5-25-1973

Effects of Speaker-Sex-Difference on Listeners' Perception of Vocal Roughness in Normal Vowel Productions

Patsy J. Phillips
Portland State University

AN ABSTRACT OF THE THESIS OF Patsy J. Phillips for the Master of Science in Speech Pathology presented May 25, 1973.

Title: Effects of Speaker-Sex-Difference on Listeners' Perception of Vocal Roughness in Normal Vowel Productions.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

Robert C. Marshall, Chairman

Robert L. Casteel

McKenzie M. Buck

The purpose of this study was to investigate the effects of speaker-sex-difference on listeners' perception of vocal roughness in the vowel [æ] produced by normal male and female speakers. In a previous investigation by Wendahl (1963) it was found that when listening to two synthesized vowels, of equal aperiodicity, judges tended to rate the lower pitched vowel as being more vocally rough. If this is true for listeners' perception of human vowel productions as well then it might be advantageous for voice clinicians,

when making vocal roughness assessments, to regard male and female speakers as two separate populations in view of the inherent pitch differences between the sexes.

In this current investigation, pairs of vowels produced by normal adult male and female speakers were presented to 10 speech pathologists (5 males and 5 females). Each vowel pair contained one male and one female production of the vowel [æ] which had been assigned equal roughness ratings in a previous judging task. The 50 vowel pairs contained 10 pairs of vowels at each of five roughness rating levels. The 10 judges were required to listen to each of the 50 pairs and to make a forced-choice selection of the most vocally rough production within each pair.

The findings in this study revealed that for the 50 vowel pairs the judges selected the vowels produced by males as being more vocally rough a significantly greater proportion of the time. With respect to the five roughness rating levels, judges chose the male produced vowels as being rougher a significantly greater proportion of the time at rating levels one, three and five but illustrated no significant preference between the sexes at rating levels two and four. Further analysis revealed that the five male judges selected the vowels produced by males as being the rougher a significantly greater proportion of the time for all 50 pairs at each of the five roughness rating

levels. The five female judges, on the other hand, illustrated no significant preferences between the sexes for the 50 vowel pairs. They did show a significant preference for the males at rating level one, a significant preference for the females at rating level two but no significant at rating levels three, four and five. In addition, male judges illustrated substantially greater inter-judge agreement and intra-judge reliability for this judging task than did the female judges.

EFFECTS OF SPEAKER-SEX-DIFFERENCE ON LISTENERS' PERCEPTION OF VOCAL ROUGHNESS IN NORMAL VOWEL PRODUCTIONS

by

PATSY J. PHILLIPS

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

MASTER OF SCIENCE IN SPEECH:

EMPHASIS SPEECH PATHOLOGY AND AUDIOLOGY

Portland State University
1973

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of Patsy J. Phillips presented May 25, 1973.

Robert C. Marshall, Chairman

Robert L. Casteel

McKenzie M. Buck

APPROVED:

Robert W. Vogelsang, Head, Department of Speech

David T. Clark, Dean of Graduate Studies and Research

May 25, 1973

ACKNOWLEDGMENTS

The author wishes to express her gratitude to Dr. Robert C. Marshall, director of this study, for his invaluable guidance, support and patience throughout the planning and completion of the investigation. Appreciation is also expressed to the members of the thesis committee, Dr. McKenzie Buck and Dr. Robert L. Casteel, for their constant encouragement, and cooperative efforts to expedite the completion of this study.

Additional acknowledgment is made to Dr. Jack R. Hegrenes, Child Development Rehabilitation Center, University of Oregon Medical School, for his invaluable assistance in the statistical analysis of the data. Thanks are also extended to the thirteen people who served as judges and to the 150 individuals who served as subjects.

The author also wishes to express her gratitude to her husband Bob and three children for their untiring patience, encouragement and cooperation throughout this period of graduate study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Practicing clinicians readily recognize the clinical significance of making perceptual judgements of vocal quality aberration and fluctuation. Such judgements, when accurately and reliably made, may be employed to reflect the patient's voice improvement (or lack of such) with treatment and to alert the medical specialists of the possibility of existing or continuing laryngeal pathology. Most of the literature pertaining to voice disorders indicates that vocal quality is the principal parameter of concern to the voice clinician. Although clinicians utilize different terms to describe an aberrant vocal quality (hoarseness, harshness, raspiness, huskiness, etc.) these quality disturbances can be conveniently grouped under the general term of vocal roughness (Third Regional Workshop on the Rehabilitation Codes and Communicative Disorder 1967). This term is useful also in that it can be used to describe a quality inherent in the normal voice (but to a lesser degree) as well as in the pathological voice (Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Whitehead, 1970; Hanson, 1970).

The voice clinician will have in his caseload both

male and female clients who manifest rough vocal quality. It is interesting, however, that when speaking of vocal quality aberrations most clinicians tend to regard their male and female patients as a single population. This tendency is not observed in the assessment and treatment of patients of both sexes with respect to disturbances in other vocal parameters. It is generally accepted that the mean pitch and loudness levels of the adult male voice is significantly lower and louder than that of the female voice (Boone, 1971; Brodnitz, 1968; Fairbanks, 1960; Fisher, 1966; Green, 1966; Murphy, 1964). For this reason the two sexes are generally viewed as two distinct populations with regard to these parameters.

While the pitch and loudness levels of the male and female voices are distinctly different one must realize that these parameters are intimately related to the perception of vocal quality. This suggests perhaps that males and females might be regarded as two populations with respect to vocal quality. Support for the view that sex difference may play an important part in the assessment of aberrant vocal quality was initially offered by Wendahl (1963) who used a laryngeal analog to generate complex acoustic stimuli which varied randomly around median frequencies of 100 Hz and 200 Hz. He found that very slight cycle-to-cycle frequency variations (as little as plus or minus 1 Hz) were perceived as rough and that the

same frequency variations around a median frequency of 100 Hz received higher (more rough) roughness ratings than the same frequency variations around a median frequency of 200 Hz. Wendahl hypothesized that if the vowel productions of male and female voices of equal aperiodicity were rated for vocal roughness, the male voice would be rated as the rougher of the two on the basis of its lower pitch level. Some limited experimental support for this hypothesis has been offered by Sansone and Emanuel (1970) and Lively and Emanuel (1970) who found that listeners vocal roughness ratings of normal speakers' vowel productions encompassed a greater range of severity for male than for female subjects. Wendahl's original hypothesis has yet to be clinically validated, however.

I. STATEMENT OF THE PROBLEM

With the exception of the studies by Sansone and Emanuel (1970) and Lively and Emanuel (1970), the literature reveals a lack of information with respect to the effects of sex difference on perceptual judgements of vocal roughness in the human voice. Information with respect to the influence of the speaker's sex upon vocal roughness assessment could be valuable to the practicing clinician who must rely on his perceptual skills to make such assessments. Accordingly the principle purpose of this study was to investigate the effects of speaker-sex-

difference on listeners' perceptual judgements of vocal roughness on the vowel [æ] produced by normal male and female speakers.

