70-2322

LIVELY, Mary Ann, 1943-
SPECTRAL NOISE LEVELS AND ROUGHNESS SEVERITY RATINGS FOR NORMAL AND SIMULATED ROUGH VOWELS PRODUCED BY ADULT FEMALES.

The University of Oklahoma, Ph.D., 1969
Speech Pathology

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SPECTRAL NOISE LEVELS AND ROUGHNESS SEVERITY RATINGS FOR NORMAL AND SIMULATED ROUGH VOWELS PRODUCED BY ADULT FEMALES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

MARY ANN LIVELY

Oklahoma City, Oklahoma

1969

SPECTRAL NOISE LEVELS AND ROUGHNESS SEVERITY

RATINGS FOR NORMAL AND SIMULATED ROUGH

VOWELS PRODUCED BY ADULT FEMALES

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

The author wishes to express her sincere appreciation to the director of this study, Dr. Floyd W. Emanuel, for his invaluable guidance, counsel, and criticism during the planning and execution of this investigation. Appreciation is also expressed to the members of the dissertation committee, Dr. Donald T. Counihan, Dr. Walter L. Cullinan, Dr. Marjorie R. Duffy, and Dr. John F. McCoy, for their helpful suggestions during the course of this study.

Additional acknowledgment is made to Dr. John F. McCoy of the Department of Biostatistics and Epidemiology, University of Oklahoma Medical Center, for his assistance and advice concerning the statistical analysis of the data, to Dr. Frank E. Sansone and Sister Lourdes Leal for their able assistance in the collection of the experimental data, and to the individuals who participated in this study as subjects and judges.

Sincere thanks are also due to Miss Diane J. Willis, Mr. and Mrs. W. E. Duffy, and to my parents, Dr. and Mrs. C. E. Lively, for their constant encouragement, understanding, and moral support throughout this period of graduate study.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF ILLUSTRATIONS vii

Chapter

I. INTRODUCTION 1

II. REVIEW OF THE LITERATURE 5

Qualitative Features of Vocal Roughness 5
Roughness as a Voice Quality Disturbance 5
Perceptual Evaluation of Roughness 8
Acoustic Features of Vocal Roughness 10
Acoustic Wave Features 10
Spectrographic Features 16

III. DESIGN OF THE INVESTIGATION 22

Research Questions 22
Subjects 23
Speech Sample 23
Instrumentation 24
Description 24
Signal System 24
Audio Recording System 24
Wave Analyzing System 25
Playback System 27
Calibration System 27
Calibration 29
Audio Recording System 29
Wave Analyzing System 31
Procedures 32
Recording Procedure 32
Rating Procedure 33
Spectral Analysis Procedure 35

IV. RESULTS AND DISCUSSION 37

Results 37
Ratings 37
Spectral Noise Levels 40
Spectral Noise Level and Roughness Rating Relationships 48
Discussion 58

V. SUMMARY AND CONCLUSIONS 74

TABLE OF CONTENTS--(Continued)

Page

BIBLIOGRAPHY 79

APPENDIXES 85

A. Instructions to Subjects 85
B. Instructions to Judges 88
C. Percentages of Inter- and Intra-Judge Roughness Rating Agreement 91
D. Summary of Analysis of Variance and Duncan's New Multiple Range Test 93
E. Regression of Median Roughness Ratings on S-1 (100 Hz to 2600 Hz) SNL Means 95

LIST OF TABLES

Table		Page
1.	Median Roughness Ratings for Each Normal and Rough Vowel Production	38
2.	Average Median Roughness Ratings for Each Normal and Rough Vowel Production	39
3.	Normal and Rough Vowel Spectral Noise Level (SNL) Means and Standard Deviations for Twenty Female Subjects, and Differences Between the Normal and Rough Vowel SNL Means (SNLDs)	43
4.	Correlation Coefficients for Mean Spectral Noise Levels and Roughness Severity Ratings for Each Test Vowel	54
5.	Correlation Coefficients for the Multiple Regression Between Spectral Noise Levels in Each 100-Hz Spectral Section from 100 Hz to 2600 Hz and Roughness Ratings for Each Test Vowel	55
6.	Median Roughness Ratings for Eleven Judges, Roughness Ratings Estimated by the Regression Equation, and Residuals for Twenty Subjects' Normal and Rough Productions of the Vowel /ʌ/	57
7.	Percentage of Inter-judge Roughness Rating Agreement ± 1 Scale Value for Two Hundred Vowel Productions	92
8.	Percentage of Intra-judge Roughness Rating Agreement ± 1 Scale Value for Two Ratings of Fifty Vowel Productions	92
9.	Summary of the Analysis of Variance for S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means for Normal Vowels	94
10.	Duncan's New Multiple Range Test for Differences Among Normal Vowel S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means	94

LIST OF ILLUSTRATIONS

Figure Page

1. Simplified Diagram of the Audio Recording System 26

2. Simplified Diagram of the Wave Analyzing System 28

3. Simplified Diagram of the Calibration System 30

4. Spectrum of a Normal /æ/ 41

5. Spectrum of a Rough /æ/ 41

6. Noise Levels in Each 100-Hz Spectral Section Averaged over Twenty Female Subjects for Normal and for Rough Productions of the Vowel /ɑ/ 47

7. Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means and Median Roughness Ratings over Eleven Judges for Twenty Female Subjects' Normal and Rough Productions of the Vowel /u/ 49

8. Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means and Median Roughness Ratings over Eleven Judges for Twenty Female Subjects' Normal and Rough Productions of the Vowel /i/ 50

9. Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means and Median Roughness Ratings over Eleven Judges for Twenty Female Subjects' Normal and Rough Productions of the Vowel /ʌ/ 51

10. Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means and Median Roughness Ratings over Eleven Judges for Twenty Female Subjects' Normal and Rough Productions of the Vowel /ɑ/ 52

11. Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means and Median Roughness Ratings over Eleven Judges for Twenty Female Subjects' Normal and Rough Productions of the Vowel /æ/ 53

12. Regression of Median Roughness Ratings over Eleven Judges on Spectral Segment S-1 (100 Hz to 2600 Hz) Spectral Noise Level Means for Each Test Vowel 96

SPECTRAL NOISE LEVELS AND ROUGHNESS SEVERITY RATINGS FOR NORMAL AND SIMULATED ROUGH VOWELS PRODUCED BY ADULT FEMALES

CHAPTER I

INTRODUCTION

Vocal roughness is a common, perceptually-delineated voice quality disturbance encompassing harshness, hoarseness, and raspiness (68). Though normal-speaking subjects can simulate roughness at will (6, 58), this aberrant quality is often the first and most apparent symptom of laryngeal disease (29, 30, 31, 41, 49, 77, 84). Studies of its physiology suggest that roughness is associated with abnormal vocal fold movements in phonation (29, 30, 40, 44, 78, 82, 83); thus, it may reflect any of several conditions which cause disturbances in laryngeal function. For example, roughness may be a symptom of benign laryngeal neoplasms (49, 77), vocal nodules (56, 77), contact ulcers (50, 77, 84), defects in vocal mechanism innervation (46, 49, 57), transient or prolonged laryngeal infections (49, 77, 84), laryngeal malignancies (31, 49, 77, 84), and many other pathologies. Roughness may also reflect an emotional disorder in the absence of vocal mechanism pathology (2, 14, 25, 41, 75, 76).

Individuals presenting vocal roughness commonly require voice therapy and/or medical treatment to achieve a healthy vocal mechanism and a socially adequate voice. In such instances, the clinician's assessment of the voice disturbance is likely to be critical to the development of an effective rehabilitative treatment program. Because of the current limited availability of more objective procedures, the clinical evaluation of vocal roughness may be predicated primarily on the clinician's perception. It is generally recognized, however, that clinicians vary in their ability to evaluate vocal roughness reliably. The clinician's prior training and experience, his auditory acuity, the testing environment, and the voice sample obtained are but a few of the factors which may influence the perceptual evaluation of a rough vocal quality. To facilitate more detailed and meaningful clinical voice evaluations than are presently practical, information is needed regarding the critical relationships between perceived vocal roughness and associated physiologic and acoustic phenomena.

Because the larynx is not readily accessible to observation and because vocal fold movements in phonation are rapid, the direct assessment of laryngeal function in rough phonation requires specialized instrumentation and measurement techniques. In spite of the technical difficulties associated with data collection, investigators employing high-speed cinematography (44, 78, 82), stroboscopy (26, 67), and pneumotachography (29, 82) have contributed information regarding the physiology of vocal roughness. In contrast to physiologic voice assessment, the acoustic voice signal is easily sampled, but quantitative acoustic data useful in the clinical evaluation of vocal roughness are currently limited. Recently, however, investigations have been completed which suggest that further study of its acoustic features may contribute to the understanding of vocal roughness and provide data which are clinically useful.

Investigations of the acoustic waves and spectra of synthesized complex sounds and human phonations (12, 30, 40, 47, 58, 79, 80, 81, 82, 83) have helped to define relationships between perceived vocal roughness and acoustic voice features. Spectrographic investigations of rough phonation are of particular interest. Several studies (30, 47, 58, 82, 83) have suggested that the level of inharmonic or noise components in a vowel is related to the roughness of the vowel as it is perceived by listeners. Specifically, the elevation of spectral noise components tends to be associated with an increase in perceived roughness. On the basis of his observation of this relationship between spectral noise and roughness, Nessel (47) indicated that hoarseness can be "... defined and differentiated when using a suitable method of frequency analysis." More recently, Isshiki, Yanagihara, and Morimoto (30) and Yanagihara (82, 83) have related the degree of perceived vowel hoarseness to differences in the intensity and frequency location of noise components in sonagrams of the vowels. Four types of hoarseness were defined on this basis. The relationships between perceived roughness and spectral noise levels for vowels are, however, as yet incompletely defined. Previously, instrumentation limitations largely precluded detailed quantitative measures of spectral noise components for vowels produced normally or with vocal roughness. In a recent study of the phonations of adult males, however, Sansone (58) demonstrated that the level of noise components could be measured in narrow-band (3-Hz) spectra of normal and simulated rough productions of selected vowels. He reported, moreover, that the measures of vowel spectral noise levels were highly correlated with quantified listener judgments of vowel roughness. Similar data for female speakers are currently unavailable, but would appear useful. Because their voices tend to differ from those of males (18, 25, 34, 37, 39, 52), data for female subjects are relevant to a somplete description of the acoustic features of vocal roughness. Further, such data may aid in the development of new techniques for the assessment of vocal roughness. It was the purpose of this investigation, therefore, to study, for adult female speakers, the spectral noise levels associated with normal and rough vowel productions and the relationships between the vowel spectral noise levels and perceived roughness.

CHAPTER II

REVIEW OF THE LITERATURE

Clinical interest in vocal roughness has produced extensive descriptive information regarding the qualitative features of roughness but relatively little quantitative data regarding its physiologic and acoustic correlates. Recent instrumentation developments have facilitated objective investigations of the acoustic correlates of roughness, but information regarding the relationships which obtain between measures of vowel spectral noise and perceived vowel roughness is as yet unavailable for adult female speakers.

The purpose of this investigation was to assess quantitatively noise components in narrow-band (3-Hz) spectra of normal and simulated rough vowels phonated by adult females and to study possible relationships between vowel spectral noise levels and judgments of vowel roughness. The literature reviewed as background for the present study is presented under two major headings: (a) Qualitative Features of Vocal Roughness and (b) Acoustic Features of Vocal Roughness.

Qualitative Features of Vocal Roughness

Roughness as a Voice Quality Disturbance

Voice quality disturbances may be distinguished from other communication problems on the basis of the criterion recommended by Fairbanks. Fairbanks said (16, p. 202):

> The test of the existence of a voice quality disorder is whether or not the quality that is heard is independent of phonemes or, in other words, whether or not the phenomenon heard can be superimposed upon a good example of a voiced sound.

The further differentiation of one voice quality disorder from another is usually predicated on perceived differences among qualities. As Van Riper (75) observed, however, the terms used by writers to specify different quality disturbances "are as numerous as adjectives." The existence of overlapping and ambiguous terminology regarding voice disorders has led to confusion among both clinicians and researchers. Such confusion may be attributed in part to the infrequency with which the validity of clinically observed quality differences is tested. For example, it is commonly assumed that some listeners, because of extensive clinical experience, can validly perceive fine differences among similar voice quality disturbances which less sophisticated listeners do not perceive clearly. Were researchers to assume the validity of such perceived quality disorders solely on the basis of the presumed "authority" of the listener, however, they might well search for acoustic and physiologic correlates of voice disturbances which lack a valid and reliable perceptual existence. There appears to be a need in research, therefore, for care in specifying the voice quality studied and for avoidance of a facile acceptance of conventional or "authoritative" differentiations of voice qualities.

The quality disturbances "harshness" and "hoarseness," for example, are among those frequently differentiated in textbook descriptions of voice disorders (14, 16, 17, 32, 76). Though most authorities seem to agree that both are associated with phonatory phenomena, i.e., they represent the perception of acoustic features produced by the laryngeal sound generator (14, 17, 32, 44), descriptions of the perceptual features of these quality disturbances vary. Curtis (14), Fairbanks (16, 17), Van Riper and Irwin (76) and others have described harsh quality as an unpleasant, noisy, rasping sound which is associated with excessive strain and hypertension of the throat and laryngeal muscles. Hard glottal attack is also mentioned as a characteristic of harsh voices (14, 76). Hoarseness, on the other hand, has been defined by several writers (3, 32, 75, 76) as a voice quality disturbance which combines the features of breathiness and harshness. Moore (43) has attempted to divide hoarseness further into three distinct types: "dry," "wet," and "rough."

Thurman (69) found, however, that when connected speech samples from speakers with clinical voice disorders were presented to sophisticated, trained listeners for classification of the perceived quality disturbance, there was a spread of classification judgments for most samples. Moreover, he reported comparatively little agreement among judges on a single classification for a given voice sample. When the voice samples were presented to a selected group of the trained judges for a second classification, a mean over judges of only 51% of the samples received a classification identical to the first judgment. Thurman noted that the greatest confusion was associated with the terms "harsh" and "hoarse," although listener confusion was also evident between such terms as "hoarse" and "breathy," and "harsh" and "strident." These findings seem to support the use of a descriptive term which encompasses these inconstantly differentiated quality disturbances.

In recognition of the need for a term specifying such a general category of quality disorders, "vocal roughness" has been defined (53, 68) as an impairment of voice function encompassing harshness, hoarseness, raspiness, and similar perceptually-delineated voice abnormalities. Rapid acceptance of this term by researchers investigating voice quality (12, 58, 81) lends support to the assumption that it is both meaningful and useful.

Perceptual Evaluation of Roughness

The research assessment of perceived voice quality disturbances commonly involves the use of rating scale procedures. Early studies employing such procedures demonstrated that various voice quality disorders could be reliably scaled for severity on the basis of listener judgments of the disturbance. Perceptual voice quality scaling has thus been useful in more recent studies investigating acoustic features associated with the perception of vocal roughness.

Where roughness scaling has been employed, the sample presented to listeners for rating, the particular rating procedure employed, and the qualifications of the judges have varied considerably across studies. For example, listener judgments of roughness have been obtained for synthesized complex waves (12, 79, 80, 81), isolated vowels (11, 54, 55, 58, 82, 83), CVC syllables (54, 55), and connected speech samples (6, 40, 59, 60, 69). A variety of rating procedures including paired-comparisons (12, 79, 80, 81) and five- and seven-point equal-appearing intervals scales (11, 55, 58, 59, 60) have been employed. Various four-point ranking scales have also been used. Lieberman (40) and Yanagihara (82, 83), for example, had their judges rate voices on a four-point scale such as: (1) normal, (2) slightly hoarse, (3) moderately hoarse, and (4) severely hoarse. The individuals serving as listeners in previous studies have differed with respect to background and professional experience. Typically, judges for investigations of vocal roughness have been selected from one of four categories: undergraduates in general speech courses (12, 79, 80, 81), trained speech pathologists including graduate students (11, 54, 55, 58, 60, 69), college teachers of speech (6, 7), or otolaryngologists (82, 83).

Generally, previous investigations suggest that listeners are able to evaluate vocal roughness reliably. Sherman and Linke (60), for example, obtained a Pearson *r* of .97 when median scale harshness values from one group of judges were correlated with medians of the ratings from a second group of judges for ninety connected speech samples. Sansone (58) obtained a Pearson *r* of .96 when medians of his judges' first and second roughness ratings of a "reliability sample" of fifty vowels were compared. Rees (55) reported a Pearson *r* of .90 for median scale harshness ratings from repeated ratings of 100 syllables.

The validity of perceptual voice ratings is seldom questioned because voice quality is by definition what the listener perceives it to be. Sherman (59), however, has indicated that judgments of harshness, i.e., roughness, in connected speech may be confounded by the presence of misarticulations or other quality deviations. The effect of misarticulations on roughness ratings may be minimized, however, when isolated, sustained vowels are evaluated. Other procedures also help to assure the validity of the roughness ratings. Because listeners may be unable to differentiate among certain voice qualities, e.g., as between harshness and hoarseness (69), the use of a more general perceptual category may enhance judgment validity. Listener confusion regarding the quality to be judged may be minimized when voice samples are rated for roughness because roughness encompasses voice qualities which tend to be difficult to differentiate.

As a further procedure, speech samples representative of the experimental samples to be evaluated may be presented to the judges at the beginning of the judgment session and practice in scaling roughness may be provided (55, 58, 60). This serves a dual purpose: the examples presented define rough voice quality operationally and provide the judges with training in scaling the specific quality disturbance under investigation. In addition, anchor stimuli may be presented as examples of the range of quality disturbance which the judges will be asked to evaluate (55, 58, 59, 60).

The controlled listening conditions employed in most investigations also help to assure valid roughness ratings by minimizing extraneous and distracting stimuli. The high percentages of inter-judge roughness rating agreement (± 1 scale value) reported by Sansone (58) lend support to the assumption that vocal roughness in vowels can be rated validly because they indicate that trained judges tend to agree among themselves regarding the presence and severity of this voice disturbance.

Acoustic Features of Vocal Roughness

Acoustic Wave Features

Investigations of vocal mechanism function in subjects with rough voices have suggested that a rough vocal quality is associated with abnormal aperiodicity in the vocal fold vibratory pattern. Similar disturbances in the phonatory acoustic waves of individuals presenting roughness have also been noted. In his analysis of the fundamental frequency characteristics of harsh vocal quality, Bowler (6, 7) examined oscillographic recordings of connected speech containing both harsh and nonharsh segments. He reported that the most striking feature of the harsh portions was the occurrence of "frequency breaks," i.e., abrupt changes in the periods of consecutive cycles. These frequency breaks occurred in both upward and downward directions on the frequency scale and were typically one octave in extent. In no instance did the segments perceived as nonharsh contain these atypical frequency characteristics. In addition, harsh segments evidenced relatively low mean fundamental frequencies and a wider range of fundamental frequency values than nonharsh segments. Coleman (11), in his study of sustained vowels produced by pathologically hoarse subjects, did not find "pitch" or "frequency breaks" as large as one octave. However, he did find aperiodic cycle-to-cycle frequency variations of less than one octave which he termed "voice breaks." Such voice breaks were prominent in the waveforms of his subjects' phonations, and their presence was closely associated with perceived hoarseness severity. Coleman found only slight differences between the median fundamental frequencies of normal and hoarse voices. Shipp and Huntington (61) found that "voice breaks" were infrequent in the acoustic waves produced by subjects presenting laryngitic hoarseness. When present, however, such breaks were said to contribute greatly to the perception of hoarseness. In contrast to Bowler's (6, 7) findings, Shipp and Huntington found a severely restricted range of fundamental frequencies for their hoarse subjects and no significant difference between either mean or median fundamental frequencies for hoarse and normal voices.

