# PERCEPTUAL AND ACOUSTIC ANALYSES OF SELECTED VOICE AND RESONANCE QUALITIES

By

ELIZABETH LOUISE ALLEN

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1972

# ACKNOWLEDGMENTS

The author wishes to thank her chairman, Dr. Harry Hollien, for his generous expenditure of time and talents during the course of this study and throughout the author's graduate career at the University of Florida.

Grateful acknowledgment is made to the author's supervisory committee consisting of Drs. Moore, Paige, Dew, Brown, Keister and Ramey who have offered much constructive guidance. The author also wishes to thank her other friends in the Communication Sciences Laboratory and elsewhere who have helped and encouraged her along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

PERCEPTUAL AND ACOUSTIC ANALYSES OF
SELECTED VOICE AND RESONANCE QUALITIES

By

Elizabeth Louise Allen

December, 1972

Chairman: Harry Hollien
Major Department: Speech

Studies were undertaken to ascertain if a set of customary and deviant voice and resonance qualities could be differentiated, using perceptual and acoustic analyses. Additional studies were carried out to determine if the results of those comparisons could be related to both sung and spoken phonations, or only to spoken phonations.

Sustained samples of back, breathy, nasal, stident and customary spoken qualities were used in the first studies. In the second studies samples of customary sung with vibrato, customary sung without vibrato and customary spoken qualities were used. All samples were recorded by five female speakers who had completed at least one year of training in singing. Each quality was produced on the vowels /ɑ/ and /i/ under two conditions. In the first condition the speaker was given a list of the quality names and asked to produce phonations which best represented each name. In the second condition a recording of the seven qualities was played for the speaker, and she was asked to match each stimulus quality as closely as possible.

Twenty listeners who were experienced in evaluating voices were chosen from the fields of experimental phonetics, speech pathology, singing and psycholinguistics. These listeners were asked to categorize the randomized phonations as sung or spoken in one set and as back, breathy, customary, nasal or strident in the other set.

Results of the perceptual studies were collated and analyzed to determine how reliably and consistently the samples were categorized as the qualities they were intended to be. Those samples that were consistently judged as they were intended were then subjected to spectral and jitter analyses.

The results indicate that 1) customary spoken and the four deviant qualities could be categorized significantly above chance; 2) samples of breathy and nasal qualities exhibited spectral characteristics which paralleled results from other studies, and specific formant frequencies were somewhat higher than those reported elsewhere; 3) breathy and nasal qualities could be differentiated on the basis of jitter factor, and both deviant qualities had higher jitter factors that were higher than those for customary spoken quality; 4) jitter factors for customary spoken on /ɑ/ were identical to those reported for males in another study; and 5) those samples of sung phonations without vibrato which were identified as sung quality had spectra similar to sung with vibrato and somewhat different from spoken quality.

It is concluded that there are customary and deviant voice and resonance qualities which can be shown to be perceptually and acoustically different. However, the relationships between the customary spoken and deviant qualities cannot be extended to comparisons between customary sung and the deviant spoken qualities, since sung and spoken qualities are both perceptually and acoustically different.

# CHAPTER I

## INTRODUCTION TO THE PROBLEM

For centuries those who have wanted to describe the voice have spoken of its quality (or its timbre, color, resonance or tone color, to name the major alternative terms). First singers and singing teachers, then experimental phoneticians, speech pathologists and psycholinguists have found themselves faced with the difficulties of communicating about this attribute of voice production.

Voice quality as a perceptual entity.--In all fields, there is a consensus that voice quality is the perceptual result of an acoustic signal which is generated by aerodynamic and physiologic activity in the larynx and vocal tract. The author agrees with this definition. Further, she believes that voice quality can be considered to be a relatively slowly changing attribute upon which speech characteristics, and particularly vowel quality, are overlayed. Thus, when a speaker changes from the vowel quality of /i/ to that of /u/, the listener perceives /i/ and /u/, but the percept of voice quality will not be changed.

The register in which the utterance is produced, however, can have an effect on the voice quality. Each of the registers also has a characteristic long-term quality, and if, for instance, a speaker's voice was perceived as breathy in modal register phonations, it might be judged as rough when he spoke in pulse register.[1] Also, extreme changes in frequency and intensity may effect a voice quality change.

In addition to modifications which can be caused by other aspects of speech production, voice quality can be changed while these other aspects are held constant. For example, a speaker whose voice is perceived as breathy may modify his quality so that it is perceived as nasal, without varying register, frequency, intensity or phoneme. Whether the change in voice quality is an intentional shift or the result of some disorder of the phonational system, it can still be considered as a quality which is abnormal, or deviant from the quality the speaker has habitually used. In this regard, a voice disorder is considered to be any voice quality which is chronic and 1) which causes the individual discomfort, 2) is physically injurious, or 3) detracts from his ability to verbally communicate effectively with others. A voice quality disorder is not assumed to be synonymous with deviant quality or abnormal quality, since these latter terms can describe any change, no matter how brief, from an expected norm.

---

1. For a comprehensive discussion of vocal registers, see Hollien (1971).

Voice quality as a vocal attribute.--The author accepts the general statement of speech pathologists and others (Scripture, 1906; Russell, 1931; Van Riper and Irwin, 1958; Fairbanks, 1960; Greene, 1964; Murphy, 1964; Moore, 1971; Perkins, 1971) that voice quality, along with pitch and loudness can be treated as linear continua--from low to high or soft to loud. Voice quality, on the other hand, appears to be a complex continuum.

Although speech pathologists can generally speak of a continuum from disorder to adequacy, and singing teachers can work with the continuum from adequacy to excellence, such a continuum is not sufficient to fully describe the attribute of voice quality. There are also many qualities which can occur at different levels of the continuum of excellence. For instance, nasality (hypernasality) can be either a quality disorder or a deviation from normal or customary voice quality. Indeed, singers strive to achieve a "ringing" quality, by adding what has been described as nasal resonance to their normal voice quality.[2] Although one cannot speak of a certain quantity of breathiness or a certain amount of nasality, for instance, it is possible, through magnitude estimation procedures

---

2. Although the classical approach to a "brilliant" or "ringing" quality is reportedly effected by the addition of nasal resonance (Stanley, 1945; Fields, 1947), Wooldridge (1954) and Vennard (1964) have studied phonations of singers, and they report no physiologic evidence of nasal coupling.

to get a relative rating of the degree to which a given quality is perceptually evident. Moreover, some amount (or degree) of one voice quality can also coexist with some amount of another voice quality, such as the perception of an extremely nasal voice quality which is also somewhat breathy. It is not surprising that the voice quality area has received less research attention than, for instance, the more easily accessible pitch and loudness attributes.

Confusion of concept and of terminology.--With such a complex and multi-dimensional attribute, it is also not surprising that much confusion exists when attempts are made to describe it. For example, there are several authors who categorize their concepts of voice quality into what can be considered loosely as "functional" components. Stanley (1945), Arnold (1957) and Trager (1958) are representative of these writers. Stanley, for instance, relates voice quality to intonation [an acoustic or perceptual characteristic], vocal fold length [a physiologic characteristic] and voice type [a perceptual characteristic]. It may be obvious to the reader that such a diverse list of (cross-modal) characteristics will not be useful in understanding voice quality beyond the most superficial level. The definition of voice quality as the perceptual result of acoustic, physiologic and aerodynamic activity appears to be both more inclusive and more realistic.

Another point of confusion arises among those authors in vocal music and in psycholinguistics, at least, who feel it necessary to separate passive and active constituents of voice quality; that is, some anatomic and physiologic components unique to the individual are considered to be separate from those physiologic components which can be modified.[3] The concept may be useful, but it has not yet been tested. Other authors, such as Stanley (1939), Michel (1964) and Judson and Weaver (1965), consider voice quality and its alternative terms as synonymous and do not attempt to separate a unique passive component from a modifiable one.

## The Problem

Recently, experimental phoneticians have developed reasonably complete descriptions of vocal pitch and loudness and their acoustic, physiologic and aerodynamic correlates in normal speakers. Such descriptions are now needed for the attribute of voice quality. However, one of the difficulties in assessing voice quality has been

---

3. Some authors consider voice quality to be the unique characteristic with some alternative term used for the modifiable quality component (Witherspoon, 1925; Sapir, 1927; Franca, 1959; Abercrombie, 1967; Crystal, 1969). Preetorius (1907) and Trager (1958) maintain that voice set or timbre is unique to the individual and voice quality is the modifiable characteristic. It is often difficult to determine in which context the term "voice quality" is being used.

the lack of either an absolute or a relative baseline to which quality deviations can be contrasted. The characteristics of normal or customary voice quality might be the appropriate frame of reference against which to contrast a range of quality deviations.

The present investigation has been made in an effort to establish an approach to the study of the voice quality of normal speakers. In order to provide a common frame of reference, the characteristics of customary spoken quality were used as the baseline against which quality deviations were contrasted.

Customary spoken quality is defined operationally as that pervasive physiologic set and its acoustic result which is ordinarily used by a given individual. Its effect can be perceived as pervasive across phonemes in connected speech. Specific abnormal voice qualities (which could be described at least subjectively) were chosen as the quality deviations to be compared to the customary voice quality samples. For this reason it was necessary to review what is known about specific voice qualities.

## Voice and Resonance Qualities Used in this Study

Although it has been convenient in the preceding discussions to write of laryngeal and vocal tract effectors as voice quality determinants, it will now be necessary to separate the two and refer to both voice and resonance qualities.

Since most voice and resonance qualities are still understood only subjectively, many quality names associated with other sensory systems have been borrowed in order to describe what the ear perceives.[4] Murphy (1964), for instance, lists 64 quality names which he found in eight books on voice disorders. There are undoubtedly other names which could be added by workers in other fields.

The purpose of this study was to investigate customary spoken quality and non-pathologic deviations from that quality. Therefore, the investigator chose to test specific abnormal qualities that were 1) producible by normal speakers; and 2) representative of a wide range of the quality variation available to the normal speaker. Using these criteria, four quality deviations were chosen: back, breathy, strident and nasal. The following discussion presents what is known and what is suspected about each of these qualities.

Nasality.--This resonance quality is considered to be the result of pervasive nasal coupling. It has also been described by Greene (1964) and Luse et al. (1964) as the

---

4. Brown (1958) suggests that although an attribute, or quality, is associated primarily with one sensory receptor system--as color is associated with vision--that a quality can be considered as inter-sensory, and, for instance, one can speak of a bright voice quality.

possible result of some excessive "tension" in the laryngeal or pharyngeal wall area. Indeed, in his study of nasality in singing, Vennard (1964) concluded that what singing teachers describe subjectively as nasal resonance does not involve the addition of nasal coupling. Further, he quotes Paget's (1930) conclusion " . . . that a part, at least, of the so-called nasal quality . . . is probably due to a constriction of some part of the pharynx" (p. 96). Vennard therefore implies (although he never quite states it) that pharyngeal constriction may result in the percept of nasal resonance or quality. Fant and Erikson (1972) also consider that the effective stiffening of the pharyngeal walls results in the perception of nasality.

