Perceptual Effects of Perturbation and Additive Noise

Richard E. Nero
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Speech Pathology and Audiology Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/1089
PERCEPTUAL EFFECTS OF PERTURBATION AND ADDITIVE NOISE

by

Richard E. Nero

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Speech Pathology and Audiology

Western Michigan University
Kalamazoo, Michigan
June 1990
PERCEPTUAL EFFECTS OF PERTURBATION
AND ADDITIVE NOISE

Richard E. Nero, M.A.
Western Michigan University, 1990

This study investigated the perceptual effects of varying glottal waveshape perturbation, duty cycle perturbation, jitter, signal-to-noise ratio perturbation, and shimmer at three levels of perceived magnitude (high, medium, and low). Voice signals were synthesized using a modified version of Klatt's (1980) formant synthesizer. Listeners rated signal dysphonia using an A-B dissimilarity procedure. Multidimensional scaling analysis suggested that signals judged low in magnitude of dysphonia were perceived to be most similar. Medium-level and high-level magnitude signals were judged to be respectively less similar. Signals varying in duty cycle and fundamental frequency perturbation were perceived to be very similar while those signals varying in waveshape and amplitude perturbation were consistently judged to be most different from all other signals within each level of perceived severity. Implications for the development of objective measures of vocal quality are discussed.
I wish to express a special acknowledgement and sincere appreciation to Dr. James M. Hillenbrand, director and chairman of this thesis, for his encouragement, direction, advice, and support throughout the course of my study. His seemingly never-ending patience during the two years of our professional relationship was and always will be an inspiration to me. I likewise express sincere appreciation to Dr. Robert L. Erickson, Dr. John M. Hanley, and Dr. Richard Pippen who served as members of my thesis committee. Their knowledge, criticisms, suggestions, encouragement, and friendships have been extremely valuable to my personal growth as well as to every aspect of the preparation and completion of this endeavor.

Appreciation also goes to the Department of Speech Pathology and Audiology and Western Michigan University for their financial support during the course of my master's program. This project was partially funded by grant number 5-R01-N523703 from the National Institutes of Health to Western Michigan University.

I will take this opportunity to again thank all my friends who gave me support through the trying times and celebrations of completing this project. I will not name
Acknowledgements - Continued

names; you know who you are and I thank you dearly. I have truly grown from knowing you.

Last, but not least, my deepest gratitude and appreciation are extended to my two loving parents, Bruce and Sally; and to my two sisters, Debra and Joyce, for their never-ending love, support, sacrifice, and encouragement that has made the attainment of another goal possible.

Richard E. Nero
INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700 800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Perceptual effects of perturbation and additive noise

Nero, Richard E., M.A.

Western Michigan University, 1990

Copyright ©1990 by Nero, Richard E. All rights reserved.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................... ii
LIST OF TABLES .................................... vi
LIST OF FIGURES .................................... vii

CHAPTER

I. INTRODUCTION TO THE INVESTIGATION ........... 1
   Purpose of the Investigation ............. 1
II. REVIEW OF THE LITERATURE .................... 5
   Acoustic Research on Dysphonia ........... 7
      Fundamental Frequency Perturbation ... 7
      Amplitude Perturbation ............... 14
      Additive Noise ..................... 17
      Intercorrelations Among Parameters .. 21
   Methodological Issues .................. 23
   Multidimensional Scaling .............. 25
III. DESIGN ....................................... 28
IV. METHODS ..................................... 30
    Stimuli ................................ 30
       Fundamental Frequency Perturbation ... 32
       Amplitude Perturbation ............... 33
       Additive Noise ..................... 33
       Duty Cycle Perturbation ............ 34
       Glottal Waveshape Perturbation .... 35
## Table of Contents - Continued