Lieberman (40) measured small, rapid variations in the durations of successive cycles, or "pitch perturbations," in oscillographically recorded acoustic waves produced by speakers with normal and pathological larynges. He found that perturbations of less than 0.5 ms were typical of isolated vowels phonated normally. However, perturbations for mildly and moderately rough phonations generally exceeded those for normal phonations. Lieberman noted that, when hoarseness was severe, the acoustic wave of phonation became markedly aperiodic and individual cycles within it were not discernable, thus preventing measurement of pitch perturbations. A "perturbation factor," indicating the percentage of occurrence in the acoustic wave of perturbations equal to or greater than 0.5 ms, was computed for each speaker. This factor was found to be sensitive to the size and location of laryngeal growths, provided the growths did not interfere with vocal fold closure. When a growth prevented complete vocal fold adduction, the acoustic waveform of phonation was "filled in" and the perturbation factor could not be determined. Connected speech samples produced by the clinically hoarse subjects were rated for hoarseness by a panel of listeners. Four categories were utilized in the rating: (1) normal, (2) slightly hoarse, (3) moderately hoarse, and (4) extremely hoarse. Lieberman found that the average ratings of the speech samples did not relate meaningfully to the underlying laryngeal pathology. Moore and Thompson (44) computed correlation coefficients to relate the perturbation factors and hoarseness ratings reported by Lieberman and found a "moderate positive correlation." To examine the relationship between pitch perturbations and the periodicity of vocal fold movements, Lieberman used a sound-synchronized high-speed camera to photograph the vocal folds of a normal male speaker during phonation. The sample of phonation was simultaneously recorded on magnetic tape. A comparison of the subject's acoustic and glottal waveforms indicated that perturbations in the acoustic wave reflected irregularities in the pattern and periodicity of vocal fold movements. On the basis of his findings, Lieberman suggested that measurement of pitch perturbations in phonation might be a useful diagnostic procedure for the early detection of laryngeal pathologies. In a later study, Moore and Thompson (44) found that random variations in the length of adjacent cycles characterized the acoustic waves of their two hoarse subjects' phonations. The periods of successive cycles were considerably more variable for the subject presenting the more severe hoarseness.

The studies reviewed above suggest that random variations in the periods of successive cycles in the voice wave are associated with the perception of vocal roughness. Additional information regarding the relationships between such frequency variations and judged roughness has been contributed through study of acoustic analogs of phonation. To investigate the degree of signal aperiodicity required for listener judgments of roughness, Wendahl (79, 80, 81) 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 $\pm$ 1 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 (12) confirmed the finding that random cycle-to-cycle frequency variations (jitter) in a synthesized complex wave were related to perceived signal roughness.

Random variations in the amplitude of successive cycles of the acoustic wave have also been associated with perceived vocal roughness. In an early study, Moore and von Leden (45) noted that amplitude variations in successive glottal waves were characteristic of abnormal phonations. Coleman (11) later reported "amplitude breaks" in his hoarse subjects' phonations. To investigate more thoroughly the effects of cycle-to-cycle amplitude variation (shimmer) around a median amplitude, Wendahl (80, 81) synthesized complex waves containing shimmer and presented them to judges for rating of perceived roughness. He found that an increase in signal shimmer was associated with listener judgments of an increase in signal roughness. Apparently, a systematic assessment of the relationships between acoustic intensity variations and the roughness of human phonations has not been made.

While jitter and shimmer in synthesized complex signals appear to be related to perceived signal roughness (12, 79, 80, 81), only a few relationships between acoustic wave features and vocal roughness in human phonation have been clearly established. The studies of Coleman (11) and Lieberman (40) suggest that acoustic features relating to perceived roughness in phonation may not be readily identified from inspection of the acoustic wave envelope alone.

The duration of the signal may also affect perceived roughness. Sherman and Linke (60) reported that high vowels, which are relatively short in duration, are perceived as less harsh than low vowels, which are relatively long in duration. Brubaker and Dolpheide (8) also suggest that signal duration may affect perceived hoarseness. To provide more quantitative data regarding relationships between stimulus duration and perceived vocal roughness, Coleman and Wendahl (12) synthesized complex acoustic stimuli which contained both jitter and periodic components. The duration of jitter segments within the total stimulus, as well as the amount of jitter around a median frequency, could be varied. As the duration of jitter segments increased from .16 to .80 seconds in a signal of finite length, more severe roughness was perceived by the listeners. However, a trading relationship between the duration and the amount of jitter in a signal was also revealed. For example, a stimulus containing large cycle-to-cycle frequency variations within a short jitter segment was judged as less rough than a stimulus containing a jitter segment of longer duration and smaller jitter excursion.

Several studies (8, 11, 55, 58, 59, 60, 61) have investigated the relative roughness of various vowels. In general, vowel roughness appears to be related to relative tongue height in vowel production. For example, Sherman and Linke (60), in their study of listener judgments of harshness severity for different vowels, found that high vowels were perceived as less harsh than low vowels. Brubaker and Dolpheide (8), Rees (55), and Sansone (58) have reported similar findings. In contrast, Coleman (11) found no significant correlation between perceived hoarseness and the high-versus-low classification of vowels. He did note, however, that /i/ ranked lowest and /æ/ ranked highest in perceived hoarseness, although differences among the vowels were slight.

To generalize, identifiable features of the acoustic voice signal have been related to the perception of vocal roughness and to its relative severity. There are, however, few clearly defined relationships between features of the acoustic wave envelope and vocal roughness. This suggests that inspection of the acoustic wave envelope alone may not fully reveal the acoustic correlates of rough voice quality. A more detailed analysis of the acoustic spectra of rough phonations may be useful.

Spectrographic Features

While studies concerned with the spectral characteristics of vocal roughness have been few, the results of previous investigations suggest that further spectrographic study of rough voice quality is needed for a more complete understanding of the acoustic features of vocal roughness. As early as 1941, Carhart (9) utilized a manually tunable heterodyne analyzer to study the spectra of tones produced by a model larynx. For several vibratory conditions of the model, predominantly inharmonic spectra were obtained. Because the resulting auditory stimuli closely approximated clinical hoarseness, he speculated that the perception of hoarseness may be related to inharmonics in the acoustic spectrum.

With the advent of more refined instrumentation, including the widely used automatic heterodyne analyzer commonly known as the Kay Sonagraph, more detailed investigation of the spectral characteristics of complex acoustic stimuli was possible. Thurman (69) utilized a Sonagraph to make a wide-band filter analysis of vowels produced by individuals with various voice quality disorders. The purpose of his study was to establish phonographically recorded scales of severity for breathy, nasal, hoarse, harsh, thin, and strident voice quality disturbances. Moreover, he sought to determine if the type and severity of quality disturbance perceived by listeners could be related to specific acoustic features observable in sonagraphic records. He attempted to measure formant frequency locations, formant bandwidths, formant amplitudes, and the level of inharmonics in vowel sonagrams. Although his listeners categorized and scaled the various voice qualities perceptually, Thurman noted that differentiation of different voice quality types, determination of the degree of voice disturbance, and measurement of inharmonic energy levels were impossible from his sonagraphic data. Formant frequencies tended to vary from their normal locations in the sonagrams of pathological phonations, but such changes were not consistent within any type of deviant voice quality. Moreover, no relationship between the amount and direction of formant frequency shift and perceived severity of voice quality disturbance could be demonstrated. For example, formant frequency shifts for both F_1 and F_2 occurred in the sonagrams of hoarse, harsh, and breathy voices, but their occurrence was not consistent across all samples. In addition, Thurman reported that the presence or absence of inharmonic partials in hoarse vowels could not be determined from his sonagraphic records.

The findings of Laguaite and Waldrop (38) suggest that changes in spectrographic features may be related to perceived severity of voice quality deviation. In their investigation of patients' voices before and after therapy, they observed that changes toward normalcy in deviant voices seemed to be related to spectral changes in vowels. They reported that spectral changes rather than changes in fundamental vocal frequency accompanied improved vocal quality.

Recently, sound spectrographic analyses have yielded more specific information regarding the acoustic properties of vocal roughness. Isshiki, Yanagihara, and Morimoto (30) and Yanagihara (82, 83) investigated harmonic and noise components in the spectra of sustained vowels phonated by subjects with laryngeal pathologies. The vowel recordings were presented to three otolaryngologists who rated them for slight, moderate, and severe hoarseness. Sonagrams and amplitude sections were made from the recordings of the subjects' phonations. The sonagrams were classified into four categories on the basis of the frequency region and intensity of the spectral noise components, since the range and level of spectral noise appeared to vary with perceived hoarseness severity. A correlation coefficient of .65 between the spectrographic type and perceived severity of hoarseness was obtained. In slight hoarseness, noise components were found to be mixed with the harmonics in the formant regions, particularly in the second and third formants. As the severity of hoarseness increased, noise components began to appear in the high frequency region above 3000 Hz. In the most severe hoarseness, the harmonics in the main formant ranges were totally obscured or replaced by elevated noise components. Yanagihara (82) also observed a relationship between the degree of abnormality in the spectrographic findings and the extent 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 supplement his findings for human phonations, Yanagihara (83) synthesized hoarseness by mixing recorded normal vowels with band-pass filtered noise. Again, as the noise components intruded in the formant ranges and as the high frequency harmonic components were obscured by noise, the severity of perceived hoarseness increased. On the basis of these results, Yanagihara suggested that the major spectrographic features of hoarseness include: (a) noise components in the main formant of each vowel; (b) high frequency noise components above 3000 Hz; and, (c) loss of high frequency harmonic components.

Using a "sound-frequency spectrograph of high selection," Nessel (47) compared frequency-by-amplitude spectra of sustained vowels produced by hoarse speakers to those of normal speakers. The spectra of hoarse vowels were characterized by a reduction of harmonic energy below 5000 Hz and substitution of noise components which were modulated by the vowel formants. Additional noise components were also evident in the upper frequency range of the spectra above 5000 Hz.

In the Kay Sonagraph, the bandwidth of the "narrow" filter is 45 Hz. It appears possible that filter bandwidths this wide may obscure spectral information important to the perception of roughness. Further, the magnitude of noise components evident in the spectra of rough vowels and relationships between spectral noise levels and perceived roughness severity cannot be determined easily from sonagraphic records (69). To overcome these limitations, Sansone (58) recently employed a graphic wave analyzer to produce very narrow-band (3-Hz) frequency-by-amplitude spectra of the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/, produced normally and with simulated vocal roughness by adult males. Spectral noise components in the vowel productions were quantified by measuring in dB SPL the lowest observable peak of energy in each 100-Hz spectral section from 100 Hz to 8000 Hz of each vowel spectrum. Spectral noise level measures were then compared to judges' ratings of vowel roughness. Sansone found that, although both normal and rough vowels were characterized by measurable spectral noise, rough productions evidenced higher spectral noise levels than normal vowel productions. For all vowel productions, spectral noise was most prominent in the lower spectral frequencies and tended to decrease in the higher frequencies. For each test vowel, spectral noise level means were found to be highly correlated with median roughness ratings for that vowel. A multiple linear regression equation was found to predict, with small residuals, each vowel production's median roughness rating from its 100-Hz section spectral noise levels from 100 Hz to 2600 Hz. Sansone's findings for males indicate that noise components in the spectra of rough and normal vowels can be quantified and that vowel spectral noise levels, particularly in the low frequency regions (100 Hz to 2600 Hz), are highly related to perceived vowel roughness.

It appears possible that relationships between acoustic spectra and vocal roughness which are found for male speakers may not obtain exactly for female speakers. Sex-associated fundamental vocal frequency differences among speakers are known to affect differentially spectral features of vowels phonated normally (5, 18, 23, 37, 52). It has also been suggested that fundamental vocal frequency differences between the sexes may affect relationships between judged vocal roughness and acoustic spectral features of phonation. Wendahl (79), for example, found that synthesized complex acoustic stimuli containing large frequency variations (± 10 cycles) around a median fundamental frequency of 200 Hz were judged to be less rough than stimuli containing small frequency variations (± 2 cycles) around a median fundamental of 100 Hz. On the basis of these results, Wendahl hypothesized that two speakers presenting equal cycle-to-cycle aperiodicity in phonation but widely different fundamental frequencies, e.g., a male and a female, would not be judged equally rough. Specifically, he suggested that the male voice would be judged as more deviant.

In summary, the results of previous spectrographic investigations suggest that suprafundamental energy distribution is different in vowels produced by speakers presenting vocal roughness and by speakers presenting no noticeable voice quality disturbance. Specifically, the elevation of noise components is reported to be a spectral feature of vocal roughness. Quantitative data regarding the magnitude of noise components in narrow-band (3-Hz) spectra of rough vowels are available for adult male but not for adult female speakers. Further investigation is needed to clarify the relationships between vowel spectral noise and the perception of vowel roughness for females. This study was designed to obtain such objective data for adult female subjects.

CHAPTER III

DESIGN OF THE INVESTIGATION

The intent of the present study was to investigate vowel spectral noise levels and relationships between the vowel noise levels and judges' ratings of vowel roughness. Twenty normal-speaking adult females individually phonated five selected vowels both normally and with simulated vocal roughness at one intensity. Each vowel production was recorded on magnetic tape for further analysis. The productions were then re-recorded in random order and presented to a panel of eleven trained judges who rated each for roughness. Medians of the judges' ratings provided an index of each production's roughness. The recording of each vowel production was also analyzed to produce a narrow-band (3-Hz) frequency-by-amplitude acoustic spectrum. To provide a quantitative index of vowel spectral noise, the lowest observable peak of energy in each of seventy-nine successive 100-Hz spectral sections from 100 Hz to 8000 Hz was measured in each vowel spectrum. The research questions and the methods employed in this study are discussed in the following sections.

Research Questions

The following research questions regarding the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/ were investigated for adult female speakers:

1. What is the relative roughness of the vowels produced normally and with simulated vocal roughness?

2. What are the spectral noise features of normal and of rough productions of the vowels?

3. What are the relationships between spectral noise levels and judges' ratings of roughness for each of the vowels?

Subjects

Twenty normal-speaking female adults, selected primarily on the basis of their ability to perform the experimental task, served as subjects in this investigation. Each subject produced selected vowels under both normal and rough phonatory conditions. Thus, each was available to serve as her own control. The investigation was limited to adult females to provide homogeneity of the subject sample with respect to vocal pitch. Each potential subject was evaluated by a trained speech pathologist to insure that those selected presented normal voice quality and speech. Subjects ranged in age from twenty-two to thirty-one years. The age range was thus limited to preclude variations in voice and speech associated with adolescence or advanced age.

Speech Sample

The speech sample for this study was composed of the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/ individually sustained by each subject at one intensity. Subjects produced each of the vowels first normally and then with simulated vocal roughness. Each production was sustained for seven seconds at 75 dB SPL ($\pm$ 1 dB) re: 0.0002 dyne/cm^2 at a mouth-to-microphone distance of six inches. This intensity level was selected after preliminary trials indicated that it was a comfortable level for production of both normal and simulated rough vowels. The vowels selected represent various positions on the traditional vowel triangle (35) and permitted analysis of the findings with respect to tongue height and placement within the oral cavity. This analysis was of interest because previous studies (8, 55, 58, 59, 60) suggest that vowel roughness may be related to tongue height during vowel production. Isolated vowels sustained at one intensity provided samples suitable for narrow-band acoustic spectral analysis.

This study investigated vocal roughness simulated by normal-speaking subjects because it was considered advantageous to exercise close control of differences among speakers in the two phonatory conditions. Moreover, findings reported by Bowler (6) and Sansone (58) suggest that judges generally do not distinguish perceptually between simulated and clinical vocal roughness. Thus, it was thought that the data regarding simulated vocal roughness obtained in this study might be useful in understanding clinical vocal roughness.

Instrumentation

The instrumentation utilized in this investigation included: (a) a signal system; (b) an audio recording system; (c) a wave analyzing system; (d) a playback system; and (e) a calibration system.

Description

Signal system. Subjects were signalled to initiate and terminate test vowel phonation by the illumination of two panel lights controlled by a simple electro-mechanical cam timer which was activated by the experimenter.

Audio recording system. The audio recording system consisted of: (a) a sound level meter (General Radio, Type 1551-C) with an attached non-directional piezoelectric ceramic microphone (General Radio, PZT Type 1560-P3); (b) a magnetic tape recorder (Ampex, Model AG 440); and (c) a monitoring amplifier (Bruel and Kjaer, Type 2603).

Its design specifications indicated that the frequency response of the PZT microphone was flat ($\pm$ 1 dB) from 20 Hz to 8000 Hz when at a 70° angle of incidence to the sound source. The sensitivity of the microphone was -60.3 dB re: 1 v/microbar. The sound pressure level at the PZT microphone was indicated by the sound level meter with an average signal-to-noise ratio of at least 66 dB in octave bands from 20 Hz to 10,000 Hz. The magnetic tape recorder had a flat frequency response ($\pm$ 2 dB) from 40 Hz to 12,000 Hz with a signal-to-noise ratio of at least 65 dB when operated at a tape speed of 15 ips.

In data collection, the output of the sound level meter was led directly to the input of the tape recorder. The output of the recorder was led to the monitoring amplifier which served as a vocal-intensity-monitoring meter. Subjects observed the monitoring amplifier's calibrated voltmeter and adjusted their vocalizations to the experimentally required intensity. A simplified diagram of the audio recording system is presented in Figure 1.

Wave analyzing system. The experimental vowels were reproduced from tape loops by the tape recorder described above and were introduced as complex electrical signals into a graphic wave analyzer (General Radio, Type 1910-A) for spectrum analysis. The graphic wave analyzer was composed of (a) a wave analyzer (General Radio, Type 1900-A), (b) a graphic level recorder (General Radio, Type 1521-B), (c) a drive unit (General Radio, Type 1521-P10-B), and (d) a link unit (General Radio, Type 1900-P3). The drive and link units mechanically coupled the wave analyzer to the graphic level recorder to permit automatic recording of the level of components in the complex electrical signal under analysis. The move-

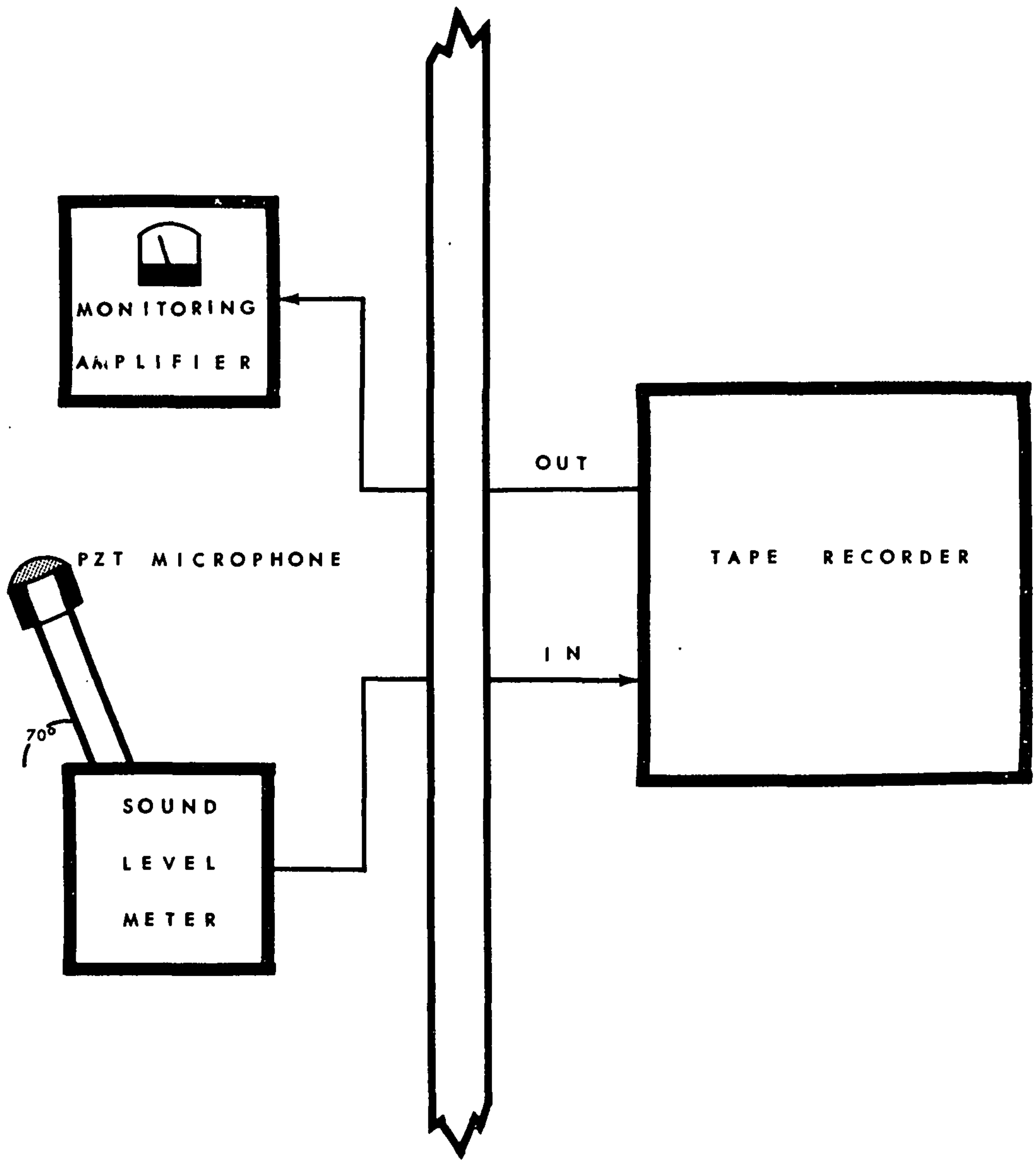

Figure 1.—Simplified diagram of the audio recording system.

ment of the chart paper in the recorder was synchronized with the wave analyzer's frequency-tuning dial.