As Fant (1960) has pointed out, nasalization is difficult to study acoustically. Variations in speaker, phonetic context and type and degree of nasal coupling, as well as differing experimental procedures, have given rise to somewhat different acoustic descriptions. Characteristics of nasalization which are most consistently reported include 1) a weakening of the first formant (Smith, 1951; Delattre, 1954; House and Stevens, 1956); 2) an additional resonance in the vicinity of 250 Hz (Curtis, 1942; Delattre, 1954; Hattori *et al*., 1958; Fujimura, 1962); 3) an additional formant at about 1000 Hz (Joos, 1948 Smith, 1951); 4) a weakening of the third formant (Smith, 1951; Delattre, 1954; House and Stevens, 1956; Dickson, 1962); 5) diffuse spectral energy (House and Stevens, 1956; Hattori _et al_., 1958; Dickson, 1962); and 6) an over-all decrease in vowel intensity (House and Stevens, 1956; Hattori _et al_., 1958). It was expected that phonations in the present study would exhibit at least some of these acoustic characteristics; however, since the other studies used male speakers, spectral data in the present study were not expected to correspond closely to the specific spectral values for males.

Further, Fant _et al_. (1972) noted that in vowel nasalization the frequency of the first nasal formant was dependent in large part on the vowel uttered. For high front vowels this formant occurred above the first oral formant and for low back vowels it occurred below the first oral formant. In 1960 Fant suggested that two formants (poles) which are close together tend to reinforce one another in amplitude, and their apparent center frequencies are shifted toward one another. In addition, in a complex system, such as that formed when there is nasal coupling, the occurrence of an additional zero near an original pole will decrease the amplitude of the pole and shift the apparent center frequency away from the zero. Such effects are to be observed in nasalized vowels. It is to be expected, then, that the spectral effects of nasality will be more easily identified on some vowels than on others.

In perceptual studies of degree of nasalization, it has been found that, in physiologically normal speakers, the sustained high vowels /i/ and /u/ are judged most nasal and the low vowels /ɑ/ and /ɔ/ are judged least nasal (Carney and Sherman, 1971). Using the reverse approach, House and Stevens (1956) adjusted synthesized vowel samples until they were judged as nasal. In their results /i/ and /u/ required very little nasal coupling to be judged as nasal, while /ɑ/ and /ɔ/ required much coupling to be judged as nasal. From these studies it seems apparent that any perceptual evaluation of nasality should include both front and back vowels.

Breathiness.--Breathy quality is considered a voice quality, whose percept is caused apparently by excessive and turbulent air flow through incompletely adducted vocal folds (Moore, 1971). Although breathiness is a frequently mentioned quality, very little is known about its acoustic characteristics.

Strevens (1960) found that the spectral components of the consonant /h/ ranged in frequency from about 400 Hz to about 6500 Hz and had energy peaks " . . . of intensity . . . so marked as to suggest a multi-formant vowel" (p. 43). He noted from five to seven spectral peaks in the frequency range up to 7000 Hz, and observed more formants for women than for men. One major peak was found in the vicinity of 100 Hz and another in the vicinity of 1700 Hz. On the basis of extrapolations from the main findings of his study and other informal investigation, Strevens suggested that the voiced /ɦ/ would have a spectrum similar to /h/, but with less energy in the higher partials. He also suggested that the burst portion of stop consonants would exhibit spectra similar to their homorganic fricatives. Fant et al. (1972), on the other hand, examined spectra of the burst (aspirated) portions of stop consonants, and found the additional spectral peaks could be explained by a coupling of the subglottal system for both voiced and voiceless stops. That is, the burst spectra can be related to the /h/, rather than to the homorganic fricative. The additional formants they described appear similar to those found by Strevens for /h/.

On the basis of these findings, it would seem reasonable to assume that breathiness would have a spectrum similar to that of /ɦ/. That is, in breathiness there would be extra formants--perhaps around 1000 Hz and 1700 Hz--caused by subglottal coupling, and spectral noise might be evident throughout the frequency range up to about 6500 Hz.

Stridency.--Stridency can be considered as a voice quality, at least for the present. In strident phonations the vocal folds may exhibit longer closed times than in customary phonations produced by normal speakers, according to Hamlet (1972). This possibility was suggested as a result of informal observations of ultrasonic data which she had collected incident to another study.

There are apparently several possible alternates to the term "stridency," including raucous (Abercrombie, 1967), pinched (Russell, 1931), shrill (Franca, 1959), and white throaty (Stanley, 1945). Subjective descriptions of these terms indicate that they probably refer to the same sort of perceived quality--one which exhibits excessive constriction somewhere in the larynx and vocal tract. Thus, it may be possible that stridency may also be a resonance quality.

The most common alternate to strident, however, is the term "metallic." One or both of these terms are used by such writers as Murphy (1964), Vennard (1967), Moore (1971) and Perkins (1971). Perkins presented spectrograms of one speaker producing /ɑ/ with various productions, among which were phonations labelled "optimal" quality and "strident" quality. A subjective evaluation of the spectrograms indicates considerable spectral energy throughout the 2-8 kHz range for the strident sample, while the optimal sample has a stronger F1, a narrow but strong F2 just below 2000 Hz and a wide energy band at about 2500-3800 Hz. However, Perkins did not make any attempt to match frequency or intensity among his samples. Therefore, these spectral changes cannot be assumed to be the result of a quality deviation.

In summary, very little can be predicted from the limited information now available about strident quality.

However, it might be expected that strident phonations would show much acoustic energy throughout an extensive frequency range, and that the additional spectral energy might be the result of long closed-times of the vocal folds and, perhaps, increased pharyngeal stiffness.

Back quality.--Back quality can be considered a resonance quality, since it appears to result from modifications to the pharynx. Moore (1972) recalled a client whom he judged as having an extreme back quality. It was found that the client had enlarged lingual tonsils which caused a noticeable posterior displacement of the epiglottis. Back quality is also referred to as pharyngealized or hollow quality (Abercrombie, 1967) or dark throaty quality (Stanley, 1945).

On the basis of the scant information available on this quality, the only expectation was that, since a quality referred to as back quality was caused by a posteriorlization of the epiglottis in the region of the tongue root, its spectral characteristics may be pervasively similar to those for the vowel /ɑ/.

In summary, it is evident that only one of the four quality deviations chosen for this study has been extensively investigated. For that quality, nasality, there are conflicting data, and male speakers only were used in the research. Although breathy quality has not been studied acoustically, some assumptions can be made about its spectral characteristics based on research on aspiration. There have been no rigorous investigations of either strident or back qualities.

## Purpose

The purpose in this set of investigations is to ascertain if back, breathy, nasal and strident qualities can be perceptually and acoustically differentiated from each other and from customary spoken quality. The specific questions to be asked include:

1. Can (intended) samples of several different voice and resonance qualities be discriminated reliably from one another by trained listeners?

2. Do all speakers exhibit similar acoustic characteristics for each of the identified qualities?

3. Can the qualities be differentiated from one another on the basis of their acoustic characteristics?

## Sung and Spoken Voice Qualities

In experimental phonetics, speech pathology, and psycholinguistics, workers are interested in the speaking voice, but those in vocal music are interested in the singing voice. Since the costumary quality and the four deviant qualities were to be produced as spoken utterances, it was important to establish if the deviant qualities could be contrasted to customary sung phonations as well as to customary spoken phonations. If the presence of vibrato[5]

---

5. Vibrato is a slow, quasi-periodic overlay on the fundamental frequency and intensity of phonation.

is the only difference between customary spoken and sung phonations, then it would be appropriate to compare customary sung quality and the deviant qualities, although the latter qualities were intended to be spoken. If, however, customary sung and customary spoken phonations were perceptually and acoustically differentiable qualities, then the abnormal spoken qualities could not be considered as deviations from customary sung quality.

Sundberg (1970) carried out research on spoken and sung utterances of male singers; he did so in order to ascertain any differences between the two productions. However, he did not stipulate whether the sung phonations were produced with or without vibrato; therefore, it is assumed that his subjects did use vibrato on their sung phonations. In the first part of his study, he compared the acoustic characteristics of spoken quality and sung quality (with vibrato) on each of several vowels. He found that sung vowels were different from spoken vowels, since in singing 1) F2 was lowered in all but the back vowels, 2) F3 was raised for back vowels and lowered for front vowels, 3) F4 and F5 were lowered for all vowels, and 4) the frequency difference between F3 and F4 was less for all vowels (see Figure 1). Using x-rays of his subjects and analog synthesis of the configuration changes found for sung vowels, he concluded that the changes in formant relationships were due to the expansion of the ventricles and piriform sinuses

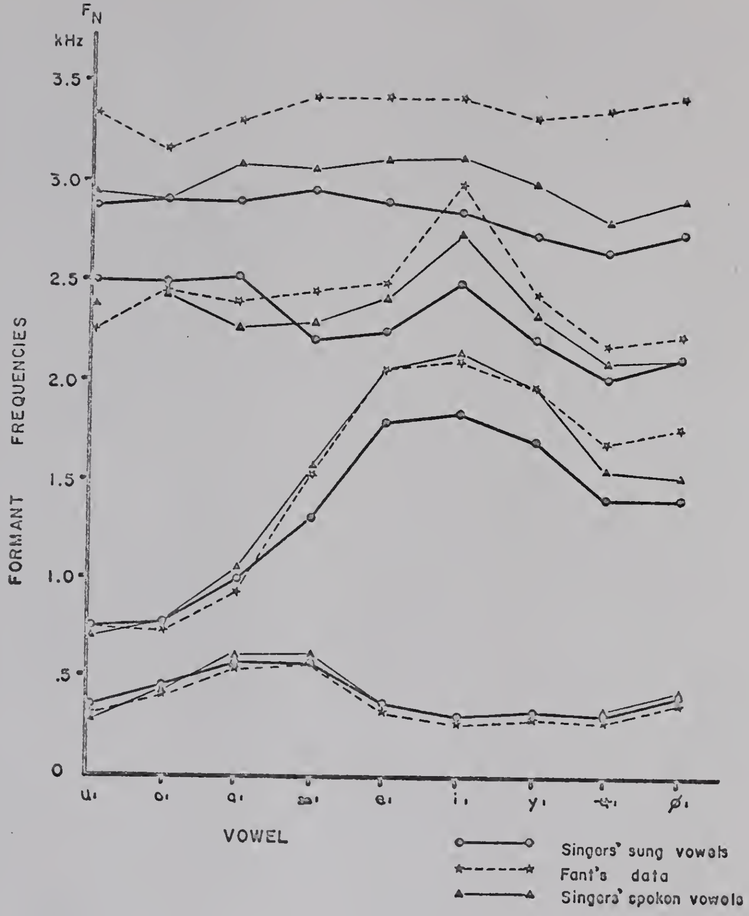


Figure 1. Mean values of formant frequencies of long Swedish vowels. Filled circles and triangles indicate data for Sundberg's study of sung and spoken phonations from four male singers. Stars indicate values for Fant's study of spoken voice in untrained subjects (from Sundberg, 1970).

and a general lowering of the larynx in singing. Frøkjaer-Jensen (1968) essentially replicated and extended the acoustic portion of the Sundberg study, and reported similar results, with the exception of any significant movement of F3 between spoken and sung phonations produced at the same frequency.