### CHAPTER
- **Subjects** ................................... 36
- **Instrumentation** ........................... 37
- **Experimental Procedures** ................... 37
  - Scaling of Overall Severity ............. 37
  - Dissimilarity Ratings ................... 38
  - Data Analysis ............................. 40
- **V. RESULTS** ................................. 42
  - Overall Severity Estimation ............ 42
  - Dissimilarity Ratings ................... 42
  - Intersubject Analysis Reliability ....... 42
  - 3-Dimensional Analysis ................. 45
  - Intrasubject Analysis ................... 45
- **VI. DISCUSSION** ............................ 48
  - Scaling of Overall Severity ............ 48
  - Dissimilarity Ratings ................... 51
  - Perception of Jitter and Shimmer ...... 51
  - Fundamental Frequency Perturbation and
    Duty Cycle Perturbation Perception .... 53
  - Limitations of the Present Study ....... 54
  - Proposed Future Studies ............... 55
- **VII. SUMMARY AND CONCLUSIONS** ......... 58
- **BIBLIOGRAPHY** .............................. 61
LIST OF TABLES

1. Stimulus Values for 13-Step Jitter Continuum ............... 32
2. Stimulus Values for 13-Step Shimmer Continuum ............... 33
3. Stimulus Values for 13-Step Additive Noise Continuum .......... 34
5. Stimulus Values for 13-Step Waveshape Continuum ............. 36
6. Stimuli Number Used for Each Severity Level During the MDS Experiment .......... 40
7. Comparison of Goodness-of-Fit Values for One to Four Dimensional Solutions for Intersubjectiveval Data Set A .......... 44
8. Individual Listener Weights and Goodness-of-Fit Values for Three Dimensional Solution .......... 47
9. Values of Perturbation at Which Increases in Perception of Dysphonia of Each Acoustic Parameter Lessened With Continued Increases in Perturbation .......... 49
LIST OF FIGURES

1. Rosenberg's Waveform ................................ 31
2. Magnitude Estimate as a Function of Stimulus Number ................. 43
3. Multidimensional Scaling Results .................... 46
CHAPTER I

INTRODUCTION TO THE INVESTIGATION

Purpose of the Investigation

There is a longstanding need in the field of speech pathology to develop accurate measures to evaluate voice disorders. Methods are needed that will reliably quantify the overall severity of dysphonia as well as characterize the qualities of the disordered voices. Speech pathologists are generally quite reliable in categorizing a voice as dysphonic or normal, but the qualitative and quantitative description of the dysphonia is much less dependable (Jensen, 1965). A reliable classification system is lacking in part because the acoustical principles of interaction of potentially relevant variables have not been delineated. Disordered voices are often described as "hoarse," "raspy," "thin," "metallic," "dark," or "heavy." The acoustical or physiological phenomena associated with these descriptors have been a matter of constant debate. Without such reliable and objective data, we have a limited understanding of the physiological basis of the voice disorder. Without this knowledge, the most efficient treatment strategies to manage dysphonia are clearly a
matter of debate, preference, and opinion.

Judgments of the degree of dysphonia clearly depend on properties of the acoustic signal. Therefore, development of a common objective scale to measure or interpret such vocal characteristics requires both an understanding of the acoustic attributes of the dysphonic voice and development of a procedure for quantifying these attributes. Most acoustic research attempts to develop acoustical measurements that will (a) aid in detecting laryngeal pathologies, and (b) correlate well with perceived degree and/or perceived type of dysphonia. Measures of a variety of acoustical perturbations have received particular attention.

A great deal of research has addressed cycle-to-cycle variations in fundamental frequency (jitter) and vocal intensity (shimmer). Although some aperiodicity is expected in a normal voice (Horii, 1979), extreme departures from periodicity may occur in a disordered voice (Davis, 1981; Horii, 1980, 1982; Kioke, 1973; Lieberman, 1961). Additional studies have investigated the amount of non-harmonic sound in the signal (signal-to-noise ratio) (Horii, 1979; Moore & Thompson, 1965; Wendahl, 1966a, 1966b).