The analyzer's frequency range was from 0 Hz to 54,000 Hz, with frequency accuracy to 50,000 Hz of ½% of its frequency dial reading plus 5 Hz. When used in its 3-Hz bandwidth mode, the instrument functioned as a continuously tunable narrow-band filter with the intensity of frequency components in a complex signal at least 30 dB down at ± 6 Hz and at least 60 dB down at ± 15 Hz from center frequency. The analyzer's signal-to-noise ratio was at least 75 dB.

The voltage output of the wave analyzer was proportional to the intensity of the frequency components in a 3-Hz band of the complex signal under analysis and served as an electrical input to the graphic level recorder. The recorder was equipped with an 80-dB input potentiometer designed for accuracy within ± 1% of full-scale decibel value. The level recorder's output was proportional to the logarithm of changes in its input and, hence, was linear in decibels. A simplified diagram of the wave analyzing system is presented in Figure 2.

Playback system. The playback system used for presentation of the recorded vowels to judges for rating of roughness severity included a dual-channel magnetic tape recorder (Ampex, Model 354), with a flat frequency response (± 2 dB) from 40 Hz to 12,000 Hz at a tape speed of 15 ips, an amplifier (Sherwood, Model S9900A), and a loud-speaker (Altec, Model 844A).

Calibration system. Components employed in instrument calibration included a pure tone oscillator (Hewlett-Packard, Model ABR 200) which drove a loud-speaker (Altec, Model 844A), a sound level meter (General Radio, Type 1551-C), a pulse generator assembly (Tektronix, 160

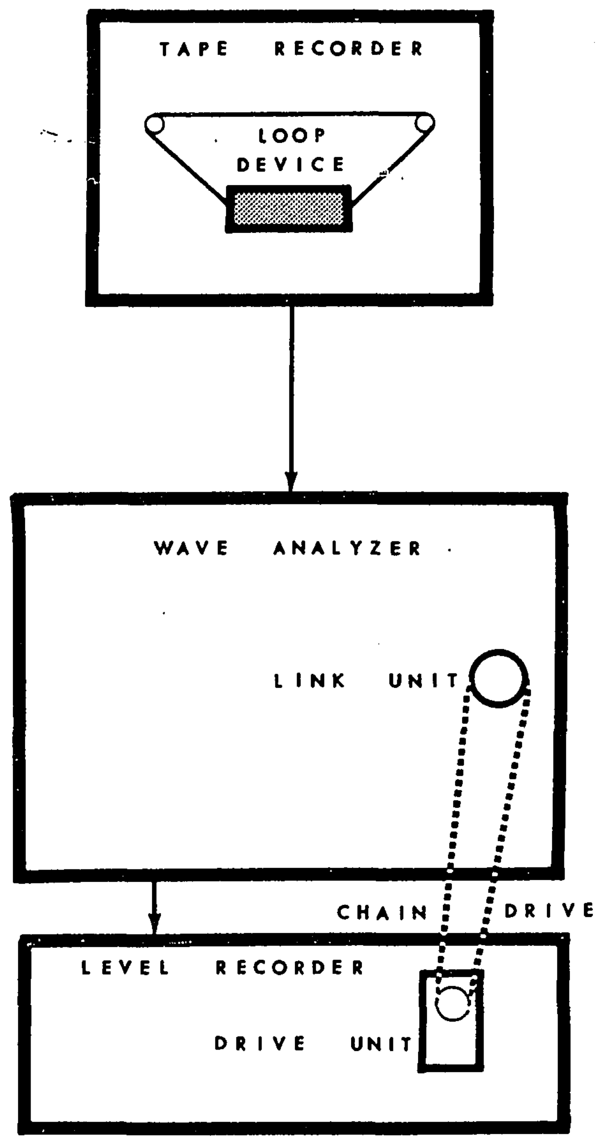

Figure 2.--Simplified diagram of the wave analyzing system.

Series), and a manufacturer-calibrated condenser microphone assembly (Bruel & Kjaer, Type 2603). A simplified diagram of the calibration system is presented in Figure 3.

Calibration

Audio recording system. Prior to the collection of data, the magnetic tape recorder was checked and aligned by an audio engineer. The vocal-intensity-monitoring section of the audio recording system was calibrated to indicate when the subject's vocal intensity reached the required intensity of 75 dB SPL. The monitoring amplifier's voltmeter was used as the subject's intensity indicator. To calibrate this meter, a 1000 Hz reference tone produced by the oscillator was led to the loud-speaker. The sound level meter PZT microphone was placed at a 70° angle of incidence to and two feet in front of the loud-speaker in an acoustically-isolated room. The intensity of the pure tone was adjusted until it produced a 75 dB SPL sound level meter deflection. The output of the sound level meter was connected directly to the input of the tape recorder, and the recorder was adjusted for a -2 dB deflection of its VU meter in response to the 75 dB SPL input. The output of the recorder was led to the monitoring amplifier and the amplifier's input potentiometer was adjusted for a 15 dB deflection of its voltmeter in response to the 75 dB SPL input. This deflection on the amplifier's voltmeter was marked with an easily visible arrow to indicate the level each subject was required to maintain during experimental vowel production. The reference tone was then recorded and played back to adjust the audio recorder's reproduce level to match its record level. Thus, vowel phonations producing a 75 dB SPL indication on the vocal-intensity-monitoring voltmeter produced a

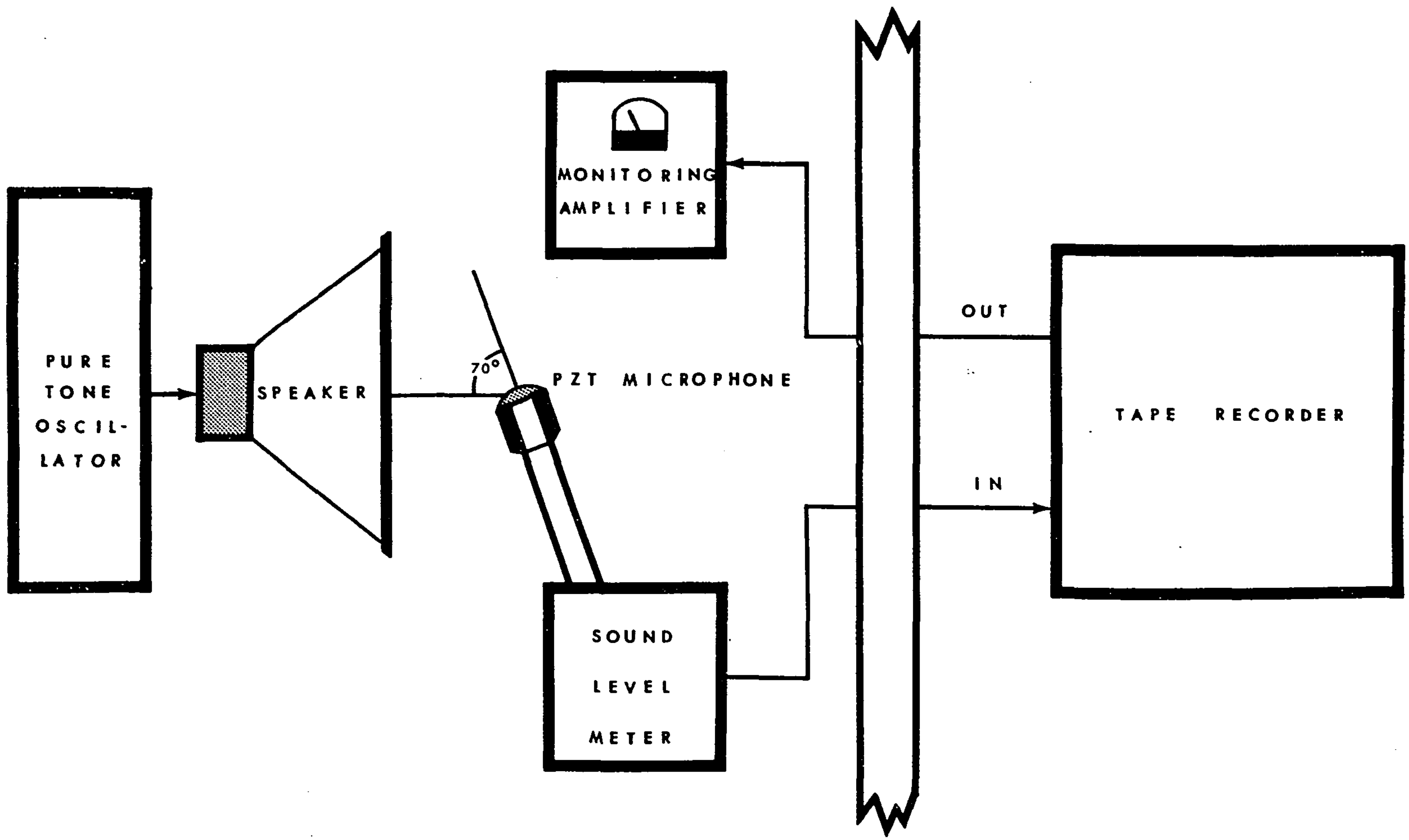

Figure 3.—Simplified diagram of the calibration system.

-2 dB deflection on the recorder's record VU meter. When recorded and played back, the vowels produced a -2 dB deflection on the recorder's reproduce VU meter.

The frequency response of the PZT microphone used in this study was reported to be flat ($\pm$ 1 dB) from 20 Hz to 8000 Hz. Immediately before and after collection of the experimental data, the PZT microphone frequency response was checked against the flat ($\pm$ .5 dB from 20 Hz to 10,000 Hz) response of a calibrated condenser microphone and was found to be within the manufacturer's specifications.

Wave analyzing system. Before each use, the graphic wave analyzer was adjusted for minimal carrier frequency intensity at low frequencies and checked for frequency analysis accuracy within design specifications for the equipment. After this initial adjustment, intensity calibration was effected by introducing a recorded 75 dB SPL 1000-Hz reference tone into the wave analyzer from the tape recorder. The gain of the analyzer and the pen excursion of the graphic level recorder were adjusted for a 75 dB SPL indication on the graph paper.

To check the frequency calibration of the wave analyzer and coupled graphic level recorder, a pulse train of known repetition rate produced by the pulse generator assembly was introduced into the graphic wave analyzing system. Accurate plotting of the fundamental and harmonics of the pulse train indicated satisfactory frequency calibration of the system from 0 Hz to 8000 Hz. To assure stability of frequency calibration, a daily check was made of the graphic wave analyzer's response to a series of reference tones of known frequency produced by the pure tone oscillator. The frequency response of the coupled audio recording and wave analyzing systems, excluding the PZT microphone, was checked and found to be flat ($\pm$ 2 dB) from 50 Hz to 12,000 Hz.

Procedures

The experimental procedures in this study included: (a) recording of the subjects' productions of the test vowels, (b) presentation of the recorded vowel productions to judges for roughness rating, and (c) derivation of frequency-by-amplitude vowel spectra.

Recording Procedure

All vowel samples were recorded in an acoustically-isolated, two-room testing suite with a low ambient noise level at the Speech and Hearing Center, University of Oklahoma Medical Center. The test room contained the subject's chair, the sound level meter with its attached PZT microphone, the vocal-intensity-monitoring amplifier, and the signal lights to indicate the beginning and end of test vowel phonation. The adjoining control room contained the magnetic tape recorder and the cam timer which controlled signal light timing.

Each subject was first familiarized with the experimental procedures and was then seated in the examination chair. The chair's headrest was adjusted vertically for comfort and a headstrap was employed to minimize changes in the subject's position with respect to the microphone during recording. The microphone was placed at a 70^{o} angle of incidence to and six inches in front of the subject's mouth. The monitoring amplifier was positioned to allow the subject to observe readily the intensity of her phonations. The investigator remained in the test room with each subject throughout the recording session to monitor the intensity of vowel productions and to cue the subject with printed cards bearing the vowel to be phonated. A copy of the instructions read to the subjects is presented in APPENDIX A.

After being familiarized with the speech material, the subject practiced phonating each vowel at 75 dB SPL while observing the monitoring amplifier's voltmeter. The subject also practiced timing her phonations with the signal lights until she was able to sustain each vowel for seven seconds while maintaining the required intensity. Upon completion of the training, the rough and normal experimental vowel productions were recorded. For each subject, the order of vowels was randomized within normal and within rough phonatory conditions. The test vowels were produced first normally and then with simulated vocal roughness. This procedure eliminated from normal productions the influence of vocal abuse associated with roughness simulation. Each vowel phonation was carefully monitored by the investigator. If the subject did not produce the appropriate vowel, did not maintain the required intensity, or did not suitably effect vocal roughness, the trial was repeated until an acceptable performance was achieved.

Rating Procedure

The two hundred rough and normal productions were randomized by means of tape dubbing for presentation to judges. Eleven judges, all graduate students in speech pathology, independently assessed the recorded vowel samples for roughness. The judgments were made in an acoustically-isolated room with the judges seated in a semicircle facing the loudspeaker. The recorder used to reproduce the vowels was located in an adjoining control room. An intercom system between the two rooms enabled the judges to indicate if they wished a particular vowel sample repeated. The judges were instructed to listen to each vowel and to rate indepen-

dently the degree of [illegible]
appearing intervals scale [illegible]
represented most severe [illegible]

Prior to the [illegible]
nary rating of all vowel [illegible]
senting "1" and two [illegible]
These vowels were played [illegible]
began to provide them [illegible]
roughness rating scale. [illegible]
sented in APPENDIX B.

The listening [illegible]
in length. The speech [illegible]
of fifty vowels each [illegible]
final series of vowels [illegible]
those presented earlier [illegible]
bility. Median scale [illegible]
computed.

A Pearson r was [illegible]
judges' first and second [illegible]
sample. An r of .98 was [illegible]
rating agreement within 1 [illegible]
productions were computed. [illegible]
Judge 1. Percentages of inter-[illegible]
scale value for the experimental [illegible]
computed. The lowest percentage, [illegible]
tings of Judge 7 were compared [illegible]
of these procedures are presented [illegible]

presented in APPENDIX A.

After being familiarized with the speech material, the subject practiced phonating each vowel at 75 dB SPL while observing the monitoring amplifier's voltmeter. The subject also practiced timing her phonations with the signal lights until she was able to sustain each vowel for seven seconds while maintaining the required intensity. Upon completion of the training, the rough and normal experimental vowel productions were recorded. For each subject, the order of vowels was randomized within normal and within rough phonatory conditions. The test vowels were produced first normally and then with simulated vocal roughness. This procedure eliminated from normal productions the influence of vocal abuse associated with roughness simulation. Each vowel phonation was carefully monitored by the investigator. If the subject did not produce the appropriate vowel, did not maintain the required intensity, or did not suitably effect vocal roughness, the trial was repeated until an acceptable performance was achieved.

Rating Procedure

The two hundred rough and normal productions were randomized by means of tape dubbing for presentation to judges. Eleven judges, all graduate students in speech pathology, independently assessed the recorded vowel samples for roughness. The judgments were made in an acoustically-isolated room with the judges seated in a semicircle facing the loudspeaker. The recorder used to reproduce the vowels was located in an adjoining control room. An intercom system between the two rooms enabled the judges to indicate if they wished a particular vowel sample repeated. The judges were instructed to listen to each vowel and to rate independently the degree of roughness perceived in each. A five-point equal-appearing intervals scale in which "1" represented least severe and "5" represented most severe roughness was used.

Prior to the listening session, the investigator made a preliminary rating of all vowel productions. Four vowel productions, two representing "1" and two representing "5" on the rating scale, were selected. These vowels were played several times to the judges before actual rating began to provide them with a common reference for the extremes of the roughness rating scale. A copy of the instructions to judges is presented in APPENDIX B.

The listening session was approximately two and one-half hours in length. The speech material to be rated was presented in five series of fifty vowels each with ten minute rest periods between series. The final series of vowels was composed of productions selected randomly from those presented earlier and were included to evaluate intra-judge reliability. Median scale values of the judges' ratings for each vowel were computed.

A Pearson $\underline{r}$ was then computed to relate the median values of the judges' first and second ratings of the fifty vowels in the reliability sample. An $\underline{r}$ of .98 was obtained. Percentages of intra-judge roughness rating agreement within 1 scale value for two ratings of the fifty vowel productions were computed. The lowest percentage, 96%, was obtained for Judge 1. Percentages of inter-judge roughness rating agreement within 1 scale value for the experimental sample of two hundred vowels were also computed. The lowest percentage, 92%, was obtained when the vowel ratings of Judge 7 were compared to those of Judges 1 and 3. The results of these procedures are presented in APPENDIX C. The intra- and inter-judge reliability indicated by these data appeared adequate for this investigation.

Spectral Analysis Procedure

Tape loops were constructed from the magnetic tape recordings of each rough and normal vowel produced by each subject. The loops were two seconds in duration (tape speed of 15 ips) and were constructed from a central portion of the vowel recording displaying a uniform intensity of 75 dB SPL ($\pm$ 1 dB) as monitored from the recorder's VU meter. Initial and terminal vowel inflections were omitted. The vowel loops were played individually into the graphic wave analyzer to obtain a 3-Hz bandwidth frequency-by-amplitude spectrum of each vowel. The analyzer was operated at a paper speed of 0.5 inches per minute and a writing speed of 20 inches per second for recording the vowel spectra. These settings insured adequate resolution of data analyzed in the 3-Hz bandwidth mode and minimized writing stylus overshoot. The time required to produce the spectrum of an individual vowel production under the described conditions was thirty-two minutes.