CHAPTER II

REVIEW OF THE LITERATURE

Investigation of the laryngeal mechanism indicates that for the normal voice the motion of the vocal folds is relatively periodic (Timcke, vonLeden and Moore, 1958), whereas the vocal fold vibratory pattern for rough voice evidences aperiodic movements (Moore and Thompson, 1967). Abnormality has also been noted in the acoustical wave spectra of individuals presenting vocal roughness. A number of studies report the existence of random variations in the periods of successive cycles of productions among subjects with deviant vocal quality. Such frequency variations have been found to be highly correlated with listeners' perception of vocal roughness. For example, in an analysis of the fundamental frequency characteristics of harsh vocal quality, Bowler (1964) employed an oscillograph to examine recordings of harsh and non-harsh connected speech samples. He reported that the most striking feature in the harsh portions of connected speech samples was the presence of "frequency breaks." These frequency breaks occurred in both upward and downward directions of the frequency scale and typically encompassed a range of one octave. In no instance did the segments perceived as non-harsh contain

these frequency break characteristics. In addition, the harsh segments were found to have lower mean fundamental frequencies than the non-harsh segments. Coleman (1960) evaluated sustained vowels produced by pathologically hoarse subjects but failed to find frequency breaks as large as those reported by Bowler. He did, however, identify aperiodic cycle-to-cycle frequency variations of less than one octave which he termed "voice breaks." These voice breaks were prominent in the waveform of the subjects' phonations, and their presence was closely associated with the degree of perceived hoarseness. In a study of laryngitic hoarseness, Shipp and Huntington (1965) also found voice breaks to be less frequent than those reported by Bowler. When present, however, such breaks were felt to contribute greatly to the perception of hoarseness. Also in contrast to Bowler's (1964) findings, Shipp and Huntington noted a more restricted range of fundamental frequencies for their hoarse subjects and failed to find significant differences between either mean or median fundamental frequencies for hoarse and normal voices. Lieberman (1963) and Michel (1964) have also suggested that the fundamental frequencies of abnormally rough speakers are within the modal range for the subject's age and sex.

Other studies of disturbances evidenced in the phonatory acoustic waves, suggest that the duration of cycle-to-

cycle frequency variations also affect perception of vocal roughness. Cooper, Peterson and Fahringer (1957) found that when period variations were eliminated from synthesized speech samples listeners perceived the sample to be mechanical and unnatural. This finding is consistent with Lieberman's (1967) observation that pitch perturbations (small, rapid variations in the duration of successive cycles) are apparently essential cues to natural speech quality. He found that in normally produced vowels the perturbation factor is small in comparison to those in rough vowels. Perturbations of less than 0.5 ms were typical of normally phonated isolated vowels; for mildly and moderately rough phonations, however, the perturbation factor generally exceeded that of normal phonation. Michel (1964) points out that a wave is aperiodic when there is ". . . a lack of recognizable repeating wave-forms." He determined the amount of time that the phonatory acoustic wave was aperiodic in relation to the total phonation time for standardized passages of connected speech spoken with harsh, vocal fry, and normal vocal qualities. The total time of each sample was first measured by phonellographic records of the signal. Subsequently, the amount of aperiodicity, or that portion of the total signal which lacked recognizable re-appearing cycles was determined and a ratic computed by dividing the time of aperiodicity by total phonation time. This made it possible to specify the proportion of total phonation time

a phonatory signal was characterized by aperiodicity. Michel concluded that normal sustained vowel phonations are aperiodic approximately two percent of the time, while harsh phonations are aperiodic approximately seventeen percent of the time.

Additional information, supporting the premise that random variations in the periods of successive cycles in the voice wave are associated with perception of vocal roughness, has been contributed through study of acoustic analogs of phonation. To investigate the degree of signal aperiodicity required for listener judgements of roughness, Wendahl (1963, 1966a, 1966b) employed an electrical laryngeal analog to generate complex acoustic stimuli which varied randomly in frequency around a median frequency. He reported that slight frequency variations, as small as plus one cycle around a median frequency of 100 Hz, caused the signal to be perceived as rough. As the frequency variation around the median frequency increased, listeners perceived an increase in signal roughness. In a later study, Coleman and Wendahl (1967) provided more quantitative data regarding the relationship between stimulus duration and perceived vocal roughness. They found cycle-to-cycle frequency variations, which they called "jitter," in a synthesized complex wave were related to perceived signal roughness. They also synthesized complex acoustic stimuli which contained both aperiodic and periodic components so

that the duration of the aperiodic segments within the total stimulus, as well as the amount of aperiodicity around a median frequency, could be varied. As the duration of the aperiodic segment increased from .16 to .80 seconds in a signal of finite length, more severe roughness was perceived by the listeners. A trading relationship between the duration and the amount of aperiodicity in the signal was also revealed. In other words a stimulus containing large cycle-to-cycle frequency variations, within a short aperiodic segment, was judged less rough than a stimulus containing an aperiodic segment of longer duration and smaller aperiodic excursion.

Recently, sound spectrographic analyses have yielded additional information regarding the acoustic properties of vocal roughness. A number of investigations have provided data on the suprafundamental energy distribution within the wave envelope and how this energy distribution relates to perceived roughness. Isshiki, Yanagihara, and Morimoto (1966) and Yanagihara (1967a, 1967b), in studying harmonic and noise components in the spectra of sustained vowels phonated by subjects with laryngeal pathologies found that noise components were mixed with the harmonics in the formant regions for speakers evidencing slight hoarseness. This was particularly evident in the second and third formants. As the severity of hoarseness increased, the noise components tended to appear in the high frequency

region above 3000 Hz. Yanagihara (1967a) also observed a relationship between the degree of spectral noise abnormality and the magnitude of cycle-to-cycle variations in the shape, amplitude and periodicity of the glottal area waves as measured by ultra-high speed cinematographic analysis. To support his findings for human phonations, Yanagihara (1967b) synthesized hoarseness by mixing recorded normal vowels with band-pass filtered noise. He found that as the noise components intruded into formant ranges and as the high frequency harmonic components became obscured by noise, the severity of perceived hoarseness increased. More recently Sansone and Emanuel (1970) and Whitehead (1970), using a constant bandwidth wave analyzer, obtained 3 Hz bandwidth frequency-by-amplitude acoustic spectra of vowels produced by adult males, both normally and with simulated vocal roughness. They found that for all vowel productions, spectral noise was most prominent in the lower spectral frequencies and tended to decrease in the higher frequencies. Similar results were obtained by Lively and Emanuel (1970) who studied the spectral noise levels associated with normal and simulated rough vowel productions of adult females, and by Hanson (1970) who studied the phonations of adult male subjects with pathologically rough voices. The results of these investigations tend to support the data reported by Isshiki, Yanagihara and Morimoto (1966) and Yanagihara (1966a, 1966b).