Thus, acoustic differences were found between sung and spoken phonations, but it is not possible to ascertain if the differences were due to a change in quality or to the addition of vibrato. Certainly, the laryngeal (and pharyngeal?) adjustments required to produce vibrato could also effect other spectral changes.

### Purpose

This investigation was conducted in order to ascertain if customary sung phonations without vibrato differ acoustically from customary sung phonations with vibrato, and, more importantly, if the sung phonations without vibrato differ acoustically from customary spoken quality. In addition, a perceptual study was made to determine if there were cues to differentiate the sung and spoken utterances. In this part of the experiment, then, the following questions were asked:

1. Can trained listeners reliably differentiate between customary spoken and customary sung phonations?

2. Are customary sung and spoken phonations acoustically differentiable?

3. Are there pervasive acoustic differences between sung phonations with and without vibrato?

# CHAPTER II

# PROCEDURE

In this study of voice and resonance qualities in normal speaking subjects, two investigations were carried out. The intent of one of the experiments was to determine if there are perceptual and acoustic differences between sung and spoken phonations. In the other investigation customary spoken quality and four deviant voice and resonance qualities were tested for perceptual and acoustic differentiability.

## Subjects

Subjects were five adult females whose phonational ranges were within three semitones of one another and whose spectra for customary spoken, or normal, voice quality were as similar as possible to each other and to that of the experimenter. It was considered important to use subjects with similar voices in order to have homogeneous sets of samples. In addition, all subjects had completed at least one year of formal training in singing. It was considered important to use subjects with some training in singing, on the assumption that such training is in part a quality modification procedure. Thus, those with training in singing might be better able to vary voice quality.

Phonation Samples

Customary sung phonations with and without vibrato and customary spoken phonations were elicited from the speakers for use in the first set of investigations. For the second set of investigations samples of back, breathy, nasal and strident qualities also were elicited. The customary spoken phonations, which were used for the two investigations, were simply rerecorded so that they could be used in both experiments. Further, acoustic and perceptual differences between front and back vowels had been noted in the literature and, since each phonation was to be sustained on a given vowel, the vowel quality effect would be as pervasive as the voice or resonance quality effect. Therefore, all qualities in these studies were produced both on the vowels /ɑ/ and on /i/.

Samples of the seven phonated qualities, produced on each of two vowels, were obtained in two ways. First, the speakers were presented with a list of the quality names (customary sung with vibrato, customary sung without vibrato, customary spoken, breathy, strident, back and nasal) and asked to produce phonations which they felt were most representative of each of the names. All the qualities were produced first on the vowel /ɑ/, then on the vowel /i/. This procedure (Condition 1) was used in order to obtain samples of each quality name which reflected the speakers' concept of the quality, rather than the experimenter's concept.

For Condition 2 the speakers were asked to listen to recorded samples of each of the qualities (presented in a different order from those in Condition 1), and they were asked to match each sample as closely as possible. This condition was used in an attempt to obtain the most homogeneous sets of samples from the different speakers.

Three basic constraints were placed on the speakers' phonational samples.

1. As was noted earlier, frequency extremes might cause modifications of quality. For this reason, the frequency of all phonations was maintained constant at about 200 Hz.

2. Since a change in register is known to affect quality, a single register was also required. Although the modal register in females has not been as well investigated as it has in the male, it was assumed that at the relatively low fundamental frequency which was to be maintained by the speakers, their phonations would remain constant in a register which was modal for their speech productions.

3. Extremes in vocal intensity also might cause modifications of the intended quality; therefore, in preliminary sessions the subject was asked to maintain a comfortable but constant voice intensity. However, both the speakers and the experimenter noted that breathy and strident qualities differed so markedly in intensity

that it was very difficult to produce both phonations at the same intensity level and still maintain the intended quality. Accordingly, constant vocal effort was substituted as an alternate to equal vocal intensity. Wright and Colton (1971) found that their subjects could give both consistent magnitude estimations of the vocal effort of their own productions and consistent magnitude productions at the magnitude estimation levels they had chosen. Based on the consistent results found by Wright and Colton, it can be assumed that speakers are able to replicate a given vocal effort level. The speakers in the present study, then, were asked to maintain a constant comfortable effort level for all samples.

To compensate for the resultant intensity differences among the qualities for the different speakers, the gain level on the tape recorder was adjusted so that all samples were recorded at a constant sound pressure level on the reference V-U meter.

Recording Procedures

First, a stimulus tape, to be used for the second recording condition, was prepared by the experimenter. Specifically, sustained samples of customary spoken, customary sung with vibrato, customary sung without vibrato, breathy, strident, back and nasal qualities were recorded on /ɑ/, then on /i/, under the constraints of constant frequency (200 Hz), register and comfortable vocal effort level. The recordings were made with Shure lavalier microphone and a Nagra IV-S tape recorder in an IAC 1200 sound-treated room. A 2 kHz tone from a Data Pulse waveform generator was recorded on the second channel. The tape was then audited by two judges from speech pathology and experimental phonetics to verify that all criteria had been met.

After the stimulus tape had been completed, the speakers were recorded under the same conditions as those observed in preparation of the experimenter's stimulus tape. Each subject recorded the first paragraph of Fairbanks' Rainbow Passage. Then, for Condition 1, she was presented with the list of quality names and was asked to produce her most representative sample of each quality, first on /ɑ/, then on /i/. For Condition 2 the experimenter's tape was played, and the speaker was asked to match each sample as closely as possible. She was allowed to listen to each sample and practice until she felt that her production was a good match with the sample on the stimulus tape.

During both recording conditions the speaker judged the effort level and whether the quality produced was representative of what was intended. The investigator monitored and judged the remaining criteria.

## Perceptual Experiments

Tape preparation.--When the samples had been collected from each of the subjects, center portions of the utterances were isolated for use in the perceptual assessment. An electronic switch with zero crossing and variable rise time was used to extract a 1350 msec center portion (50 msec rise and decay) from each sample. The resultant samples were rerecorded to obtain two presentations of each sample. The master tape consisted of three parts: 1) recordings of each subject reading the short passage; 2) all samples of customary sung and customary spoken qualities presented in random order; and 3) all samples of breathy, strident, back, nasal and customary spoken qualities presented in random order. A six-second silence was placed between individual samples and a short tone followed by a ten-second silence was placed at the end of each ten samples.

Listeners.--Twenty listeners (10 males and 10 females), experienced in evaluating voices, were chosen from the areas of experimental phonetics, singing, speech pathology, and psycholinguistics at the University of Florida. It was judged that, since the specific qualities under investigation are germane to workers in all of these fields, their experience would allow them to make reliable judgments.

Categorization procedure.--One or two listeners at a time were placed in the sound-treated room and given verbal instructions about the categorization tasks (see Appendix A). They were asked to become acquainted with the speakers' voices during the readings of the Rainbow

Passage. Then, on response checklists (see Appendix B), they were to categorize the first set of samples into "Sung" and "Spoken" and those in the second set of samples into one of the five quality categories. After the listeners were instructed, the perceptual tape was presented for their evaluation.

Perceptual analysis.--Results from the checklists from the twenty listeners were compiled. From these data a test/retest reliability measure was made for each listener by comparing his responses on the two presentations of each original sample. Further, confusion matrices were prepared for each of the two experiments. Finally, an analysis of variance was computed. Since the cell values to be used were proportions of "correct" responses, they were first subjected to an arcsine transformation to remove the inherent relationship between mean and variance of the distribution.

Acoustic Analyses

Sample selection.--Listeners' categorizations were not right or wrong, as such; they were only agreements or disagreements with the speakers' concepts and subsequent productions of the seven qualities. Consequently, those samples that were both intended and perceived as the same qualities should have acoustic characteristics representative of those qualities. For this reason, only those samples which were perceived as they were intended were subjected to acoustic analysis. In the first study agreement between a minimum of 11 of the 20 listeners (55% agreement) was required for that sample to be selected for acoustic analysis. For the second study agreement of 8 to 19 listeners (42% agreement of categorization) on each of the two presentations was necessary. Both percentages reflect perceptual agreement well above chance (51.3% and 21.2% respectively).

Spectral analysis.--Three narrow-band sections of each selected sample were made on a Kay Sonagraph (Model 6061B). All apparent eigenfrequencies[6] were identified. The first three formants were located for all samples for a given speaker, and they were essentially normalized by comparing them to the values for customary spoken samples for that speaker, using the scale factor $k_n$. This measure is the difference between the formant frequency of the test sample and the frequency of the same formant of the referent, expressed as a percent of the referent formant frequency.

Jitter analysis.--Jitter, that is, the cycle-to-cycle random frequency variation of a phonation, is probably affected by some laryngeal modifications. Therefore, it seemed appropriate to determine the amounts of jitter

---

6. The term "eigenfrequency" is used to indicate any region of spectral energy concentration, whether it is commonly accepted as a formant, or not.

present in the selected customary spoken phonations and deviant quality phonations to determine if the ones considered as voice quality deviations could be differentiated on the basis of amount of jitter.

In this analysis an oscillographic record is made of simultaneous productions of the phonation sample and a high frequency reference signal. At least twenty consecutive cycles of phonation are selected for analysis. Then, the period for each cycle is determined by counting the corresponding reference cycles, and period values are converted to frequencies. The jitter, or mean frequency variation, from the fundamental frequency, can then be calculated. The final reported values, called jitter factors, are frequency-adjusted; that is, the jitter is divided by the mean fundamental frequency for that sample and expressed as a percent. The procedure is more fully described in a manuscript by Hollien et al. (n.d.).

In the present study the selected phonation samples and the accompanying 2 kHz reference tone were input to an oscilloscope and filmed with a Fastax camera at 1500 frames per second. The film was then developed, twenty-four cycles of each phonation sample were measured and jitter factors were calculated.

# CHAPTER III

# RESULTS AND DISCUSSION

In a study of the customary and deviant voice and resonance qualities of normal speakers, recorded samples of seven qualities were obtained from each of five speakers. These qualities included customary sung with vibrato, customary sung without vibrato, customary spoken, back, breathy, strident and nasal phonations. One perceptual categorization was made for the sung and spoken samples, and a second perceptual categorization was made for customary spoken and the four deviant qualities. Further, selected samples were subjected to spectral and jitter analysis, in order to determine if there are specific acoustic characteristics which discriminate the different qualities from one another.

## Comparison of Sung and Spoken Samples

The samples of customary sung with and without vibrato were compared with customary spoken phonations, in order to determine if there are perceptual and acoustic differences between the two types of phonation.

