocal Parameters Among Black and White College Females" (1975). *Dissertations and Theses*. Paper 2241.
<https://doi.org/10.15760/etd.2238>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

AN ABSTRACT OF THE THESIS OF Barbara Kuhl Jacobs for the Master of
Science in Speech Communication presented July 24, 1975.

Title: A Comparative Study of the Relative Incidence of Various Vocal
Parameters Among Black and White College Females.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

Mary E. Gordon, Chairperson

Robert C. Casteel

Jean McMahon

Evelyn I. Crowell

The purpose of this study was to make a comparative investigation of the voice quality of black and white women. A review of the literature revealed considerable research has been conducted in the area of black-white language differences in the United States, but very little research has been conducted regarding comparisons of voice quality of Blacks and Whites. The suggestion that certain voice qualities are more characteristic of Blacks or Whites when compared as

groups led the writer to pursue the present investigation. One suggestion of particular interest was that Blacks were judged more often to have hoarse voices. Two separate experiments were designed to consider the following questions:

- 1) Do the voices of a group of black women differ significantly from those of a group of white women on any parameter assessed by a) measurements made by the investigator using the Fairbanks Musical Scale Pitch Level Comparison (Fairbanks, 1960), or b) scaled value ratings made by ten judges on the Jewish Hospital Voice Profile (Wilson, 1972)?
- 2) Can ethnic group identity be ascertained according to perception of voice quality more often than expected by chance?

Indexical information on the subjects was collected for the following variables: 1) age, 2) length of longest sustained phonation, 3) optimum pitch, 4) modal pitch, 5) relationship of modal pitch to optimum pitch, 6) lowest possible pitch, 7) highest possible pitch, and 8) pitch range in semitones. The data analysis revealed no statistically significant differences between black and white women regarding these variables.

In the experiment involving the JHVP, fifty samples of women's voices (twenty-five Blacks and twenty-five Whites) were submitted to listener analyses of voice quality by ten women judges, two of whom were trained speech pathologists and eight were in training at the graduate level in speech pathology. The voice samples were heard by reverse audio tape playback to eliminate dialectical differences. The validity of the ratings of the JHVP is questionable, based on the poor inter-judge reliability. Recognizing this qualification, the ratings were nevertheless submitted to data analysis with the following re-

sults: Statistically significant differences observed were that black women were more often rated as too low-pitched during speech, hyponasal, and more "severe," while white women were more often rated as too soft and too breathy. No difference in the incidence of perceived hoarseness was found.

In the second experiment, brief segments of each of the fifty women's voices were presented by reversed playback in a randomized order for a forced listener choice of ethnic identity: Black or White. The results indicated such identity could not be made at a greater than chance level.

A COMPARATIVE STUDY OF THE RELATIVE INCIDENCE
OF VARIOUS VOCAL PARAMETERS AMONG
BLACK AND WHITE COLLEGE FEMALES

by

BARBARA KUHL JACOBS

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

MASTER OF SCIENCE
in
SPEECH COMMUNICATION

Portland State University
1975

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of Barbara Kuhl
Jacobs presented July 24, 1975.

Mary E. Gordon, Chairperson

Robert C. Casteel

Joan McMahon

Evelyn I. Crowell

APPROVED:

Robert W. Vogelsang, Head, Department of Speech Communication

Richard B. Halley, Dean of Graduate Studies and Research

November 26, 1975

ACKNOWLEDGMENTS

Bouquets to Mary E. Gordon, a woman of uncommon patience and good sense and good humor. Her untiring assistance was crucial to the reporting of this study.

Many thanks to the other members of my thesis committee for their participation: Ms. Evelyn I. Crowell, Ms. Joan McMahon, and Dr. Robert C. Casteel.

My gratitude to the fifty beautiful women who acted as subjects. Meeting them was the most enjoyable part of this exercise.

Sincere gratitude also to those who provided invaluable technical assistance: Michael Fairchild, Ted Clarke, Lynn Gibson, and Leslie Grueger.

Lastly, loud songs of praise to some very good people who rejuvenated me from time to time and to Baba Ram Dass and Lord Martin Cecil, who were light in the storm.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS	iii
LIST OF TABLES.	vii
LIST OF FIGURES	viii
 CHAPTER	
I INTRODUCTION.	1
Statement of Purpose.	4
II REVIEW OF THE LITERATURE.	5
Voice Quality: A Potpourri	5
Sources of Voice Quality Variability.	6
Acoustical, Physiological, and Perceptual Nature of Voice	8
Sound Production and Transmission Resonance and the Vocal Tract Resonators	
Analysis of Voice Quality	13
Reverse Playback of Audio Tape Recordings	
Hoarse Voice Quality.	19
Perceptual Definition of Hoarseness	
The Physiology of Hoarseness	
Relationship of Hoarseness to Pathology	
Measurement of Hoarseness	
Black-White Comparisons of Speech, Language, and Voice	25

CHAPTER	PAGE
III METHODS AND PROCEDURES.	27
Selection of Participants	27
Speaker-Subjects	
Judges	
Data Collection	28
Physical Setting	
Recording of Personal Information	
Audio Recording of Voice Samples	
Recording of Voice-Related Information	
Audio Tape Recordings	30
Reverse Audio Tape Recordings	
Training Tape	
Tape I	
Tape II	
Judging the Voice Tapes	32
Directions to Judges	
Rating Forms	
Data Analysis	33
IV RESULTS AND DISCUSSION.	34
Results	34
Indexical Information	
Tape I	
Tape II	
Discussion.	44
V SUMMARY AND IMPLICATIONS.	50
Summary and Conclusions	50
Implications.	52
SELECTED BIBLIOGRAPHY	54
APPENDICES.	57

	PAGE
A Subject Information Form.	57
B Text Read by Subjects	58
C <u>Fairbanks Musical Scale Pitch Level Comparison.</u>	59
D Key to Numerical Coding of Pitch Scale for Purposes of Data Analysis	60
E Rating Instructions to Judges as Transcribed from Audio Tape Recordings.	61
F Rating Form for Tape II	63

LIST OF TABLES

TABLE		PAGE
I	Black-White Group Differences on Indexical Information.	36
II	Intra-Judge Reliability on the <u>Jewish Hospital Voice Profile</u>	37
III	Instances of Judgments of Normalcy (Rating of "1"). . .	38
IV	Inter-Judge Reliability on the <u>Jewish Hospital Voice Profile</u> : Analysis of Variance.	38
V	Inter-Judge Agreement on the Scale of Resonance (JHVP).	39
VI	Ratings on the <u>Jewish Hospital Voice Profile</u>	40
VII	Severity Ratings on the <u>Jewish Hospital Voice Profile</u> .	42
VIII	Severity Ratings: Clinical Significance.	43
IX	Correct Judgments of Ethnic Group	45

LIST OF FIGURES

FIGURE		PAGE
1	Vibratory particle movement which occurs in sound transmission	9
2	A comparison of complex waves (d_1 and d_2) whose shapes differ due to the phase relationships of the components (after Denes and Pinson, 1963).	11
3	<u>Jewish Hospital Voice Profile</u> (from Wilson, 1972). . .	15

CHAPTER I

INTRODUCTION

Each person has a distinctive and unique voice quality. The layman usually regards voice quality as "how the voice sounds" or perhaps as the "melody" or "tone" of the voice. The term "voice quality" assumes more particular meaning for the speech pathologist because of his extensive training in the perceptual analysis of various types of voices. Training in voice analysis involves learning about the physical and acoustical nature of sound production and sound transmission, particularly of the human voice. An understanding of the inter-relatedness of these aspects and their effect upon consequent auditory perceptions by the listener allows the voice clinician to make sophisticated judgments regarding what factors may be operating to produce a particular quality. While it is not within the speech pathologist's realm to diagnose medical voice pathologies nor to determine etiology, he may make valuable clinical observations contributing to a consequent medical diagnosis and useful in planning vocal reeducation and management.

The speech pathologist's internal criteria for assessing voice quality will tend to differ from those of the average layman. Arbitrary limits of acceptability and normalcy of voice are subjectively established by both layman and speech pathologist, but it is likely the standards of the speech pathologist will be more conservative.

In other words, the speech pathologist will accept less divergence in voice quality from his internal criteria of normalcy than will the average layman before he regards the voice as deviant. This seemingly hypercritical evaluation results from the speech pathologist's awareness that consistent voice deviancies, even when mild, may be symptomatic of an underlying organic or functional pathology. He has been trained to be suspicious of even slight deviations of voice quality if they are heard consistently; whereas, the layman tends to ignore such variations as long as they are not unpleasant to the listener.

While a person's voice quality is in many ways highly individualistic, voice quality traits are subject to the influence of social and/or cultural factors. It has been suggested that there is often a discernible difference between the voice quality of black and white American speakers, even when dialectical differences are ignored. Interest in the topic of possible differences in voice quality between Blacks and Whites was further aroused by a personal experience as an instructor in a university course entitled "Voice and Diction." During a period of three terms, seven black students (four women and three men) were enrolled in the course. Of these seven, four (two females and two males) were consistently rated as having hoarse voices. This incidence of hoarseness (57.1 per cent overall, 50.0 per cent for females, and 66.6 per cent for males) was considerably higher than the incidence of hoarseness among the other students, all of whom were white. Admittedly, it would be premature to regard such an informal observation as significant due to possible judge bias, the small sample size, and the possibility that a "Voice and Diction"

class is a selective factor for enrollment of a group tending toward a greater incidence of voice disorders. These factors could have contributed to a false representation of the true incidence of hoarse voice quality among black students.

On the other hand, it is possible the perceptions, and hence judgments, of deviant voice quality among the black speakers were valid in terms of the judges' criteria, while the voices would have been regarded as within normal limits relative to the speakers' own ethnic and cultural standards. In other words, the black ethnic group may have different criteria for assessing normalcy and acceptability of voice quality than do a group of trained voice judges who are white. In fact, each of the black students rated as having a hoarse voice maintained his voice was normal in his judgment. One student expressed the opinion that the "huskiness" and somewhat low pitch accompanying her alleged hoarseness were characteristics typical of many speakers of her ethnic group, especially of women. She maintained that to purposely change her voice would be to compromise her ethnic identity.

Because of the possible implications of voice disorders to general health and well-being, a situation deserving serious consideration is created if one segment of the population is observed to have a significantly higher and unexplained incidence of such a disorder. This writer, therefore, desired to begin exploring the matter of whether there did, indeed, appear to be a trend toward more hoarseness (as judged by trained voice judges) among Blacks.

Statement of Purpose

The purpose of this investigation was to compare various aspects of the voice quality of black women and white women. Specifically this investigation was designed to determine whether ten speech pathologists trained in voice analysis would perceive voice quality differences between a group of twenty-five black females and a group of twenty-five white females.