To determine the level of test room and instrumental system noise present during collection of the experimental data, recordings of test-chamber noise were made at various times during the day. Tape loops constructed from these recordings were analyzed to produce 3-Hz bandwidth room noise spectra. The high peak of energy in each 100-Hz spectral section from 100 Hz to 8000 Hz was measured in each spectrum. Low noise levels were evident at all frequencies throughout the total spectral frequency range. The average noise level from 100 Hz to 8000 Hz for the room noise spectra was -3 dB SPL. There was negligible variation in system noise at different times during the day as evidenced by similar low noise levels in the spectra of all test-chamber noise recordings.

As a quantitative index of vowel spectral noise levels, the lowest observable peak graphic level recorder stylus marking in each 100-Hz section of each vowel spectrum was measured in dB SPL. Seventy-nine measures, one for each successive 100-Hz section from 100 Hz to 8000 Hz, were obtained from the spectrum of each vowel. Stylus marking overlap, in some instances, may have precluded measurement of the true low peak in a 100-Hz section; however, measurement of the lowest observable peak provided a numerical index of vowel spectral noise levels.

To determine the reliability of the spectral analysis procedure, three consecutive spectra were made from one vowel tape loop. Spectral noise levels averaged over the frequency range 100 Hz to 8000 Hz did not vary more than $\pm$ .2 dB across the three spectra. Differences among noise level means for comparable 1000-Hz segments of the spectra ranged from $\pm$ .4 dB to $\pm$ 1.2 dB. Thus, the vowel spectrum analysis procedure appeared to be sufficiently reliable for this study.

CHAPTER IV

RESULTS AND DISCUSSION

Results

This study investigated spectral noise levels and judges' ratings of roughness for selected vowels. Twenty normal-speaking adult females individually produced each of the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/ both normally and with simulated vocal roughness at one intensity. Randomized tape recordings of each vowel production were rated for roughness on a five-point equal-appearing intervals scale by a group of eleven trained judges. The recording of each production was also analyzed to produce a 3-Hz frequency-by-intensity spectrum of its acoustic components. As an index of vowel spectral noise levels, the lowest observable peak of energy in each 100-Hz section, from 100 Hz to 8000 Hz, of each vowel spectrum was measured in dB SPL. The spectral noise levels and medians of the roughness ratings for each vowel production were then related.

Ratings

Table 1 presents the median of the eleven judges' roughness ratings for each of the five vowels produced normally and with simulated roughness by each of the twenty subjects. This table shows that a higher median scale value was obtained for each vowel produced with simulated

TABLE 1

MEDIAN ROUGHNESS RATINGS FOR EACH NORMAL AND ROUGH VOWEL PRODUCTION

	Vowels									
	/u/		/i/		/ʌ/		/ɑ/		/æ/	
Subject	N	R	N	R	N	R	N	R	N	R
1	1.11	4.95	1.29	5.00	1.05	2.92	1.19	4.42	1.29	5.00
2	1.00	4.81	1.19	5.00	1.11	4.95	1.00	4.06	1.11	4.95
3	1.19	4.89	1.11	5.00	1.71	4.81	2.00	4.81	1.81	5.00
4	1.00	4.71	1.11	4.95	1.05	4.89	1.00	4.59	1.29	5.00
5	1.00	4.29	1.59	4.08	1.29	4.89	1.05	4.81	1.29	4.89
6	1.89	4.89	1.81	4.59	1.95	4.95	1.29	4.06	2.00	3.29
7	1.19	3.60	1.89	3.94	1.94	3.60	2.00	3.42	2.05	4.40
8	1.00	4.00	1.59	3.89	1.05	2.95	1.11	3.75	1.71	4.81
9	1.11	3.19	1.11	3.95	1.81	4.71	1.19	2.08	1.92	2.75
10	1.42	3.59	1.86	2.29	1.42	4.29	1.05	2.60	2.00	4.40
11	1.00	4.89	1.19	4.89	1.71	3.42	1.00	2.75	1.86	3.62
12	1.05	4.11	1.42	4.00	1.89	4.81	1.42	4.71	1.71	4.11
13	1.42	4.14	1.81	3.38	1.19	3.29	2.00	3.86	1.75	3.88
14	1.19	3.29	1.19	3.11	1.11	3.60	1.05	2.29	1.29	3.81
15	1.11	4.71	1.29	4.59	1.89	4.59	1.05	4.08	1.05	3.60
16	1.05	3.42	1.19	3.25	1.00	3.71	1.00	2.42	1.11	2.92
17	1.11	4.95	1.71	3.92	1.59	3.25	2.11	3.94	1.89	4.20
18	1.05	3.60	1.11	3.42	1.05	3.86	1.05	4.33	1.00	3.59
19	1.00	3.00	1.11	2.75	1.05	3.80	1.71	3.00	1.05	3.06
20	1.00	2.42	1.11	3.86	1.42	3.89	1.05	2.06	1.89	4.40

vocal roughness than for its normal counterpart, indicating that the subjects were uniformly successful in simulating a voice quality judged to be rough. The range of median scale values for normal vowel productions was from 1.00 to 2.11, while the range for rough productions was from 2.06 to 5.00. The greater range for rough productions was expected because the degree of roughness simulated by the subjects was not controlled. Table 2 presents median roughness ratings, averaged over the twenty subjects, for normal and for rough productions of each vowel.

TABLE 2

AVERAGE MEDIAN ROUGHNESS RATINGS FOR EACH
NORMAL AND ROUGH VOWEL PRODUCTION

	Average Median Roughness Rating	
Vowel	Normal	Rough
/u/	1.14	4.07
/i/	1.38	3.99
/ʌ/	1.41	4.06
/ɑ/	1.32	3.60
/æ/	1.55	4.08

Inspection of the average median ratings for normal productions reveals that the high vowel /u/ is rated least rough and the low vowel /æ/ is rated most rough. The scale value separation between the extremes is .41. The average ratings for the remaining vowels /i/, /ʌ/, and /ɑ/ are between those for /u/ and /æ/ and are distributed over a limited range. Considering rough productions, the averages for the low vowels /ɑ/ and

/æ/ represent the extremes of the distribution of ratings, with /ɑ/ rated less rough than /æ/. The scale value separation between rough vowel extremes is .38. With the exception of /ɑ/, the average median ratings for the rough productions of each vowel differ maximally by only .09 scale value.

Spectral Noise Levels

To illustrate the frequency-by-amplitude acoustic spectra obtained in this investigation, examples of a rough and a normal spectrum are presented in Figures 4 and 5. These spectra are for the vowel /æ/ as produced by Subject 10. For both rough and normal /æ/ productions, harmonic and noise components tend to be most prominent in the lower frequency regions and to diminish toward higher frequencies. The normal /æ/ spectrum presented in Figure 4 is characterized by prominent harmonics toward the lower spectral frequency range and by relatively low-level noise components between the identifiable harmonics. The highest spectral noise levels are evident in formant locations where harmonic amplitudes are highest. In the high frequency range, the harmonics are obscured by noise. A comparison of the rough and normal spectra reveals several differences. A feature of the rough /æ/ spectrum presented in Figure 5 is the elevation of noise components at all frequencies from 0 Hz to 8000 Hz. In addition, most of the harmonic partials have been obscured. A decrease in amplitude of the harmonics which remain identifiable in the very low frequency range of the rough spectrum can also be observed. Spectral features similar to those for /æ/ were evident in the spectra of all test vowels.

Spectral noise levels in each vowel production were estimated by

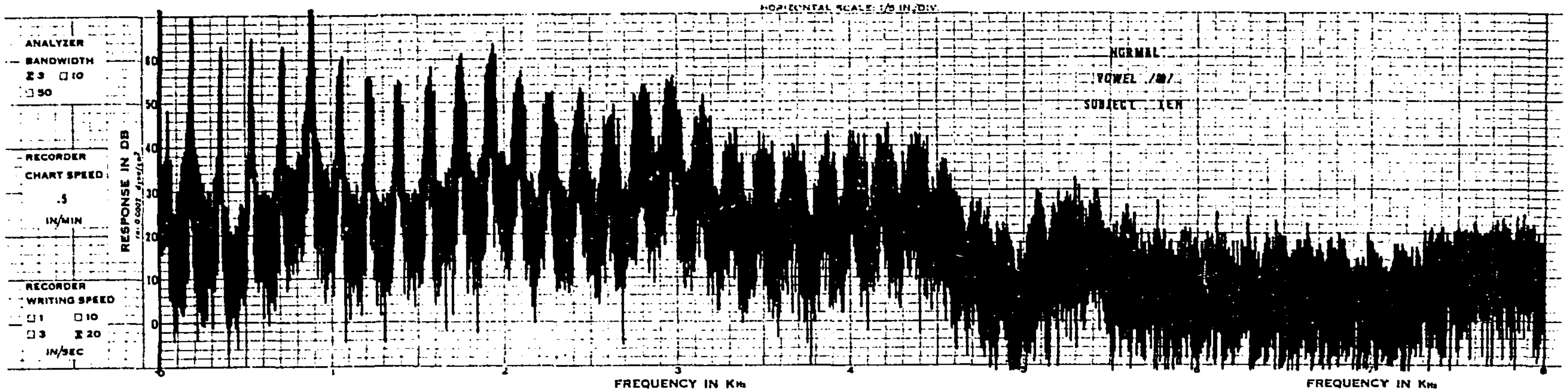

Figure 4.—Spectrum of a normal /æ/.

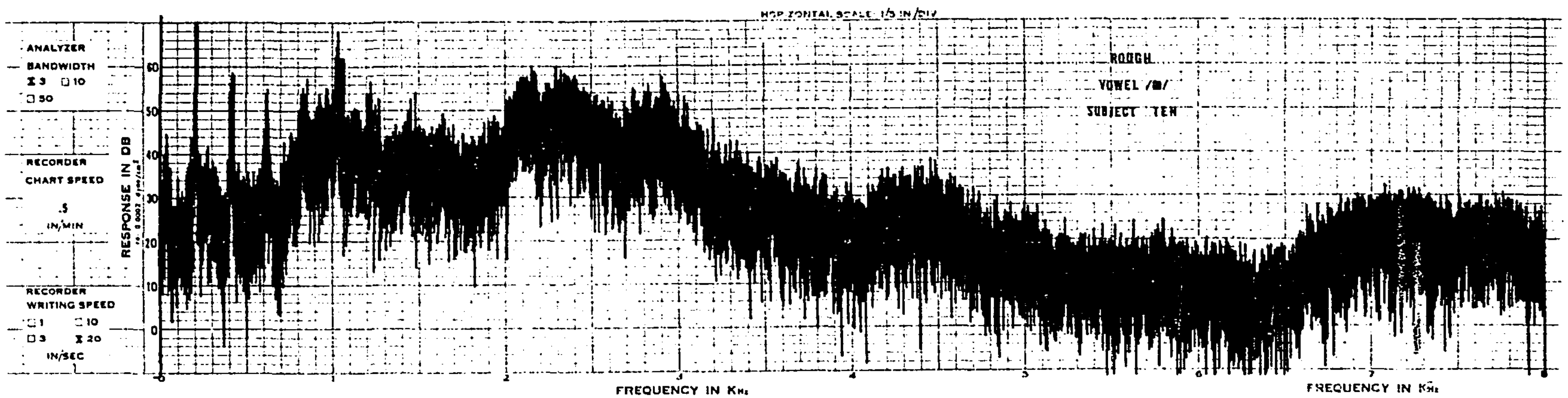

Figure 5.—Spectrum of a rough /æ/.

measures in dB SPL of the lowest observable peak of energy in each 100-Hz spectral section from 100 Hz to 8000 Hz. Selected functions of the spectral measures also considered for each vowel spectrum were: (1) the mean of spectral noise measures from 100 Hz to 2600 Hz; (2) the mean of measures from 2600 Hz to 5100 Hz; (3) the mean of measures from 5100 Hz to 8000 Hz; (4) the mean of measures from 100 Hz to 5100 Hz; and (5) the mean of measures from 100 Hz to 8000 Hz. These functions were selected because they differ in the extent to which they include the formant frequencies of the vowels (18, 52) and because noise levels associated with these frequency ranges may relate differently to perceived vowel roughness (47, 58, 82, 83).

To facilitate presentation of the spectral findings, the total spectral frequency range studied (100 Hz to 8000 Hz) is referred to as the TSR; a spectral noise level is referred to as an SNL; and, segments of the total spectral frequency range (TSR) are referred to as SSs. The spectral segments (SSs) studied are referred to as segment one (S-1), 100 Hz to 2600 Hz; segment two (S-2), 2600 Hz to 5100 Hz; segment three (S-3), 5100 Hz to 8000 Hz; and segment four (S-4), 100 Hz to 5100 Hz.

Spectral noise level (SNL) means and standard deviations for normal and rough vowels are presented in Table 3. The means are over all subjects, over the TSR, and, separately, over each SS. It can be seen in Table 3 that the mean SNLs for rough productions of each test vowel exceed those for its normal productions. This trend is observable in each SS as well as in the TSR. When vowel /i/ is omitted from consideration, the order of the vowels with respect to their SNLs is the same within each SS and the TSR for both normal and rough productions. Excluding /i/, the rough and normal vowels could be ranked with respect to increasing mean

TABLE 3

NORMAL AND ROUGH VOWEL SPECTRAL NOISE LEVEL (SNL) MEANS AND STANDARD DEVIATIONS FOR TWENTY FEMALE SUBJECTS, AND DIFFERENCES BETWEEN THE NORMAL AND ROUGH VOWEL SNL MEANS (SNLDS)

Vowel	Normal SNL Mean	Standard Deviation	Rough SNL Mean	Standard Deviation	SNLD
Spectral Segment S-1 (100 Hz to 2600 Hz)					
/u/	13.0	3.2	31.6	5.9	18.6
/i/	10.6	3.0	26.7	5.2	16.1
/ʌ/	24.9	2.7	38.3	3.0	13.4
/ɑ/	22.8	2.6	35.9	5.2	13.1
/æ/	27.8	2.2	39.1	3.2	11.3
Spectral Segment S-2 (2600 Hz to 5100 Hz)					
Vowel	Normal SNL Mean	Standard Deviation	Rough SNL Mean	Standard Deviation	SNLD
/u/	4.7	4.6	23.5	7.8	18.8
/i/	21.5	4.3	36.4	5.0	14.9
/ʌ/	15.9	4.0	30.3	4.1	14.4
/ɑ/	13.7	4.5	28.2	6.0	14.5
/æ/	21.3	4.6	35.2	5.7	13.9
Spectral Segment S-3 (5100 Hz to 8000 Hz)					
Vowel	Normal SNL Mean	Standard Deviation	Rough SNL Mean	Standard Deviation	SNLD
/u/	- 0.3	3.8	13.3	8.5	13.6
/i/	9.5	4.8	21.1	7.8	11.6
/ʌ/	8.5	5.8	20.6	6.5	12.1
/ɑ/	6.5	4.3	17.0	5.9	10.5
/æ/	12.0	5.1	22.2	7.2	10.2

TABLE 3--Continued

Spectral Segment S-4 (100 Hz to 5100 Hz)					
Vowel	Normal SNL Mean	Standard Deviation	Rough SNL Mean	Standard Deviation	SNLD
/u/	8.9	3.6	27.6	6.5	18.7
/i/	16.1	3.1	31.5	4.7	15.4
/ʌ/	20.4	2.7	34.3	3.0	13.9
/ɑ/	18.2	2.8	32.1	5.3	13.9
/æ/	24.6	2.5	37.2	3.9	12.6

Total Spectral Range (100 Hz to 8000 Hz)					
Vowel	Normal SNL Mean	Standard Deviation	Rough SNL Mean	Standard Deviation	SNLD
/u/	5.8	3.0	22.8	6.5	17.0
/i/	13.9	2.5	28.0	5.0	14.1
/ʌ/	16.4	3.0	29.7	3.5	13.3
/ɑ/	14.3	2.6	27.0	5.0	12.7
/æ/	20.4	2.5	32.2	4.8	11.8

SNLs: /u/, /ɑ/, /ʌ/, and /æ/. This order obtaines with no reversals for the TSR and for each SS. The SNL means for the vowel /i/ vary considerably more in absolute and in relative magnitude across spectral segments than those for the other vowels. The variations in the means for /i/ across the SSs tend to be similar, however, for both normal and rough productions. With respect to the SNL means for the other vowels, the mean for /i/ is lowest in relative magnitude in S-1 (100 Hz to 2600 Hz), highest in relative magnitude in S-2 (2600 Hz to 5100 Hz), and between the extremes in the remaining SSs and the TSR.

Within spectral segment S-1 (100 Hz to 2600 Hz), normal vowel productions could be ranked with respect to increasing mean SNLs: /i/, /u/, /ɑ/, /ʌ/, and /æ/. A randomized complete-block analysis of variance was employed to determine whether significant differences existed among the normal vowel S-1 SNL means. A significance level of .05 was selected for this analysis. The obtained F value of 179.08 (df = 4, 76) was statistically significant at the .05 level, indicating significant differences among the S-1 mean SNLs of normal vowels. A Duncan's New Multiple Range Test was employed to locate the differences among the means detected by the analysis of variance. A .05 significance level was also set for this analysis. The results of this test indicated that all normal vowel S-1 SNL means were significantly different at the .05 level. A summary of these statistical analyses is presented in APPENDIX D.

Additional trends may be observed in Table 3. For rough and for normal productions, a decrease in SNL means is evident from S-1 to S-3 for all test vowels except /i/. From S-1 (100 Hz to 2600 Hz) to S-2 (2600 Hz to 5100 Hz), the decrease is approximately eight dB. From S-2 to S-3 (5100 Hz to 8000 Hz), the decrease is approximately seven dB for normal productions and eleven dB for rough productions.

Differences between the normal and rough vowel SNL means are also presented in Table 3. It can be seen that the SNLDs for each vowel tend to be similar in magnitude across each of the SSs and the TSR. The vowel /u/ is characterized by the largest SNLD and /æ/ by the smallest in each SS and the TSR. The differences between SNLs for rough and for normal productions of the test vowels are illustrated for the vowel /ɑ/ in Figure 6. Figure 6 presents a plot of the individual SNLs in each 100-Hz spectral section of the TSR averaged over the twenty subjects, with rough and normal productions of the vowel plotted separately. The average SNLs for each 100-Hz spectral section of the rough productions of /ɑ/ exceed those of the normal productions throughout the TSR. Differences in SNL means between rough and normal productions of each vowel, similar to those for /ɑ/, were observed for all test vowels.

Table 3 also presents standard deviations for rough and normal vowel SNL means. For both normal and rough productions, the least SNL variability is associated with S-1 (100 Hz to 2600 Hz) and the TSR (100 Hz to 8000 Hz), while the greatest variability is associated with S-2 (2600 Hz to 5100 Hz) and S-3 (5100 Hz to 8000 Hz). For normal and for rough productions, no test vowel is characterized by consistently smaller standard deviations than any other. Within each SS and the TSR, the standard deviations for the rough vowel productions exceed those for their normal counterparts. The greater variability associated with SNLs for rough productions probably reflects the fact that the degree of vowel roughness simulated by the subjects was not controlled.