Perceptual categorization

Twenty listeners were asked to decide if these samples were sung or spoken. The original samples had been re-recorded, so that there were two presentations of each sample to be judged. Therefore, reliability measures could be made, comparing the two presentations.

Reliability ratings, calculated for each listener, appear in Table 1 as percentages of the total number of matches. Reliability ranged from 61.7% to 80.0% with a mean of 69.2%. Since there were no precedents for an expected reliability on these sorts of tasks, it was decided that so long as the individual ratings clustered closely--as they did in this study--the listeners' responses would be accepted. However, the reliability scores for the listeners, individually and as groups, are not high enough to allow good predictions to be made, even though the listeners were reliable at a level above chance ($\alpha = .05$). With these scores in mind, the perceptual results must be viewed with caution.

Individual listeners categorized the 120 sung and spoken quality samples as they were intended a mean of 64.2% of the time, with the range of responses between 51.2% and 70.0%. All response levels were significantly above chance. In Table 1 the number and percent "correct" responses are listed for each listener. Although mean response levels for females were slightly higher than those for males in this investigation, there was no significant difference between

Table 1. Number and percent "correct" responses made by listeners in a comparison of sung and spoken phonational samples, listed by sex of listener (percent reliability, based on two presentations of each sample)

| Listeners | Number "Correct" Responses | Percent "Correct" Responses | Percent Reliability |
|---|---|---|---|
| Males | | | |
| 1 | 71 | 51.2 | 73.3 |
| 2 | 81 | 67.5 | 61.7 |
| 3 | 79 | 65.8 | 65.0 |
| 4 | 62 | 51.7 | 63.3 |
| 5 | 77 | 64.2 | 65.0 |
| 6 | 77 | 64.2 | 65.0 |
| 7 | 73 | 60.8 | 61.7 |
| 8 | 72 | 60.0 | 80.0 |
| 9 | 75 | 62.5 | 68.3 |
| 10 | 80 | 66.7 | 75.0 |
| Mean | 74.7 | 62.2 | 67.8 |
| Females | | | |
| 1 | 79 | 65.8 | 80.0 |
| 2 | 77 | 64.2 | 71.7 |
| 3 | 78 | 65.0 | 65.0 |
| 4 | 77 | 64.2 | 70.0 |
| 5 | 82 | 68.3 | 78.3 |
| 6 | 78 | 65.0 | 63.3 |
| 7 | 84 | 70.0 | 76.7 |
| 8 | 74 | 61.7 | 66.7 |
| 9 | 81 | 67.5 | 65.0 |
| 10 | 77 | 64.2 | 68.3 |
| Mean | 78.7 | 65.6 | 70.5 |

groups. Many singers, at least, claim the ability to empathize with the voice production of other singers. Therefore, it might be expected that the female listeners would be better able to empathize with the female speakers used in this study. However, since response levels for all speakers were relatively low (although above chance), and since female listeners responded only slightly better than males, the present study does not support this claim. It is possible, however, that the listeners' abilities to empathize were diminished by the absence of the perceptual cues of onset and offset of phonation.

Listeners' responses were also examined by field of expertise. In Table 2 the same results are listed by field, in order to test the generalizability of the results to workers in the various disciplines. In this study none of the means differed significantly. It is interesting to note that listeners with extensive backgrounds in singing were not able to categorize the samples in this investigation better than the listeners from the other fields. One would have expected that they would have been more familiar with the singing voice, and, therefore, would have been able to distinguish the samples more easily. There are two possible explanations for this finding. First, neither field of expertise nor years of experience seem to facilitate decision-making in this sort of task. This explanation seems quite feasible, since Colton (1969) also found his

Table 2. Number of "correct" responses made by listeners in a comparison of sung and spoken phonational samples, listed by listeners' field of expertise

| Field of Expertise | Number "Correct" Responses |
|---|---|
| Experimental Phonetics | |
| 1 | 79 |
| 2 | 62 |
| 3 | 78 |
| 4 | 77 |
| 5 | 77 |
| 6 | 73 |
| 7 | 72 |
| Mean | 74.0 |
| Vocal Music | |
| 1 | 71 |
| 2 | 79 |
| 3 | 81 |
| 4 | 77 |
| Mean | 77.0 |
| Psycholinguistics | |
| 1 | 77 |
| 2 | 82 |
| 3 | 75 |
| 4 | 78 |
| Mean | 78.0 |
| Speech Pathology | |
| 1 | 84 |
| 2 | 74 |
| 3 | 80 |
| 4 | 81 |
| 5 | 77 |
| Mean | 79.2 |

listeners from experimental phonetics, speech pathology and vocal music, and even naive listeners, did not differ significantly in their ability to discriminate or categorize modal and falsetto register qualities. Second, it is possible that the task was less familiar to listeners from the field of singing than for those from the other disciplines. In the unfamiliar environment of a sound-treated booth and with only center portions of utterances to attend to, the perceptual abilities of this subclass of listeners may have been degraded.

A confusion matrix was constructed, based on the 800 possible "correct" responses for each quality studied (20 samples, 2 presentations of each, and 20 listeners), comparing intended quality and perceived quality responses. As can be seen in Table 3, samples sung with vibrato were categorized as they were intended 83.9% of the time, customary spoken samples were "correctly" categorized 56.4% of the time (still above chance), but samples sung without vibrato were categorized at chance level. It has been suggested earlier that customary sung quality without the cue of vibrato is perceptually the same as customary spoken quality. If this is so, the results that have been obtained were expected. Eighty-four percent of the appropriate responses were made for customary sung samples with vibrato. Since vibrato is a readily perceived phenomenon, it seems reasonable that listeners were cued by its presence.

Table 3. Listeners' confusions for each of the intended qualities

| Perceived Qualities | Intended Qualities: Sung with Vibrato | Sung without Vibrato | Spoken |
|---|---|---|---|
| Sung | (83.9%)* | (50.0%)* | 43.6% |
| Spoken | 16.1% | 50.0% | (56.4%)* |

*Percentages indicate percent "correctly" categorized by listeners. For each intended quality 800 responses were possible.

Customary sung samples without vibrato were appropriately categorized 50.0% of the time and customary spoken samples were categorized as such 56.4% of the time. Since the listeners had to choose between "Sung" and "Spoken" categories on this task, it was to be expected that customary spoken samples might be identified (as opposed to any sung sample) more easily than samples of customary sung quality which did not have vibrato.

Finally, an analysis of variance (Table 4) indicated that confusions in this investigation were due to the different speakers, qualities and vowels. Post hoc t-tests showed that sung samples with vibrato were confused significantly less than the other two qualities; samples produced on /i/ were confused less than those on /ɑ/; and samples by Speakers 2, 3 and 5 were confused less frequently than the other two speakers in this investigation.

The significantly better categorizations of customary sung with vibrato was also obvious in the confusion matrix. Again, this finding was to be expected. However, the significant effect of the vowels was not expected, and no explanation for the difference can be offered. The effect of the different speakers indicates either that the subjects were not similar enough to produce homogenous samples of the qualities, or that the assumption that the homogeneity of customary spoken productions is not sufficient to insure homogeneity on the other qualities.

Table 4. Analysis of variance summary table for comparisons among sung and spoken phonations

| Source | Error Term | SS | df | MS | F |
| --- | --- | --- | --- | --- | --- |
| Speaker (S) | 60 | 1.6666 | 4 | 0.4167 | 13.88* |
| Quality (Q) | 8 | 5.3385 | 2 | 2.6693 | 11.73* |
| Vowel (V) | 4 | 0.3339 | 1 | 0.3339 | 9.60* |
| Condition (C) | 4 | 0.0465 | 1 | 0.0465 | 0.64 |
| CV | 4 | 0.0004 | 1 | 0.0004 | 0.01 |
| CQ | 8 | 0.0993 | 2 | 0.0496 | 0.27 |
| VQ | 8 | 0.2279 | 2 | 0.1139 | 0.64 |
| CVQ | 8 | 0.1837 | 2 | 0.0918 | 1.35 |
| R(SCVQ) | | 1.8017 | 60 | 0.0300 | |

*Indicates significance at the .05 level.

Acoustic analyses

The samples of sung and spoken phonations which were perceived as they were intended by 11 of the 20 listeners (55% agreement) were considered to be representative of the intended qualities. Therefore, those samples were subjected to spectral and jitter analyses, in order to determine if there were apparent acoustic correlates to the three qualities.

Spectral analysis.--Three narrow-band sections were made for each selected sample on the Kay Sonagraph. Measurements of the eigenfrequencies of the three sections were averaged, and the first three formant center frequencies were located. Since formant frequency values for samples from Condition 1 and Condition 2 of the recording procedure were well within measurement error (100 Hz) for each of the selected samples, the results for these two conditions were combined. The formant data for sung and spoken samples appear in Table 5. For those four values where only one sample was available, a (1) appears next to the first formant frequency value.

Sung formant frequencies were generally lower than those for customary spoken quality. In addition, the $k_n$ values for the samples--the difference between the sung and referent spoken quality values--are reported in this table. The trends for shifts in formant frequencies for sung samples are easily observable from these values.

Table 5. Formant frequencies and $k_n$ values for selected phonation samples on the vowels /ɑ/ and /i/ (results for three subjects reported)

| Formant | | Sung with Vibrato | | Quality Sung without Vibrato | | Customary Spoken |
|---|---|---|---|---|---|---|
| | | Hz | $k_n$ | Hz | $k_n$ | Hz |
| Vowel /ɑ/ | | | | | | |
| F1 | S1 | 800[1] | -3.03 | 750[1] | - 9.09 | 825 |
| | S2 | | | | | 915 |
| | S3 | | | | | 925 |
| | *PB | | | | | 850 |
| F2 | S1 | | | 1250 | + 2.46 | 1220 |
| | S2 | | | | | 1300 |
| | S3 | | | | | 1650 |
| | *PB | | | | | 1220 |
| F3 | S1 | 2575 | -7.21 | 2620 | - 5.58 | 2775 |
| | S2 | | | | | 2805 |
| | S3 | | | | | 2600 |
| | *PB | | | | | 2810 |
| Vowel /i/ | | | | | | |
| F1 | S1 | | | | | 380[1] |
| | S2 | 465 | -5.10 | 400 | -18.37 | 490 |
| | S3 | | | | | 430[1] |
| | *PB | | | | | 310 |
| F2 | S1 | | | | | 2630 |
| | S2 | 2400 | -8.60 | 2490 | - 5.14 | 2625 |
| | S3 | | | | | 2480 |
| | *PB | | | | | 2790 |
| F3 | S1 | | | | | 2980 |
| | S2 | 3280 | +7.54 | 3240 | + 6.23 | 3050 |
| | S3 | | | | | 3170 |
| | *PB | | | | | 3310 |

*Peterson and Barney (1952) mean values for female subjects

Two tentative observations can be made from these data. First, the formant values for sung with and sung without vibrato are all within 100 Hz of one another. Although only a few selected samples can be compared, the results indicate that these phonations of sung quality with and without vibrato can be considered as having essentially the same spectral characteristics, and that the only acoustic effect of the vibrato is quasi-periodic variation in fundamental frequency and intensity, which was noted earlier.