The essential questions posed in this investigation were:

- 1) Do the voices of a group of black women differ significantly from those of a group of white women on any parameter assessed by a) measurements made by the investigator using the Fairbanks Musical Scale Pitch Level Comparison (Fairbanks, 1960), or b) scaled value ratings made by ten judges on the Jewish Hospital Voice Profile (Wilson, 1972)?
- 2) Can ethnic group identity be ascertained according to perception of voice quality more often than expected by chance?

The data obtained from this study cannot be regarded as conclusive in any sense because of the restricted sample size and lack of controls in selecting the sample. However, observed trends may lead to further investigation into the more pertinent aspects of the general question of voice quality differences between Blacks and Whites.

CHAPTER II

REVIEW OF THE LITERATURE

The following survey of reported research reviews the general nature of voice quality; its acoustical, physiological, and perceptual correlates; and methods of voice quality analysis. Research concerned with hoarse and rough voice quality is given special consideration.

Voice Quality: A Potpourri

The sensory experience of listening to a voice is determined largely by the simultaneous summation of many features such as pitch, loudness, and complex time relationships. A single dimension may influence perceptions more at any one time, but each contributes to the overall impression. It should be noted that learned mental "sets" may also be variables which influence perceptions. For example, people in the Midwest section of this country consistently speak with more nasal resonance than residents in other sections, this feature actually being a part of their accent. This phenomenon results in the development of a different perceptual reference "set" regarding the characteristic of resonance, which affects the way those speakers will perceive resonance in others. The same principle could hold true for other groups of people and other voice quality characteristics.

Based on one's perception of another's voice, one may make conscious or unconscious inferences about that person. Laver (1968) has

discussed the types of inferences commonly made. For example, during a telephone conversation with a previously unknown person, one might infer biological information regarding size and physique, sex and age, and general medical state. One might also infer psychological information such as personality-type and present emotional state. Further inference might regard social information such as regional origin, social status, and profession or occupation.

Voice quality is surely a complicated matter, considered from both the standpoints of production and perception. Part of the complexity is due to the fact that each person has a highly individualistic voice quality. Because of this uniqueness, one can recognize acquaintances almost as readily auditorially as visually; and police make identifications by "voice prints," graphic representations of the speech spectrum. On the other hand, each person can voluntarily, or by habit, modulate his voice in ways to express affective meaning and emotion in accordance with learned cultural associations. One may observe there are many ways in which voice quality is individualistic and others in which voice quality tends to follow cultural norms.

Sources of Voice Quality Variability

Laver (1968) pointed out there are basically two sources of variation of voice quality. First is the anatomical and physiological foundation of a speaker's equipment "which determines the width of the potential range of operation for any voice quality feature." These determinants are based on heredity, sex, age, and health, and as such are beyond the speaker's control. An example is that the longer and

thicker a speaker's vocal folds, the lower his pitch will be (Zemlin, 1968).

The second source of variation is the long-term habitual muscular adjustments or settings of the speaker's larynx and supralaryngeal tract which restrict voice quality features to a more limited range of operation. These are more directly accessible to speaker control, although particular speaking habits may be so firmly established that the speaker is unaware of them. Laver (1968) divided the learned "settings" of the larynx into three subcategories: phonation types (such as breathy, falsetto, and ventricular), pitch ranges, and loudness ranges. For settings of the supralaryngeal vocal tract, four subcategories were listed (Laver, 1968): modification of the longitudinal dimension (dependent on jaw and tongue position), modification of the latitudinal dimension (dependent on width of the pharynx and oral cavity), modification of the muscular tension of the pharyngeal walls and tongue, and nasalization. These supralaryngeal features are involved in resonance modifications of the emitted laryngeal tone; whereas, the settings of the larynx influence the production of the laryngeal tone initially.

Hecker (1971) essentially presented the same classification system for general sources of speaker variability, although his labels for the categories were "organic" and "learned" differences.

Acoustical, Physiological, and Perceptual
Nature of Voice

In order to understand the acoustical, physiological, and perceptual aspects of vocal production, a brief review of speech sound production is presented below.

Sound Production and Transmission

The transmission of sound through any medium is a physical phenomenon involving wave motion. In human voice production, the medium for sound transmission is simply the air being exhaled from the lungs, in a modification of the process of respiration. Molecules of this air are set into vibratory motion by the vibrating vocal folds at the level of the larynx. According to the myoelastic-aerodynamic theory, the vocal folds are caused to vibrate by the air stream being emitted from the lungs and trachea (Zemlin, 1968). Both vocal folds vibrate in synchrony in an undulating wave-like motion along both horizontal and vertical planes, approximating along their free medial margins in the closed phase of vibration. The rapidity with which the folds execute a complete to and fro motion is referred to as the frequency of vibration and is calculated in cycles per second, usually called Hertz (Hz). The frequency of vibration determines the perceived pitch; and the greater the frequency, the higher the perceived pitch. The degree of the excursion of the vibrating vocal fold (distance moved from an original position of rest) is called the amplitude and is determined by the amount of energy expended. The amplitude determines the perceived intensity, or loudness of the sound; and the

greater the amplitude, the greater the perceived loudness.

To explain cyclical vibratory motion more completely, the course of one complete vibration or cycle will be described in detail. This motion is illustrated graphically in Figure 1. When a molecule (representative of any mass) is set in motion by a force impinging upon it, it moves away from its position of rest (Point A in Figure 1) to a

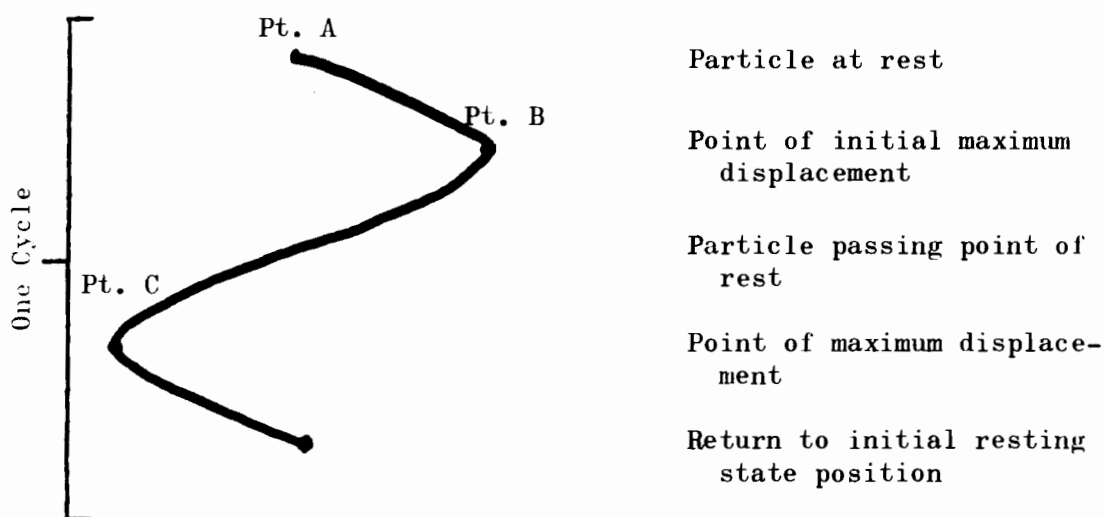


Figure 1. Vibratory particle movement which occurs in sound transmission.

degree proportionate to the amount of energy expended. In voice production, pressure (from the energy source) is determined by the rate of air flow through the larynx and the resistance applied to this exhaled air flow. After reaching its maximum point of displacement (Point B in Figure 1), the air molecule tends to "spring back" and return to its place of rest, due to the inherent elastic nature of matter. Due to another inherent property of matter, inertia, it travels beyond the previous position of rest on its return journey, being displaced in the opposite direction to the same extent as the

initial displacement (Point C). From this second displacement, the air molecule springs back again toward the "rest" position, but again inertia forces it beyond in the direction of original displacement. This cyclical action, or vibration, continues until the friction caused by collisions with other air molecules dissipates the energy to the point that there is no longer sufficient energy to sustain the motion (Denes and Pinson, 1963).

Bodies of matter generally vibrate in a complex fashion, with numerous segmental vibrations being algebraically summed to the fundamental frequency of vibration (Denes and Pinson, 1963). Perceived voice pitch is directly determined by the number of times the vocal folds vibrate each second (the fundamental frequency), but the segmental vibrations (partials or overtones) play an important part in the determination of tonal quality. The complex speech sound spectrum may be reduced by analysis to its respective frequency components of varying amplitude. Figure 2 (Denes and Pinson, 1963) illustrates how, in two instances, the same components may add together to form differing waveforms due to differences in phase relationships, referring to the time of onset of the vibration. The sound energy relationships of the two separate waveforms would be different at any one time, producing a different quality, but the perceived pitch would be the same since the perception is determined by the lowest, or fundamental, frequency.

A sound whose cycles recur with predictable regularity is called periodic. A sustained vocal tone is usually periodic and thus is perceived as having a pleasant, tonal quality. Aperiodicity (irregu-