Most research on voice disorders has been carried out using naturally produced human voices. The disadvantage of this approach is that it is not possible to control just
one parameter at a time, making it difficult to interpret which of many potential co-varying acoustic parameters affect a particular perceptual attribute of a voice. Synthesized signals, however, allow the researcher to vary any one parameter in discrete and controlled quantities and thus investigate simple effects or interactions among parameters.

The present study represents an attempt to address these concerns by studying the perceptual characteristics of synthetically generated vowel signals. The five parameters chosen for the study were glottal waveshape perturbation, duty cycle perturbation, fundamental frequency perturbation (jitter), additive noise, and amplitude perturbation (shimmer). This study is one step toward learning how these acoustic parameters combine to affect a listener's overall impression of vocal quality. The present study examines the univariate relationships between each of the acoustic dimensions and the perception of dysphonic vocal quality.

In this study, listeners were presented with two tasks. First, subjects listened to individual synthesized speech signals with varying degrees of fundamental frequency perturbation (jitter), amplitude perturbation (shimmer), signal-to-noise (S/N) ratio, duty cycle perturbation, and glottal waveshape perturbation. Subjects were asked to rate the signals on the amount of perceived dysphonia. The
second task entailed rating the amount of dissimilarity between pairs of signals presented to them. Multidimensional scaling (MDS) was used to identify the perceptual dimensions underlying the listeners' judgments of dissimilarity between the synthesized voice signals. MDS algorithms represent stimuli as points on a spatial map such that the distance between the points represents the degree of perceived dissimilarity between individual pairs of stimuli. Researchers have found MDS to be a valuable tool for investigating the underlying dimensions used to judge voices as similar or different (Kempster, 1984; Murry, Singh, & Sargent, 1977).
CHAPTER II

REVIEW OF THE LITERATURE

The acoustical and physiological characteristics of the human voice have been studied to understand how the natural voice is produced, to better diagnose and evaluate the pathological voice, and to improve the naturalness of synthesized speech. The literature on voice production abounds with terminology to describe vocal deviations and voice quality (Ogilvie, 1942; Van Riper, 1947). Terms and classification systems tend to be vague and have not been consistently used in either research or clinical management in speech pathology. Jensen (1965) found that what one listener may consider an obvious example of hoarseness, another may categorize quite differently. In the same study, Jensen also discovered that an individual may not label and rate the severity of a voice disorder consistently over time. He warns that care must be taken that diagnostic labels are based on various observations other than subjective perceptions and are reported as such. Jensen concluded by stating that there is a critical need for determining the co-varying relationships of voice production (normal and disordered) which exist along perceptual, acoustical, and physiological dimensions. When
such relationships are established, there should be adequate information to demonstrate certain related invariant acoustical and physiological manifestations of a vocal event that individuals agree on perceptually.

Clinical diagnosis of voice presently tends to follow the definitions and terminology proposed by Fairbanks (1960). He defined voice quality as that attribute of tone which is determined by the composition of the sound wave, and which allows us to discriminate between two sounds that are alike in pitch, duration and loudness, but yet are different. This vocal characteristic is determined by the way in which the laryngeal tone is generated and the effects of transmission of the tone through the vocal tract. Disruptions in either of these functions result in a voice quality deviation. Fairbanks classified these vocal anomalies as (a) harshness, (b) breathiness, (c) hoarseness (all due to defects of tone generation) and (d) nasality (a defect of transmission).

Normal voice is characterized by synchronous, quasi-periodic vibration of the true vocal folds. The sound produced in the larynx is modulated by a series of resonators of the supraglottic vocal tract and emerges as the unique, distinctive voice of an individual. A disordered voice may result from a laryngeal pathology or from a functional (i.e., nonorganic) disturbance in normal voice production. The normal voice has not been fully inves-
tigated, and much less is known about abnormal voice production.