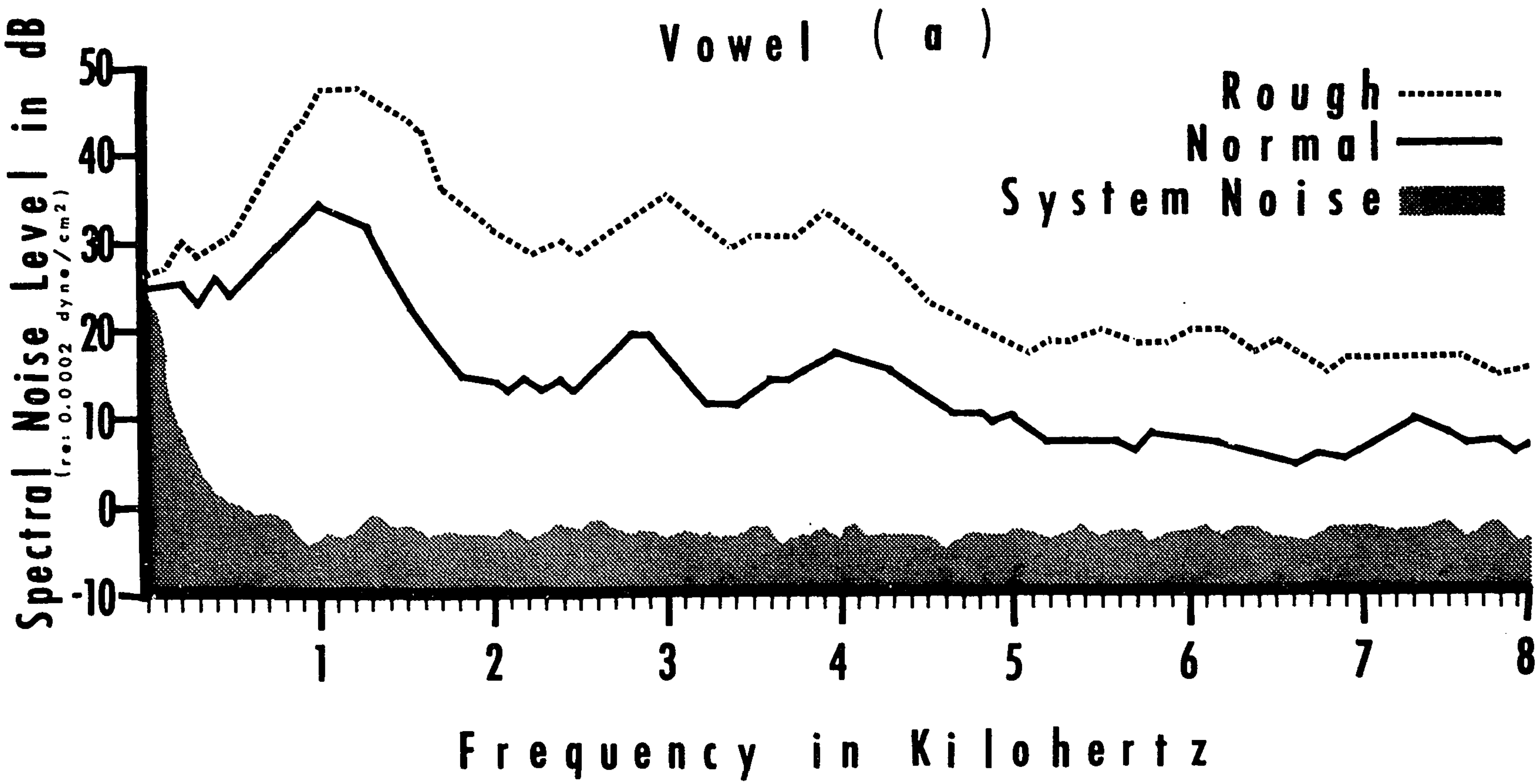

Figure 6.—Noise levels in each 100-Hz spectral section averaged over twenty female subjects for normal and for rough productions of the vowel /ɑ/.

Spectral Noise Level and Roughness Rating Relationships

One objective of the present study was to explore the relationships between vowel spectral noise levels and roughness severity ratings. To investigate these relationships, scatter diagrams of vowel SNL means and median roughness ratings were plotted. The SNL means for the TSR and each SS were considered separately in these plots. All the diagrams suggested a positive relationship between mean SNLs in each SS and the TSR and median roughness ratings for each of the test vowels. In general, as the roughness of each vowel increased, its spectral noise level tended to increase. This relationship was most evident, however, in the low frequency SSs where data point scatter tended to be less than in the higher frequency SSs. Data point scatter was greatest in the diagrams for S-3 (5100 Hz to 8000 Hz). Mean spectral noise levels and perceived roughness appeared to be most directly related when S-1 SNL means were compared to median roughness ratings. The scatter diagrams for S-1 (100 Hz to 2600 Hz) for each test vowel are presented in Figures 7 through 11. The data points in the diagram for each vowel represent SNLs averaged over S-1 and median roughness ratings for each subject's rough and normal vowel productions.

To explore further the degree of association between mean spectral noise levels and perceived roughness for the vowels, a correlation statistic (Pearson r) was employed. A significance level of .05 was selected for this correlation. Table 4 presents correlation coefficients indicating the degree of association between mean SNLs and median roughness ratings for each test vowel. Because the coefficients obtained for S-4 (100 Hz to 5100 Hz) and the TSR (100 Hz to 8000 Hz) are not statisti-

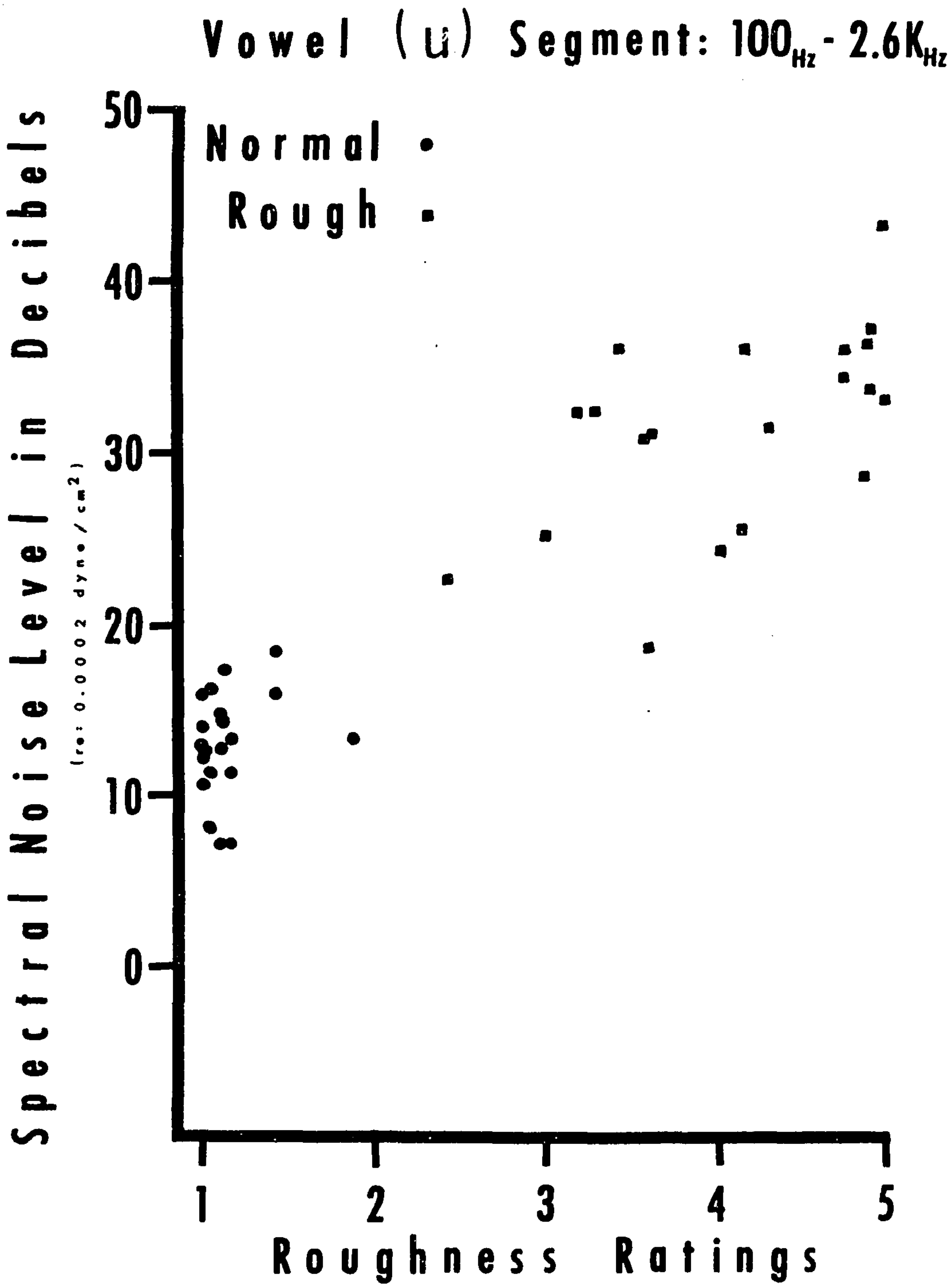

Figure 7.—Spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means and median roughness ratings over eleven judges for twenty female subjects' normal and rough productions of the vowel /u/.

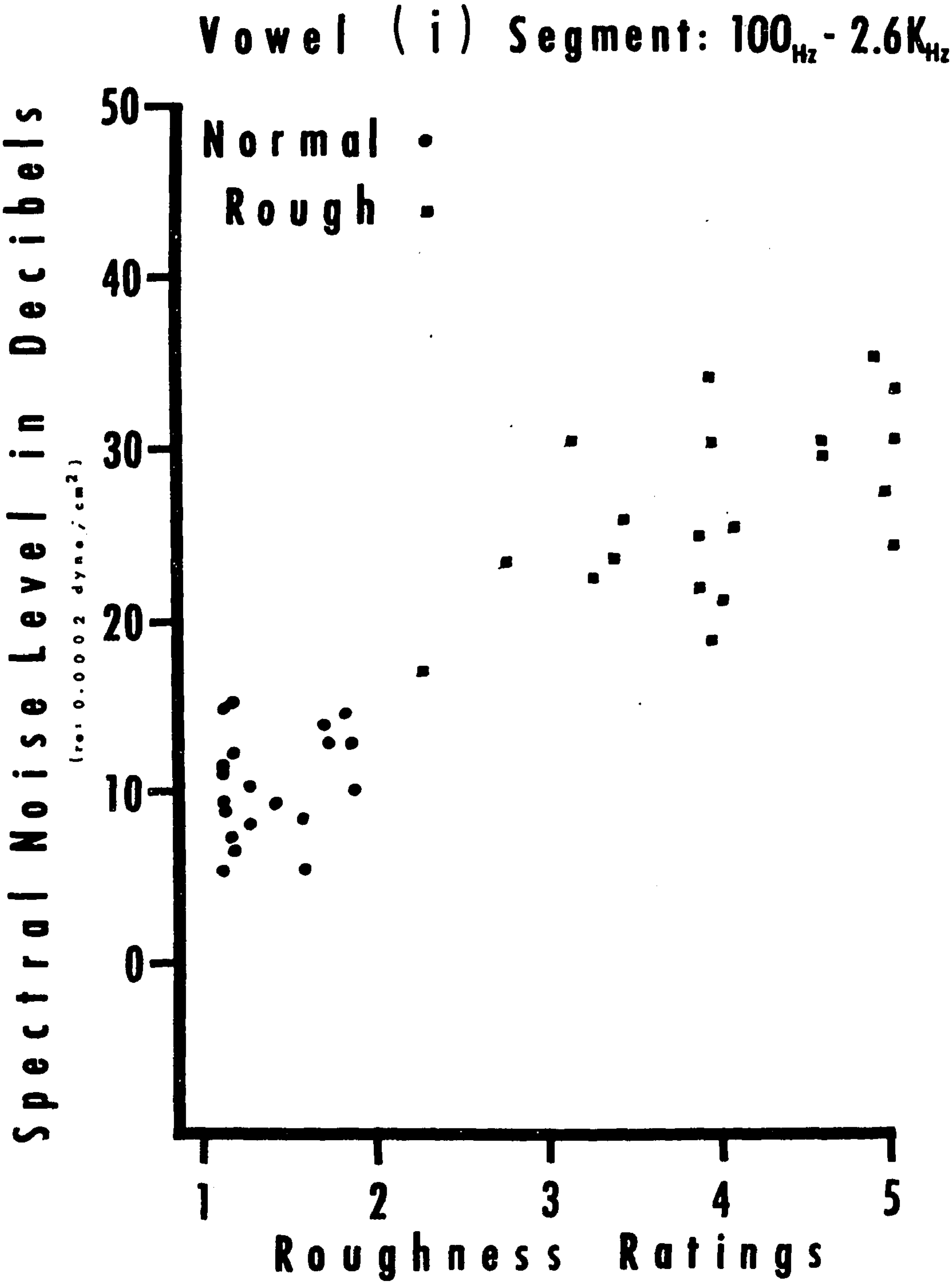

Figure 8.—Spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means and median roughness ratings over eleven judges for twenty female subjects' normal and rough productions of the vowel /i/.

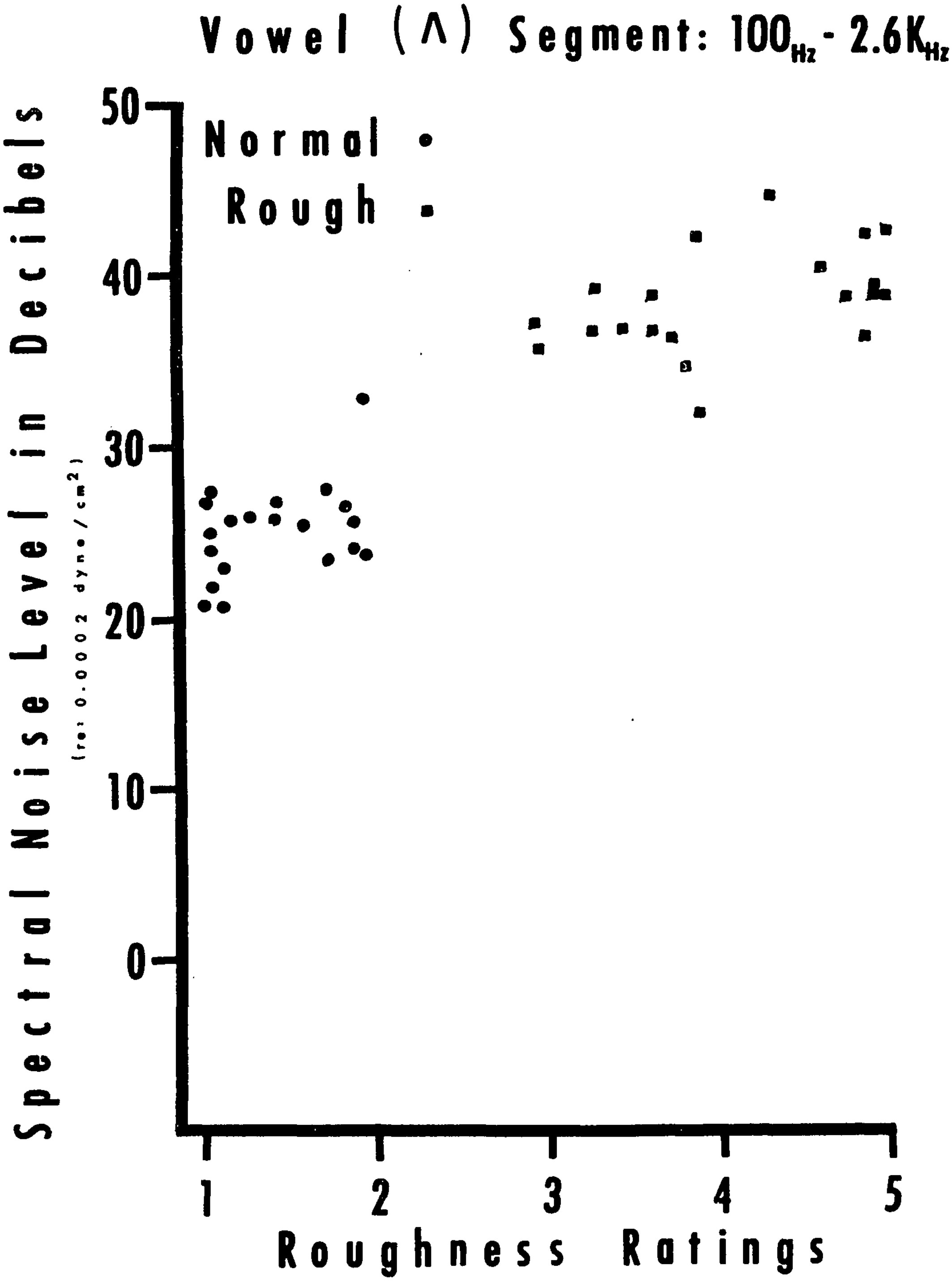

Figure 9.—Spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means and median roughness ratings over eleven judges for twenty female subjects' normal and rough productions of the vowel /ʌ/.

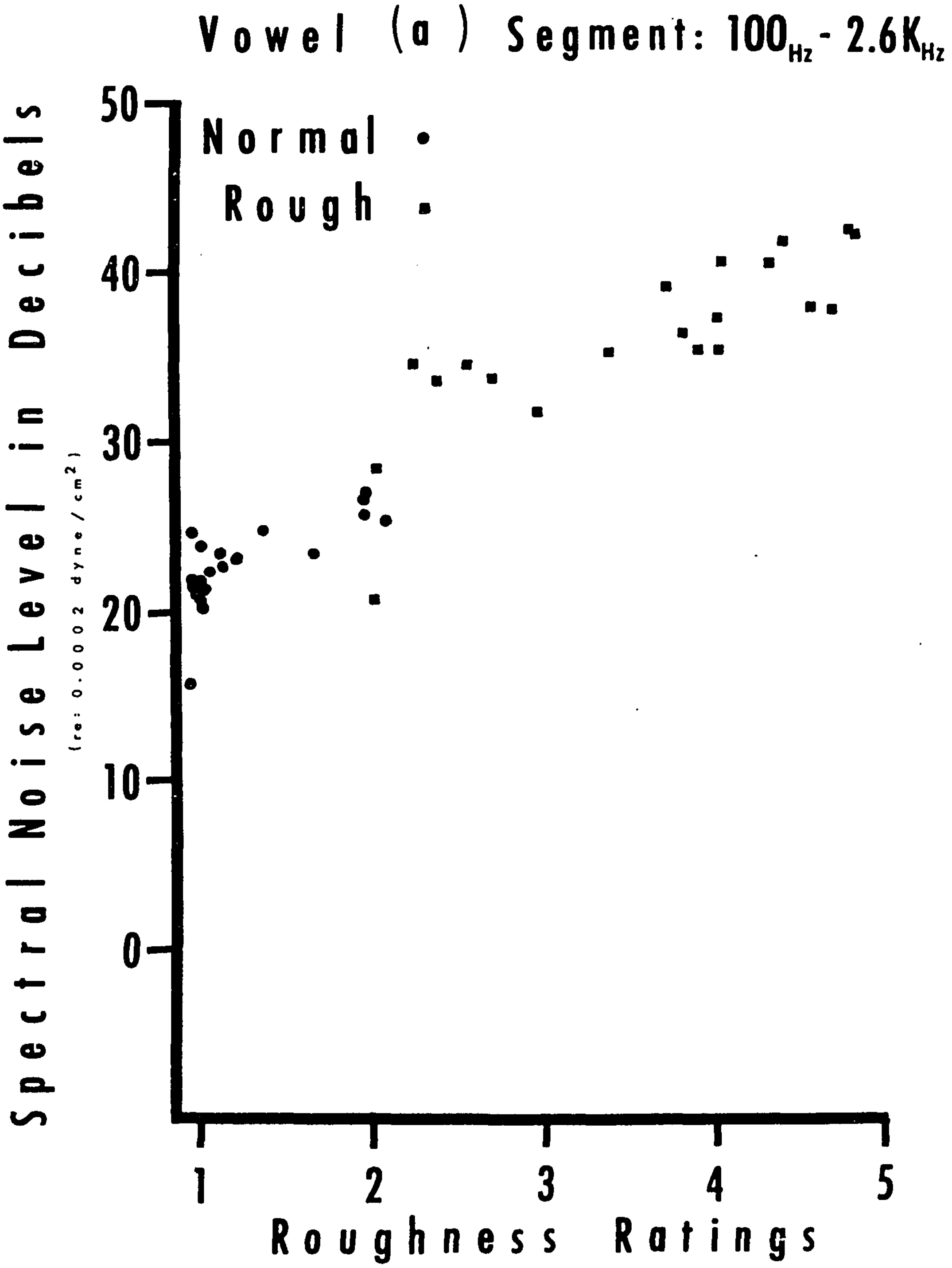

Figure 10.—Spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means and median roughness ratings over eleven judges for twenty female subjects' normal and rough productions of the vowel /a/.