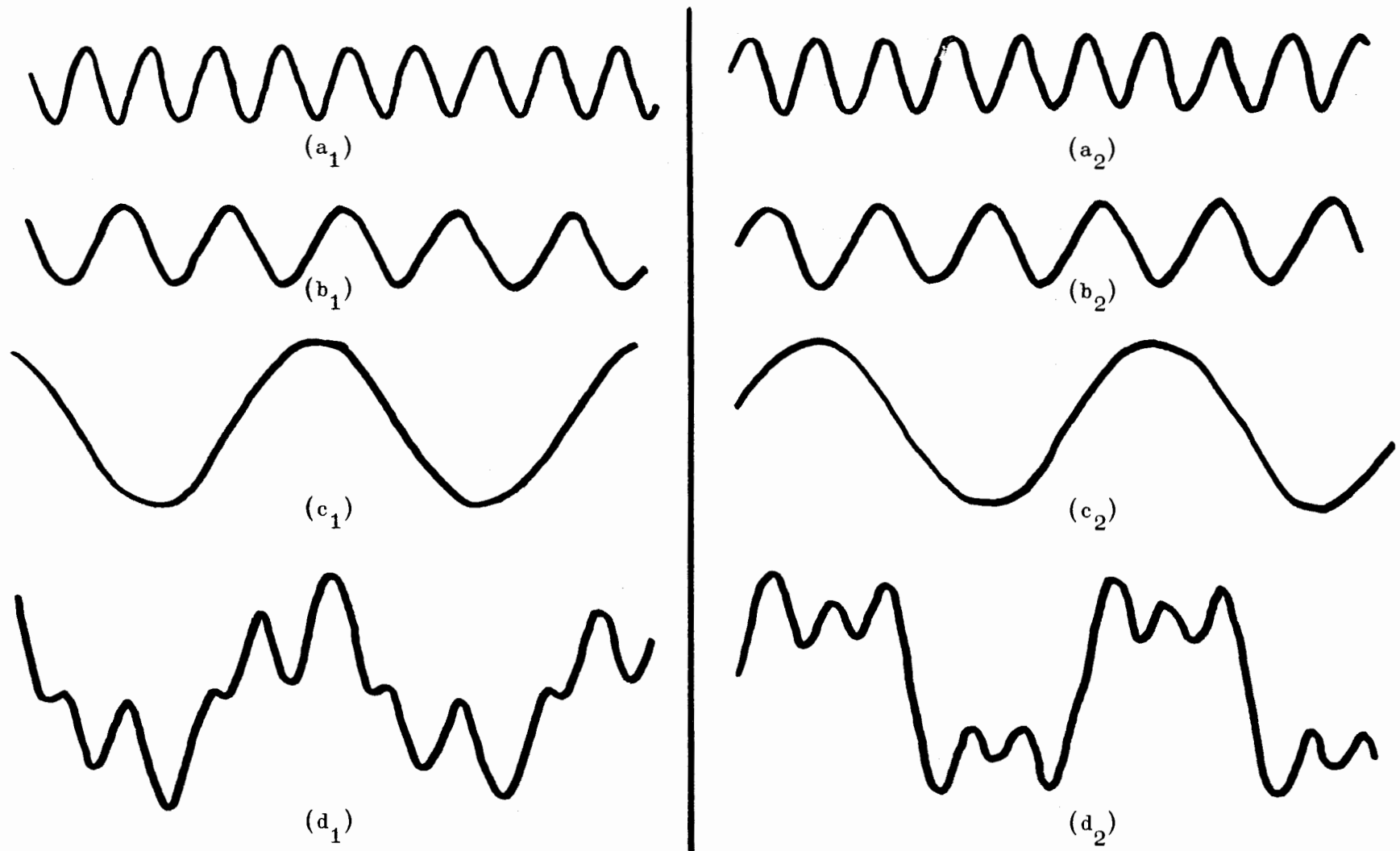


Figure 2. A comparison of complex waves (d_1 and d_2) whose shapes differ due to the phase relationships of the components (after Denes and Pinson, 1963).

lar vibration) may occur because of asynchronous action of the vocal folds or because successive vibrations differ in frequency or amplitude. A "noisiness" is then created which contributes to a perception of harshness by a listener. (See the section on hoarse voice quality which follows in this chapter.) Breathiness may also result from malfunction at the level of the larynx. If the vocal folds do not completely approximate during the closed phase of vibration, due to physical obstructions or to improper physiological action, there will be an abnormal escape of air which will be perceived as breathiness.

Resonance and the Vocal Tract Resonators

The characteristics of the vocal tone as it is generated at the larynx have been essentially described. As the tone follows the air passageway as the air is emitted, it is influenced by the supraglottal vocal tract (pharynx, oral cavity, and nasal cavity) which acts as a series of coupled resonators (Denes and Pinson, 1963). A partially enclosed space containing air acts as a resonator by serving to respond to an applied tone by reinforcing or damping it. The vocal tract has many resonators, and although the effects are ultimately combined, they act independently. Each resonator is optimally responsive to a particular frequency called its natural frequency. The vocal tract resonator's characteristics, and hence the final tone, can be changed by modifying the horizontal and/or vertical vocal tract dimensions, muscular tension of the vocal tract walls, and coupling or uncoupling of the nasal tract (Laver, 1968). The supralaryngeal structures, then, also contribute to the perception of quality.

The most common voice deviancies related to resonance are hypernasality and hyponasality (Moore, 1971). Hypernasality is the voice quality heard when sounds other than those normally nasalized are resonated in the nasal cavity. Hyponasality results from lack of nasal resonance on normally-nasalized sounds (m, n, ŋ). The common cause is obstruction of the nasal passageway, such as by pathological growths or mucous congestion.

Analysis of Voice Quality

Various ways of describing and differentiating voice quality have evolved from emphasizing different aspects of vocal production. A review of subjective methods follows.

Perkins (1971) suggested that multitudinous terms have arisen in attempts to describe voice quality because it is not an entity which lies along a simple continuum, as do pitch and loudness. He employed four behavioral dimensions for specifying nonlinguistic quality. The behavioral dimensions, apparently referring to physiological function are voicing, vocal constriction (of both laryngeal and supraglottic areas), vocal mode (light, heavy, and pulsated), and vocal focus (oral, head, or chest) (Perkins, 1971).

Wilson (1971) has also proposed a subjective method of vocal analysis, based upon a listener's perception of the voice and with reference to a physiological model (Wilson, 1971). The Jewish Hospital Voice Profile (JHVP) provides for rating voices on the following dimensions: laryngeal openness, pitch, resonance, intensity (loudness), and use of vocal variety. A rating along a continuum is assigned

separately for each of these dimensions. Wilson maintains that inferences about physiological function can be made on the basis of these ratings, and in fact, in his training tape (1972) which provides examples of each rating, describes correlations between each rating and its assumed physiological referent. As this rating form was selected as a tool of evaluation for this investigation, it will be described in detail here. The adaptation of the JHVP form used in this investigation may be found in Figure 3. The measurement of sustained phonation as found on the original form was omitted, but it was included on the Subject Information Form (Appendix A).

Wilson has described his rating form in an audio tape recording which includes examples of each disorder. The continuum for laryngeal cavity openness extends from a -4 (open) to a +3 (closed). A -4 indicates open laryngeal function, i.e., there is no closure of the vocal folds to prevent a free flow of air from the lungs. The only sound heard is the noise produced by forcing air against the articulators. A -3 indicates the folds are slightly abducted but they do not approximate at midline. In this case a whisper is heard. A -2 is assigned when the vocal folds have been set into a vibrating pattern in which less time is spent in the approximation phase of vibration than in the opening and closing phases. Breathiness is then perceived.

A rating of 1 represents normal laryngeal function; appropriate time is spent in the opening, closing, and closed phases of vocal fold vibration. The +2 rating indicates excessive time is being spent in the closed phase of vibration due to excessive tension. The listener has the impression that there is great tension in the area of the

VOICE PROFILE

NAME: _____ AGE: Between 18 and 30 years SEX: F DATE: 5-75

Constant _____ Variable _____ Voice Rating: 1 2 3 4 5 6 7

LARYNGEAL CAVITY			RESONATING CAVITY			INTENSITY		
PITCH			NASALITY					
high			hypernasal			-2	1	+2
B			C			soft		loud
+3			+4					
+2			+3					
			+2			VOCAL RANGE		
A open	-4	-3 -2 1 +2 +3 closed	1			-2	1	+2
			-2			monotone		variable
			hyponasal					pitch
			low					

Comments: _____

X = Primary feature / = Secondary feature / = Noted feature

int. = Intermittent feature

Figure 3. Jewish Hospital Voice Profile (from Wilson, 1972).

vocal folds, and a tendency to increase loudness may also be noted. A +3 rating is used to describe pathological function in which the vocal folds "come together in a tight closure and are blown rapidly apart in a rather explosive fashion." The listener perceives the extreme laryngeal tension which prohibits an ongoing flow of vibration and often hears excessive breathiness following the sudden explosion of the tightly abducted vocal folds.

A rating of +2, -2 indicates hoarseness, a condition in which breathiness and tension are both present in measurable degree. In fact, any combination of tense and breathy components will contribute to some perception of hoarseness.

The continuum for pitch extends from a -3 to a +3. A rating of 1 indicates normal habitual pitch level. Ratings of +2 and -2 represent inappropriately high and low pitch levels, respectively, although the deviation is not great enough for the speaker to lose sexual identity. The +3 and -3 ratings apply in instances where the deviation of pitch level (too high or too low, respectively) is so exaggerated that the speaker does in fact lose sexual identity. The voice, at this point, has become "socially demeaning."

The continuum of nasalization ranges from a -2 to a +4. Wilson stated the balance between nasality and orality is actually what is being assessed. A -2 resonance rating refers to an inability of the vocalized air to travel through the nasal passage. The speaker will probably sound as if his nasal passageway is blocked by congestion. Normal balance between nasality and orality receives a rating of 1.

A +2 rating is applied to instances of "assimilation nasality" in which vowels are inappropriately nasalized when they immediately precede and/or follow a normally-nasalized sound (m, n, ŋ). A rating of +3 indicates consistent nasalization of all of the vowel sounds, but the consonants are reasonably well articulated. In comparison, a +4 rating indicates the consonant sounds as well as the vowels are being nasalized, causing them to be distorted. Nasal escape of air may be heard. The velum may be wholly or partly absent, or it may have insufficient musculature to attain adequate approximation to the posterior pharyngeal wall.

A three-point continuum is used for the dimension of intensity. A voice which is perceived as too soft receives a -2 and a voice that is perceived as too loud receives a +2. Normal intensity is represented by a rating of 1.

Vocal range, also summarized by a three-point continuum, reflects the subjective impression of the speaker's use of his or her pitch range. A -2 indicates that too little variety is used, with the consequent perception of a monotonous voice. Normal variety is expressed as 1, and a +2 rating is assigned when there is uncontrolled pitch variation or overexaggerated use of pitch variety.

On the JHVP form there is a provision for indicating whether the predominant quality as rated occurred on a constant or variable (intermittent) basis. There is also a severity rating scale for assessing relative severity of any deviation detected in the overall analysis. The severity scale, as employed in this investigation, has seven points, where 1 is least severe and 7 is most severe.

Reverse Playback of Audio Tape Recordings

A method known as backward or reverse playback of audio tape recordings has been employed by some investigators as a type of experimental control in studies of voice quality judgments. The semantic content of the speech is lost since the words are rendered unintelligible. Similarly, any distinctive manner of articulation, including dialect, would presumably be eliminated.

Sherman (1954) was an early evaluator of the method of performing voice quality analyses using audio recorded speech heard by reverse playback. She reported that scale value judgments of both nasality and harshness appeared to be more valid than judgments made on speech played forwards. Manning (1957) found ratings of nasality tended to be more severe when speech stimuli were played backwards, but the relative severity of the compared speech samples remained the same.

It must be noted that the question of how reverse playback may affect voice quality has not been fully explored experimentally. It can be observed that the altered time relationships created by backwards playing of a series of transient sounds (such as speech) affect the quality. Risset and Matthews (1969) noted that piano music played backwards has a different quality than when played forwards. They suggested that backward-played speech may likewise be more distant from the expected sound than would be generally assumed on the basis of simple spectral composition. Nevertheless, this method appeared to offer the most viable solution for the control of the variable of dialectical differences between black and white speakers in the present

investigation.