Dysphonic voices have been described as hoarse, rough, breathy, metallic, thin, or harsh in quality. Much research in the past several decades has been devoted to relating physiological, acoustical, and perceptual differences in normal and abnormal voices. Research has concluded that extreme departures from periodicity may occur in a disordered voice (Davis, 1981; Horii, 1980, 1982; Kioke, 1973; Lieberman, 1961). A multivariate approach is clearly needed, however, since no one acoustic variable has been found that consistently detects a disordered sounding voice (Hecker & Kreul, 1971; Lieberman, 1963; Sansone & Emanuel, 1970; Yanagihara, 1967).

Acoustic Research on Dysphonia

Fundamental Frequency Perturbation

In an early study attempting to determine a physiological basis for voice quality, Lieberman (1961) investigated jitter (cycle-to-cycle variations in fundamental frequency) during connected speech, and found an almost continuous fluctuation within the samples of normal speech from six native speakers of English. In 86% of the cases, period duration did not remain constant across 3-period samples. Adjacent periods were shown to have a difference
in magnitude of 0.6 msec 20% of the time and greater than 1.0 msec 15% of the time. The magnitude difference increased with decreasing pitch until the pitch period reached 6.0 msec. Beyond 6.0 msec, the difference was found to be independent of the period duration. Lieberman concluded that the output of the glottis overshoots with alternately long and short periods which are highly correlated and that this perturbation is greater in the lower frequencies. Jitter has been likewise found in Swedish, British English (Risberg, 1962), and Japanese speakers (Saito, Kato, & Teranishi, 1958; Sugimoto & Hiki, 1962), thus revealing that jitter is not language specific.

Von Leden, Moore, and Timcke (1960) and Lieberman (1963), based upon data indicating that pathological larynges vibrate differently, thought it was reasonable to expect pitch perturbations from pathological voices to differ from those of normal voices if these measures really reflect irregularities in the vibratory pattern of the vocal cords. In both studies, results indicated that pitch perturbations reflect variations in the periodicity of vocal cord vibration in both normal and pathological larynges. Longer fundamental periods were found to have greater perturbations in both normal and abnormal larynges. Lieberman’s perturbation factor (percent perturbations greater than or equal to 0.5 msec that occurred) was found to be sensitive to the size and location of small growths.
It was concluded, however, that this technique is not very sensitive to very large growths. Very large growths near the vocal cords were characterized by perturbation factors greater than normal larynges, as was expected. However, large growths on the vocal cords had characteristically smaller perturbation factors than the normal larynges exhibited when these growths interfered with normal closure of the vocal folds.

Hecker and Kreul (1971) verified the inadequacy of Lieberman's perturbation factor for diagnosing laryngeal pathologies, but did find evidence supporting a directional perturbation factor. A directional perturbation factor is defined as the percentage of the total number of differences for which there is a change in the algebraic sign. This emphasizes the algebraic sign rather than the magnitude of the difference between adjacent glottal pulse intervals. Comparing the speech of speakers with laryngeal cancer with that of speakers who had normal larynges, Hecker and Kreul reasoned that the directional perturbation factor appears to be sensitive to sources of vibratory instability other than growths on the larynx. The growths were those that exist when the vocal folds are invaded by cancer. They further stated that "these results suggest that the directional perturbation factor may be more reliable than Lieberman's perturbation factor for detecting a malignant growth in the larynx" (p. 1281).
Studies of roughness associated with naturally produced voices have produced results that reveal a direct correlation between levels of measured jitter and perceived roughness. Lieberman (1963) studied 23 speakers of American English with pathologic larynges and nine speakers of American English with normal larynges. Subjects read an utterance as a statement, as a question, and as a confidential communication. Results indicated that greater jitter is associated with more severely hoarse than with mildly hoarse productions (see also Moore & Thompson, 1965). The perturbation factor measure of this study was found to be directly correlated with small growths on the vocal folds that were smaller than those associated with advanced cases of cancer of the larynx. The measure, however, was not found to be sensitive to large growths.