A further comparison between the studies by Sansone and Emanuel (1970), Lively and Emanuel (1970), Whitehead (1970), and Hanson (1970) reveals that vowel roughness exists in the normal voice, but to a lesser degree than in either simulated vowel roughness or vowels produced by pathologically rough voices. Sansone and Emanuel obtained median roughness ratings for vowels produced by males first normally and then with simulated roughness. Based on a five point equal-appearing-interval scale the median roughness ratings ranged from 1.19 to 1.69 and 3.58 to 4.26 respectively. Whitehead obtained median roughness ratings, using the same roughness scale for vowels produced normally, with vocal fry and with simulated roughness by adult males. His obtained median roughness ratings ranged from 1.56 to 2.11, 4.48 to 4.76 and 4.08 to 4.39 respectively. Hanson's obtained median roughness ratings for vowels and connected speech samples produced by pathological male voices ranged from 2.60 to 3.53. Lively and Emanuel obtained median roughness ratings of 1.14 to 1.55 and 3.60 to 4.08 for normal and simulated rough vowels respectively, produced by females. In each of these studies the simulated and pathologically rough vowel productions consistently received higher median roughness ratings than their normal counterparts. When a comparison is made between the overall median roughness ratings reported for the normal male speakers by Sansone and Emanuel (1970) and Whitehead (1970)

to those reported by Lively and Emanuel (1970) for the normal female speaker it is apparent that the vowels produced by the male speakers were rated more severely than the vowels produced by the female speakers. Lively and Emanuel (1970) report that a direct comparison, based on sex difference, can not be made between their study and that of Sansone and Emanuel (1970) because the roughness ratings assigned to normal productions may have been influenced by the degree of simulated roughness for the vowels in each study. She suggests that the roughness associated with normal vowels may be different for the two sexes because of vocal pitch differences between the sexes.

It is generally agreed that sex associated fundamental frequency differences are due primarily to differences in the length and mass of the vocal folds (Hollien, 1960; 1962; Murphy, 1964; Fisher, 1966; Green, 1966; Brodnitz, 1968; Moore, 1971; Boone, 1971). The males vocal folds average 18 mm in length while the females average 10 mm in length. The average fundamental frequency for the female is approximately 200 cps while that for the male is approximately 125 cps (Fisher, 1966; Brodnitz, 1968; Boone, 1971). Sex related pitch differences appear to be an accepted fact. The effect of these pitch differences on the perception of quality disorders is not clear, however. As stated earlier, Lively and Emanuel (1970) point out that perception of

roughness for vowels produced normally by females may be different than those produced normally by males because of pitch differences. It will be recalled that Wendahl (1963) in his laryngeal analog study reported slight frequency variations (as small as plus or minus one Hz) were perceived as rough and that frequency variations around a median frequency of 100 Hz was rated by the judges in his study as being more rough than the same frequency variations around a median frequency of 200 Hz. The inferences made by Lively and Emanuel, and Wendahl not only points out the importance of considering the speaker's pitch level when assessing vocal roughness but also that it might be beneficial to the voice clinician to regard males and females as two separate populations when assessing the severity of vocal quality disorders. It has been the purpose of this chapter to review the literature on vocal roughness and to provide background information for the present investigation.

CHAPTER III

METHODS AND PROCEDURES

It has been hypothesized by Wendahl (1963) that if a male and a female with equal aperiodicity in their voices were judged for vocal roughness, the male would be judged as being the rougher due to his lower pitch level. Wendahl based this hypothesis on the results of his study in which he employed synthesized vowel stimuli and not on human voices. It was the purpose of the present study to investigate the effects of Speaker-Sex-Difference (SSD) on listeners' perception of vocal roughness in normal male and female productions of the vowel [æ].

I. RESEARCH QUESTION

The following research question was investigated regarding the effect of SSD on listeners' perceptual judgements of vocal roughness:

> When male and female productions of the vowel [æ], having previously received the same roughness rating, are paired and presented to judges, will the male vowel production be selected as the more vocally rough?

Subjects. The subjects in this study included 150 adults, 75 females and 75 males. Subjects were all students at Portland State University, Portland, Oregon, majoring in

a variety of fields. Criteria for subject selection were (1) that subjects be between 18 and 45 years of age and (2) have no present or past history of voice, speech or language disorders. Subjects were asked if they were currently or had in the past been to an Otolaryngologist due to a voice problem or received voice, speech or language therapy from a speech pathologist. The investigator was not concerned whether the subject was a smoker or suffered with allergies. The primary concern was that subjects had normal sounding voices at the moment and no previous history of a voice disorder.

Each of the 150 subjects produced the isolated vowel [æ] for approximately 3 seconds duration at a comfortable pitch level. When producing this vowel each subject sat before an Unidyne III microphone (Model 545) placed at a 70 degree angle to his mouth and six inches in front of his mouth. Each subject visually monitored the intensity of his phonation within a plus or minus 2 dB by means of a VU meter of an Ampex magnetic dual-channel tape recorder (Model AG-500). Prior to subjects actually producing the vowel sample, the experimenter explained and demonstrated the procedure and allowed the subject several practice trials. All subjects' vowel productions were recorded for later analysis by the Ampex tape recorder. These 150 recorded vowel productions served as the voice samples for this investigation. The voice samples were dubbed onto two

additional tapes which were used in the two judging tasks in this investigation. The two tapes will be referred to as Listening Tape 1 (LT-1) and Listening Tape 2 (LT-2). The purpose of these tapes and the methods used in preparing them are as follows:

Listening Tape 1. LT-1 contained all 150 vowel productions. Female subjects' productions of [æ] made up the first 75 samples on the tape; male subjects' productions of [æ] made up the second 75 vowel samples on the same tape. LT-1 was played for three experienced speech pathologists each of whom made vocal roughness ratings for each of the 150 vowel productions. Ratings of vocal roughness were made on a five point equal-appearing-interval scale with number one being designated as the least rough and number five being designated as the most rough. Because the rating of vocal roughness for normal speakers has proven to be a difficult perceptual task (Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Whitehead, 1970) judges were given practice in rating both male and female vowel samples before actually beginning the rating task. During the practice session and the actual rating task judges were permitted to hear each vowel sample as many times as they wished and to mutually discuss their ratings of the various vowels if they felt it necessary. The practice session took twenty minutes. The rating task took one

hour. All vowel roughness ratings were made in a quiet room with the judges seated in front of an Ampex speaker (Model AG-500) through which the vowel stimuli were played. The judges were given the opportunity for a rest period at fifteen minute intervals.