Hoarse Voice Quality

Perceptual Definition of Hoarseness

In the past, there have not been universally accepted definitions of voice disorders; the situation has been characterized by the use of many terms which have not denoted the same meanings to all, such as "breathy," "husky," "harsh," "hoarse," "throaty," and "metallic" (Moore, 1971). In fact, with specific reference to hoarseness, Palmer (1956) has noted there have been more discussion and debate around the problem of definition and description of this classification than around the etiology and intervention. He reported that the answers regarding definition have centered around "acoustical terminology, physiological, neurological, mechanical, and psychological." Many common observations have been reported which contribute to a composite sketch of hoarseness.

Darley (1964) indicated hoarseness is a combination of the characteristics of harshness and breathiness. Boone (1971), in discussing these components, explained that harshness is caused by an overadduction of the vocal folds (probably creating an aperiodicity of laryngeal vibration), while breathiness is caused by "an audible escape of air as the approximating edges of the glottis fail to make optimum contact." Fairbanks (1960) noted the harsh element predominates in some voices, while the breathy element is predominant in others. Similarly to Darley, Moore (1971) observed that hoarseness had both components of roughness and noisiness. The term "roughness"

has been used to classify certain voice quality aberrations, including hoarseness (Third Regional Workshop on the Rehabilitation Codes and Communicative Disorders, 1967). "Harshness" appears to remain prevalent in the area of speech pathology, whereas the term "roughness" was found in the literature of acousticians.

The Physiology of Hoarseness

Several investigators have contributed to an understanding of what occurs physiologically during hoarse voice production. Isshiki and von Leden (1964) reported the noise component of a hoarse voice is evidently attributable to "imperfect modulation of the air stream at the glottis." Palmer (1956) indicated that when hoarse voice quality is produced, the laryngeal mechanism has failed, to some extent, in the approximation, tension, or vibration of the vocal cords.

Normal approximation of the vocal folds can be impeded by the presence of swelling or tumors (Moore, 1971). According to Moore (1971), abnormal tension may be produced in an effort to force the pitch level lower than normal, a condition commonly associated with harshness. It has also been suggested that laryngeal hypertension when concomitant with hoarseness is used as a compensatory effort to attain adequate vocal fold approximation in the presence of a pathological condition (Isshiki and von Leden, 1964).

Regarding the relationship of vibration patterns and hoarseness, Moore (1971) suggested that two vocal folds vibrating at different rates may have the effect of generating irregularly spaced pulses which create a vocal roughness. Coleman (1971) expanded on the matter

of vibratory patterns by stating that during vocal roughness a rapidly changing open-closure relationship at the glottis, in addition to asynchrony or phase differences in the vibratory patterns of the folds, are commonly observed. Paralysis could affect normal approximation, tension, or vibration of the vocal folds. There are, then, various types of interferences which may occur to contribute to hoarse voice quality production.

Relationship of Hoarseness to Pathology

Moore (1971) stated that hoarseness "always means that the vocal cords are not functioning normally." Hoarseness usually results from laryngeal pathology, according to Darley (1964). In fact, as Isshiki and von Leden (1964) pointed out, hoarseness is often the only symptom of laryngeal disease. The hoarseness occurs because of imperfect closure of the glottis due to edema, vocal nodules, contact ulcers, and other laryngeal lesions which have developed on the margins of the vocal folds as a result of vocal abuse (Isshiki and von Leden, 1964).

Measurement of Hoarseness

Isshiki and von Leden (1964) listed three types of investigations which have been attempted to measure hoarseness qualitatively: physiological, visual, and acoustic.

Physiological Measurements. Isshiki and von Leden (1964) reported results of one type of physiological measurement, i.e., aerodynamic analysis, in which they measured the rate of emitted air flow during phonation. In patients with hoarse, loud voices, they found

that a low flow rate was indicative of high glottal resistance or muscular hyperfunction. This finding indicated the measurement is a valuable adjunct to other physiological studies for an understanding of laryngeal behavior.

Visual Measurements. Visual inspection of the larynx can be accomplished through direct or indirect laryngoscopy (Darley, 1964). In this manner, it is possible to visually search the laryngeal area for pathological lesions which may be contributing to hoarseness.

Acoustical Measurements. Wendahl (1963) reported that a review of earlier investigations had shown general agreement that harsh voice quality is the result of aperiodic vibration of the vocal folds. Lieberman (1963) analyzed "pitch perturbations" which he described as small but rapid variations in the fundamental periodicity (frequency) of normal connected speech. Such pitch perturbations may reflect "variations in the glottal periodicity, variations in the glottal waveshape, or changes in the vocal tract configuration." The relative duration of the perturbations served to differentiate which factor was at variance. Lieberman suggested the magnitude of the pitch perturbation factor might be useful in the detection of laryngeal disease since the differences between the perturbation factors of the normal and pathologic larynges are proportional to the size of pathological growths.

Nearly ten years later, Iwata (1972) reported further instrumental sophistication had led to the development of correlograms, i.e., graphic representations of the pitch perturbations for successive fun-

damental periods. The correlograms indicated different patterns for normal voices as compared to voices of patients with various laryngeal pathologies, such as chronic laryngitis, neoplasms, and paresis.

Iwata suggested correlograms may be able to provide an objective measure for diverse groups of laryngeal diseases.

In his studies Wendahl (1963, 1966a, 1966b) found perceptions of harshness were related to small cycle-to-cycle variations of both pitch and loudness. Coleman (1971) later found that varying the shape of the sound wave (corresponding to the open-closed relationship of the vocal folds) contributed to perceptions of roughness. Employing an electrical laryngeal analog machine, referred to as LADIC, some acousticians have artificially simulated speech. Each sound wave cycle, when represented successive glottal excitations (or vibrations), was subject to control. Wendahl (1963) conducted an experiment using frequency variations of from 1 to 10 cycles per second imposed on a median frequency. The term employed to describe this acoustical phenomenon is "jitter." Listener judgments of the degree of perceived harshness were found to be directly related to the degree of variance from the median frequency, i.e., the greater the imposed frequency variations, the greater the perceived harshness. Coleman (1969) found that when the same frequency variations were imposed on median tones of different frequencies, the stimuli of the lower median frequency were judged to be rougher. This conclusion earlier had been predicted by Wendahl (1963) and would tend to suggest that people with lower pitched voices will be judged as rougher when compared to other voices having equal frequency variations. Wendahl (1966a) later demonstrated that

the magnitude of the jitter products in an auditory signal is also directly related to listener perceptions of harshness, i.e., the louder the jitter stimuli, the greater the perceived hoarseness.

Wendahl (1966a, 1966b) also studied the effect of applying random amplitude (loudness) variations to successive glottal impulses, as simulated by LADIC. It was found that the effect of this distortion of the auditory signal (referred to as "shimmer") could also be scaled on a dimension of auditory roughness in a systematic relationship.

Coleman (1971) was interested in studying the effect of waveform variations of adjacent glottal impulses upon perceptions of roughness. Employing LADIC, he found that such variations, obtained by varying the duty cycle (simulating the glottal open quotient), did have a direct positive relationship upon perceptions of roughness. Perceptions of roughness were even greater when the waveform changes were superimposed on a standard jitter program.

The subject of spectral noise levels and their relationship to perceptions of roughness was studied by Emanuel and his associates in several investigations. Initially, Emanuel and Sansone (1969) found, through sound spectrum analyses, that noise levels were greater in simulated rough vowels than in normally-produced vowels. Sansone and Emanuel (1970) later found evidence suggesting the perception of roughness may be predicated primarily on the relative amplitude of the harmonic and inharmonic (noise) components. Elevated spectral noise levels and diminished harmonic components were associated with an increase in perceived roughness. Lively and Emanuel (1970) expressed the relationship between harmonic and inharmonic spectral energy by

the ratio H/I. It was hypothesized that this relationship is inversely related to perceived vowel roughness.

The findings of Emanuel, Lively, and McCoy (1973) reiterated that the vowel spectral noise levels and mean roughness ratings are highly related, both when men's and women's voices were judged together and when judged separately. A psychological variable, however, presumably was responsible for the fact that some male productions of vowels tended to be rated slightly rougher when judged with the productions of women than when judged separately.

Black-White Comparisons of Speech, Language, and Voice

Much has appeared in the literature during the past ten years regarding black dialect. Most of the discussions have included, in one form or another, comparisons of black speech to standard American English. Two main positions regarding the historical relationships between the two language systems have arisen. The first assumes they both were derived from a British dialect (McDavid and McDavid, 1951), while the second assumes Black dialect is derived from a creole-based system more like the Caribbean creoles (Bailey, 1965). According to Baratz (1973), regardless of the origin of differences between the two language systems, a child will learn to speak the language of the people around him, and both black and white children learn to speak according to well-ordered, highly-structured, and highly-developed language systems. In other words, as Burling (1970) has stated, it is early childhood experiences and not race or ethnicity per se, which determines how a child will speak. An illustration of this point was

provided by an informal study which Labov reported in 1973. He recorded the speech of Blacks raised in middle-class white neighborhoods and Whites raised in southern black areas and asked listeners to identify whether the speakers were Blacks or Whites. The result was that none of the judges reliably classified the speakers according to race.

Differences between the grammatical constructions and pronunciation of black and white speakers have been documented by Labov (1973). Differences have also been noted in the suprasegmental events of speech such as tonal inflection, prosody (rhythm), and loudness. Loman (in von Raffler Engel, 1972) has found that Black English has "sharper pitch variations" than Standard English. In a limited investigation of black and white children in a southern city, von Raffler Engel (1972) found time-relationship differences in one specific consonant-vowel combination. The consonant/vowel ratio, based on duration in milliseconds, was 0.7285 for the black subjects and 0.1481 for the white subjects. In other words, the consonant was of shorter duration than the vowel in both instances, but it was considerably longer for the black speakers than for the white speakers. Another observation arising from the same study was that the rate of black speech was slightly slower in the sample (von Raffler Engel, 1972).

There apparently have been relatively few investigations regarding black-white differences of the phonological aspects of speech, and none were found to be reported with reference to voice quality comparisons per se.

CHAPTER III

METHODS AND PROCEDURES

Selection of Participants

Speaker-Subjects

Twenty-five black women comprised the experimental group and twenty-five white women served as control subjects in this investigation. All were students at Portland State University during Winter Term, 1975. The chronological ages of the subjects ranged from eighteen to thirty years. Each subject indicated her voice was "normal," i.e., representative of her everyday voice. In this way, women with temporary vocal abnormalities due to the "common cold" or other factors were excluded from the study. The subjects were randomly selected from the halls and dining room in Smith Memorial Center at Portland State University during school hours and asked to participate voluntarily in a thesis study in the Speech and Hearing Sciences Program.