Hollien, Michel, and Doherty (1973) suggested that when vocalizations other than sustained phonation are used to examine the cycle-to-cycle variations in frequency, other phenomena become important. The perturbations may be due to voluntary and/or learned phonatory behavior associated with meaningful speech patterns produced by the speakers or natural prosodic differences. Hollien et al. suggested the use of sustained vowels to avoid this differential loading of the glottis. Sustained vowels reduce the variability due to learned speech patterns, and eliminate the differential loading of the subject's glottis.
related to changes in vocal tract configuration. Although Hollien et al. did not compare jitter measurements of sustained vowel phonations with those of connected speech, the authors were able to develop a reliable method to measure vocal jitter.

Sustained vowels and connected speech while reading were compared by Kitajima, Tanabe, and Isshiki (1975). Sustained [a] vowels and phrases (e.g., "blue sea") from thirteen men with laryngeal cancer and 20 men and women with normal larynges were contrasted. Kitajima et al. stated, "It is commonly observed that the voice shows more roughness during speech than in sustained phonation. In case of the minor lesion of the vocal cord, roughness is only noticed during speech" (p. 31). They concluded that the only way to detect slight perturbations was during reading speech samples. Hollien et al. counter this interpretation, however, and support the theory that connected speech contains more jitter than sustained phonation because of the differential loading of the glottis that occurs during connected speech and thus is not a valuable diagnostic procedure.

Smith, Weinberg, Feth, and Horii (1978) found that listeners were able to provide reliable judgments of the severity of perceived vocal roughness of sustained vowel samples produced by esophageal speakers. The magnitude of vocal jitter present in the vowels was substantially larger.
than that in normal speakers or speakers with laryngeal disorders. It was concluded, however, that mean fundamental frequency, mean jitter, or jitter ratio did not serve as useful predictors of the perceived severity of vocal roughness for esophageal voices.

Murray and Doherty (1980) found that fundamental frequency perturbation of sustained [a] phonation was the most useful parameter for separating groups of normal and laryngeal cancer speakers. Although the subject population was small (5 normal speakers and 5 with laryngeal cancer), Murray and Doherty concluded that jitter measurements during sustained vowel phonation more consistently differentiated between the subject groups than did jitter measurements during conversation or reading. The significance of fundamental frequency perturbation was followed by directional perturbation factor, and magnitudinal perturbation factor.

One of the difficulties of studying voices that are produced naturally is the inability to control variation in specific acoustic dimensions in the voice. Of particular interest to investigating the perceptions of disordered voice quality in naturally produced voice samples is the inability to control the range of variation on particular acoustic dimensions and the degree of intercorrelation among individual acoustic parameters. The intercorrelation is particularly important because several
studies have reported significant intercorrelations among individual acoustic parameters such as jitter, shimmer, and additive noise (Davis, 1976; Deal & Emanuel, 1978; Heiberg-er & Horii, 1982; Hillenbrand, 1987; Horii, 1980; Kempster, 1984; Yumoto, Sasaki, & Okamura, 1984).

These two problems, however, have been directly addressed by studying synthetically generated voices. Studies using synthesized voice signals have shown that a minimum amount of jitter is required for a synthesized voice to sound natural (Cooper, Peterson, & Fahringer, 1957; Gill, 1961; Holmes, 1962; Horii, 1979; Rozsypal & Millar, 1979; Schroeder, 1961).

Using a sustained sawtooth waveform signal and 536 listeners, Wendahl (1963) reported a direct relationship between the degree of judged roughness and frequency differences between successive cycles for a given median frequency value. Wendahl also noted that voices with lower fundamental frequencies were judged more rough than a higher fundamental frequency voice even if the amount of jitter was held constant (see also Coleman, 1969).

Wendahl (1966a) corroborated his 1963 findings in a subsequent investigation. Subjects were instructed to listen to pairs of stimuli and indicate which stimulus in each pair sounded more rough. Wendahl concluded that "rather small frequency variations in repetition rates will be judged as rough and it is the contention of the writer
that investigators analyzing the phonations of individuals with harsh or hoarse voices should look carefully for these small variations in glottal time patterns" (p. 30).