Following the rating of the 150 vowel samples, a second listening tape was prepared which contained the voice samples employed for the second judging task. The purpose and method of preparing LT-2 was as follows:

Listening Tape 2. LT-2 contained 100 selected vowel samples (50 female and 50 male) and was constructed in the following fashion:

1. Those vowel samples on which judges had agreed in their ratings within one scale value in LT-1 were assigned single number ratings. For example, if a particular vowel had received ratings of 4, 3 and 4 respectively, by the three judges, the assigned rating was 4. If a vowel received respective ratings of 3, 4 and 3 the assigned rating was 3.
2. Male and female vowel productions having the same assigned roughness rating, were then arranged in pairs to form 50 pairings each containing an [æ] produced by a male speaker and an [æ] produced by a female speaker. For example, a vowel assigned

a roughness rating of 1 from a female speaker was paired with a vowel assigned a rating of 1 from a male speaker. This procedure continued until ten vowel pairings for each of the five roughness rating levels had been constructed. Male and female vowel samples for each of the ten pairs at each of the five roughness rating levels were selected in a random fashion. When a vowel sample was drawn from the pool it was returned to the pool to insure that each sample had the opportunity of being selected an equal number of times. At four of the roughness rating levels (levels 1 and 5 for females and levels 2 and 5 for males) fewer vowel samples went into the pool. The reason for this was that judges on the original rating task did not agree in their vocal roughness ratings (within one scale value) an equal number of times at all rating levels.

3. The 50 vowel pairings were made by dubbing the selected vowels from the original data tape onto LT-2. Precautions were taken to insure that male and female vowel productions of a particular pair had an equal chance of being first or second in a pair. Each of the vowel pairings were separated by approximately a one second interval. The

order of presentation of the 50 pairings was determined randomly, without regard to vocal roughness rating levels (See Appendix B).

4. Ten of the 50 vowel pairings were randomly selected and placed after the 50 original pairs at the end of LT-2. These ten pairings were utilized to assess judges' reliability for the perceptual judging task described in the following section.

5. Ten speech pathologists listened to the 50 male-female vowel pairs of LT-2. Five speech pathologists heard LT-2 on one occasion with the five additional speech pathologists hearing LT-2 some two weeks later. Both judging sessions took place in the same room with the five listeners seated in a semi-circular fashion in front of an Ampex loud speaker (Model AG-500) through which the vowel pairs were presented. The judging task for the ten speech pathologists involved making a forced-choice response in which they selected the vowel sample in each of the 50 pairs, they perceived as being the more vocally rough of the two. Specific instructions given to the ten judges were as follows:

> You will hear two speakers produce the same vowel. After the second speaker has finished, circle the number 1 or 2 according

> to the speaker whose voice you perceive as being the most vocally rough. Do the same for the other pairs of vowel productions. Do not leave any space blank--guess if you have to--but circle only one number for each space.
>
> The vowel productions may vary according to parameters other than roughness; however, you are asked to ignore these variations. As you are asked to determine which production you perceive to be the most vocally rough, there are no right or wrong decisions. Therefore, make your decision independent of the person sitting next to you. Do not discuss your decisions during the judging session. You may hear each pair as many times as you wish. Are there any questions?

Prior to beginning the actual judging task the ten judges were given practice in making forced-choice judgements. Ten pairs of vowel samples not used in the actual judging task, were played for the judges. Judges made forced-choice responses to these vowel pair samples and then mutually discussed their rationale for selecting one vowel of a pair as being rougher than the other. During the actual judging task judges were allowed to hear each vowel pair as often as they wished but were not permitted to discuss their choices for the 50 experimental or the 10 reliability vowel pairs. The instruction and training period took 30 minutes. The judging task took 45 minutes.

CHAPTER IV

RESULTS AND DISCUSSION

I. RESULTS

The purpose of this investigation was to determine the effects of Speaker-Sex-Difference (SSD) on listeners' perception of vocal roughness in normally produced vowels. On the basis of previous investigations by Wendahl (1963) and others (Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Whitehead, 1970) it has been suggested that when hearing two productions of the same vowel, which had received approximately equal scale value ratings for roughness severity, listeners would tend to perceive the lower pitched voice (male) as being more vocally rough than the higher pitched voice (female). To test this hypothesis an initial step in this study, previously discussed in Chapter III, involved the rating by three judges of 150 productions of the vowel [æ] (75 female and 75 male) on a five point equal-appearing interval scale. Table I gives the roughness ratings assigned by the three judges to each of the 150 vowel samples.

Table II shows the degree of agreement reached by the three judges for samples where they agreed 100 percent

TABLE I

THREE JUDGES' ROUGHNESS RATINGS ON A FIVE POINT SCALE FOR EACH MALE AND FEMALE VOWEL PRODUCTION

Female Sample	Judges 1 2 3	Female Sample	Judges 1 2 3	Female Sample	Judges 1 2 3
1	4 2 4	26	3 3 2	51	4 4 2
2	3 3 2	27	4 3 4	52	5 3 5
3	2 2 3	28	2 3 2	53	3 2 3
4	4 3 3	29	4 3 4	54	4 1 4
5	4 4 4	30	4 2 3	55	3 2 4
6	5 4 2	31	3 2 2	56	3 3 3
7	4 2 3	32	2 2 2	57	4 3 2
8	3 3 4	33	1 1 3	58	5 4 5
9	2 2 2	34	3 1 4	59	3 4 4
10	2 2 2	35	4 3 2	60	3 4 2
11	2 3 2	36	2 1 3	61	3 2 3
12	1 3 3	37	4 4 2	62	2 2 3
13	4 5 5	38	3 2 3	63	4 4 4
14	1 4 4	39	5 3 5	64	5 5 5
15	3 2 2	40	3 2 4	65	2 2 3
16	3 2 3	41	2 2 2	66	5 4 5
17	4 4 4	42	3 2 4	67	2 3 2
18	3 2 3	43	4 4 3	68	3 3 2
19	4 3 4	44	4 2 5	69	4 4 4
20	5 3 4	45	4 1 2	70	3 2 2
21	2 3 2	46	5 2 3	71	1 1 2
22	1 2 1	47	3 4 3	72	2 1 1
23	3 2 3	48	4 4 3	73	2 1 1
24	3 2 3	49	5 2 5	74	1 2 1
25	2 3 1	50	3 3 2	75	4 3 4

TABLE I--continued

Male Sample	Judges 1 2 3	Male Sample	Judges 1 2 3	Male Sample	Judges 1 2 3
1	3 3 1	26	3 3 5	51	2 3 4
2	3 2 2	27	2 2 2	52	3 2 2
3	3 3 2	28	4 4 3	53	4 2 2
4	1 2 1	29	5 5 4	54	1 1 1
5	3 3 2	30	4 3 3	55	2 1 2
6	4 4 3	31	3 4 2	56	4 2 3
7	4 1 4	32	5 4 5	57	3 3 2
8	4 3 2	33	2 1 1	58	3 3 3
9	2 1 2	34	4 3 4	59	1 3 1
10	5 4 4	35	1 1 2	60	4 4 4
11	3 1 4	36	3 2 2	61	3 1 4
12	3 5 3	37	4 3 4	62	3 3 2
13	4 3 4	38	3 3 2	63	4 2 5
14	4 3 3	39	2 1 3	64	2 2 2
15	4 4 3	40	4 3 4	65	3 3 2
16	3 4 2	41	3 5 3	66	2 1 1
17	4 4 3	42	2 3 3	67	1 1 2
18	4 4 3	43	1 1 1	68	3 1 2
19	1 1 1	44	2 2 2	69	4 3 1
20	4 5 4	45	1 2 1	70	2 1 3
21	3 5 3	46	3 3 2	71	2 2 2
22	5 5 5	47	4 4 2	72	4 3 2
23	5 5 5	48	1 2 1	73	2 1 3
24	2 3 2	49	5 3 5	74	5 2 4
25	2 4 2	50	2 2 2	75	3 1 1