Judges

Ten females who had completed a graduate seminar in voice disorders and had participated as clinicians in the Voice Clinic at Portland State University within the preceding two years served as judges for the voice analyses. All were graduate students in speech pathology with the exception of two who were public school speech pathologists; and all were white. The judges had been introduced during the voice

disorders seminar and clinic to the use of the Jewish Hospital Voice Profile (JHVP) (Wilson, 1971). The judges were unaware of the nature of the investigation prior to their participation.

Data Collection

Physical Setting

The participation of the voice subjects occurred in two separate rooms in Smith Memorial Center at Portland State University. Both rooms were carpeted, and each housed a piano which was required to perform the pitch analyses. The rooms were relatively well acoustically insulated so that there was minimal sound interference during collection of the data.

Recording of Personal Information

By an interview procedure the investigator completed Part A of the "Subject Information Form" shown in Appendix A. Each subject's name, sex, age, birthdate, address, and ethnic classification were recorded and a number was assigned to her, according to sequential order of participation.

Audio-Recording of Voice Samples

Following the recording of personal information, each subject was instructed to read a standard passage in a book while being audio-recorded. A copy of the passage read appears in Appendix B. It was explained to each subject that she would read for forty-five seconds, at which time the investigator would signal her to stop.

Each subject read the passage while seated at a table on which a Unidyne III microphone (Model 545) was set at a 70° angle, approximately seven inches from the mouth. The microphone was connected to an Ampex magnetic dual-channel tape recorder (Model AG-500) which recorded the voice samples at a speed of fifteen inches per second for later analysis by the judges.

Recording of Voice-Related Information

Length of Sustained Phonation. This measurement from Wilson's JHVP form (1972) was included on the "Subject Information Form" as part of the indexical information obtained by the examiner. After completion of the audio recording, the investigator requested the subjects to "sing or say 'ah' at a comfortable pitch for as long as possible." The durations of three separate trials were recorded and the longest duration was noted for subsequent consideration. Times were recorded to the nearest second according to stop watch calculations.

Pitch Analyses. Following the recording of durations for sustained tones, the investigator conducted a pitch analysis (Fairbanks, 1960) for each subject with the aid of a piano. The pitch analysis form is shown in Appendix C. The Fairbanks pitch analysis yields a measurement of an individual's pitch range and estimates of both optimum and modal pitch levels. Reliability of the investigator's ability in conducting the pitch analysis was established prior to the data collection by an inter-judge agreement score of 80 per cent, obtained when independent analyses of five individuals by the investigator and a Portland State University Voice Clinic supervisor were

made. Sequential numbers were assigned to the semitones of the ascending pitch scale so that the pitch data could be presented and treated numerically; the key is presented in Appendix D.

Audio Tape Recordings

Reverse Audio Tape Recordings

Three audio tapes were prepared, one training tape and two tapes used for voice judgments. The recorded voice samples of the subjects were re-recorded as played backwards for the purpose of eliminating dialectical differences between the speakers. The reverse-recording of the voice samples was completed by a technician at the Portland State University Audio-Visual Department after the samples were arranged in the proper sequence. The tapes were recorded in their final form at $3 \frac{3}{4}$ inches per second.

Reversed playback of the audio tape recordings, a methodology described in Chapter II, was employed in this investigation. It is assumed that as the results of Manning's investigation (1957) indicated, in such a procedure all voices undergo the same relative changes and intergroup comparisons are, therefore, valid.

Training Tape

A training tape for the judges was prepared as a review of the rating procedures of the JHVP, as well as an auditory reference for backward-played speech. The training tape consisted of paraphrasing Wilson's (1972) narrative on the training tape which was prepared by him to provide voice clinicians with instructions and criteria for

rating voices on the Voice Profile. Additionally, examples of simulated voice disorders from his tape were re-recorded as they were backward-played and then incorporated into the training tape. The training tape was twenty-eight minutes in duration.

Tape I

The first tape prepared for listener analysis consisted of forty-five-second voice samples of all subjects, arranged in a randomized order determined by a table of random numbers. Each sample was identified by the number originally assigned to the respective speaker. The fifty voice samples were divided into ten groups of five voices each and were labelled groups 1 through 10. In order to assess intra-judge reliability, group 7 was repeated and labelled group 11, resulting in fifty-five total voice samples. The speakers were identified on the tape by a narrator. There was approximately five seconds of silence between speakers.

Prefacing the series of backward-played voices, brief recorded instructions were read to the judges. A transcription of these instructions is included in Appendix E. The duration of Tape I was fifty-eight minutes.

Tape II

Tape II was composed of seven-second segments of the fifty backward-played speech samples of Tape I. The order of the fifty voice samples was determined by using a random number table. Ten of the samples, randomly chosen using the random number table, were repeated as numbers 51 through 60 to assess intra-judge reliability.

Approximately a two-second pause between speakers was made.

Verbal directions to the judges were included at the beginning of this recording; a transcription of them may be found in Appendix E. The tape was twenty-eight minutes in duration.

Judging the Voice Tapes

Directions to Judges

Prior to their participation, the judges were given brief written instructions of the procedure to follow when listening to all the tapes. The judges were requested to listen to the Training Tape before beginning the voice analyses. They were then instructed to listen to the tapes in sequential order, i.e., Tapes I and II.

The judges listened to the tapes alone and at their own convenience, using a portable Sony reel-to-reel tape recorder for playback. When judging Tape I, each judge was requested to listen to the groups in a specific order that differed from every other judge in order to control for order effect.

Rating Forms

The judges used two separate rating forms, one for each of the respective experimental tapes. The JHVP was used to evaluate the forty-five-second voice samples of Tape I (Figure 3).

A separate form was provided for rating Tape II (ethnic classification). A copy of the form may be found in Appendix F.

Data Analysis

Each voice was submitted to 1) indexical measures of phonation and pitch, conducted by the investigator, 2) scaled value rating analysis of six parameters of voice production during speech, conducted by ten trained voice judges, and 3) a forced classification of black or white ethnic identity by the ten judges.

A two-tailed t-test for independent means was employed to determine if significant differences were noted between the black and white women speakers regarding the measures of pitch and phonation (Walpole, 1968).

The scaled value rating analyses (JHVP) were submitted standard test for the differences between proportions for comparison of the two speaker groups.

A one-tailed test of a single proportion was employed to determine if voices could reliably be classified as "black" or "white" at a greater than chance level.

An assessment of intra-judge reliability on the JHVP ratings was conducted by a test of difference between proportions. Inter-judge reliability for these ratings was assessed by an analysis of variance. The percentage of agreement between judges was also determined for the parameter of resonance on the JHVP.

A standard test of proportions was employed to assess intra-judge reliability for the task of choosing the ethnic identity of the speakers.

CHAPTER IV

RESULTS AND DISCUSSION

Results

The purpose of this investigation was to compare various aspects of the voice quality of black and white women. One question of particular interest was whether black women tend to have hoarse voices more often than white women. Accordingly, certain indexical information regarding phonatory attributes was collected by the investigator, and two separate experiments involving analysis of backward-played audio recorded voices by ten judges were completed. Group (black and white) comparisons were made of the data as well as tests of inter- and intra-judge reliability. These results are presented below.

Indexical Information

The investigator completed the Subject Information Form (Appendix A) and the Fairbanks Musical Scale Pitch Level Comparison (Appendix C). The variables submitted to evaluation were: 1) age, 2) length of longest sustained phonation, 3) optimum pitch, 4) modal pitch, 5) relationship of modal pitch to optimum pitch, 6) lowest possible pitch, 7) highest possible pitch, and 8) pitch range in semitones. For each variable, no statistically significant differences were observed between the black and white women when two-tailed t-tests for independent means and F-tests for differences in variance were

conducted (Walpole, 1968). The data are presented in Table I.

Tape 1

The data for each of the six parameters rated by judges on the Jewish Hospital Voice Profile (Wilson, 1971 and 1972) were submitted to standard tests for the differences between proportions (Walpole, 1968).

Intra-Judge Reliability on the JHVP. In an assessment of intra-judge test-retest reliability on five subjects, 205 of 300 possible choices were found to be in agreement (68.33 per cent). A test of difference of proportions was made between the hypothetical chance level (150 correct of 300 possible choices) and the observed level of agreement. The resultant z-score was 3.60 (statistically significant at the .001 level), indicating good intra-judge reliability. See Table II.

Inter-Judge Reliability. An analysis of variance was done treating the judges as the independent variable (Hays, 1973). The resultant F-value of 3.256 is significant at the .025 level, indicating poor inter-judge reliability. See Tables III and IV.

Because there were conflicting judgments on the scale of resonance (i.e., the same voice was sometimes judged hypernasal by most judges but hyponasal by one or two judges) further analysis of inter-judge reliability was done on that particular parameter. Table V presents the percentages of agreement among all judges. Of 5000 possible agreements for ratings, 2152 were made, resulting in 43.04 per cent agreement, indicating the inter-judge reliability was poor.