There is an apparent contradiction between the spectral results, which indicates that sung samples with and without vibrato are essentially the same, and the perceptual results, in which sung samples with vibrato were appropriately categorized 84% of the time, but sung samples without vibrato were appropriately categorized only 50% of the time. It will be remembered, however, that the samples of customary sung without vibrato that had been selected for acoustic analysis were indeed perceived as sung phonations in order to be selected.

Second, the spectral comparisons for the selected samples of sung and spoken phonations generally paralleled those found by Sundberg (1970) and Frøkjaer-Jensen (1968). It will be remembered that Sundberg found that sung samples evidenced a lowering of F2 in non-back vowels. In the present study both available sung samples (two conditions, each) on the vowel /i/ exhibited lowered F2, while F2 was slightly higher for the sung phonation on /ɑ/ than for the corresponding spoken sample. (For the sample of sung with vibrato on /ɑ/, no energy region was found in the vicinity of F2.) Also, Sundberg found that F3 was raised for back vowels and lowered for other vowels compared to spoken phonations. In the present study the reverse trend was noted. However, it is to be remembered that Frøkjaer-Jensen did not note and significant shifts of F3 in his study. The present study, then, supported the findings of Sundberg and Frøkjaer-Jensen for F2. However, specific values cannot be compared, since the other two investigations used male subjects.

Jitter analysis.--Samples of customary sung quality with vibrato were not analyzed for jitter, since the vibrato would add a systematic variability to the period of the waves which could not be factored out. The calculated jitter factors for the selected samples of customary sung without vibrato and customary spoken appear in Table 6. A comparison of the data for the vowel /ɑ/ in that table shows jitter factors for sung samples to be very similar and only slightly higher than those for spoken samples. It is conjectured that some very slight vibrato may be present in all sung samples, but that its frequency and intensity variations are too small to be perceived. The mean jitter factor for customary spoken phonations on the vowel /ɑ/ in the present study can be compared to the

Table 6. Jitter factors for selected samples of sung and spoken qualities (separated by vowel and by recording condition C)

| Subject | | Sung without Vibrato /ɑ/ | Sung without Vibrato /i/ | Customary Spoken /ɑ/ | Customary Spoken /i/ |
|---|---|---|---|---|---|
| 1 | C1 | | | .48 | .38 |
| | C2 | | | .38 | |
| 2 | C1 | | .44 | .52 | |
| | C2 | | .48 | .36 | .29 |
| 3 | C1 | | | .50 | |
| | C2 | | | .49 | .28 |
| 4 | C1 | | | | .36 |
| | C2 | | .58 | | |
| 5 | C1 | .51 | | .61 | |
| | C2 | | | .42 | .62 |
| Means | | .51 | .50 | .47 | .39 |
| Quality Means | | .50 | | .43 | |

results found in the study by Hollien et al. In that study four male speakers produced phonations on the vowel /ɑ/ at 100, 141, 200 and 282 Hz. At 100 Hz, a fairly low but comfortable level for male speakers, the mean jitter factor was .47--identical to the mean value for females in the present study. The mean for the first three frequencies was .476. It would appear that males and females exhibit very similar amounts of random variability in their customary spoken phonations, at least on the vowel /ɑ/. If speakers with some training in singing are expected to have developed more precision in vocal fold operation--and less random variation--than untrained speakers, the results of this study do not reflect the effect, since the jitter factor values for customary spoken phonations for these trained subjects were identical to the values for the untrained (in singing) subjects in the Hollien et al. study.

For the vowel /i/, however, the results for customary spoken and sung phonations do differ. For the customary sung samples the mean jitter factors for the two vowels are essentially identical, although it should be noted that there are few samples to compare. For the customary spoken quality, however, the mean jitter factor for the vowel /i/ is smaller that for /ɑ/.

It is possible, based on these results, that in customary spoken quality jitter factors are different for the different vowels for subjects not trained in speaking. However, for sung phonations (and perhaps for spoken phonations for those with voice training in speaking) singers may have the same baseline jitter for each of the vowels. That baseline jitter may be comprised of 1) vocal fold operation which is more precise than the vibratory patterns they used for spoken phonations, and 2) some quasi-periodic variation in frequency--either a perceptible vibrato or some other periodic variation which may be too slow to be perceived as vibrato. Again, it should be mentioned that these postulations are based on very limited data, and therefore are speculative.

In summary, a comparison of sung and spoken quality yields the following results:

1. Although sung samples with vibrato were perceptually differentiated from spoken quality, sung samples without vibrato were not distinguishable. The basis for this distinction appears to be the presence of vibrato, rather than some perceptual cue associated with sung quality.

2. No group of listeners--grouped either by sex or by field of expertise--was able to perceive the intended qualities significantly better than any other group of listeners.

3. Spectral characteristics of sung phonations with and without vibrato are very similar. Parenthetically, vibrato evidenced no acoustic effects beyond the quasi-periodic variation of fundamental frequency and intensity.

4. Spectral characteristics of the sung samples differed from spoken samples in a manner parallel to the differences found by other researchers for male subjects.

5. Mean jitter factor on the vowel /ɑ/ for customary spoken samples in this study is identical to the value for males in another study.

6. Mean jitter factor on the vowel /i/ is lower than for the /ɑ/ in customary spoken samples. Mean jitter factor does not vary between /ɑ/ and /i/ for customary sung samples.

## Comparison of the Five Qualities

Samples of back, breathy, nasal and strident qualities were also compared to customary spoken quality phonations, to determine if the deviant qualities were differentiable from each other and from the common referent. As in the previous study, both perceptual and acoustic analyses of the qualities were made.

### Perceptual Categorization

As in the previous study, a test-retest reliability measure was calculated for each of the twenty listeners. Results are presented in Table 7. On the basis of clustering of reliability scores, the responses of all but one of the listeners were accepted for this study. When that listener was removed, the reliability scores ranged from 47.0% to 62.0% with a mean of 55.2% (significant at the .05 level). Although the reliabilities for this study (based on five choices might be better predictors than those for the first study, they are still not high enough to be good predictors. Thus, the conclusions for these data are also tentative.

Table 7 also presents the number and percent "correct" categorizations of 200 samples of back, breathy, customary spoken, nasal and strident qualities. For this study

Table 7. Number and percent "correct" responses made by listeners in a comparison of back, breathy, customary spoken, nasal and strident qualities, listed by sex of listener (percent reliability, based on two presentations of each sample)

| Listeners | Number "Correct" Responses | Percent "Correct" Responses | Percent Reliability |
|---|---|---|---|
| Males | | | |
| 1 | 78 | 39.0 | 62.0 |
| 2 | 75 | 37.5 | 50.0 |
| 3 | 80 | 40.0 | 48.0 |
| 4 | 92 | 46.0 | 53.0 |
| 5 | 89 | 44.5 | 57.0 |
| 6 | 83 | 41.5 | 58.0 |
| 7 | 69 | 34.5 | 62.0 |
| 8 | 86 | 43.0 | 57.0 |
| 9 | 100 | 50.0 | 57.0 |
| 10 | 82 | 41.0 | 58.0 |
| Mean | 83.4 | 41.7 | 56.2 |
| Females | | | |
| 1 | 86 | 43.0 | 47.0 |
| 2 | 75 | 37.5 | ** |
| 3 | 95 | 47.5 | 54.0 |
| 4 | 88 | 44.0 | 50.0 |
| 5 | 101 | 50.5 | 53.0 |
| 6 | 91 | 45.5 | 56.0 |
| 7 | 93 | 46.5 | 59.0 |
| 8 | 100 | 50.0 | 59.0 |
| 9 | 75 | 37.5 | 48.0 |
| 10 | 82 | 41.0 | 60.0 |
| Mean | 90.1 | 45.0 | 54.0 |

samples were perceived as they were intended a mean of 40.5% of the time, with a range from 34.5% to 50.5%. Again, the response levels for all listeners were well above chance. Although the mean response level for females were again higher than those for males, there was no significant difference between groups, thus supporting the notion that empathy may not provide a good strategy for perception, at least under these experimental constraints. In Table 8 the results are listed by field. Based on a multiple t-test, listeners from psycholinguistics did identify the intended samples significantly better than workers in the other fields, but there were no other significant differences. The only apparent explanation for this finding might be that workers in this discipline had listened to a training tape of other qualities, and they had made perceptual quality judgments on a series of tapes within the last year. However, it should be noted that the qualities they were to identify on those tapes were different from the qualities in the present study.

A confusion matrix was constructed, based on 800 possible "correct" responses for each of five qualities. As can be seen in Table 9, all intended qualities were categorized as they were intended well above the chance level. Intended breathy samples were most frequently so-perceived (50.8%), intended nasal samples were perceived 49.8% of the time, back, 35.6% of the time and strident,

Table 8. Number of "correct" responses made by listeners in a comparison of back, breathy, customary spoken, nasal and strident quality phonations, listed by listeners' field of expertise

| Field of Expertise | Number "Correct" Responses |
|---|---|
| Experimental Phonetics | |
| 1 | 80 |
| 2 | 92 |
| 3 | 95 |
| 4 | 89 |
| 5 | 83 |
| 6 | 69 |
| 7 | 86 |
| Mean | 84.8 |
| Vocal Music | |
| 1 | 78 |
| 2 | 86 |
| 3 | 75 |
| 4 | 75 |
| Mean | 78.5 |
| Psycholinguistics | |
| 1 | 88 |
| 2 | 101 |
| 3 | 100 |
| 4 | 91 |
| Mean | 95.0 |
| Speech Pathology | |
| 1 | 93 |
| 2 | 100 |
| 3 | 82 |
| 4 | 75 |
| 5 | 82 |
| Mean | 86.4 |

Table 9. Listeners' confusions for each of the intended five qualities

| Perceived Qualities | Intended Qualities: Back | Breathy | Customary Spoken | Nasal | Strident |
|---|---|---|---|---|---|
| Back | (35.6%)* | 12.8% | 14.0% | 14.1% | 10.8% |
| Breathy | 9.6% | (50.8%)* | 12.0% | 10.6% | 6.5% |
| Customary Spoken | 30.4% | 15.5% | (41.8%)* | 17.4% | 21.8% |
| Nasal | 16.2% | 8.4% | 18.5% | (49.8%)* | 28.9% |
| Strident | 8.1% | 5.4% | 13.8% | 8.1% | (32.1%)* |
| Total Confusions | 64.3% | 41.1% | 58.2% | 50.2% | 68.0% |

*Values indicate percent "correctly" categorized by listeners. For each intended quality 800 responses were possible.