TABLE II

DEGREE OF AGREEMENT BY THREE JUDGES ON ROUGHNESS RATINGS OF SEVENTY-FIVE MALE AND SEVENTY-FIVE FEMALE PRODUCTIONS OF THE VOWEL [æ]

	Total Agreement		Agreement within scale values of 1		2		3	
	N	%	N	%	N	%	N	%
Males	11	14.80	37	49.33	21	28.00	6	8.00
Females	10	13.33	39	52.00	18	24.00	8	14.60
Total	21	14.00	76	50.66	39	26.00	14	9.30

and for samples where their ratings varied within one, two or three scale values. The judges agreed unanimously on 14.80 percent of the male vowel samples, on 13.33 percent of the female vowel samples and there was 14.00 percent agreement on all vowel productions. Percentages of judges rating agreement for the 150 productions, which varied within one, two and three scale values were 50.66, 26.00 and 9.30 percent respectively.

Those vowel samples upon which the three judges demonstrated 100 percent concurrence or agreement within one scale value were used to form 50 vowel pairs. Each pair contained one male and one female vowel production that had received the same roughness rating. A group of ten speech pathologists (five males and five females)

listened to each of the 50 vowel pairs and made a forced-choice selection as to which vowel within each pair they perceived to be the most vocally rough. The ten judges' selections of the most vocally rough vowel (male or female) for the 50 vowel pairs are summarized in Table III. Out of the 500 forced choices made, the judges selected the male vowel samples as being more vocally rough a significantly greater proportion of the time (X^2=33.84; df=1; p<.001).

TABLE III

TEN JUDGES' SELECTION BY SEX OF THE MOST VOCALLY ROUGH VOWELS OF FIFTY VOWEL PAIRS

Judges	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Total
Males	40	40	34	31	26	26	28	23	27	21	296
Females	10	10	16	19	24	24	22	27	23	29	204
										Total	500
296/500=0.590											
X^2=33.84;df=1;p<.001											

The ten judges selections of the most vocally rough production (male or female) at each of the five roughness rating levels are itemized in Table IV. The male vowel samples were selected as being more vocally rough a significantly greater proportion of the time at rating levels one (X^2=46.08; df=1; p<.001), three (X^2=8.00; df=1; p<.01) and five (X^2=6.48; df=1; p<.02). At levels two and four

TABLE IV

TEN JUDGES' SELECTIONS BY SEX OF THE MOST VOCALLY ROUGH VOWELS OF TEN VOWEL PAIRS AT EACH OF THE FIVE ROUGHNESS RATING LEVELS

Roughness Rating Level		Judges 1	2	3	4	5	6	7	8	9	10	Total	Percentage of Agreement	Chi-Square Analysis
1	Males	8	9	8	8	7	7	6	4	10	7	74	74/100=0.74	X^2=46.08;df=1;p<.001
	Females	2	1	2	2	3	3	4	6	0	3	26		
2	Males	8	8	5	6	6	4	5	4	2	3	51	51/100-0.510	n s
	Females	2	2	5	4	4	6	5	6	8	7	49		
3	Males	9	7	8	7	5	4	8	5	4	3	60	60/100=0.60	X^2=8.00;df=1;p<.01
	Females	1	3	2	3	5	6	2	5	6	7	40		
4	Males	7	9	6	4	5	5	4	4	5	3	52	52/100=0.520	n s
	Females	3	1	4	6	5	5	6	6	5	7	48		
5	Males	8	7	7	6	3	6	5	6	6	5	59	59/100=0.590	X^2=6.48;df=1;p<.02
	Females	2	3	3	4	7	4	5	4	4	5	41		

judges showed virtually no preference for selection of the male or female vowels as being more vocally rough and the Chi-Square values were not statistically significant.

Separate analyses were also conducted of the vocal roughness selections for the five male and five female judges in this study. Selections of the most vocally rough vowel (male or female) made by male judges and by female judges for all 50 vowel pairs are summarized in Table V. Out of the 250 forced-choice responses made, the

TABLE V

FIVE MALE AND FIVE FEMALE JUDGES' SELECTIONS BY SEX OF THE MOST VOCALLY ROUGH VOWEL PRODUCTION

Male Judges	Judges' Selections Males	Females	Female Judges	Judges' Selections Males	Females
#1	40	10	#6	34	16
#2	33	17	#7	18	32
#3	36	14	#8	24	26
#4	31	19	#9	21	29
#5	31	19	#10	28	22
Total	171	79	Total	125	125
171/250=68.50 percent			125/250=50.00 percent		
X^2=17.52;df=1;p<.001			n. s.		

five male judges selected the vowels produced by males as being more vocally rough a substantially greater proportion (68.50 percent) of the time (X^2=17.52;df=1;p<.001). Of the 250 forced-choice responses made by the five female judges, male and female produced vowels were chosen with equal frequency and the Chi-Square value was not significant (125/250=50.00 percent).

Male and female judges' selections of the most vocally rough vowels at each roughness rating level are summarized in Table VI. At roughness rating level one, both male and female judges selected male vowel productions as being more vocally rough a significantly greater proportion of the time (X^2=20.00;df=1;p<.001 for male judges; X^2=12.96;df=1;p<.001 for female judges). At roughness rating level two, male judges selected the male produced vowels as being more vocally rough a significantly greater proportion of the time (X^2=10.24;df=1;p<.01) while female judges illustrated a statistically significant preference for selection of female produced vowels as being more vocally rough (X^2=7.84;df=1;p<.01). At roughness rating levels three (X^2=19.36;df=1;p<.001), four (X^2=5.76;df=1;p<.02) and five (X^2=5.76;df=1;p<.02) male judges selected male vowel samples as being more vocally rough a significantly greater proportion of the time. At these same roughness rating levels, however, female judges showed no significant preference for selection of the male or female produced vowels as being

TABLE VI

FIVE MALE AND FIVE FEMALE JUDGES' SELECTIONS BY SEX OF THE MOST VOCALLY ROUGH VOWEL PRODUCTION AT EACH OF FIVE ROUGHNESS RATING LEVELS

Roughness Rating Level	Male Judges Roughness Selections		Male Judges Chi-Square Analysis	Female Judges Roughness Selection		Female Judges Chi-Square Analysis
1	Male Female	40 10	$X^2=20.00;df=1;p<.001$	Male Female	34 16	$X^2=12.96;df=1;p<.001$
2	Male Female	33 17	$X^2=10.24;df=1;p<.01$	Male Female	18 32	$X^2=7.84;df=1;p<.01$
3	Male Female	36 14	$X^2=19.36;df=1;p<.001$	Male Female	24 26	n.s.
4	Male Female	31 19	$X^2=5.76;df=1;p<.02$	Male Female	21 29	n.s.
5	Male Female	31 19	$X^2=5.76;df=1;p<.02$	Male Female	28 22	n.s.

more vocally rough and Chi-Square values were not significant.