TABLE I

BLACK-WHITE GROUP DIFFERENCES ON INDEXICAL INFORMATION

Variable	Mean		Difference of means	<u>t</u> -score	<u>p</u>	Standard Deviation		<u>F</u> -score	<u>p</u>
	Black	White				Black	White		
Age	21.60	21.28	.32	-.0513	NS	3.12	1.97	.39	NS
Length of longest phonation (seconds)	15.60	20.76	-5.16	.9348	NS	6.40	6.78	.89	NS
Optimum pitch	16.00	16.76	-.76	.1598	NS	2.16	1.45	.0009	NS
Modal pitch	15.00	5.44	-.44	.0994	NS	2.000	1.76	.77	NS
Relationship of modal to optimum	-.88	-1.32	.44	-.8854	NS	1.53	1.80	.73	NS
Lowest pitch	7.32	8.08	-.76	.3299	NS	2.29	2.04	.79	NS
Highest pitch	41.76	43.32	-1.56	.1264	NS	4.06	4.62	.77	NS
Pitch range	34.44	35.24	-.80	.0768	NS	4.69	5.83	.64	NS

TABLE II
 INTRA-JUDGE RELIABILITY ON THE JEWISH HOSPITAL VOICE PROFILE

Speaker number	Percentage of times judges agreed with own judgments					
	Laryngeal openness	Pitch	Resonance	Intensity	Vocal range	Severity rating
13	7	10	5	10	10	3
20	5	6	5	8	9	3
21	7	9	7	10	10	6
22	3	8	5	7	10	1
50	7	9	6	7	10	2
Totals	29	42	28	42	49	15 = 20 ₅

$205/300 = 68.33$ per cent agreement

$z = 3.6$ ($p < .001$)

TABLE III
 INSTANCES OF JUDGMENTS OF NORMALCY
 (RATING OF "1")

Vocal parameter	Judge number									
	1	2	3	4	5	6	7	8	9	10
Laryngeal openness	16	4	7	23	46	2	3	14	9	7
Pitch	39	35	24	37	32	18	18	30	38	42
Resonance	24	42	33	33	30	13	25	39	29	28
Intensity	50	48	48	48	47	46	30	47	50	49
Pitch range	48	46	46	50	41	31	29	50	49	50

TABLE IV
INTER-JUDGE RELIABILITY ON THE JEWISH HOSPITAL
VOICE PROFILE: ANALYSIS OF VARIANCE

Source	Sum of squares	df	Mean square	F-score	P
Total	10832.8	49			
Variable	7313.3	4			
Judges	1580.0	9	175.5	3.256	.025
Error	1939.5	36	43.9		

TABLE V
 INTER-JUDGE AGREEMENT ON THE SCALE
 OF RESONANCE (JHVP)

Judge Number	Percent of times judges were in agreement								
	10	9	8	7	6	5	4	3	2
1	54	32	56	56	44	32	42	50	50
2	68	50	78	50	28	56	66	60	
3	66	44	62	54	40	54	58		
4	52	42	52	42	26	54			
5	50	54	50	36	22				
6	34	20	34	36					
7	60	34	42						
8	60	52							
9	50								
10									

Laryngeal Cavity. The -4 to +3 rating scale was collapsed for data analysis to a -2 to +2 scale, so that -3's were treated as -2 ratings and +3's were treated as +2 ratings. A combination +2,-2 rating was treated as a separate rating, indicative of hoarseness. The results are presented in Table VI. The white speakers received significantly more judgments of -2 ($p < .01$). Other differences between proportions, including +2,-2 (hoarseness), were not significant.

Pitch. Results of rating the -3 (low) to +3 (high) pitch scale are presented in Table VI. Black women were more often judged to speak at abnormally low pitch levels than white women ($p < .0002$ for -3

TABLE VI
RATINGS ON THE JEWISH HOSPITAL VOICE PROFILE

Variable	Number of judgments		Probability of the judgment		Difference of Probabilities	z-score	Level of Significance
	Black	White	Black	White			
<u>A. Laryngeal openness</u>							
+2, -2	128	127	.512	.508	.004	.09	(NS)
-2	7	18	.028	.072	-.044	-2.26	.01
1	63	69	.252	.276	-.024	.61	(NS)
+2	52	36	.208	.144	.064	.59	(NS)
<u>B. Pitch</u>							
-3	25	1	.10	.004	.096	3.52	.0002
-2	76	51	.304	.204	.100	2.57	.005
1	135	181	.540	.724	-.184	-4.27	.001
+2	14	17	.056	.068	-.012	-.56	.29 (NS)
+3	0	0	0	0	0	0	-
<u>C. Resonance</u>							
-2	13	1	.052	.004	.048	3.21	.0007
1	133	173	.532	.692	-.160	-3.67	.0001
+2	61	47	.244	.188	.056	1.52	.06 (NS)
+3	36	28	.144	.112	.032	1.07	.14 (NS)
+4	7	1	.028	.004	.024	1.52	.06 (NS)
<u>D. Intensity</u>							
-2	2	14	.008	.056	-.048	-3.05	.0012
1	234	225	.936	.900	.036	1.47	.0708 (NS)
+2	14	11	.056	.044	.012	.62	(NS)
<u>E. Vocal range</u>							
-2	27	21	.108	.084	.024	.28	.3900 (NS)
1	221	229	.884	.916	.032	1.13	.1292 (NS)
+2	2	0	.008	0	.008	-	-

ratings, $p < .005$ for -2 ratings). White women were found more likely to be rated as normal for this parameter ($p < .0001$). There were no other statistically significant differences between the two groups.

Resonating Cavity. Voice ratings on the -2 (hyponasal) to +4 (hypernasal) resonance scale indicated black women were more likely to be judged as hyponasal than white women ($p < .001$) and white women were more likely to be judged normal ($p < .0001$). The results are presented in Table VI. There was also some evidence, although not statistically significant, that the black women were more likely to be judged hypernasal (+2 and +4 ratings).

Intensity. Results of the ratings of intensity (loudness) of the speakers are presented in Table VI. White women were more likely to be given a -2 (too soft) rating than were black women ($p < .005$). There were no other significant results.

Vocal Range. Results of the ratings of vocal range (variety during speech) are presented in Table VI. There were no significant differences between the two groups.

Severity Rating. Analysis of overall severity ratings assigned to all speakers (0 representing no disorder observed) yielded inconsistent results by the difference between proportions test (Walpole, 1968). The results are presented in Table VII.

Two separate comparisons of ranges into which severity ratings could fall were done. The first consideration was whether the severity ratings indicated "clinical significance" according to criteria used by Laskey (1975). The criteria she established designated ratings of 0 through 3 as nonsignificant clinically and ratings of 4 through 7 as

TABLE VII
SEVERITY RATINGS ON THE JEWISH HOSPITAL VOICE PROFILE

Variable	Number of judgments		Probability of the judgment		Difference of Probabilities	<u>z</u> -score	Level of Significance
	Black	White	Black	White			
<u>Severity Rating</u>							
0	16	35	.064	.140	-.076	-2.81	.0025
1	41	45	.164	.180	.016	.47	.31 (NS)
2	54	73	.216	.292	.076	1.95	.0256 (NS)
3	74	47	.296	.188	-.108	-2.82	.0024
4	44	36	.176	.144	-.032	-.98	.1635 (NS)
5	17	10	.068	.040	-.028	-1.38	.0838 (NS)
6	3	4	.012	.016	-.004	1.52	.0643 (NS)
7	1	0	.004	-	-.004	-	-

clinically significant. A standard test for the difference between proportions (Walpole, 1968) yielded a z-score of 1.59, indicating no statistically significant difference between the ratings of either group. See Table VIII. However, a Chi square test of independence (Walpole, 1968) indicated that when compared, black women had significantly fewer ratings in the 0-2 range and more ratings in the 3-7 range than white women ($X^2 = 18.32$, $df = 6$, $p < .005$). This indicated the black women tended to receive ratings indicating greater severity.

Tape II

Intra-Judge Reliability. Ten voices submitted to the forced Black-White classification choice were repeated as a measure of test-retest reliability. A standard test of proportions which was tested against the hypothesis that the classification would be made on the basis of pure chance yielded a z-score of 6, significant at the .0001 level. This indicates high test-retest reliability.

TABLE VIII

SEVERITY RATINGS: CLINICAL SIGNIFICANCE

	<u>Black</u>	<u>White</u>
Clinically nonsignificant (0-3 ratings)	185	200
Clinically significant (4-7 ratings)	65	50

z-score = 1.59

p = .06

Black-White Classification. When all fifty women's voices were heard and classified in a forced choice as black or white, correct choices were not made significantly above the chance level. Of the black voices, 46.4 per cent were correctly identified and of the white voices, 49.6 per cent were correctly identified. A one-tailed standard test of proportions revealed correct classification of voices according to ethnic identity was not made at a greater than chance level ($z = 1.158$ for blacks, $z = .126$ for whites). The number of correct judgments and respective percentages made by each judge is shown in Table IX. The range of correct judgments by the judges extended from 38 to 58 per cent with a mean of 48 per cent.

Discussion

The present investigation was designed to consider several parameters of vocal production of black and white women speakers in order to determine if significant differences were evident when group comparisons were made, and further, if ethnic identity could be made on the basis of voice quality alone. The results generally indicated the similarities of voice quality of the women greatly exceeded the differences, and ethnic identity could not be made on the basis of voice quality.

On all of the measurements of pitch and phonation collected by the investigator, no significant differences were observed.

For the judge ratings on the JHVP, it was shown that while the intra-judge test-retest reliability was good, the inter-judge reli-

TABLE IX
CORRECT JUDGMENTS OF ETHNIC GROUP

Judge Number	Number of correct judgments	Percentage of correct judgments
1	29	58
2	25	50
3	23	46
4	25	50
5	28	56
6	19	38
7	24	48
8	22	44
9	25	50
10	20	40

ability was not and this brings to question the validity of the ratings. The most notable instance of poor agreement was on the rating for resonance, where the same voice was rated as lacking in nasal resonance by one and sometimes two judges and as having excessive nasal resonance by other judges. Five of the judges made these disparate judgments at one time or another. All the judges received essentially the same training in the analysis procedure at Portland State University in the Program in Speech and Hearing Sciences, but it may be they had not engaged in sufficient practice to establish adequate agreement with the tool as it was devised. Moreover, the fact that they had had

no previous experience in analyzing backward-played speech may have been a factor in the validity of their judgments.

Acknowledgment of the fact that the validity of the ratings of the JHVP is, indeed, questionable, due to the poor inter-judge agreement already described, imposes substantial qualifications on any conclusions which might have been drawn from the results. Nonetheless, it is felt that cautious speculation and discussion of relationships observed in the results of the investigation is merited in the interest of deriving as much value from the investigation as possible. The writer cautions that the following discussion is premised on questionable results and, therefore, cannot be considered as factual, but rather as speculative.

Differences which were found between groups at statistically significant levels by the ten judges were that white women were rated more often as "breathy" (-2 on the scale of laryngeal openness) and "too soft" (-2 on the scale of vocal intensity). Blacks were rated significantly more often as "too low" (-2 or -3 on the scale of pitch) and as "hyponasal" (-2 on the scale of resonance), although the validity of ratings on resonance was especially poor. No significant difference was noted between the incidence of severity ratings of clinical significance and nonsignificance, although black women did receive comparatively more ratings in the 3-7 range than did white women.

It is interesting that the judges apparently assumed the voices they rated to be deviant since very few ratings of complete normalcy were given. As was previously mentioned, the judges were not informed about the nature of the investigation prior to their participation, so

they were unaware the speakers were drawn from a "normal" population. It must be noted, however, that only 23 per cent of the voices overall were judged to have "clinically significant" (Laskey, 1975) voice disorders and many of the criticisms for the voices probably would have been made only by a person trained in voice analysis.

It is not surprising the two significant findings regarding white women (breathiness and insufficient volume) accompany one another since physiologically both conditions are related to too little tension and incomplete closure of the glottis during phonation (Boone, 1971). It should be noted, however, that few ratings of breathiness and insufficient loudness were actually assigned to the white women (Table VI).