32.1% of the time. Intended back, breathy and nasal samples were most frequently confused as customary spoken quality, while intended strident and customary spoken samples were more apt to be confused as nasal quality.

Again, based on 800 possible "correct" responses for each quality, a confusion matrix was constructed, comparing intended quality with perceived quality responses. As can be seen in Table 9, all intended qualities were categorized as they were intended well above the chance level. Intended breathy samples were most frequently so-perceived (50.8%), with intended nasal sample perceived at 49.8%, back (35.6%) and strident (32.1%). Intended back, breathy and nasal samples were most frequently confused as customary spoken quality, while intended strident customary spoken samples were more apt to be confused as nasal quality.

As examination of the confusion matrix for this investigation suggests several relationships. First, breathy quality and nasal quality were least confused of the five qualities. This result may be because the samples of these qualities were most distinctive. It may also be because these qualities are more familiar to more of the listeners than are back and strident qualities.

Back and strident qualities were least easily perceived. Indeed, intended back quality was perceived as customary spoken quality at better than chance level, and intended

Table 10. Analysis of variance summary table for comparison between back, breathy, customary spoken, nasal and strident phonations

| Source | Error Term | SS | df | MS | F |
|---|---|---|---|---|---|
| Speakers (S) | 100 | 1.5533 | 4 | 0.3883 | 17.42* |
| Quality (Q) | 16 | 1.5786 | 4 | 0.3946 | 1.06 |
| Vowel (V) | 4 | 0.2242 | 1 | 0.2242 | 5.40 |
| Condition (C) | 4 | 1.0804 | 1 | 1.0804 | 9.22* |
| CV | 4 | 0.1732 | 1 | 0.1732 | 0.96 |
| CQ | 16 | 1.2824 | 4 | 0.3206 | 2.05 |
| VQ | 16 | 1.7968 | 4 | 0.4492 | 1.98 |
| CVQ | 16 | 1.3610 | 4 | 0.3403 | 2.07 |
| R(SCVQ) | | 2.2287 | 100 | 0.0223 | |

*Indicates significance at the .05 level.

strident quality was perceived as nasal at better than chance level. Again, these confusions may be due to the listeners' lack of familiarity with these two qualities. Or it may be reflective of the acoustic characteristics of back and breathy qualities. Unfortunately, there were too few samples of these qualities to allow acoustic evaluations and comparisons to the other qualities. In the case of the strident/nasal confusion, however, two conjectures can be made.

It has been noted by Greene (1964) that laryngeal and pharyngeal tension may be present in nasal quality phonations. Luse et al. (1964) pursued this possibility with cleft palate speakers when they trained their speakers to relax the pharyngeal area during speech. Following this therapy, their cleft palate speakers were judged as less nasal. If strident quality is the result of laryngeal and pharyngeal stiffening, it seems reasonable that the acoustic effects might be similar, and, more particularly, the perceptual effects might easily be confused. In addition, it is quite possible that the cues from onset and offset of phonation are important to the perception of strident quality, since one would expect an abrupt initiation and termination of phonation in the production of this quality.

The analysis of variance (Table 10) for the second investigation indicated that speakers and conditions had significantly affected listener judgments. Samples from

Speakers 1, 4 and 5 were significantly less apt to be confused than those of the other two speakers, and the first recording condition yielded samples that were less frequently confused than samples from the second recording condition.

The possible reasons for the speaker effect were discussed relative to the first study. To reiterate, either the speakers were not homogeneous enough to produce similar phonations of the qualities, or the assumption that similar customary spoken quality phonations will allow consistent productions of the other qualities is not appropriate. Since Speakers 2, 3 and 5 were least confused in the first study and Speakers 1, 4 and 5 were least confused in the second study, it seems possible that homogeneity on any of the qualities is not a guarantee that there will be similarity of production for the other qualities. Related to this issue is the finding that Condition 1 was significantly less confused than Condition 2. If the subjects were not sufficiently homogeneous, their attempts to match the qualities in Condition 2 might be less accurate (possibly because of the constraints of the unique qualities they possess?). However, it seems just as likely that the effect was the result of the difference in elicitation procedures. That is, speakers were able to produce more distinctive productions of qualities when given the quality names than they were when asked to match to the samples.

Finally, there was no precedent for a perceptual study of these multiple qualities for normal speakers, and therefore there was no way to judge the possible effect of context on the judgments made. Results of the two studies suggest that there may be a context effect on perceptions of the different qualities. For instance, if an intended breathy sample was presented once in the vicinity of nasal or strident samples and once in proximity to an extremely breathy sample, perceptions of the given sample might differ because of the different environments of presentation. The best indication of this possibility comes from a comparison of the customary spoken samples which were selected for acoustic analysis. In the first study, eight samples of customary spoken quality met the criteria for selection, and in the second study nine samples of this quality were selected. The customary spoken samples for the two studies were rerecordings of the same phonations. However, only three of the samples were common to both studies. Also, there were some samples of other qualities which were appropriately categorized by few listeners in the other presentation. This inconsistency may be because the listeners were unsure of their decisions on the given sample, or it may be the result of the context in which the sample was presented. At present there is no way to determine if either or both of these possible problems had an effect on listener perceptions.

Acoustic analysis

Samples of back, breathy, customary spoken, nasal and strident quality samples, which were perceived as they were intended by 8 of 19 listeners (42% agreement) for both presentations, were presumed to be representative of the intended qualities. Therefore, those samples were subjected to acoustic analyses. Very few of the samples of back and strident qualities met the criteria for selection. Therefore, acoustic analyses were not made of phonations of these two qualities.

Spectral analysis.--Spectral results for breathy and nasal qualities appear in Table 11. Formant frequencies for these two qualities were generally above those for customary spoken formants. On the vowel /ɑ/ two of the three nasal quality samples exhibited additional spectral energy regions at 400 Hz and 450 Hz, respectively, and an F3 was not apparent for either sample. With these exceptions the formant frequencies for breathy and nasal samples were very similar. In addition, sections of five of the breathy samples showed apparent noise up to 8000 Hz. These samples were all the ones for Speakers 1 and 5 on the vowel /i/ and the one sample for Speaker 5 on the vowel /ɑ/. The remainder of breathy samples for the vowel /ɑ/ showed apparent noise up to about 6500 Hz, and one sample for Speaker 2 on /i/ showed apparent noise up to 5000 Hz.

Table 11. Formant frequencies and $k_n$ values for selected samples of the five qualities on the vowels /ɑ/ and /i/ (results for three subjects reported)

| Formant | | Breathy Hz | Breathy $k_n$ | Quality Nasal Hz | Quality Nasal $k_n$ | Customary Spoken Hz |
|---|---|---|---|---|---|---|
| **Vowel /ɑ/** | | | | | | |
| F1 | S1 | 800 (1) | - 3.03 | (400) (1)<br>1000 | +21.20 | 825 |
| | S2 | 900 (1) | - 1.63 | (450)<br>1080 | +18.03 | 915 |
| | S3 | 975 | + 5.40 | | | 925 |
| | *PB | | | | | 850 |
| F2 | S1 | 1650 | +35.24 | 2000 | +63.93 | 1220 |
| | S2 | 1430 | +10.00 | 2115 | +62.69 | 1300 |
| | S3 | 1650 | 0.00 | | | |
| | *PB | | | | | 1220 |
| F3 | S1 | 2770 | - 0.18 | | | 2775 |
| | S2 | 2980 | + 6.24 | | | 2805 |
| | S3 | 3065 | +17.88 | | | 2600 |
| | *PB | | | | | 2810 |
| **Vowel /i/** | | | | | | |
| F1 | S1 | 425 | +11.84 | 425 | +11.84 | 380 (1) |
| | S2 | 400 (1) | -18.37 | 500 (1) | + 2.04 | 490 |
| | S3 | 450 | + 4.65 | 500 (1) | +16.28 | 430 (1) |
| | *PB | | | | | 310 |
| F2 | S1 | 2620 | - 0.38 | 2710 | + 3.04 | 2630 |
| | S2 | 2900 | + 9.48 | 2880 | + 8.85 | 2625 |
| | S3 | 2645 | + 6.65 | 2600 | + 4.84 | 2480 |
| | *PB | | | | | 2790 |
| F3 | S1 | 3050 | + 2.35 | 3280 | +10.07 | 2980 |
| | S2 | | | | | 3050 |
| | S3 | 3550 | +11.99 | 3700 | +16.72 | 3170 |
| | *PB | | | | | 3310 |

*Peterson and Barney (1952) mean values for female subjects

Special analysis of nasal quality showed the expected results. The general predicted spectral effects were noticeable only on the nasal samples produced on /ɑ/. F1 for /ɑ/ is at about 850 Hz and at 310 Hz for /i/, according to Peterson and Barney (1952). The nasal formant in this study was at 400 or 450 Hz, and would be more easily located for /ɑ/ than for /i/ where it is almost coincident with the first oral formant (Fant, 1960). Also, the reported 250 Hz nasal formant, found for male speakers in other studies, was noted at 400 and 450 Hz in this study with female speakers. This spectral shift was expected, since female speakers customarily have formants 16-20% higher than male speakers (Peterson and Barney, 1952; Fant, 1960).

Spectral analysis of breathy quality samples also followed expected patterns. These samples showed added resonances (eigenfrequencies), as might be predicted from Fant _et al_. (1972) and from Strevens (1960). Fant suggested that these additional resonant areas were the result of some subglottal coupling, since the vocal folds would not be so firmly approximated during aspiration, thus not providing the infinite impedence necessary to isolate the vocal tract from the subglottal area. The effects of the possible subglottal coupling make spectra of the samples of breathy quality phonations appear similar to nasal quality phonations, in which nasal coupling is assumed. For samples of breathy quality produced on /ɑ/, there was apparent noise (inharmonic spectral energy) throughout the frequency range up to about 6500 Hz, the same range that Strevens noted for the /h/. However, all but one sample of breathy quality produced on /i/ exhibited noise up to 8000 Hz. The experimenter can make no conjecture on the reason for this difference.

<u>Jitter analysis</u>.--The calculated jitter factors for the selected samples appear in Table 12. In all cases but one the jitter factors for breathy quality exceeded those of the other qualities for the same vowel and speakers, and jitter factors for this quality were the highest values of all the qualities. The next highest jitter factors were for nasal quality, and the smallest jitter factors related to customary spoken samples. For all qualities except breathy quality the jitter factors for the vowel /i/ are lower than those for the vowel /ɑ/.

Jitter analysis of breathy quality phonations was particularly interesting. For three of the breathy samples very high jitter factors were found. Indeed, in one case, the jitter factor for the breathy /i/ was ten times as great as the corresponding customary spoken sample. The remaining five samples of breathy quality had jitter factors that were only slightly greater than customary spoken values.