Intra-Judge Reliability. To assess judges' reliability for the second listening task (LT-2), ten vowel pairs were selected randomly from the fifty vowel pairs of LT-2 and placed at the end of LT-2. Intra-judge reliability was then computed by comparing each judges' selections of the most vocally rough vowel of the ten repeat pairs with his responses to those same pairs in LT-2. Intra-judge reliability ranged from 40 percent to 90 percent for the ten judges with a mean agreement of 73.00 percent. Intra-judge reliability for male judges ranged from 70 percent to 90 percent with a mean agreement of 80.00 percent. Intra-judge reliability for female judges ranged from 40 percent to 90 percent with a mean agreement of 66.00 percent. Table VII shows the percentages of intra-judge reliability.

TABLE VII

INTRA-JUDGE RELIABILITY

Judges									
1	2	3	4	5	6	7	8	9	10
90	90	80	70	70	90	80	60	60	40

Inter-judge Agreement. Inter-judge agreement for the ten judges ranged from 32 percent to 74 percent with

a mean agreement of 56.33 percent. Table VIII shows the percentage of inter-judge agreement for the 50 pairs of vowel samples.

TABLE VIII

PERCENTAGE OF INTER-JUDGE AGREEMENT FOR FIFTY PAIRS OF VOWEL SAMPLES

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	54	46	50	60	62	60	74	64	72
2	52	58	54	52	58	68	70	64	
3	42	50	42	48	52	58	62		
4	56	56	46	58	58	66			
5	46	50	32	48	52				
6	58	54	66	52					
7	44	54	54						
8	62	48							
9	56								

Inter-judge agreement percentages were also calculated at each of the five roughness levels. Tables IX, X, XI, XII and XIII show the percentages of inter-judge agreement at each of the five roughness rating levels. At rating level one (Table IX) judges' percentages of agreement ranged from 40 percent to 90 percent, with a mean agreement of 62.66 percent. At rating level two (Table X) judges' percentages

TABLE IX

PERCENTAGE OF INTER-JUDGE AGREEMENT BY TEN JUDGES AT ROUGHNESS RATING LEVEL ONE

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	70	80	40	40	70	70	60	70	70
2	90	90	50	70	70	60	70	70	
3	70	80	40	40	70	90	60		
4	70	80	40	40	50	50			
5	60	70	40	50	60				
6	80	70	60	50					
7	40	60	60						
8	40	40							
9	70								

TABLE X

PERCENTAGE OF INTER-JUDGE AGREEMENT BY TEN JUDGES AT ROUGHNESS RATING LEVEL TWO

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	50	20	60	50	60	60	80	70	80
2	30	20	60	70	60	90	80	70	
3	60	30	70	40	70	40	70		
4	50	40	60	70	60	60			
5	10	40	40	70	60				
6	50	40	80	50					
7	40	50	50						
8	70	40							
9	50								

TABLE XI

PERCENTAGE OF INTER-JUDGE AGREEMENT BY TEN JUDGES AT ROUGHNESS RATING LEVEL THREE

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	40	50	60	70	50	60	80	70	80
2	60	70	60	50	50	80	80	50	
3	10	40	50	80	40	30	50		
4	40	50	50	50	70	80			
5	60	50	40	30	70				
6	50	40	40	60					
7	30	40	50						
8	60	50							
9	70								

TABLE XII

PERCENTAGE OF INTER-JUDGE AGREEMENT BY TEN JUDGES AT ROUGHNESS RATING LEVEL FOUR

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	40	40	30	70	70	60	70	40	60
2	40	60	50	30	40	60	50	70	
3	30	50	20	40	30	70	60		
4	50	70	40	80	50	70			
5	40	60	30	50	40				
6	60	60	70	70					
7	50	70	40						
8	90	50							
9	40								

TABLE XIII

PERCENTAGE OF INTER-JUDGE AGREEMENT BY TEN JUDGES AT ROUGHNESS RATING LEVEL FIVE

	Judges								
Judges	10	9	8	7	6	5	4	3	2
1	70	40	60	70	60	50	80	70	70
2	40	50	50	40	70	50	70	60	
3	40	50	30	40	50	60	70		
4	70	40	40	50	60	70			
5	60	30	40	40	30				
6	50	60	80	40					
7	60	50	70						
8	50	60							
9	50								

of agreement ranged from 10 percent to 90 percent, with a mean agreement of 54.88 percent. Judges' agreement at rating level three (Table XI) ranged from 10 percent to 80 percent with a mean agreement of 54.88 percent. Agreement at rating level four (Table XII) ranged from 20 percent to 90 percent with a mean agreement of 52.44 percent. Agreement at rating level five (Table XIII) ranged from 10 percent to 80 percent with a mean agreement of 55.76 percent.

II. DISCUSSION

It has been shown that judges, when listening to two synthesized vowels of approximately equal aperiodicity, tend to rate the lower pitched vowel as being more vocally rough (Wendahl, 1963). If this is true for listeners' perception of human vowel productions as well, then it might be advantageous for voice clinicians, when making vocal roughness assessments, to regard male and female speakers as two separate populations in view of the inherent pitch differences between the sexes. The findings of this investigation indicate that when judges select the most vocally rough vowel from pairs of male and female normal vowel productions (assigned the same roughness rating) they chose the male produced vowels a significantly greater proportion of the time. This result tends to support the findings of Wendahl's study of listeners' perception of vocal roughness in synthesized vowels and the contention that SSD does affect listeners' perception of vocal roughness. This finding strongly suggests the value of regarding male and female speakers as two separate populations when making vocal roughness assessments.

The phenomenon of vocal roughness may be viewed along a continuum. Normally produced vowels can be expected to be perceived as less rough than simulated rough vowels. Simulated rough vowels will be perceived as being less

rough than pathologically rough vowels. The assignment of vocal roughness ratings to normal speakers' vowel productions, however, is a difficult perceptual task. The range of vocal roughness ratings for normal vowel productions of male and female speakers is somewhat more constricted than for simulated rough vowels (Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Whitehead, 1970) and for pathologically rough vowels (Hanson, 1970). Judges in the present study made forced-choice selections of the most vocally rough vowel (male or female) of 50 vowel pairs. Vowels within each pair had been assigned identical roughness ratings (based on a five point scale) by three judges in a prior judging session. There were ten such vowel pairs at each of the five roughness rating levels. The judges' selections at each of the five roughness levels revealed that the male produced vowels were selected as being more vocally rough a significantly greater proportion of the time at roughness rating levels one, three and five. The judges illustrated virtually no preference, however, for selections of male or female vowel samples at roughness levels two and four. Thus, when judges made forced-choice selections between male and female vowel productions they chose the male produced vowels a significantly greater proportion of the time at low (rating level one), moderate (rating level three) and high (rating level five) points on the five point equal-appearing rough-

ness scale. They displayed no selection preference, however, for intermediate levels (rating levels two and four). This tends to suggest that it might be more advantageous to rate normal vowel productions on a three point scale rather than a five, six, seven or eight point interval scale.