Blacks were rated more often as too low-pitched during speech, although measurements by the investigator revealed no group differences in various pitch measurements. This may reflect the invalidity of the judges' rating or it may indicate the black speakers lowered their voices during the tape recording of the speech samples. It is possible that during the oral reading the black women tended to use the lower part of their pitch range or perhaps tended to read monotonously, either of which could possibly contribute to a perception of being too low-pitched. It is possible the reading task could have introduced an artifact which would not have been noted had conversational speech samples been employed, i.e., the black women may have lowered their voices partly in response to the task which they may have perceived as stressful or boring. On the other hand, it may be that the black women tend to use their lower range in all speaking situations.

It is possible that hyponasality judgments occasionally made by some of the judges (which were in disagreement with the majority of judgments) were invalid, and that more of the black voices than are reflected by the statistics are, indeed, hypernasal. All but one of the ratings of hyponasality were assigned to black speakers. Had the judgments of hyponasality been changed to hypernasality, the incidence of hypernasality observed in the black speakers might have been found to be significantly greater than among the white speakers. A logical relationship between the co-occurrence of the low pitch observed among the black women and hypernasality could be posited. Both conditions could be related physiologically to generally relaxed laryngeal, pharyngeal, and velar (soft palate) musculature. The more relaxed the vocal folds, the lower the emitted tone; and a more relaxed supra-laryngeal cavity also will selectively amplify the lower pitched tones, enhancing the low pitch and possibly creating a perceptual impression of greater "resonance."

The slightly greater severity ratings assigned black speakers may be related to the fact that "disorders" of resonance, observed here to be more prevalent among black women, are considered to be more socially or physically serious disorders than the types of disorders observed to be more prevalent among white women (breathiness and soft volume).

No significant difference was found in the incidence of perceived hoarseness as judged on the JHVP (+2,-2 rating). However, it is possible that, in general, the lower pitch of some black speakers may influence listener perceptions of vocal roughness, although not reflected in the results of this particular investigation. Phillips

(1973) found that judges tended to choose male vowel productions as rougher from pairs of male and female normal vowel productions. Wendahl (1963) earlier found that in pairs of electrically simulated vowel productions of equal aperiodicity, the lower pitched vowel tended to be rated rougher.

CHAPTER V

SUMMARY AND IMPLICATIONS

Summary and Conclusions

The purpose of this study was to make a comparative investigation of the voice quality of black and white women. A review of the literature revealed considerable research has been conducted in the area of black-white language differences in the United States, but very little research has been conducted regarding comparisons of voice quality of Blacks and Whites. The suggestion that certain voice qualities are more characteristic of Blacks or Whites when compared as groups led the writer to pursue the present investigation. One suggestion of particular interest was that Blacks were judged more often to have hoarse voices. Two separate experiments were designed to consider the following questions:

- 1) Do the voices of a group of black women differ significantly from those of a group of white women on any parameter assessed by a) measurements made by the investigator using the Fairbanks Musical Scale Pitch Level Comparison (Fairbanks, 1960), or b) scaled value ratings made by ten judges on the Jewish Hospital Voice Profile (Wilson, 1972)?
- 2) Can ethnic group identity be ascertained according to perception of voice quality more often than expected by chance?

Indexical information on the subjects was collected for the following variables: 1) age, 2) length of longest sustained phonation,

3) optimum pitch, 4) modal pitch, 5) relationship of modal pitch to optimum pitch, 6) lowest possible pitch, 7) highest possible pitch, and 8) pitch range in semitones. The data analysis revealed no statistically significant differences between black and white women regarding these variables.

In the experiment involving the JHVP, fifty samples of women's voices (twenty-five Blacks and twenty-five Whites) were submitted to listener analyses of voice quality by ten women judges, two of whom were trained speech pathologists and eight were in training at the graduate level in speech pathology. The voice samples were heard by reverse audio tape playback to eliminate dialectical differences. The validity of the ratings of the JHVP is questionable, based on the poor inter-judge reliability. Recognizing this qualification, the ratings were nevertheless submitted to data analysis with the following results: Statistically significant differences observed were that black women were more often rated as too low-pitched during speech, hyponasal, and more "severe," while white women were more often rated as too soft and too breathy. No difference in the incidence of perceived hoarseness was found.

In the second experiment, brief segments of each of the fifty women's voices were presented by reversed playback in a randomized order for a forced listener choice of ethnic identity: Black or White. The results indicated such identity could not be made at a greater than chance level.

Implications

The results of this investigation did not reflect any difference in the incidence of perceived hoarseness among black and white women speakers. This particular voice quality, a commonly observed type of voice disorder, was of primary interest in the current investigation because it is often symptomatic of organic pathology (Palmer, 1956) and the investigator had observed an uncommonly high incidence of it among a small group of black students.

The results of the ratings regarding resonance were conflicting, apparently due to poor judge validity. There was apparently a difference between resonance characteristics of black and white women speakers which the judges could not identify with agreement as lack of nasality or as excessive nasality. The matter of resonance differences would require further investigation by judges thoroughly trained in analysis of resonance in order to be resolved. Furthermore, lack of training in making judgments of reverse playback recorded speech may have had a particularly confounding effect on ratings for this parameter. However, the medical implications regarding resonance are not as imposing as are those related to hoarseness, and, therefore, the matter might be without sufficient relevance to warrant further investigation.

As mentioned in Chapter II, ways of perceiving sensory phenomena are influenced by cultural factors. Since analysis of Blacks' voices by white judges may be inherently prejudicial, investigations of voice quality of Blacks by subjective analysis methods should include black judges. A simple rating scale might be devised employing easily under-

stood terms which both black and white judges could use. Ratings would be made according to each judge's individual perceptions (which would presumably be influenced by certain cultural experience) rather than standardized relative to one specific set of criteria. The judges could then be "untrained" listeners of both ethnic groups.

A second possible approach regarding judges would be to establish high inter-judge reliability for a small number of judges prior to experimental ratings. In any case, the judges should practice making ratings on backward-played speech before engaging in experimental ratings.

An informal study which may be of interest would be a survey on the attitudes and opinions of both Blacks and Whites regarding possible voice quality differences. Those comments offered by participants in the current investigation were interesting, thoughtful, provocative, and widely divergent.

A challenging area of investigation which the writer feels would provide the most objective information regarding black and white voice differences is that initiated by von Raffler Engel (1972): spectrographic analysis of prosodic factors. Further objective analyses such as his regarding the factor of time, as well as other factors especially related to stress patterns in speech, may well provide the most definitive answers regarding group differences in the speech of black and white ethnic groups.

SELECTED BIBLIOGRAPHY

- BAILEY, B. L., Toward a new perspective in Negro English dialectology. American Speech, 40, 171-177 (1965).
- BARATZ, J. C., The language of the ghetto child. In: Bentley, R. H., and Crawford, S. D. (Eds.), Black Language Reader. Glenview: Scott, Foresman and Co., 77-79 (1973).
- BOONE, D. R., The Voice and Voice Therapy. Englewood Cliffs: Prentice-Hall, Inc. (1971).
- BURLING, R., Man's Many Voices: Language in Its Cultural Context. New York: Holt, Rinehart & Winston (1970).
- COLEMAN, R. F., Effect of median frequency levels upon the roughness of jittered stimuli. J. Speech Hearing Res., 12, 330-336 (1969).
- _____, Effect of waveform changes upon roughness perception. Folia Phoniatr., 23, 312-322 (1971).
- DARLEY, F. L., Diagnosis and Appraisal of Communication Disorders. Englewood Cliffs: Prentice-Hall, Inc. (1964).
- DENES, P. B., and PINSON, E. N., The Speech Chain. Bell Telephone Laboratories (1963).
- EMANUEL, F. W., LIVELY, M. A., and MCCOY, J. F., Spectral noise levels and roughness ratings for vowels produced by males and females. Folia Phoniatr., 25, 110-120 (1973).
- EMANUEL, F. W., and SANSONE, F. E., Some spectral features of "normal" and simulated "rough" vowels. Folia Phoniatr., 21, 401-415 (1969).
- FAIRBANKS, G., Voice and Articulation Drillbook. New York: Harper & Bros. (1960).
- HAYS, W. L., Statistics for the Social Sciences. New York: Holt, Rinehart, and Winston, Inc. (1973).
- HECKER, M. H. L., Speaker Recognition: An Interpretive Survey of the Literature. American Speech and Hearing Assoc. Monograph No. 16 (1971).

- ISSHIKI, N., and VON LEDEN, H., Hoarseness: Aerodynamic studies. Arch. Otolaryng., 80, 206-213 (1964).
- IWATA, S., Periodicities of pitch perturbations in normal and pathological larynges. Laryngoscope, 802, 87-96 (1972).
- LABOV, W., Language characteristics: Blacks. In: Bentley, R. H., and Crawford, S. D. (Eds.), Black Language Reader. Glenview: Scott, Foresman and Co., 96-116 (1973).
- LASKEY, A. L., A voice incidence study--Portland, Oregon. Master's Thesis, Portland State University (1975).
- LAVER, J. D. M., Voice quality and indexical information. Brit. J. Dis. Communication, 3, 43-54 (1968).
- LAWING, N. N., Alaska Nellie. Seattle: Seattle Printing and Publishing Co. (1940).
- LIEBERMAN, P., Some acoustic measures of the fundamental periodicity of normal and pathological larynges. J. Acous. Soc. Amer., 35, 344-353 (1963).
- LIVELY, M. A., and EMANUEL, F. W., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. J. Speech Hearing Res., 13, 503-517 (1970).
- MANNING, C. C., A survey of hypernasality in children who have had adenotonsillectomy as judged through forward and reversed playback recordings. Master's Thesis, University of Washington (1957).
- MCDAVID, R. I., and MCDAVID, V. G., The relationship of the speech of American Negroes to the speech of Whites. American Speech, 26, 3-17 (1951).
- MOORE, P. T., Organic Voice Disorders. Englewood Cliffs: Prentice-Hall, Inc. (1971).
- PALMER, J. M., Hoarseness in laryngeal pathology: A review of the literature. Laryngoscope, 66, 500-516 (1956).
- PERKINS, W. H., Speech Pathology. St. Louis: The C. V. Mosby Co. (1971).
- PHILLIPS, P. J., Effects of speaker-sex-difference on listener's perception of vocal roughness in normal vowel productions. Master's Thesis, Portland State University (1973).