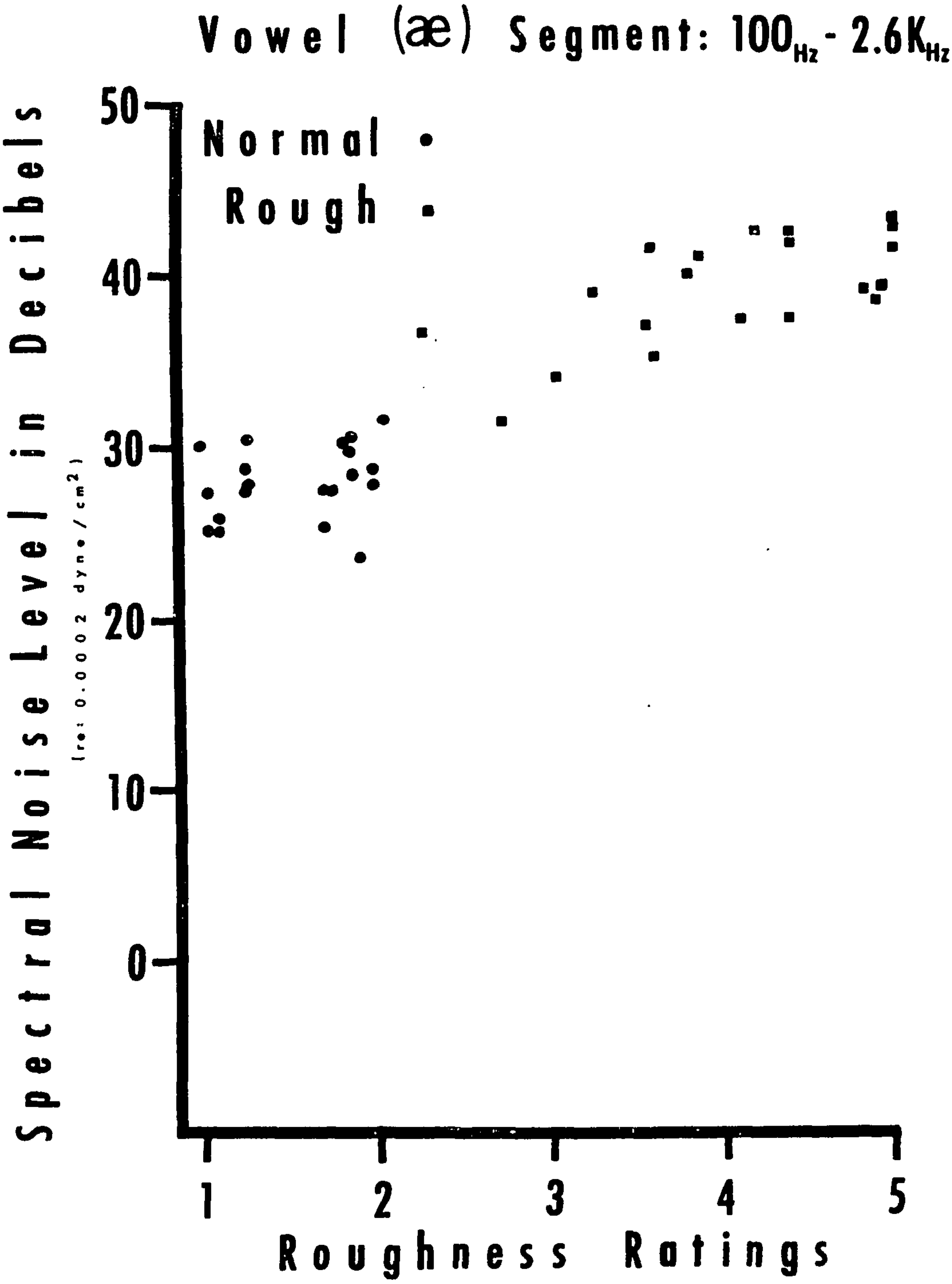

Figure 11.—Spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means and median roughness ratings over eleven judges for twenty female subjects' normal and rough productions of the vowel /æ/.

cally independent of those for S-1, S-2, and S-3, the correlation coefficients for these spectral segments were not tested for significance.

TABLE 4

CORRELATION COEFFICIENTS FOR MEAN SPECTRAL NOISE LEVELS AND ROUGHNESS SEVERITY RATINGS FOR EACH TEST VOWEL

	Correlation Coefficients*				
	S-1	S-2	S-3	S-4	TSR
Vowel	100 Hz to 2600 Hz	2600 Hz to 5100 Hz	5100 Hz to 8000 Hz	100 Hz to 5100 Hz	100 Hz to 8000 Hz
/u/	.92	.84	.69	.89	.86
/i/	.91	.87	.68	.91	.89
/ʌ/	.91	.82	.66	.89	.85
/ɑ/	.94	.86	.81	.93	.93
/æ/	.92	.85	.74	.92	.90

*All coefficients for S-1, S-2, and S-3 are significant at .05 level as determined by analyses of variance.

The coefficients for S-1, S-2, and S-3 for each test vowel are greater than .65 and are statistically significant at the .05 level of confidence. In general, the correlation coefficients tend to be highest for S-1 (100 Hz to 2600 Hz) and lowest for S-3 (5100 Hz to 8000 Hz). The coefficients for S-2, S-4, and the TSR are quite similar, ranging from .82 for /ʌ/ in S-2 to .93 for /ɑ/ in S-4 and the TSR. Considering the vowels separately, the highest correlations between mean SNLs and roughness ratings for /u/, /ɑ/, and /æ/ are associated with S-1, ranging from .92 to .94. Correlations for /i/ and /æ/, .91 and .92 respectively, are of the same magnitude for both S-1 and S-4. A plot of the regression of median roughness ratings on S-1 mean SNLs for each vowel is presented in APPENDIX E. Because the correlations for S-1 (100 Hz to 2600 Hz) were uniformly high for all vowels, a more detailed investigation of the relationship between S-1 SNL means and median roughness ratings was made.

For each vowel production, a multiple regression analysis was performed relating the SNL in each 100-Hz spectral section of S-1 (100 Hz to 2600 Hz) to the median roughness rating for that production. A significance level of .05 was selected for this analysis. Table 5 presents the multiple correlation coefficients for each vowel. The coefficients obtained in this analysis tend to be higher than those obtained when spectral segment SNL means and roughness severity ratings for each vowel

TABLE 5

CORRELATION COEFFICIENTS FOR THE MULTIPLE REGRESSION BETWEEN SPECTRAL NOISE LEVELS IN EACH 100-Hz SPECTRAL SECTION FROM 100 Hz to 2600 Hz AND ROUGHNESS RATINGS FOR EACH TEST VOWEL

Vowel	Correlation Coefficients*
/u/	.99
/i/	.98
/ʌ/	.97
/ɑ/	.97
/æ/	.99

*All coefficients significant at .05 level as determined by analyses of variance.

were compared. Table 5 shows that the multiple linear regression correlation coefficients for the vowels /u/ and /æ/ are each .99 and the coefficient for /i/ is .98. The multiple correlation coefficients for the vowels /ʌ/ and /ɑ/ are each .97. All of these coefficients were significant at the .05 level as determined by analyses of variance. The magnitude of these coefficients indicates a high degree of linear relationship between 100-Hz section SNLs in S-1 (100 Hz to 2600 Hz) and the median roughness ratings for each of the test vowels. Because the multiple correlation coefficients were uniformly high and significant, the median roughness rating for each test vowel production could be predicted from its S-1 100-Hz section SNLs. The multiple linear regression equation used for the prediction was:

$$Y = B_0 + B_1X_1 + B_2X_2 + \ldots + B_{25}X_{25}$$

where Y equals the roughness prediction, B_0 the Y intercept determined by the regression analysis, B_{1-25} the regression coefficients determined by the regression analysis, and X_{1-25} the successive S-1 100-Hz section SNLS from 100 Hz to 2600 Hz for each vowel production.

Table 6 shows judges' median roughness ratings, roughness ratings predicted by the linear model, and residuals (the observed roughness rating minus the predicted rating) for each subject's normal and rough productions of the vowel /ʌ/. Residuals for this vowel were the largest obtained and are presented to show the magnitude of the greatest residuals resulting from use of this regression equation. Examination of this table reveals that the roughness predictions for five of forty vowel productions deviate more than .50 scale value from the median roughness ratings actually obtained for those productions. Inspection of similar data for the other test vowels revealed that roughness predictions for four productions of /ɑ/ and for three productions of both /i/ and /æ/ deviated more than

TABLE 6

MEDIAN ROUGHNESS RATINGS FOR ELEVEN JUDGES, ROUGHNESS RATINGS PREDICTED BY THE REGRESSION EQUATION, AND RESIDUALS FOR TWENTY SUBJECTS' NORMAL AND ROUGH PRODUCTIONS OF THE VOWEL /ʌ/

	Normal			Rough		
Subject	Roughness Rating	Prediction	Residual	Roughness Rating	Prediction	Residual
1	1.05	0.89	.16	2.92	3.25	-.33
2	1.11	0.87	.24	4.95	4.29	.66*
3	1.71	1.84	-.13	4.81	4.28	.53*
4	1.05	1.08	-.03	4.89	4.66	.23
5	1.29	1.53	-.24	4.89	4.78	.11
6	1.95	1.56	.39	4.95	5.07	-.12
7	1.94	1.67	.27	3.60	3.45	.15
8	1.05	1.01	.04	2.95	3.74	-.79*
9	1.81	1.83	-.02	4.71	4.72	-.01
10	1.42	1.52	-.10	4.29	4.76	-.47
11	1.71	1.28	.43	3.42	3.09	.33
12	1.89	2.21	-.32	4.81	4.42	.39
13	1.19	1.29	-.10	3.29	3.37	-.08
14	1.11	0.86	.25	3.60	3.77	-.17
15	1.89	1.81	.08	4.59	4.83	-.24
16	1.00	0.95	.05	3.71	3.46	.25
17	1.59	1.54	.05	3.25	3.68	-.43
18	1.05	1.92	-.87*	3.86	3.87	-.01
19	1.05	1.20	-.15	3.80	3.67	.13
20	1.42	1.94	-.52*	3.89	3.49	.40

* Residual > .50 scale value

.50 scale value from the median roughness ratings obtained for those productions. For /u/, only one roughness prediction differed more than .50 scale value from the median roughness rating. The remaining residuals for /u/, /i/, /ɑ/, and /æ/ were relatively small.

Discussion

This study of the rough and normal vowel phonations of adult females was designed to replicate, in most respects, a similar study of adult males completed previously by Sansone (58). Except for the sex of the subject samples, the experimental design was essentially the same for both studies. With similar data available for the two sexes, it was possible to confirm that many of the findings reported by Sansone for males are equally valid for females. In the following discussion, therefore, frequent comparisons are made between the present findings and those of Sansone.

The findings for adult females in this study indicate that simulated rough vowels consistently received higher median roughness ratings than their normal counterparts, although the degree of roughness achieved for a particular vowel varied across subjects. The range of median ratings for the normal vowels phonated by females was, however, somewhat less than that reported previously for males. Sansone (58) reported normal vowel roughness ratings ranging from 1.00 to 3.00 for males as compared to a range from 1.00 to 2.11 for the females in this study. Median ratings equal to or greater than 2.00 were common among the normal vowel productions of males but were uncommon among those of females.

It may be that the roughness associated with normal vowels is different for the two sexes because of vocal pitch differences between the sexes. Wendahl (79) has previously hypothesized that male and female voices evidencing equal acoustic jitter may not be judged equally rough. Specifically, he suggested that the female voice would be judged less deviant. Sex differences in normal vowel roughness are not demonstrable on the basis of this comparison of the present data to Sansone's data, however. The roughness ratings assigned to normal productions may have been influenced by the degree of roughness simulated for the vowels in each study. Because it was not controlled, the degree of simulated vowel roughness may not have been the same in the two investigations. A further investigation is needed to determine if the range of roughness for normal vowels produced by males and by females is similar when the productions for the two sexes are judged together.

The present findings for females indicate that the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/ produced normally tend to differ in perceived roughness. The average median roughness ratings for female normal vowel productions are similar to those reported by Sansone (58) for males, although the average ratings for males tend to be slightly larger. In both investigations, normal vowels produced with higher tongue positions tended to be perceived as less rough than those with lower tongue positions. It appears, therefore, that changes in the configuration of the vocal tract associated with changes in tongue height may influence the degree of roughness associated with different normal vowels. It may be noted that tongue height *per se* is probably not the critical variable affecting vocal tract configuration. Stevens and House (65) have noted that the important parameters of vocal tract shape in vowel production are the distance of the maximum tongue constriction from the glottis, the size of the constriction formed by the tongue, and the size of the mouth opening as specified by a ratio of the cross-sectional area of the opening to the length of the front portion of the vocal tract, i.e., that portion that is more than 14.5 cm from the glottis. The judgment data for normal vowels in this study may have clinical implications because they suggest that the degree of vowel roughness which listeners judge to be within normal limits may be different for different vowels.

The findings of other investigators suggest that various rough vowels tend to differ in their relative roughness. For example, data reported by Rees (55) and Sherman and Linke (60) for harsh vowels and by Sansone (58) for rough vowels suggest that the high vowels /u/ and /i/ tend to be judged as less harsh or less rough than the low vowels /ɑ/ and /æ/. A tendency for vowels produced with simulated roughness to differ in their relative roughness was not evident in the present study, however.

In the present study, both normal and rough vowels evidenced noise components over the spectral range 100 Hz to 8000 Hz. Noise components in the spectra of rough and normal vowels have seldom been reported previously, apparently because the instruments commonly used in the acoustic analysis of vowels do not display these components clearly (58). The spectral features characterizing the present rough vowel productions were generally consistent with those reported previously. An outstanding feature in the rough vowel spectra was the elevation of noise components throughout the frequency range analyzed. Sansone (58) reported earlier that elevated noise components characterized the spectra of simulated rough vowels produced by male speakers. Elevated noise components have also been observed in frequency-by-intensity spectra (47) and sonagrams (30, 82, 83) of vowels produced by clinically hoarse subjects. It was observed in the present study that the amplitudes of identifiable harmonic components in the low spectral frequencies were diminished in the rough vowel spectra. Similar findings have been reported by Sansone (58), Nessel (47), and Yanagihara (82, 83). It appears, therefore, that both simulated vocal roughness and clinical hoarseness are associated with similar spectral features.

The observation in this study and others (30, 47, 58, 82, 83) that an increase in vowel spectral noise tends to be associated with a decrease in the level of harmonic components suggests that there is a trading relationship between vowel SNLs and harmonic amplitudes. It may be that the degree of vowel roughness perceived is primarily dependent upon the relative amplitude of harmonic and inharmonic components rather than the absolute magnitude of inharmonic components. It is hypothesized that, in general, an increase in perceived roughness occurs when the level of vowel inharmonic or noise components increases relative to the level of harmonic components.

The finding in the present study that rough vowel productions were characterized by larger mean SNLs than their normal counterparts is generally consistent with the results of previous studies (30, 47, 58, 82, 83). Yanagihara (82, 83) and others (30) reported additional noise components and elevated noise levels in the formant and higher frequency regions of sonagrams of hoarse vowels. The intensity levels of the elevated noise components could not be determined from their sonagraphic data, however. In the Sansone study (58) and in the present investigation, rough vowels evidenced higher SNL means than normal vowels. Considering each SS and the TSR separately, mean SNLs for female normal vowel productions in the present study were similar to those reported by Sansone for males. For example, considering spectral segment S-4 (100 Hz to 5100 Hz), the normal vowels produced by females evidenced the following SNL means: /u/, 8.9 dB; /i/, 16.1 dB; /ʌ/, 20.4 dB; /ɑ/, 18.2 dB; and /æ/, 24.6 dB. The S-4 SNL means for males reported by Sansone were: /u/, 5.3 dB; /i/, 17.0 dB; /ʌ/, 17.7 dB; /ɑ/, 18.9 dB; and /æ/, 23.9 dB. The standard deviations associated with normal vowel mean SNLs within each SS and the TSR were also of comparable magnitude in the two investigations. It appears, therefore, that the spectral noise level means characteristic of different vowels produced normally are similar for male and female speakers. Regarding rough vowel productions, mean SNLs for females in the present study tended to be larger than those reported by Sansone for males for each test vowel.

Comparison of the relative magnitudes of SNLs and roughness ratings for each normal and simulated rough vowel in this investigation with the findings of Rees (55) and Sansone (58) revealed interesting similarities. In her study of clinically harsh speakers, Rees reported that the vowels considered in the present study were ordered with respect to increasing harshness: /i/, /u/, /ʌ/, /æ/, and /ɑ/. The normal vowels in Sansone's study were ordered with respect to increasing roughness: /u/, /i/, /ʌ/, /ɑ/, and /æ/. With respect to increasing mean SNLs for S-1 (100 Hz to 2600 Hz), Sansone found that the normal and rough test vowels were ordered: /u/, /i/, /ʌ/, /ɑ/, and /æ/. In the present study, normal vowels were ordered with respect to increasing roughness: /u/, /ɑ/, /i/, /ʌ/, and /æ/. The normal and rough vowels for this study were ordered with respect to increasing mean SNLs for S-1: /i/, /u/, /ɑ/, /ʌ/, and /æ/. Although some reversals occur in the middle of the continuum, /u/ and /i/ tend to rank at the low end and /æ/ tends to rank at the high end of the continuum whether the vowels are ranked for harshness, roughness, or increasing S-1 mean SNLs.

The present data suggest a relationship between tongue height and vowel SNLs. In general, high vowels evidence less spectral noise and lower mean SNLs than low vowels, within both normal and simulated rough phonatory conditions. Since tongue height in vowel production is related to the overall configuration of the supraglottic cavity (5, 18, 22, 37, 65), a relationship between vowel noise levels and the configuration of the vocal tract is also suggested. It is thought that different vowels are distinguished perceptually largely on the basis of their formant frequency locations. Vowel formants occur at those frequencies which are minimally attenuated by the filter action of the vocal tract, that is, at those frequencies which correspond to the natural resonances of the vocal tract (5, 18, 22, 23, 24, 37, 39). The frequency-selective acoustic damping which produces the characteristic energy minima and, thus, indirectly the energy maxima in a vowel spectrum is thought to be directly related to the shape of the supraglottic cavity (5, 18, 20, 37, 39, 64, 65). It may be noted that the normal vowel spectra obtained in the present study evidenced relatively high noise components in those frequency regions where harmonic amplitudes were relatively high, i.e., in the formant regions. In interformant regions and high frequency regions above the formants, vowel noise levels tended to be relatively low. Regarding hoarse vowel spectra, Nessel (47) has noted that the noise spectrum which replaces the harmonic spectrum is "modulated according to the formants of the vowels." In the present study, when the "envelope" of the noise spectrum was outlined by drawing a line which connected successive 100-Hz section noise levels, the resulting noise spectrum tended to reflect the distinctive spectral pattern of energy prominances or formants of the vowel. Thus, it appears that the relative amplitudes of vowel noise components across the spectral frequency range may be influenced by the acoustic damping of the vocal tract. The presence of vowel noise components may reflect aperiodic variations in vocal fold movements and in the glottal volume-velocity wave (30, 82). Their relative amplitudes in formant and interformant ranges, however, may be determined largely by the configuration of the supraglottic cavity. It is hypothesized, therefore, that the presence and configuration of the noise spectrum observed for both rough and normal vowels is determined by an interaction of both glottic and supraglottic factors.