One of the more interesting and somewhat surprising findings was that not only did the sex of the speaker affect judges' perception of vocal roughness but that the sex of the judges also appeared to influence the perception of vocal roughness. The five male judges in this study selected male produced vowels as being more vocally rough a significantly greater proportion of the time for all 50 vowel pairs and at each of the five roughness rating levels. The five female judges, on the other hand, illustrated no sex bias with respect to vocal roughness selections for the total sample. They did show a significant preference for males at roughness level one, a significant preference for females at roughness level two, but no preference at roughness levels three, four and five. Moreover, male judges illustrated substantially greater inter-judge agreement ($\overline{X}$=65.80 percent) than did the female judges ($\overline{X}$=55.00 percent). This fact coupled with the fact that the mean intra-judge reliability for male judges (80.00 percent) was substantially higher than that for female judges (66.00 percent) may indicate that the male evaluators

of vocal roughness may be more consistent, and perhaps more sex biased, in making assessments of vocal roughness than female evaluators. One might also speculate that male listeners tend to equate low-pitched male voices with the quality roughness on a psychological basis. These interpretations must be viewed cautiously, however.

Intra-judge reliability and inter-judge agreement was substantially lower for the present study than in previous similar investigations. For example, in the studies of Sansone and Emanuel (1970), Lively and Emanuel (1970) and Whitehead (1970) reported intra-judge reliability ranges were from 92 percent to 100 percent. Inter-judge agreement ranges were from 80 percent to 100 percent. These studies differed methodologically from the present investigation, however. The judges in the cited studies were asked to rate vocal roughness on a five point scale, rather than to select by means of a forced-choice response, the rougher of two vowels. In addition reliability measures, in the aforementioned studies, were based on percentages of judges agreement within one scale value which affords some margin of disagreement. The forced-choice task imposed upon the ten judges in the present study enhanced the possibility for disagreement.

It is possible that selection of the rougher of two vowels on a forced-choice basis is a more difficult perceptual task and therefore necessitates more judge training

than when making a vocal roughness scale rating. Judge training in this investigation was much less intense than that employed in previous studies (Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Whitehead, 1970; Hanson, 1970). It was felt, however, that by providing limited judge training in the forced-choice response task, conditions might more closely parallel what transpires in the clinical situation.

The possibility that the sex of the listener might influence vocal roughness perception has yet to be explored. Only one investigator (Coleman, 1971) has reported the sex of judges used in perceptual studies of vocal roughness. The fact that the male judges in the present study showed substantially higher intra-judge reliability and inter-judge agreement points out the importance in considering and possibly controlling for listener sex in future investigations of this nature. Furthermore, it may be that males and females, when making roughness ratings, react differently to particular components of the human voice. Further research would seem in order to answer this question.

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

The purpose of this study was to investigate the effects of speaker-sex-difference on listeners' perception of vocal roughness in the vowel [æ] produced by normal male and female speakers. In a previous investigation by Wendahl (1963) it was found that when listening to two synthesized vowels, of equal aperiodicity, judges tended to rate the lower pitched vowel as being more vocally rough. If this is true for listeners' perception of human vowel productions as well then it might be advantageous for voice clinicians, when making vocal roughness assessments, to regard male and female speakers as two separate populations in view of the inherent pitch differences between the sexes.

In this current investigation, pairs of vowels produced by normal adult male and female speakers were presented to 10 speech pathologists (5 males and 5 females). Each vowel pair contained one male and one female production of the vowel [æ] which had been assigned equal roughness ratings in a previous judging task. The 50 vowel pairs contained 10 pairs of vowels at each of five roughness rating levels.

The 10 judges were required to listen to each of the 50 pairs and to make a forced-choice selection of the most vocally rough production within each pair.

The findings in this study revealed that for the 50 vowel pairs the judges selected the vowels produced by males as being more vocally rough a significantly greater proportion of the time. With respect to the five roughness rating levels, judges chose the male produced vowels as being rougher a significantly greater proportion of the time at rating levels one, three and five, but, illustrated no significant preference between the sexes at rating levels two and four. Further analysis revealed that the five male judges selected the vowels produced by males as being the rougher a significantly greater proportion of the time for all 50 pairs at each of the five roughness rating levels. The five female judges, on the other hand, illustrated no significant preferences between the sexes for the 50 vowel pairs. They did show a significant preference for the males at rating level one, a significant preference for the females at rating level two but no significant preferences at rating levels three, four and five. In addition, male judges illustrated substantially greater inter-judge agreement and intra-judge reliability for this judging task than did the female judges.

The principle findings of this investigation tend to support the findings of Wendahl's (1963) study of listeners'

perception of vocal roughness in synthesized vowels and the contention that speaker-sex-difference does affect listeners' perception of vocal roughness. These findings also strongly suggest the value of regarding male and female speakers as two separate populations when making vocal roughness assessments. Further implications indicate that it might be more advantageous to employ a three point scale rather than a five point scale, when judging normal male and female vowel productions. There was also a significant male bias displayed by male judges in this study which was not observed in the female judges. What influence the sex of the judge has on the perception of vocal roughness has yet to be explored. Male judges in the present study showed substantially higher intra-judge reliability and inter-judge agreement. This points out the importance in considering and possibly controlling listener sex in future investigations of this nature. It may be that males and females, when making roughness ratings, react differently to particular components of the human voice. Further research would seem in order to answer these questions.

II. CONCLUSIONS

Future investigations of the effects of speaker-sex-difference, on listeners' perception of vocal roughness in vowels produced by normal male and female speakers, might profit from the following alterations and additions to the

design of the present study:

1. Since the judges in this investigation displayed significant preferences for selecting the male at roughness rating levels one, three and five, future studies might employ a three point scale rather than a five point scale when assessing vocal roughness in normal vowel productions.

2. The fundamental frequencies of the subjects employed in this investigation were not assessed. Future research might profit from 1) analysis of subjects' fundamental frequency differences, on a paired comparison basis, to determine the effects of these differences on listeners' perception of vocal roughness, and 2) employ subject who's fundamental frequencies encompass the traditional male and female pitch ranges, including the overlapping frequencies of the sexes.

3. Provide more extensive training for the judging task than was employed in this study.

4. The age range for subjects in this study was from 18 to 45 years of age. A more restricted age range of 18 to 30 years of age might help to control for the variable of the aging process.

BIBLIOGRAPHY

Boone, D.R., The Voice and Voice Therapy. Englewood Cliffs, New Jersey: Prentice-Hall Inc. (1971).

Bowler, N.W., A fundamental frequency analysis of harsh vocal quality. Speech Monogr., 31, 128-134 (1964).