- RISSET, J. C., and MATTHEWS, M. V., Analysis of musical instrument tones. Psych. Today, 22, 23-30 (1969).
- SANSONE, F. E., Jr., and EMANUEL, F. W., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. J. Speech Hearing Res., 13, 489-502 (1970).
- SHERMAN, D., The merits of backward playing of connected speech in the scaling of voice quality disorders. J. Speech Hearing Dis., 19, 312-321 (1954).
- THIRD REGIONAL WORKSHOP ON THE REHABILITATION CODES AND COMMUNICATION DISORDERS. PHS Grant No. B-3676, The National Institute of Neurological Diseases and Blindness, Communication Disorders Research Training Committee (1967).
- VON RAFFLER ENGEL, W., Some phono-stylistic features of Black English. Phonetica, 25, 53-64 (1972).
- WALPOLE, R. E., Introduction to Statistics. London: The Macmillan Co. (1968).
- WENDAHL, R. W., Laryngeal analog synthesis of harsh voice quality. Folia Phoniatr., 15, 241-250 (1963).
- _____, Some parameters of auditory roughness. Folia Phoniatr., 18, 26-32 (1966a).
- _____, Laryngeal analog synthesis of jitter and shimmer: Auditory parameters of harshness. Folia Phoniatr., 18, 98-108 (1966b).
- WILSON, F. B., The voice-disordered child: A descriptive approach. Language, Speech and Hearing Services in Schools, 1 (4), 14-22 (1971).
- _____, Voice Disorders (Phonotape). St. Louis: The Jewish Hospital of St. Louis (1972).
- ZEMLIN, W. R., Speech and Hearing Science. Englewood Cliffs: Prentice-Hall (1968).

APPENDIX A

SUBJECT INFORMATION FORM

PART A. NAME _____ NO. _____
SEX _____ GROUP _____
B.D. _____ AGE _____
ADDRESS _____

PART B. "I would like to make a brief tape recording of your speech for my study. I will give you written instructions and something to read for the recording. Do you have any question?"

PART C. 1) "Now I would like for you to say 'ah' at a comfortable pitch, holding it as long as you possibly can. You will have 3 trials. Begin when you are ready."

2) "Are your voice and speech today typical? That is, may they be considered representative of your normal speech and voice?"

3) Optimum pitch _____

Modal pitch _____

Modal pitch is above
 near optimum.
 below

APPENDIX B

TEXT READ BY SUBJECTS

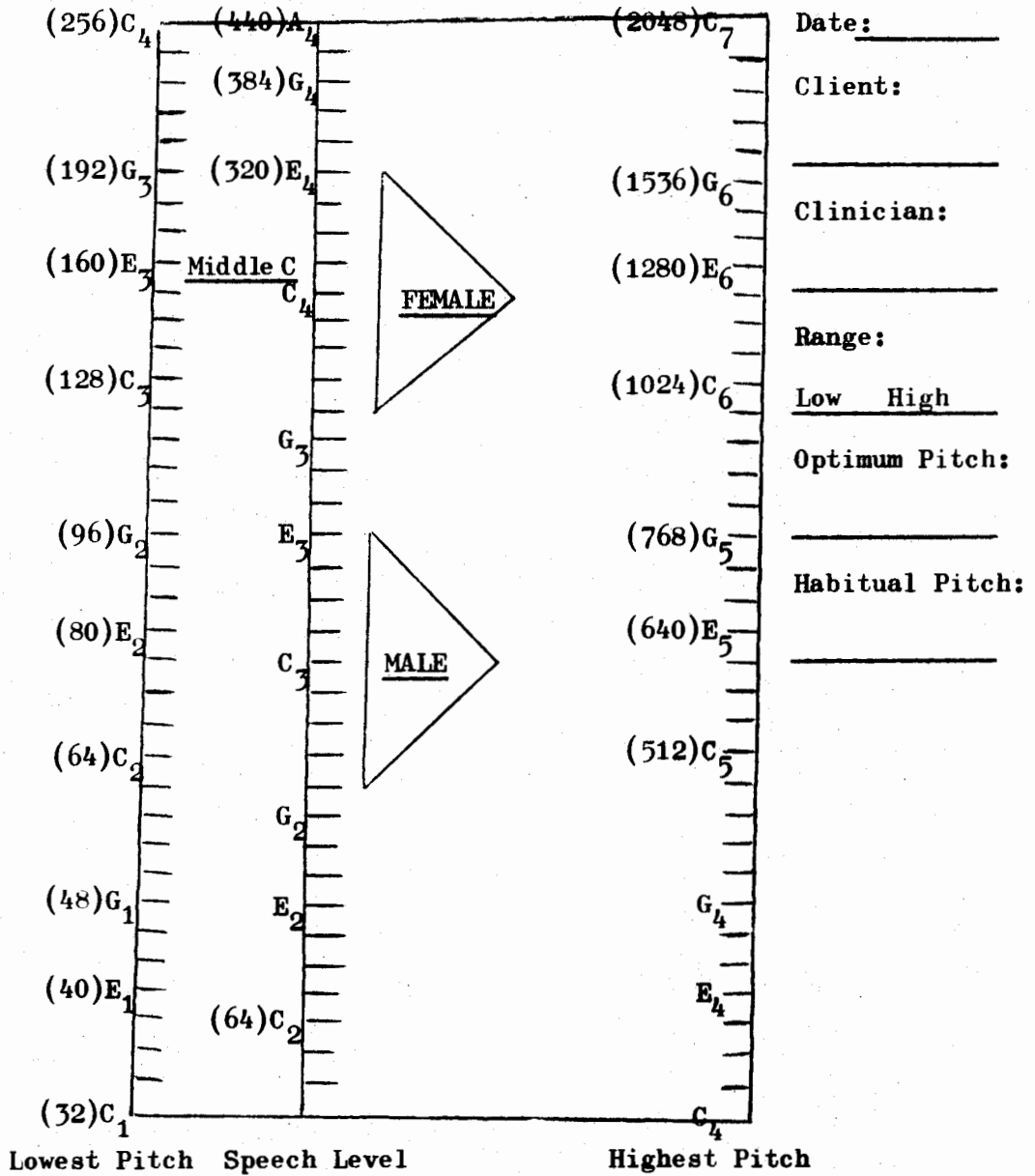
For some time I had been wanting a fur coat, to be made from animals which I planned to get myself. My trap line would have to be in the immediate locality, as business would not permit my taking the time to go far from home.

When the trapping season opened, I made sets for mink, lynx, land otter and rabbit snares, placing them along the creek and using dried fish as bait.

One day on my rounds to the traps, I saw the tracks of a fox. Following them up the valley, I came to his den. These clever animals are as silent as their shadow and as quick as light. I decided to conceal a trap for him without bait, and went to it every day for a week. One morning I found the snow trampled down and the evidence of a desperate struggle having taken place. Pieces of what had once been a beautiful red fox were scattered about and a small piece was in the trap (Lawing, 1940).

APPENDIX C

FAIRBANKS MUSICAL SCALE PITCH LEVEL COMPARISON
FROM: G. FAIRBANKS, VOICE AND ARTICULATION
DRILLBOOK (NEW YORK: HARPER & ROW, 1960)



APPENDIX D

KEY TO NUMERICAL CODING OF PITCH SCALE FOR
PURPOSES OF DATA ANALYSIS

F ^b	A ^b	E ^b	D ^b	E ^b	G ^b	A ^b	E ^b	D ^b	E ^b	G ^b	A ^b	B ^b	D ^b	E ^b	G ^b	A ^b	E ^b	D ^b	E ^b	G ^b	A ^b	B ^b	D ^b										
2	4	6	9	11	14	16	18	21	23	26	28	30	33	35	38	40	42	45	47	50	52	54	57										
F ₂	G ₂	A ₂	B ₂	C ₃	D ₃	E ₃	F ₃	G ₃	A ₃	B ₃	C ₄	D ₄	E ₄	F ₄	G ₄	A ₄	B ₄	C ₅	D ₅	E ₅	F ₅	G ₅	A ₅	B ₅	C ₆	D ₆	E ₆	F ₆	G ₆	A ₆	B ₆	C ₇	
1	3	5	7	8	10	12	13	15	17	19	20	22	24	25	27	29	31	32	34	36	37	39	41	43	44	46	48	49	51	53	55	56	

MIDDLE
C

APPENDIX E

RATING INSTRUCTIONS TO JUDGES AS TRANSCRIBED FROM AUDIO TAPE RECORDINGS

Tape I

On the following tape recording you will hear samples of fifty-five women's voices. The speakers are between the ages of eighteen and thirty years. They have been assigned numbers from one through fifty-five, and are presented randomly on this tape. The samples are approximately forty-five seconds long and they have been arranged so that you will hear them by reverse playback. You are asked to complete a Wilson Voice Analysis for each speaker. Please complete all parts of the analysis form and remember to give a severity rating on the 1-7 scale when a disorder is observed. The order that you are to listen to the voices in is unique. Your particular order is indicated on your written directions and also by the order of your analysis forms which have been numbered by the speaker number. Please follow this order with care. Essentially the voices have been divided into eleven groups of five voices each and they are labelled Group 1, 2, 3, 4, and so forth, up through Group 11. You will be responsible to find the groups in the proper order on the tape. There is colored leader tape between groups, and each voice as well as group number is announced beforehand. You may replay voices if you feel you must, but bear in mind that prolonged listening will tend to produce

listening fatigue. You may take breaks as necessary.

Tape II

This is listening Tape III (sic). On the following tape recording, you will hear brief samples of sixty female voices. The speakers are between the ages of eighteen and thirty years. Some of the speakers are white women and some are black women. On the form you have, please indicate whether you believe the speaker is black or white by circling the B for black or the W for white. Make the best choice you can, but always indicate either black or white. There are approximately five seconds between speakers.

APPENDIX F

RATING FORM FOR TAPE II

Circle B or W as directed by tape instructions.

- | | | |
|---------|---------|---------|
| 1. B W | 21. B W | 41. B W |
| 2. B W | 22. B W | 42. B W |
| 3. B W | 23. B W | 43. B W |
| 4. B W | 24. B W | 44. B W |
| 5. B W | 25. B W | 45. B W |
| 6. B W | 26. B W | 46. B W |
| 7. B W | 27. B W | 47. B W |
| 8. B W | 28. B W | 48. B W |
| 9. B W | 29. B W | 49. B W |
| 10. B W | 30. B W | 50. B W |
| 11. B W | 31. B W | 51. B W |
| 12. B W | 32. B W | 52. B W |
| 13. B W | 33. B W | 53. B W |
| 14. B W | 34. B W | 54. B W |
| 15. B W | 35. B W | 55. B W |
| 16. B W | 36. B W | 56. B W |
| 17. B W | 37. B W | 57. B W |
| 18. B W | 38. B W | 58. B W |
| 19. B W | 39. B W | 59. B W |
| 20. B W | 40. B W | 60. B W |