The relationships considered above may help to explain why vowel noise levels and SNL means for discrete spectral frequency ranges tend to differ for different vowels produced at the same intensity. As an example, the vowel /i/ may be considered. With respect to the mean SNLs of other vowels, the SNL mean for /i/ was lowest in relative magnitude in S-1 (100 Hz to 2600 Hz), highest in relative magnitude in S-2 (2600 Hz to 5100 Hz), and between the extremes in the remaining SSs and the TSR. If, as suggested above, vowel spectral noise is modulated according to the formants, these variations in the relative magnitude of the SNL means for /i/ might be expected on the basis of data presently available regarding average formant frequencies of vowels produced by adult females. According to the formant frequency averages for adult females reported by Peterson and Barney (52), S-1 (100 Hz to 2600 Hz) contains both F_1 and F_2 for all the vowels considered in the present study except /i/. Since S-1 includes only F_1 for /i/, the vowel /i/ evidences less spectral energy (both harmonic and inharmonic) below 2600 Hz than the other vowels and, thus, its S-1 SNL mean tends to be relatively small. Moreover, S-2 (2600

Hz to 5100 Hz) contains F_2 for the vowel /i/ but only formants higher than F_1 and F_2 for the remaining vowels. Fant (18) has noted that formants higher than F_2 evidence considerably lower amplitudes than F_1 and F_2. This might help to explain why the S-2 SNL mean for /i/ is higher than those for the other test vowels.

Other similarities between the harmonic spectra and noise spectra of normal and rough vowels may be considered. According to current theory (5, 18, 21, 22, 23, 33, 37), most of the spectral energy for vowels is located in the low frequency end of the spectrum. Harmonic components diminish in amplitude toward higher spectral frequencies because of the vocal tract's frequency-selective acoustic damping, i.e., its transfer function (5, 18). In the present study, the vowel spectra evidenced relatively high spectral noise levels in low frequencies and diminished noise levels in the high frequencies. Similarly, the SNL means of all test vowels except /i/ tended to decrease in magnitude from S-1 (100 Hz to 2600 Hz) to S-2 (2600 Hz to 5100 Hz) to S-3 (5100 Hz to 8000 Hz).

When the pattern of vowel SNL variability in each of the spectral segments studied by Sansone (58) was compared to the SNL variability observed in the present study, certain similarities were apparent. In each spectral segment, the SNL variability associated with normal and rough vowels was similar for both male and female speakers. For example, considering spectral segment S-1 (100 Hz to 2600 Hz), the SNL means for normal and rough vowel productions respectively evidenced the following standard deviations in this study: /u/, 3.2 dB and 5.9 dB; /i/, 3.0 dB and 5.2 dB; /ʌ/, 2.7 dB and 3.0 dB; /ɑ/, 2.6 dB and 5.2 dB; and /æ/, 2.2 dB and 3.2 dB. The standard deviations for normal and rough vowel S-1 SNL means reported by Sansone were: /u/, 2.4 dB and 6.2 dB; /i/, 2.8 dB and 5.0 dB; /ʌ/, 3.4 dB and 3.6 dB; /ɑ/, 2.6 dB and 3.7 dB; and /æ/, 3.2 dB and 2.8 dB. In both investigations, the least vowel SNL variability was associated with spectral segment S-1 (100 Hz to 2600 Hz) for both rough and normal vowel productions. The vowel formants which are thought to be important cues in the perception of vowel identity are located primarily in the lower spectral frequencies (5, 18, 22, 23, 33, 34, 37, 39, 51, 52). As Sansone has suggested, it may be that the subjects tended to control the distribution of acoustic energy in the low spectral frequencies in order to preserve vowel identity. This would tend to restrict SNL variability in this spectral region. Sansone also found that, among the normal vowels, the vowel /u/ tended to evidence the least SNL variability. On this basis, he suggested that normal /u/ SNL measures in S-1 might provide a standard to which similar data for rough speakers might be compared clinically. In the present study, however, no individual vowel consistently evidenced less SNL variability than the other test vowels. It appears that the S-1 SNL data for any one or all of the five vowels considered in these studies could provide a useful clinical standard to which clinically rough productions might be compared.

The present data suggest that differences between the spectral noise levels of normal and rough productions of each test vowel do not vary greatly across the analyzed frequency range. Several investigators (30, 47, 82, 83) have previously suggested that the relative increase in spectral noise which accompanies an increase in perceived hoarseness, i.e., roughness, is particularly evident in the frequency range above 3000 Hz. The present findings and those of Sansone (58), however, indicate that the spectral noise level differences between rough and normal productions of the test vowels were similar for both high and low spectral frequency ranges. This relationship may be difficult to observe in spectra produced with some instruments including those which amplify acoustic energy in high and low frequency ranges differently.

The present findings concerning the relationships between vowel spectral noise levels and listener judgments of vowel roughness were compatible with the results of previous investigations. In general, the present study indicated that increases in the level of spectral noise were accompanied by increases in perceived vowel roughness. Sansone (58), Yanagihara (82, 83) and others (30, 47) have also found a positive relationship between vowel spectral noise and perceived roughness or hoarseness. Yanagihara (82, 83) and others (30) have suggested that sonagrams of hoarse vowels can be classified into four types on the basis of the frequency region and intensity of elevated vowel noise components. In his study of vowels produced by clinically hoarse speakers, Yanagihara (83) obtained a correlation coefficient of .65 between four types of vowel sonagrams and judges' rating of vowel hoarseness. Correlations between vowel SNLs in 100-Hz sections of S-1 (100 Hz to 2600 Hz) and vowel median roughness ratings obtained in this study and in a previous study (58) were higher. Sansone (58) reported multiple correlation coefficients of .98 for the vowels /u/, /i/, /ʌ/, and /æ/ and .97 for /ɑ/. In the present study, the obtained correlation coefficients were similarly high. The multiple correlation coefficients were .97 for the vowels /ʌ/ and /ɑ/, .98 for /i/, and .99 for /u/ and /æ/. Moreover, because the severity of vowel roughness appears to vary along a continuum, it would seem desirable to study its acoustic correlates in ways which provide measurement of the relevant features on a continuous rather than a discrete scale. In this regard, the measures of noise in very narrow-band frequency-by-amplitude vowel spectra made in this study and in Sansone's study would appear to offer advantages over the classification of vowel sonagrams.

The findings of the present investigation and those of Sansone (58) indicate that the relationship between vowel SNLs in the low spectral frequencies and listener judgments of vowel roughness is nearly linear for the range of roughness studied. In both studies, multiple correlation coefficients between 100-Hz section SNLs in S-1 (100 Hz to 2600 Hz) and roughness ratings were high for all vowels considered. A multiple linear regression equation was employed to predict roughness ratings for individual productions of each test vowel from each production's S-1 100-Hz section SNL measures. Residuals indicating the difference between the observed and predicted roughness ratings for each production were small. If the relationship between vowel SNLs and roughness ratings is linear outside of the range of roughness investigated in this study, a multiple regression equation may be employed to predict listener judgments of vowels characterized by extreme spectral noise levels.

In most previous investigations (30, 47, 82, 83) of relationships between spectrographic features and perceived severity of hoarseness, the acoustic features of hoarseness were studied over a wide frequency range, e.g., 80 Hz to 8000 Hz. Sansone (58) reported that, for his male subjects, high correlations between median roughness ratings and vowel spectral noise levels were obtained when only a portion of the spectral range, e.g., 100 Hz to 2600 Hz, was considered. The present study also revealed high correlations between low frequency (100 Hz to 2600 Hz) SNLs and roughness severity ratings for female speakers. It appears, therefore, that acoustic information relating to vowel roughness may be redundant in the vowel spectrum. A question may be raised, however, regarding the relative importance of acoustic information contained in various spectral frequency ranges to the perception of vowel roughness. It may be that listeners judge vowel roughness primarily on the basis of their perception of acoustic relationships which obtain in only a limited segment of the total frequency range analyzed. Possibly, the range of greatest importance to vowel roughness perception is that which includes the vowel formants.

An underlying goal of many acoustic investigations of vocal roughness is to define and examine measurable acoustic correlates of perceived roughness which might provide objective indices of the degree of voice quality disturbance useful in the clinical evaluation and rehabilitation of speakers presenting rough voice. Lieberman (40), for example, has suggested that measures of pitch perturbation, or small, rapid variations in the durations of successive cycles of the acoustic wave, may provide such an objective clinical index. The present data support Sansone's (58) conclusion that measures of noise levels in vowel spectra may be similarly useful. Although both types of analyses result in objective measures which tend to correlate with perceived roughness, it appears that measures of vowel spectral noise may offer certain advantages. Lieberman noted that pitch perturbations could not be measured accurately when hoarseness was severe because the acoustic wave became "filled in" and individual cycles within it were not discernable. In contrast, spectral noise measures similar to those for isolated vowels in this study and in Sansone's study would appear feasible even for severely hoarse speakers. Moreover, it may be that spectral noise measures reflect the acoustic variables relevant to vowel roughness perception more completely than measures of pitch perturbation.

A unified concept of vocal roughness which organizes and interrelates the pertinent research and clinical data is presently unavailable. It appears, however, that a tentative theory of vowel roughness might be evolved through elaboration of existing concepts regarding normal vowel phonation. The results of physiologic and acoustic investigations, including the following, could provide the empirical basis for such a theory. Several investigations (44, 45, 78) have indicated that aperiodicities in the vocal fold vibratory pattern are associated with rough voice. Marked cycle-to-cycle variations in the shape, amplitude, and periodicity of the glottal area wave have also been reported for rough voices (40, 78, 82). Studies of synthesized complex waves have revealed that increases in acoustic jitter and shimmer are associated with increases in perceived signal roughness (12, 79, 80, 81). Rapid, random variations in the periods of successive cycles have also been found in the acoustic waves of rough voices (6, 11, 40, 44). Moreover, these cycle-to-cycle acoustic variations have been found to reflect variations in glottal periodicity (40, 44). Spectrographic investigations (30, 47, 58, 82, 83), including the present study, have indicated that elevated noise and diminished harmonic components are associated with perceived vowel roughness. A relationship between these spectrographic features of roughness and cycle-to-cycle changes in the glottal area function has also been observed (82). The following represents an attempt to conceptualize possible interrelationships among these physiologic and acoustic factors in vocal roughness.

It is thought that the force directly active in producing the acoustic voice wave is that imparted to the supraglottic air column by puffs of air emitted through the glottis and, further, that the volume-velocity wave for these puffs is dependent primarily upon glottal resistance and subglottic pressure (5, 8, 21, 22, 23, 28, 37, 39, 72, 73). It seems reasonable to assume that the acoustic voice wave, which is generated when the supraglottic air is set into vibration by the glottally-emitted air puffs, reflects in its features the characteristics of the glottal volume-velocity wave as well as the acoustic damping of the vocal tract (18, 21, 22, 23). As others have suggested (21, 22, 29, 30, 43, 82, 83), irregular glottal area changes in phonation may disturb the modulation of the air flow at the level of the glottis causing turbulence in the expiratory air flow. It is hypothesized, therefore, that aperiodic variations in vocal fold movements effect turbulence in the glottal puffs and, thus, aperiodic variations in the period, amplitude, and/or configuration of the glottal volume-velocity wave and the related acoustic voice wave. Existing acoustic theory (37) would seem to predict that all aperiodic acoustic wave features would be manifested spectrographically as inharmonic or noise components distributed over a broad frequency range. Thus, it appears reasonable to suggest that vowel spectral inharmonics or noise components are derived from aperiodic variations in the period, amplitude, and/or configuration of the phonatory acoustic wave and that these acoustic wave variations, in turn, relate to the underlying aperiodic variations in glottal air flow and in vocal fold movements. Similarly, harmonic components in vowel spectra may be considered to reflect periodic features in the acoustic voice wave, the glottal volume-velocity wave, and the vocal fold vibratory pattern. Because the vocal acoustic wave is a single-valued function of time, an increase in its aperiodic features is thought to be accompanied by a decrease in its periodic features for a given time segment of phonation. Thus, it appears that the level of vowel spectral inharmonics would be directly related to the presence of aperiodic features in the acoustic voice wave. Conversely, the amplitudes of vowel harmonics would be inversely related to the presence of acoustic wave aperiodicities. Available empirical data (30, 58, 82, 83) suggest that both diminished spectral harmonic components and elevated noise components are associated with an increase in perceived vowel roughness. Further, the observation by Fant (18) and others (5, 21, 23) that vocal fold phonatory movements are normally quasi-periodic is consistent with the present concept of vocal roughness and may be interpreted to predict the presence of low-level inharmonic components in the spectra for normal vowels.

It may be hypothesized further that the information essential to the perception of vowel roughness is the relationship between harmonic and inharmonic spectral energy. One measure of this relationship which might be considered is a ratio of harmonic to inharmonic energy (H/I). Other investigators (30) have also recognized this possibility. The present data and that of Sansone (58) suggest that the H/I ratio for discrete spectral frequency ranges tends to diminish from low to high frequencies for both normal and rough vowels. It is also hypothesized that, in general, the H/I ratio is inversely related to the degree of perceived vowel roughness. The high degree of linear relationship between spectral noise levels and perceived roughness demonstrated in the present study and in Sansone's study might be interpreted to indicate that spectral noise levels are an analog of the H/I ratio for vowels produced at a constant intensity. Finally, it is suggested that when the H/I ratio diminishes below some as yet undefined critical level or range of levels, which probably differs for different vowels, the quality of the vowel sample tends to be perceived as abnormally rough.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate spectral noise levels (SNLs) in narrow-band (3-Hz) spectra of normal and simulated rough vowels produced by adult female speakers, and possible relationships between the SNLs and perceived vowel roughness. The need for this investigation became evident when Sansone's previous study (58) of adult males indicated that quantitative measures of vowel noise levels in narrow-band acoustic spectra were highly correlated with listeners' perception of vowel roughness. This study sought to determine if similar relationships held for adult females. Because most features of Sansone's experimental design were replicated in this study, comparisons of the present findings for females to those for Sansone's males were possible.

Twenty normal-speaking adult females served as subjects for this investigation. The subjects individually produced each of the vowels /u/, /i/, /ʌ/, /ɑ/, and /æ/ first normally and then with simulated vocal roughness at one intensity. Each production was sustained for seven seconds at 75 dB SPL at a mouth-to-microphone distance of six inches. The vowel productions were recorded on magnetic tape and were presented in random order to eleven trained judges for roughness rating. Each judge rated the vowels for roughness on a five-point equal-appearing intervals scale in which "1" represented least severe and "5" represented most severe roughness. Anchor stimuli representing examples of the rating scale extremes were presented to the judges at the beginning of the judgment session. The median of the eleven judges' ratings was computed as an index of each vowel production's roughness. The recording of each production was also analyzed to produce a narrow-band (3-Hz) frequency-by-amplitude spectrum of its acoustic components from 0 Hz to 8000 Hz. This analysis was made from two-second tape loops constructed from a central portion of each vowel production evidencing a uniform intensity (75 dB SPL ± 1 dB). As a quantitative index of vowel spectral noise levels, the lowest observable peak of energy in each of seventy-nine successive 100-Hz sections from 100 Hz to 8000 Hz was measured in dB SPL in each vowel spectrum.

With regard to the findings, this study revealed generally that the normal vowel SNLs and the relationships between normal and rough vowel SNLs and judged roughness for females are not unlike those previously reported by Sansone (58) for males. Specifically, in this study, the median roughness ratings obtained for each test vowel indicated that each simulated rough production was judged more rough than its normal counterpart. The average median roughness ratings for normal vowel productions indicated that high vowels tended to be rated less rough than low vowels. For the normal vowel productions, the high vowel /u/ was rated least rough and the low vowel /æ/ was rated most rough. A similar order with respect to rough vowel productions was not apparent.

The present findings revealed that both normal and rough vowel productions evidenced noise components above system noise levels over the spectral range 0 Hz to 8000 Hz. Spectral noise levels tended to be higher for rough than for normal productions of each vowel when both were phonated at the same intensity. For both normal and simulated rough vowel productions, spectral noise appeared to be most prominent in lower spectral frequencies and decreased toward high frequencies. Harmonic amplitudes for rough productions of each vowel tended to be somewhat diminished with respect to those for normal productions. For both normal and rough vowel productions, spectral noise levels tended to be relatively high in the vowel formant ranges and relatively low in interformant and higher frequency ranges. Individual normal test vowel spectral noise levels averaged over S-1 (100 Hz to 2600 Hz) were all significantly different. For both normal and rough vowels, an increase in mean spectral noise levels appeared to be associated with changes in vocal tract configuration related to decreasing tongue height in vowel production. Within spectral segment S-1 (100 Hz to 2600 Hz), normal and rough vowel productions were ranked with respect to increasing mean spectral noise levels: /i/, /u/, /ɑ/, /ʌ/, and /æ/. For both normal and rough productions, inter-subject spectral noise level variability for each vowel was less for SNLs averaged over S-1 (100 Hz to 2600 Hz) than for SNLs averaged over the other spectral segments studied. For each test vowel, SNL variability was greater for rough than for normal productions. Differences between the SNLs of normal and rough productions of each test vowel did not vary greatly across the analyzed frequency range (100 Hz to 8000 Hz).

A primary objective of this study was to explore relationships between vocal spectral noise levels and listener judgments of vowel roughness. For each test vowel, spectral noise levels averaged over the total spectral range (100 Hz to 8000 Hz) and separately over each of the spectral segments correlated highly with the median roughness rating for that vowel. That is, as the mean spectral noise level of a vowel production increased, its median roughness rating tended to increase. Individual vowel SNLs averaged over spectral segment S-1 (100 Hz to 2600 Hz) tended to correlate more highly with the median roughness rating for each vowel than SNLs averaged over the other spectral segments. High ($\geq .97$) and significant ($p < .05$) multiple correlation coefficients were obtained between each test vowel's S-1 (100 Hz to 2600 Hz) 100-Hz section spectral noise levels and the median of judges' ratings of the vowel's roughness. A multiple linear regression equation predicted, with small residuals, each vowel production's median roughness rating from its S-1 (100 Hz to 2600 Hz) 100-Hz section spectral noise levels.

Some possible interrelationships among a number of physiologic and acoustic factors in vocal roughness were also considered. In brief, it was hypothesized (a) that the amplitude of vowel spectral harmonics tends to be inversely related to the level of the inharmonics for a given time segment of vowel phonation, (b) that the perception of roughness in a sustained vowel is predicated upon the relationship between harmonic and inharmonic energy levels in the vowel's spectrum, e.g., H/I, rather than upon the level of spectral noise alone, and (c) that the level of vowel spectral noise or inharmonics is directly related to aperiodic features in the vocal acoustic wave, in the volume-velocity wave of glottal air flow, and, finally, in the vocal fold movements which modulate the glottal air flow.

This study indicates generally that the investigation of vowel roughness by narrow-band acoustic spectrography reveals relationships of basic importance to the understanding of this voice quality. Additional investigations are needed, however, to provide data relevant to the hypotheses developed in this investigation. In particular, further information is needed regarding the relative importance to perceived roughness in human phonation of various types and degrees of acoustic signal aperiodicity, i.e., jitter, shimmer, and configurational variations. Moreover, additional data are needed regarding relationships between aperiodic features in the vocal fold vibratory pattern, the glottal volume-velocity wave, and the acoustic voice wave and the spectral features associated with roughness. Empirical investigation of the hypothesis that the relationship between harmonic and inharmonic spectral energy is of primary importance to the perception of vowel roughness would seem to be central to the further development of a coherent theory of vocal roughness. The results of such study might also bear importantly on the understanding of acoustic features associated with other voice qualitiss.