Table 12. Jitter factors for selected samples of the five qualities

| Subject | Customary Spoken /ɑ/ | Customary Spoken /i/ | Breathy /ɑ/ | Breathy /i/ | Nasal /ɑ/ | Nasal /i/ |
|---|---|---|---|---|---|---|
| 1 (C1) | .48 | .38 | 2.98 | 3.82 | | |
| (C2) | .38 | | .54 | 1.49 | | .46 |
| 2 (C1) | .52 | | .72 | | .67 | 1.18 |
| (C2) | .36 | .29 | | .65 | .50 | |
| 3 (C1) | .50 | | | | | .32 |
| (C2) | .49 | .28 | | | | .30 |
| 4 (C1) | | | | .36 | | |
| (C2) | | | | | | |
| 5 (C1) | .61 | | .56 | .95 | .53 | .35 |
| (C2) | .42 | .62 | | .72 | | |
| Means for Vowels | .47 | .39 | 1.20 | 1.53 | .57 | .55 |
| Means for Qualities | .43 | | 1.36 | | .56 | |

This broad and possibly bimodal distribution suggests the possibility of two different strategies of producing breathy quality. In the case of those samples with high jitter factors the full length of the vocal folds may be incompletely approximated, and the lack of complete closure may result in less controlled, more random vibratory motion. Those breathy samples with relatively low jitter factors may be produced with the anterior portions of the folds well approximated but with a glottal chink in the vicinity of the arytenoid cartilages. Thus, the vibrating portions could operate with less randomness, and still the turbulent air passage at the posterior portion of the folds would provide the effect of breathy quality.

In summary, the study of the perceptual and acoustic comparisons of customary spoken quality and back, breathy, strident and nasal qualities had the following results:

1. All five qualities were perceptually categorized well above chance level ($\alpha = .05$).

2. One group of listeners, those from the field of psycholinguistics, were able to categorize the samples significantly better than the other groups. There was no significant difference by sex.

3. Although the five qualities were "correctly" categorized significantly better than chance level, confusions of customary spoken for intended back quality and nasal for intended strident quality were also significant.

4. Spectral characteristics for breathy and nasal qualities had the expected patterns, according to the literature, and specific spectral values were shifted upward for the female subjects. Although back and strident qualities were "correctly" categorized significantly above chance, there were too few samples of these qualities to allow acoustic evaluations to be made.

5. Jitter factors for nasal quality were consistently slightly higher than those for customary spoken quality. Mean jitter factors for breathy quality were much higher than those for spoken quality.

6. The jitter factors for breathy quality could be divided into two groups--those which were moderately high and those that were extremely high.

# CHAPTER IV

## CONCLUSIONS

The purpose of the present study was to investigate customary and deviant voice and resonance qualities in the normal speaking voice. There was also a need to determine if customary spoken and customary sung qualities were differentiable, since, if they represented the same basic quality, spoken deviant qualities could also be compared to the sung phonations and results could be generalized to customary quality and deviant qualities.

The research questions, then, were:

1. Can (intended) samples of several different voice and resonance qualities be discriminated reliably by trained listeners?

2. Do all speakers exhibit similar acoustic characteristics for each of the identified qualities?

3. Can the qualities be differentiated from one another on the basis of acoustic characteristics?

4. Can trained listeners reliably differentiate between customary spoken and customary sung phonations?

5. Are customary sung and spoken phonations acoustically differentiable?

6. Are there pervasive acoustic differences between sung phonations with and without vibrato?

These studies were undertaken in order to establish basic information from which the study of voice quality can proceed. Because this was an exploratory investigation many constraints were placed on both production and perception procedures. For instance, speakers were required to maintain constant register, frequency and vocal effort levels during their productions of the samples, and recording gain was controlled to minimize the perceptual effects of the changing intensity that resulted. Loudness differences were still perceived by the listeners, and it is thought that these are the result of the varying arrangements of spectral energy for the qualities that were investigated. However, another effect was noted by some listeners, who found some samples seemed to be produced closer to the microphone than others. This cannot be the result of the speakers' physical proximity to the microphone, but it is quite likely that the effect was the result of the recording gain adjustment that was made in order to equalize intensity levels for all samples and all apeakers. Because of the wide variance in intensity which was noted for the deviant qualities (especially breathy and strident qualities) the experimenter chose to control this variable in an effort to minimize the perceptual effects of extremes of intensity. However, it is possible that the differences in loudness for these qualities are important perceptual cues to the identities of the qualities.

In addition, only the center portions of the sustained utterances were chosen for perceptual evaluation, since these were the portions which were to be acoustically analyzed. Important perceptual cues, particularly for strident quality, may have been eliminated by this constraint, and for some listeners the elimination of onset and offset of phonation may have been distracting enough to degrade their performance on all perceptual decisions.

Another aspect of the study may well have affected the perceptual results obtained. Since this was an exploratory investigation of the whole area of voice quality, both customary and several deviant voice and resonance qualities were chosen for evaluation, and several speakers were asked to produce what they felt were representative examples of those qualities. When these samples were then randomized and presented to listeners, the possibility of what might be considered as forward and backward masking effects was introduced into the perceptual investigation.

However, if the investigation of these samples of the several qualities (which reflected the concepts of the qualities the speakers had) was a weakness of the study in one sense, it was also a strength, since it allowed the comparison of the several qualities and the selection of those samples which were most distinctive perceptually. If gain control of sample recording and the use of center portions of the samples degraded the results of the perceptual tasks, it must be noted that both of these constraints were deemed necessary to control the stimuli as carefully as possible in this initial investigation. It is important to remember that despite any difficulties introduced by these constraints, listeners were still able to categorize the phonational samples of customary and deviant spoken qualities well above the chance level. In short, it is felt that the results obtained are representative of the most carefully controlled phonational examples of these qualities that could be obtained at this time.

Since the experimenter wished to utilize samples obtained by matching phonations to her recordings of the qualities, it was necessary to use female subjects in the investigation. Therefore, the acoustic results of this study become more important, since they supply some comparative data to parallel studies which have been made of male speakers for nasal quality and for the comparison between customary sung and spoken qualities.

Samples were selected for acoustic analysis if they were perceived in the same way by several listeners. It is interesting to note that in no case were the two presentations of a particular sample perceived by the requisite number of listeners as one quality and intended by the speaker as another quality. These selected samples, then, represented the distinctive qualities, according to the consensus of speakers and listeners. The number of samples that could be investigated was small, but there was reason to believe that the qualities to be investigated were validly represented. That spectral results paralleled either some previous results or expectation, and that mean jitter factor for customary spoken /ɑ/ was the same as that reported for males, substantiates this belief.

Therefore, it can be concluded that listeners can discriminate between back, breathy, nasal, strident and customary spoken qualities at a level well above chance. In addition, breathy, nasal and customary spoken qualities can be differentiated on the basis of acoustic analyses. Spectral results for nasal phonations and for aspiration were predictive of the spectral results in the present study, and the spectral characteristics for nasal and breathy qualities are similar to each other and differentiable from customary spoken. Finally, breathy quality can be clearly differentiated from customary spoken and nasal qualities.

In the comparison between sung and spoken qualities it is concluded that sung phonations with vibrato can be perceptually distinguished from spoken phonations, and that some samples of sung phonations without vibrato can also be distinguished from spoken samples. On the basis of the selected samples, sung phonations with and without vibrato were found to be spectrally similar, and both were different from customary spoken spectra. Therefore, sung and spoken phonations cannot be considered as the same quality, and all deviant qualities for the present study are comparable only to customary spoken quality.

It is evident from the present investigation that there is a great deal to be learned about voice quality. Several possibilities for further research are suggested by these studies, among them:

1. Studies similar to the present ones, but with:
   a. Investigation of amplitude and bandwidth, as well as formant center frequencies;
   b. Male speakers;
   c. Experimenter control of the qualities produced;
   d. A range of qualities for the singing voice;

2. Perceptual studies using paired or ABX comparisons of customary and deviant spoken qualities;

3. Perceptual studies to establish generally accepted names for the abnormal qualities, probably using identification procedures, rather than categorizations;

4. Perceptual and acoustic studies of entire phonations, as well as center portions of phonations (these would require additional acoustic analyses of onset and offset of phonation);

5. Physiologic and aerodynamic studies of these qualities;

6. Studies of the possible strategies for production of the qualities.

The development of acoustic, and physiologic and aerodynamic patterns for the present set of abnormal qualities may then permit studies of how much a given phonation deviates from customary quality, and toward what quality (or qualities). In addition, studies may be made of the allophonic limits of a given quality.

A large and challenging area of research can now be pursued in an effort to better define the parameters of phonation in normal speakers. Such a program of research should be of value to workers in all disciplines who are concerned with describing the human voice.

APPENDICES

## APPENDIX A

## INSTRUCTIONS TO LISTENERS

ON THESE TAPES YOU WILL HEAR NORMAL SPEAKING SUBJECTS PRODUCE SOME DIFFERENT VOICE AND RESONANCE QUALITIES, SUSTAINED EITHER ON /ɑ/ OR ON /i/. THE SAMPLES YOU WILL HEAR WILL BE ONLY THE CENTER PORTIONS OF THE UTTERANCES--THAT IS, YOU WILL NOT HEAR THE SPEAKERS START OR STOP THE UTTERANCES. AFTER EACH SAMPLE IS PRESENTED, THERE WILL BE A 6-SECOND PAUSE. DURING THAT PAUSE PLEASE DO TWO THINGS--FIRST, DECIDE WHICH QUALITY NAME BEST IDENTIFIES THE SAMPLE AND MARK YOUR RESPONSE SHEET IN THE APPROPRIATE COLUMN. THEN, ON THE RIGHT SIDE OF THE RESPONSE SHEET THERE IS A PLACE FOR YOU TO INDICATE HOW CONFIDENT YOU ARE OF YOUR CHOICE. 1 = VERY UNSURE: 5= VERY SURE. PLEASE WRITE THE APPROPRIATE NUMBER IN THE RIGHT-HAND COLUMN.

IN ORDER TO ACCLIMATE YOU TO THE VOICES TO WHICH YOU WILL BE LISTENING, YOU WILL FIRST HEAR EACH SUBJECT READ A SHORT PASSAGE. LISTEN TO THE VOICES.

IMMEDIATELY FOLLOWING THE READINGS, SUSTAINED SPOKEN AND SUNG SAMPLES WILL BE PRESENTED. THE SPEAKERS WERE ASKED TO PHONATE BOTH CUSTOMARY SPOKEN AND CUSTOMARY SUNG SAMPLES; THE CUSTOMARY SUNG SAMPLES WERE PRODUCED BOTH WITH AND WITHOUT VIBRATO. THERE ARE 120 SAMPLES. AT THE END OF EACH 10 SAMPLES YOU WILL HEAR A TONE, FOLLOWED BY A 10-SECOND PAUSE. IF YOU WANT TO STOP AT ANY TIME, NOTIFY ME.

(PRESENT SAMPLES 1-120)

THE REST OF THE SAMPLES ARE PRODUCTIONS OF SIMULATED QUALITIES BY THE SAME SPEAKERS. THEY INCLUDE: BACK, BREATHY, CUSTOMARY SPOKEN, NASAL AND STRIDENT QUALITIES. PLEASE PROCEED AS IN THE FIRST SET OF SAMPLES. THERE ARE 200 SAMPLES IN THIS SET.