Coleman, R.F., Some acoustic correlates of hoarseness. Master's Thesis, Vanderbilt University (1960).

_____, Effect of waveform changes upon roughness perception. Folia Phoniatr., 23:314-322 (1971).

Coleman, F. and Wendahl, R., Vocal roughness and stimulus duration. Speech Monogr., 34, 85-92 (1967).

Cooper, F., Peterson, G. and Fahringer, G., Some sources of characteristic vocoder quality. J. Acous. Soc. Amer., 29, 183 (1957).

Fairbanks, G., Voice and Articulation Drillbook (2nd Ed.). New York: Harper and Brothers (1960).

Fisher, H.B., Improving Voice and Articulation. Boston: Houghton Mifflin Company (1966).

Green, M.C.L., The Voice and Its Disorders (2nd Ed.). Philadelphia: J.B. Lippincott Co. (1964).

Hanson, W., Vowel spectral noise levels and roughness severity ratings for vowels and sentences produced by adult males presenting abnormally rough voice. Ph.D. Dissertation, University of Oklahoma (1970).

Hollien, H., A study of some laryngeal correlates of vocal pitch. J. Speech Hearing Res., 3, 52 (1960).

Isshiki, N., Yanagihara, N. and Morimoto, M., Approach to the objective diagnosis of hoarseness. Folia Phoniatr., 18, 393-400 (1966).

Lieberman, P., Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. Acous. Soc. Amer., 35, 344-355 (1963).

Lively, M.A., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. Ph.D. Dissertation, University of Oklahoma (1969).

_____, and Emanuel, F.W., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. J. Speech Hearing Res., 13, 503-517 (1970).

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

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

_____, and Thompson, C.L., Comments on the physiology of hoarseness. Arch. Otolaryng., 81, 97-102 (1965).

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

Sansone, F.E., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. Ph.D. Dissertation, University of Oklahoma (1969).

_____, and Emanuel, F.W., Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. J. Speech Hearing Res., 13, 489-502 (1970).

Shipp, T. and Huntington, D., Some accustic and perceptual factors in acute-laryngitic hoarseness. J. Speech Hearing Dis., 30, 350-359 (1965).

Third Regional Workshop on the Rehabilitation Codes and Communicative Disorders. PHS Grant No. B-3676, The National Institute of Neurological Diseases and Blindness, Communicative Disorders Research Training Committee (1967).

Timcke, R., von Leden, H. and Moore, P., Laryngeal vibrations: measurements of the glottic wave, Part 1. The normal vibratory cycle. Arch. Otolaryng., 68, 1-19 (1958).

Wendahl, R., Laryngeal analog synthesis of harsh voice quality. Folia Phoniatr., 15, 241-250 (1963).

_____, Some parameters of auditory roughness. Folia Phoniatr., 18, 26-32 (1966).

_____, Laryngeal analog synthesis of jitter and shimmer. Folia Phoniatr., 18, 98-108 (1966).

Whitehead, R.L., Some spectrographic and perceptual features of normal, vocal fry, and simulated abnormally rough vowel phonations. Ph.D. Dissertation, University of Oklahoma (1970).

Yanagihara, N., Hoarseness: investigation of the physiological mechanisms. Ann. Oto. Rhino. Laryng., 76, 472-488 (1967).

_____, Significance of harmonic changes and noise components in hoarseness. J. Speech Hearing Res., 10, 531-541 (1967).

APPENDIX A

JUDGES

JUDGES

Robert W. Blakely Ph.D.; Professor, Director of Division of Speech Pathology and Audiology, Crippled Children's Division, University of Oregon Medical School, Portland, Oregon.

Robert H. English D.Ed.; Professor and Director of Program in Speech and Hearing Science, Portland State University, Portland, Oregon.

Warren H. Fay Ph.D.; Associate Professor of Speech Pathology, Crippled Children's Division and Pediatrics, University of Oregon Medical School, Portland, Oregon.

Mary E. Gordon M.S.; Clinical Supervisor, Portland State University, Portland, Oregon.

Nancy R. Heisley Ph.D.; Volunteer, University of Oregon Medical School, Portland, Oregon.

Joseph H. Hopkins M.S.; Speech Therapist, Parkrose Public Schools, Portland, Oregon.

Jane P. Lazere M.S.; Private Practice and Consultant, Providence Hospital, Portland, Oregon.

Robert C. Marshall Ph.D.; Chief, Audiology and Speech Pathology, Veteran's Administration Hospital, Portland, Oregon.

Norma C. McAleer M.S.; Speech Clinician, Beaverton School District #48, Beaverton, Oregon.

Marie T. Rau M.S.; Speech Pathologist, Portland Veteran's Administration Hospital, Portland, Oregon.

JoAnn V. Smithpeter M.S.; Speech Pathologist, Private Practice, Portland, Oregon.

Paul F. Ventura Ph.D.; Audiologist, Kaiser Foundation Hospital, Self-Employed Speech and Hearing Clinic, Portland, Oregon.

Curtis Weiss Ph.D.; Assistant Professor of Speech Pathology, University of Oregon Medical School, Portland, Oregon.

APPENDIX B

LT-2 PAIR PRESENTATION ORDER

Order	Roughness Rating	Pair Number	Sex and Sample Number	
1	1	9	M 67	F 72
2	1	6	M 33	F 71
3	5	43	M 29	F 66
4	4	36	F 27	M 10
5	2	18	F 21	M 36
6	5	50	M 23	F 64
7	4	35	F 19	M 28
8	1	4	F 71	M 43
9	1	10	M 54	F 73
10	4	37	M 17	F 19
11	3	27	M 5	F 68
12	2	11	M 36	F 32
13	4	40	F 29	M 40
14	4	33	F 59	M 13
15	2	20	M 50	F 9
16	4	39	F 63	M 18
17	1	3	F 73	M 33
18	2	14	M 24	F 3
19	3	25	F 56	M 38
20	4	34	M 60	F 17
21	5	42	M 22	F 64
22	5	49	F 58	M 22
23	3	23	F 26	M 14
24	5	41	F 13	M 29
25	2	19	F 67	M 71
26	3	26	M 3	F 23
27	5	46	F 64	M 29
28	1	7	F 22	M 19
29	2	15	F 11	M 64
30	5	44	F 58	M 29
31	4	38	M 6	F 37
32	2	13	F 41	M 71
33	3	21	M 30	F 56
34	5	45	F 13	M 23
35	4	31	F 63	M 15
36	3	28	M 58	F 18
37	1	8	M 35	F 74
38	3	29	F 16	M 38
39	5	47	M 32	F 58
40	3	22	M 27	F 53
41	4	32	M 34	F 5
42	2	12	F 15	M 24
43	5	48	M 22	F 66
44	3	30	M 62	F 4
45	1	1	F 22	M 4
46	3	24	F 50	M 42
47	1	5	F 74	M 4
48	2	17	M 44	F 65
49	2	16	M 71	F 10
50	1	2	M 45	F 72