BIBLIOGRAPHY

1. Anastasi, A. _Psychological Testing_. (2nd Ed.) New York: The Macmillan Company (1961).

2. Arnold, G.E. Vocal rehabilitation of paralytic dysphonia: II. Acoustic analysis of vocal function. _Arch. Otolaryng._, _62_, 593-601 (1955).

3. Berry, M. and Eisenson, J. _Speech Disorders: Principles and Practices of Therapy_. New York: Appleton-Century-Crofts, Inc. (1956).

4. Black, J.W. Natural frequency, duration, and intensity of vowels in reading. _J. Speech Hearing Dis._, _14_, 216-221 (1949).

5. Bogert, B.P. and Peterson, G.E. The acoustics of speech. In L. Travis (Ed.) _Handbook of Speech Pathology_. New York: Appleton-Century-Crofts, Inc. (1957).

6. Bowler, N.W. A fundamental frequency analysis of harsh vocal quality. Ph.D. Dissertation, Stanford University (1957).

7. ____________ A fundamental frequency analysis of harsh vocal quality. _Speech Monogr._, _31_, 128-134 (1964).

8. Brubaker, R.S. and Dolpheide, W.R. Consonant and vowel influence upon judged voice quality of syllables. _J. acoust. Soc. Amer._, _27_, 1000-1002 (1955).

9. Carhart, R. The spectra of model larynx tones. _Speech Monogr._, _8_, 76-84 (1941).

10. Cherry, C. _On Human Communication: A Review, A Survey, and A Criticism_. (2nd Ed.) Cambridge, Mass.: M.I.T. Press (1966).

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

12. Coleman, F. and Wendahl, R. Vocal roughness and stimulus duration. _Speech Monogr._, _34_, 85-92 (1967).

13. Curry, E.T. A vocal frequency analysis in voice dysfunction. _Eye, Ear, Nose, and Throat Monthly_, _32_, 518-520 (1953).

14. Curtis, J.F. Disorders of voice. In W. Johnson et al. (Eds.) Speech Handicapped School Children (Revised Ed.). New York: Harper and Brothers (1956).

15. Denes, P.B. and Pinson, E.N. The Speech Chain: The Physics and Biology of Spoken Language. Bell Telephone Laboratories, Inc., Baltimore, Maryland: Waverly Press, Inc. (1963).

16. Fairbanks, G. Voice and Articulation Drillbook. New York: Harper and Row, Inc. (1940).

17. ____________ Voice and Articulation Drillbook. (2nd Ed.) New York: Harper and Row, Inc. (1960).

18. Fant, C.G.M. Acoustic Theory of Speech Production. 'S-Gravenhage, The Hague, The Netherlands: Mouton and Company (1960).

19. ____________ Descriptive analysis of the acoustic aspects of speech. Logos, 5, 3-17 (1962).

20. ____________ On the predictability of formant levels and spectrum envelopes from formant frequencies. In. I. Lehiste (Ed.) Readings in Acoustic Phonetics. Cambridge, Mass.: M.I.T. Press (1967).

21. Flanagan, J.L. Some properties of the glottal sound source. J. Speech Hearing Res., 1, 99-116 (1958).

22. ____________ Speech Analysis, Synthesis, and Perception. New York: Academic Press, Inc. (1965).

23. Fletcher, H. Speech and Hearing in Communication. Princeton, N.J.: D. Van Nostrand Co. (1953).

24. Gray, G.W. and Wise, C.M. The Bases of Speech. New York: Harper and Brothers (1959).

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

26. Heymann, O. Die stroboskopie in dienste der laryngoskopie. Arch. Ohren-usw. Heilk., 136, 116 (1933) (Cited by Luchsinger and Arnold).

27. House, A.S. and Fairbanks, G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. acoust. Soc. Amer., 25, 105-113 (1953).

28. Isshiki, N. Regulatory mechanism of voice intensity variation. J. Speech Hearing Res., 7, 17-29 (1964).

29. Isshiki, N. and von Leden, H. Hoarseness: aerodynamic studies. Arch. Otolaryng., 80, 206-213 (1964).

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

31. Jackson, C.L. and Norris, C.M. Cancer of the larynx. Ca.-A Cancer Journal for Clinicians, 12 (1962).

32. Johnson, W., Darley, F.L. and Spriestersbach, D.C. Diagnostic Methods in Speech Pathology. New York: Harper and Row (1963).

33. Joos, M. Acoustic phonetics. Lang. Suppl., 24, 1 (1948).

34. Judson, L.S.V. and Weaver, A.T. Voice Science. (2nd Ed.) New York: Appleton-Century-Crofts, Inc. (1965).

35. Kenyon, J.S. American Pronunciation. Ann Arbor, Michigan: George Wahr (1940).

36. Koenig, W., Dunn, H.K., and Lacy, L.Y. Sound spectrograph. J. acoust. Soc. Amer., 18, 19-49 (1946).

37. Ladefoged, P. Elements of Acoustic Phonetics. Chicago: University of Chicago Press (1966).

38. Laguaite, J. and Waldrop. W. Acoustic analysis of fundamental frequency of voices before and after therapy. Folia Phoniatr., 16, 183-192 (1964).

39. Lieberman, P. Intonation, Perception, and Language. Cambridge, Mass.: M.I.T. Press (1967).

40. ____________ Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. acoust. Soc. Amer., 35, 344-353 (1963).

41. Luchsinger, R. and Arnold, G. Voice-Speech-Language, Clinical Communicology: Its Physiology and Pathology. Belmont, California: Wadsworth Publishing Co. (1965).

42. McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, Inc. (1962).

43. Moore, P. Voice disorders associated with organic abnormalities. In L. Travis (Ed.) Handbook of Speech Pathology. New York: Appleton-Century-Crofts, Inc. (1957).

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

45. Moore, P. and von Leden, H. Dynamic variation of the vibratory pattern in the normal larynx. Folia Phoniatr., 10, 205-238 (1958).

46. Morrison, L.F. Recurrent laryngeal nerve paralysis. Ann. Oto. Rhino. Laryng., 61, 567-592 (1952).

47. Nessel, E. Über das tonfrequenzspektrum der pathologisch veränderten stimme. Acta Otolaryng. Suppl. 157, 3-45 (1960).

48. Operating Instructions: Type 1900-A Wave Analyzer and Type 1910-A Recording Wave Analyzer, General Radio Company, West Concord, Mass. (1964).

49. Palmer, J.M. Hoarseness in laryngeal pathology: review of literature. Laryngoscope, 66, 500-516 (1956).

50. Peacher, G. Contact ulcer of the larynx, Part III. Etiological factors. J. Speech Dis., 12, 177-178 (1947).

51. Peterson, G.E. Parameters of vowel quality. J. Speech Hearing Res., 4, 10-29 (1961).

52. Peterson, G.E. and Barney, H.L. Control methods used in a study of the vowels. J. acoust. Soc. Amer., 24, 175-184 (1952).

53. Proceedings of the Workshop on Nomenclature of Communicative Disorders. PHS Grant No. B-3676, Bethesda, Maryland (1962).

54. Rees, M. Harshness and glottal attack. J. Speech Hearing Res., 1, 344-349 (1958).

55. ________ Some variables affecting perceived harshness. J.Speech Hearing Res., 1, 155-168 (1958).

56. Rubin, H. and Lehrhoff, I. Pathogenesis and treatment of vocal nodules. J. Speech Hearing Dis., 27, 150-161 (1962).

57. Rullan, A. Associated laryngeal paralysis. Arch. Otolaryng., 64, 207-212 (1956).

58. 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).

59. Sherman, D. The merits of backward playing of connected speech in the scaling of voice quality disorders. J. Speech Hearing Dis., 19, 312-321 (1954).

60. Sherman, D. and Linke, E. The influence of certain vowel types on degree of harsh voice quality. J. Speech Hearing Dis., 17, 401-408 (1952).

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

62. Snidecor, J.C. Speech Rehabilitation of the Laryngectomized. Springfield, Illinois: Charles C. Thomas (1962).

63. Steel, R.G. and Torrie, J.H. Principles and Procedures of Statistics. New York: McGraw-Hill Book Company (1960).

64. Stevens, K.N. and House, A.S. An acoustical theory of vowel production and some of its implications. J. Speech Hearing Res., 4, 303-320 (1961).

65. ______________________________ Development of a quantitative description of vowel articulation. J. acoust. Soc. Amer., 27, 484-493 (1955).

66. Stout, B. Harmonic structure of vowels in singing in relation to pitch and intensity. J. acoust. Soc. Amer., 10, 137-146 (1938).

67. Tarneaud, J. Séméiologie stroboscopique des maladies du larynx et de la voix. Paris: Jaurès (1937) (Cited by Luchsinger and Arnold).

68. 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).

69. Thurman, W. The construction and acoustic analysis of recorded scales of severity for six voice quality disorders. Ph.D. Dissertation, Purdue University (1954).

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

71. ____ Laryngeal vibrations: measurements of the glottic wave, Part II. Physiologic variations. Arch. Otolaryng., 69, 438-444 (1959).

72. Van den Berg, J.W. Myoelastic aerodynamic theory of voice production. J. Speech Hearing Res., 1, 227-244 (1958).

73. __________________ Subglottic pressures and vibrations of the vocal folds. Folia Phoniatr., 9, 65-71 (1957).

74. Van Dusen, R. A laboratory study of the metallic voice. J. Speech Dis., 6, 137-140 (1941).

75. Van Riper, C. Speech Correction: Principles and Methods. (3rd Ed.) Englewood Cliffs, N.J.: Prentice-Hall, Inc. (1954).

76. Van Riper, C. and Irwin, J.V. Voice and Articulation. Englewood Cliffs, N.J.: Prentice-Hall, Inc. (1958).

77. Von Leden, H. The clinical significance of hoarseness and related disorders. J. Lancet, 78, 50-53 (1958).

78. Von Leden, H., Moore, P., and Timcke, R. Laryngeal vibrations: measurements of the glottic wave, Part III. The pathologic larynx. Arch. Otolaryng., 71, 26-45 (1960).

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

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

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

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

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

84. Zinn, W. The significance of hoarseness. Ann. Oto. Rhino. Laryng., 54, 136-138 (1945).

APPENDIX A

Instructions to Subjects

Instructions to Subjects

In this experiment you will phonate five vowel sounds, at first normally and then while simulating vocal roughness, into the microphone. The vowel sounds you are to produce are the underlined sounds in the words printed on the cards: /i/ as in bee, /u/ as in boot, /ʌ/ as in hut, /ɑ/ as in hot, and /æ/ as in cat. You are not to say the entire word, but only the vowel sound that is underlined. The cards will be held so you can see them easily during recording. I will also say each vowel immediately before you speak it.

You should say the vowel sounds loudly enough so that the needle on the meter will peak at the red mark. You will be given two signals from the signal lights. The amber light will come on briefly, indicating that you are to begin to phonate and to peak the needle of the meter steadily at the mark. When the red light comes on, you are to continue to keep the needle steadily at the mark as long as the red light is on. Be very careful to keep the needle on the meter at the mark. Some of the sounds are weak sounds and will have to be spoken loudly to peak at the mark. Some of the sounds are strong sounds and will not have to be spoken as loudly to peak the needle at the mark. You will be given an opportunity to practice peaking the needle on the vowel sounds before actually making the recording.

Produce vocal roughness by phonating while "making your throat tight." A "tight throat" occurs on the initiation of a cough. If you have trouble making your throat tight, start to cough, hold your laryngeal structures in that posture, and phonate. If you wish, I will demonstrate vocal roughness for you. When you are simulating vocal roughness, be sure to avoid producing "glottal fry." I will indicate to you if you produce "glottal fry." If you do, we will re-record the vowel. I will also indicate to you if you are not producing the vowel printed on the card. Sometimes while simulating vocal roughness, the vowel is distorted. If you do not produce the vowel, we will re-record. Are there any questions?

APPENDIX B

Instructions to Judges

Instructions to Judges

You are asked to listen to 250, seven-second sustained vowel samples produced by adult females. The samples are comprised of the vowels /i/, /u/, /ʌ/, /ɑ/, and /æ/, and represent a range of vocal productions from smooth to rough. The vowel samples will be presented to you one at a time, and you are to judge each in relation to a five-point scale of severity of vocal roughness. Make your judgments on the basis of the severity of vocal roughness perceived.

Each vowel is to be rated on a scale of equal-appearing intervals with scale values from "1" to "5." Scale value "1" represents <u>least</u> severe vocal roughness and "5" represents <u>most</u> severe. Do not attempt to rate vowel samples between any two scale points. The vowel samples may vary according to parameters other than roughness; however, you are asked to ignore these variations. Restrict your attention to the degree of roughness perceived.

The vowels to be judged will be presented to you in random order. There will be a short interval between productions and each will be preceded by a number announcement.

You are to judge each of the vowel samples in relation to the five-point scale of severity of vocal roughness. Record on your response sheet the scale value from "1" to "5" you think each production should be assigned. As you are asked to scale <u>your perceptions</u> of the severity of vocal roughness, there are no right or wrong scale values. Thus, a scale value you record for a vowel may not be the scale value the person sitting next to you records for that same vowel. For this reason, be sure to make your judgments independently. Record the scale value assigned to each vowel to the right of its number on your response sheet. You may hear each vowel production to be judged as many times as you wish. Notice that you will start at the top of a column and work down. Be sure to record a judgment for every vowel sample. Leave no blank spaces. The vowels will be presented in five segments of fifty vowels each with a short rest period following each segment. These instructions will be presented again at the beginning of each segment. Are there any questions?

APPENDIX C

Percentages of Inter- and Intra-Judge Roughness Rating Agreement

TABLE 7

PERCENTAGE OF INTER-JUDGE ROUGHNESS RATING AGREEMENT
±1 SCALE VALUE FOR TWO HUNDRED VOWEL PRODUCTIONS

	Judge									
Judge	11	10	9	8	7	6	5	4	3	2
1	94	99	98	96	92	98	97	99	98	95
2	100	99	99	99	100	97	98	97	95	
3	96	100	99	96	92	98	100	99		
4	98	99	99	95	94	96	98			
5	98	100	99	98	95	99				
6	96	99	99	97	94					
7	99	96	99	99						
8	98	97	97							
9	98	99								
10	98									

TABLE 8

PERCENTAGE OF INTRA-JUDGE ROUGHNESS RATING AGREEMENT ±1 SCALE
VALUE FOR TWO RATINGS OF FIFTY VOWEL PRODUCTIONS

Judge										
1	2	3	4	5	6	7	8	9	10	11
96	100	98	100	100	98	100	100	98	100	100

APPENDIX D

Summary of Analysis of Variance and Duncan's New Multiple Range Test

TABLE 9

SUMMARY OF THE ANALYSIS OF VARIANCE FOR S-1 (100 Hz to 2600 Hz) SPECTRAL NOISE LEVEL MEANS FOR NORMAL VOWELS

Analysis of Variance				
Source of Variation	df	ss	ms	F
Subjects	19	234.36	12.33	1.93
Vowels	4	4586.65	1146.66	179.08 *
Error	76	482.36	6.35	
Additivity	1	2.13	2.13	0.33
Residual	75	480.23	6.40	

* $p < .05$

TABLE 10

DUNCAN'S NEW MULTIPLE RANGE TEST FOR DIFFERENCES AMONG NORMAL VOWEL S-1 (100 Hz to 2600 Hz) SPECTRAL NOISE LEVEL (SNL) MEANS

Vowels	/i/	/u/	/ɑ/	/ʌ/	/æ/
SNL Means	10.6	13.0	22.8	24.9	27.8

Note: All means are significantly different at the .05 level.

APPENDIX E

Regression of Median Roughness Ratings on S-1 (100 Hz to 2600 Hz) SNL Means

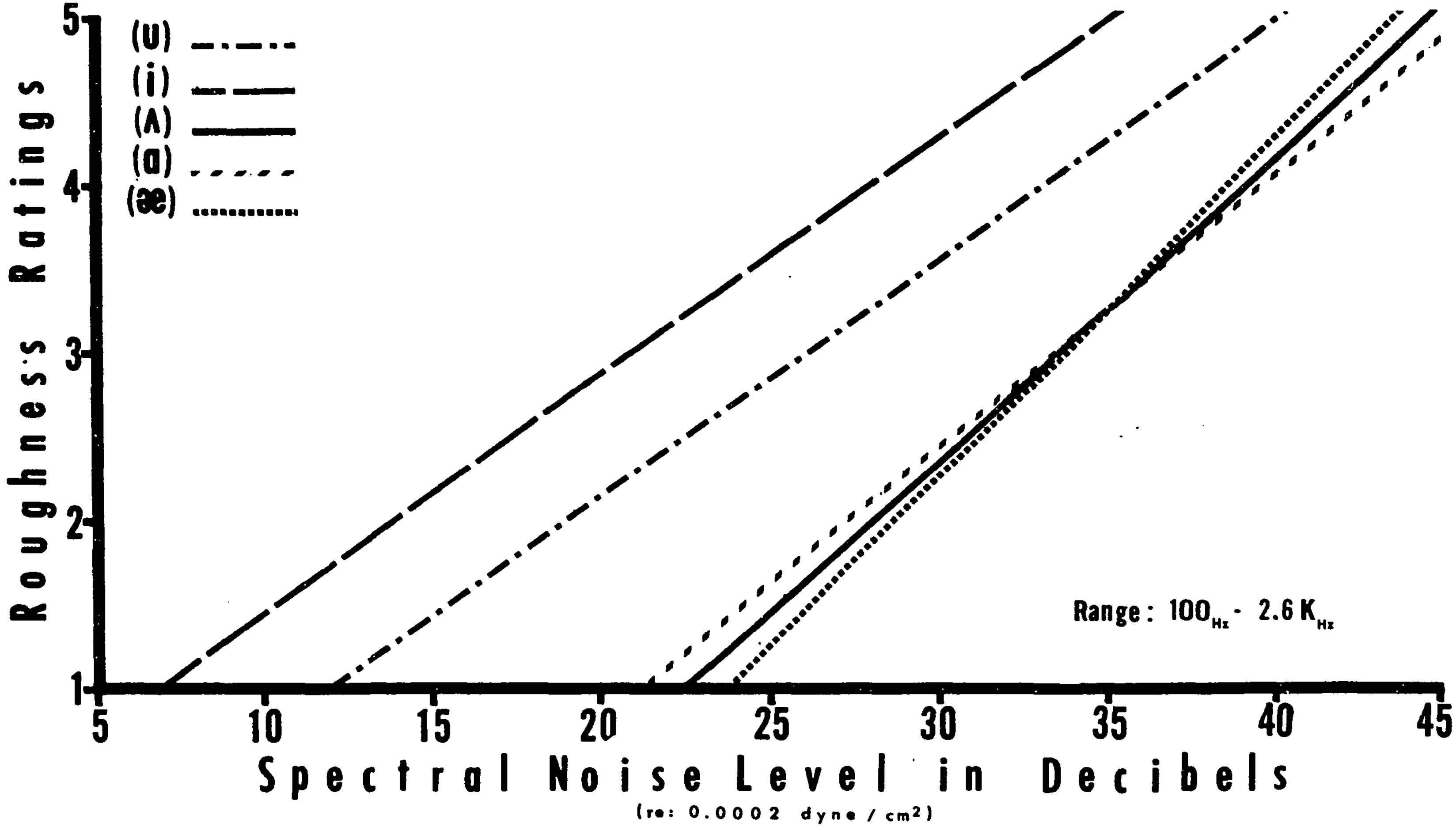

Figure 12.--Regression of median roughness ratings over eleven judges on spectral segment S-1 (100 Hz to 2600 Hz) spectral noise level means for each test vowel.