(PRESENT SAMPLES 1-200)

APPENDIX B

LISTENER RESPONSE SHEET

VOICE AND RESONANCE QUALITY STUDY

NAME ______________________________ YEARS OF EXPERIENCE ____

MAJOR AREA OF VOICE TRAINING/EXPERIENCE (Circle appropriate area)

Vocal Music    Speech Pathology    Linguistics    Communication Sciences

| Sample Number | Sung | Spoken | Confidence Rating 1 2 3 4 5 |
|---|---|---|---|
| 1 | ____ | ____ | ____ |
| 2 | ____ | ____ | ____ |
| 3 | ____ | ____ | ____ |
| 4 | ____ | ____ | ____ |
| 5 | ____ | ____ | ____ |
| 6 | ____ | ____ | ____ |
| 7 | ____ | ____ | ____ |
| 8 | ____ | ____ | ____ |
| 9 | ____ | ____ | ____ |
| 10 | ____ | ____ | ____ |
| 11 | ____ | ____ | ____ |
| 12 | ____ | ____ | ____ |
| 13 | ____ | ____ | ____ |
| 14 | ____ | ____ | ____ |
| 15 | ____ | ____ | ____ |
| 16 | ____ | ____ | ____ |
| 17 | ____ | ____ | ____ |

APPENDIX B (CONTINUED)

| Sample Number | Back | Breathy | Customary | Nasal | Strident | Confidence Rating 1 2 3 4 5 |
|---|---|---|---|---|---|---|
| 1 | ____ | ____ | ____ | ____ | ____ | ____ |
| 2 | ____ | ____ | ____ | ____ | ____ | ____ |
| 3 | ____ | ____ | ____ | ____ | ____ | ____ |
| 4 | ____ | ____ | ____ | ____ | ____ | ____ |
| 5 | ____ | ____ | ____ | ____ | ____ | ____ |
| 6 | ____ | ____ | ____ | ____ | ____ | ____ |
| 7 | ____ | ____ | ____ | ____ | ____ | ____ |
| 8 | ____ | ____ | ____ | ____ | ____ | ____ |
| 9 | ____ | ____ | ____ | ____ | ____ | ____ |
| 10 | ____ | ____ | ____ | ____ | ____ | ____ |
| 11 | ____ | ____ | ____ | ____ | ____ | ____ |
| 12 | ____ | ____ | ____ | ____ | ____ | ____ |
| 13 | ____ | ____ | ____ | ____ | ____ | ____ |
| 14 | ____ | ____ | ____ | ____ | ____ | ____ |
| 15 | ____ | ____ | ____ | ____ | ____ | ____ |
| 16 | ____ | ____ | ____ | ____ | ____ | ____ |
| 17 | ____ | ____ | ____ | ____ | ____ | ____ |
| 18 | ____ | ____ | ____ | ____ | ____ | ____ |
| 19 | ____ | ____ | ____ | ____ | ____ | ____ |
| 20 | ____ | ____ | ____ | ____ | ____ | ____ |
| 21 | ____ | ____ | ____ | ____ | ____ | ____ |
| 22 | ____ | ____ | ____ | ____ | ____ | ____ |
| 23 | ____ | ____ | ____ | ____ | ____ | ____ |

BIBLIOGRAPHY


Abercrombie, D. (1967). Elements of General Phonetics (Aldine Publishing Company, Chicago).

Arnold, G. E. (1957). "Morphology and Physiology of the Speech Organs," in L. Kaiser, ed., Manual of Phonetics (North Holland Publishing Company, Amsterdam).

Brown, R. (1958). Words and Things (The Free Press, New York).

Carney, P. J. and Sherman, D. (1971). "Severity of nasality in three selected speech tasks," J. Speech Hear. Res. 14, 396-407.

Colton, R. H. (1969). Some Acoustic and Perceptual Correlates of the Modal and Falsetto Registers, Ph.D. Dissertation, University of Florida.

Crystal, D. (1969). Prosodic Systems and Intonation in English (Cambridge University Press, Cambridge).

Curtis, J. F. (1942). An Experimental Study of the Wave-composition of Nasal Voice Quality, Ph.D. Dissertation, The University of Iowa.

Delattre, P. (1954). "Les attributs acoustiques de la nasalité vocalique et consonantique" ("Acoustic attributes of vowel and consonant nasality,") Studia Linguistica 8, 103-109.

Dickson, D. R. (1962). "An acoustic study of nasality," J. Speech Hear. Res. 5, 103-111.

Fairbanks, G. (1960). Voice and Articulation Drillbook (Harper and Brothers, Publishers, New York).

Fant, G. (1960). Acoustic Theory of Speech Production (Mouton and Company, The Hague).

Fant, G., and Erikson, Ylva (1972). "Analysis of nasal consonants," paper presented at the 83rd Meeting of the Acoustical Society of America.

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (1972). "Spectral effects of subglottal cavity system," paper presented at the 83rd Meeting of the Acoustical Society of America.

Fields, V. A. (1947). Training the Singing Voice (King's Crown Press, New York).

Franca, L. (1959). Manual Bel Canto (Coward-McCann, Inc., New York).

Frøkjaer-Jensen, Borge (1968). "Comparisons between the vowel formant frequencies in speech and song," Annual Report of the Institute of Phonetics, University of Copenhagen, ARIPUC 3, 115-127.

Fujimura, O. (1962). "Analysis of nasal consonants," Journal of the Acoustical Society of America 34:12, 1865-1867.

Greene, M. C. L. (1964). The Voice and Its Disorders (J. B. Lippincott Company, Philadelphia).

Hamlet, Sandra (1972). Personal Communication.

Hattori, S., Yamamoto, K., and Fujimura, O. (1958). "Nasalization of vowels in relation to nasals," Journal of the Acoustical Society of America 30, 267-274.

Hollien, Harry (1971). "Vocal registers: A new look at an old problem," paper presented at the Collequim Medicorum Thearti, Buenos Aires, Argentina.

Hollien, H., Michel, J., and Doherty, E. T. "A method for analyzing vocal jitter in sustained phonation," manuscript, n.d. (submitted to J. of Phonetics).

House, A. S., and Stevens, K. N. (1956). "Analog studies of the nasalization of vowels," J. Speech Hear. Res. 21, 218-232.

Joos, M. (1948). "Acoustic Phonetics," Language 24, 1-136.

Judson, L. S. V., and Weaver, A. T. (1965). Voice Science (Appleton-Century-Crofts, New York).

Luse, E., Heisse, J., and Foley, J. (1964). The vocal approach in the correction of cleft palate speech," Folia Phoniatrica 16, 123-129.

Michel, J. F. (1964). Vocal Fry and Harshness, Ph.D. Dissertation University of Florida.

Michel, J. F. (1965). "Reliability of phonational quality classification," paper presented at American Speech and Hearing Association Meetings.

Moore, Paul (1971). Organic Voice Disorders (Prentice-Hall, Inc., Englewood Cliffs, New Jersey).

Moore, Paul (1972). Personal Communication.

Murphy, A. T. (1964). Functional Voice Disorders (Prentice-Hall, Inc., Englewood Cliffs, New Jersey).

Paget, R. (1930). Human Speech (Harcourt, Brace, New York).

Perkins, W. H. (1971). Speech Pathology (The C. V. Mosby Company, St. Louis).

Peterson, G. E., and Barney, H. . (1952). "Control methods used in a study of the vowels," Journal of the Acoustical Society of America 24:2, 175-184.

Preetorius, C. (1907). The Tone Placed and Developed (Faulkner-Ryan Company, Chicago).

Russell, G. O. (1931). Speech and Voice (The Macmillan Company, New York).

Sapir, E. (1927). "Speech and Personality," in N. N. Markel, ed., Psycholinguistics (The Dorsey Press, Homewood, Illinois).

Scripture, E. W. (1906). Researches in Experimental Phonetics (Carnegie Institution of Washington, Washington, D.C.).

Smith, S. (1951). "Vocalization and added nasal resonance," Folia Phoniatrica 3, 165-169.

Stanley, D. (1939). The Science of Voice (Carl Fischer, Inc., New York).

Stanley, D. (1945). Your Voice (Pitman Publishing Corporation, New York).

Strevens, P. (1960). "Spectra of fricative noise in human speech," Language and Speech 3, 32-49.

Sundberg, J. (1970). "Formant structure and articulation of spoken and sung vowels," Folia Phoniatrica 22, 28-48.

Trager, G. L. (1958). "Paralanguage: A first approximation," Studia Linguistica 13, 1-12.

Van Dusen, C. F. (1941). "A laboratory study of metallic voice," J. Speech Hear. Dis. 6, 137-140.

Van Riper, C., and Irwin, J. W. (1958). Voice and Articulation (Prentice-Hall, Englewood Cliffs, New Jersey).

Vennard, W. (1964). "An experiment to evaluate the importance of nasal resonance in singing," Folia Phoniatrica 16, 146-153.

Vennard, W. (1967). Singing: The Mechanism and the Technique. (Carl Fischer, New York).

Witherspoon, H. (1925). Singing (G. Schirmer, Inc., New York).

Wooldridge, W. B. (1954). The Nasal Resonance Factor in the Sustained Vowel Tone in the Human Voice. Ph.D. Dissertation, Indiana University.

Wright, H. N., and Colton, R. H. (1971). "Some parameters of vocal effort," paper presented at the 80th Meeting of the Acoustical Society of America.

## BIOGRAPHICAL DATA

Elizabeth Louise Allen completed a Bachelor of Arts degree in English and Speech at Linfield College in June, 1958. She was graduated with a Master of Arts in vocal music at the University of Iowa in June 1960. During post-master's training at the same university from 1960-1963, she was a teaching assistant in vocal music. She then pursued a career of professional singing and voice teaching. During that time she was Instructor of Music at Henderson College, Arkadelphia, Arkansas (1963-1966) and Assistant Professor of Music at Louisiana State University (1966-1968).

In August, 1969, Ms. Allen was accepted as a V.A. predoctoral trainee at the Speech Research Laboratory, V.A. Hospital, San Francisco, California. She was awarded an NIH predoctoral fellowship in September, 1970, to study experimental phonetics in the Communication Sciences Laboratory at the University of Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry Hollien

Harry Hollien, Chairman
Professor of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

G. Paul Moore

G. Paul Moore
Professor of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald Dew

Donald Dew
Associate Research Professor
of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Arnold Paige

Arnold Paige
Associate Professor of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

W. Samuel Brown

W. Samuel Brown
Assistant Professor of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope, and quality, as a dissertation for the degree of Doctor of Philosophy.

Elwood Keister

Elwood Keister
Professor of Music

This dissertation was submitted to the Department of Speech in the College of Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1972

Dean, Graduate School