In another study with synthesized nonspeech signals, Wendahl (1966b) presented subjects with signals varying in fundamental frequency perturbation. Three jitter conditions were superimposed on an inflected model tone. The model tone featured a rising inflection from 100 to 135 Hz. After a 100 msec. steady state frequency of 100 Hz, the amplitude was modulated from -30 to 0-dB. When the signal reached full power, one period of a 101 Hz tone was introduced. With each consecutive cycle, the fundamental frequency was increased by one cycle until a 135 Hz tone was achieved. This inflected tone was abruptly ended at full amplitude. The three experimental jitter conditions were superimposed on the model after full signal amplitude had been reached. Using this technique, Wendahl found that the sawtooth wave signals varying in synthesized jitter correlated well with perceived roughness.

Amplitude Perturbation

Shimmer, cycle-to-cycle variability in pitch pulse amplitude, has also been investigated as a correlate of roughness. Studies have correlated shimmer with perceived vocal roughness as discussed above with jitter. Kitajima and Gould (1976) reported that measured amplitude perturba-
tion was significantly larger for 57% (20 of 35) of a group of dysphonic subjects than that for a control group of 25 normals. Measurements were based on successive differences from a three-point moving average. The shimmer values of five dysphonic speakers were found to overlap the critical regions of the normal speakers. The authors stated that this overlap was not unexpected as the subjects with polyps were selected without regard to the size or shape of the polyps. It was concluded that vocal shimmer, as measured in this study, may be a useful parameter to differentiate normal and dysphonic voices.

To investigate how indices of shimmer and jitter are related to acoustic noise levels and to roughness ratings over a range of vocal roughness, Deal and Emanuel (1978) band-pass filtered samples of naturally produced sustained vowels through a wave analyzer. The filter bandpass was centered at the frequency of the first harmonic of each sustained vowel sample. The fundamental frequency component was thus isolated and measurements of pitch and amplitude variations were completed for all samples of normal, simulated rough, and clinically diagnosed pathological voices. Positive correlations between the variability indices and the simulated rough or clinically hoarse vowel productions were found. These findings support the hypothesis that increases in vowel acoustic wave variability (jitter, shimmer, or both) are associated with
increases in vowel spectral noise levels and perceived vowel roughness. A nonsequentially measured amplitude variability index, used as an index of amplitude perturbation, was also found to provide a better index of perceived roughness than the index of nonsequential measure of pitch perturbation. Deal and Emanuel concluded that because of the linear relationship between spectral noise and perceived roughness, vowel spectral noise levels may provide a more clinically useful indicator of vowel roughness than the shimmer and jitter indices derived from the filtering procedures employed in this study.

Horii (1980) further investigated the potential role of shimmer as an indicator of roughness using sustained vowels. Vowel phonations were produced by thirty-one adult males who were free from any known speech or hearing problems. Results revealed an overall average shimmer of 0.39 dB with an upper extreme of 0.98 dB in these sustained phonations. Horii concluded that shimmer greater than 0.98 dB can be considered too large for mid-segments of sustained vowels produced by normal young adult males at comfortable Fo and intensity levels.

Using synthesized nonspeech signals in a study designed to investigate the relationship between shimmer and listener judgments of auditory roughness, Wendahl (1966b) found that the perception of shimmer is correlated to perceived roughness as is the perception of jitter.
Results were also in agreement with the previously reported data, i.e., that increases in shimmer within the acoustic signal are accompanied by listener judgments of mounting roughness (Deal & Emmanuel, 1978).

Additive Noise

The acoustic by-product of turbulence generated at the glottis is referred to as additive noise. A number of studies have reported that noise measurements correlate with subjective ratings of dysphonia, and that noise levels in dysphonic voices tend to be higher than those in normal voices (Deal & Emanuel, 1978; Emanuel & Sansone, 1969; Kojima, Gould, Lambiasse, & Ishiki, 1980; Lively & Emanuel, 1970; Sansone & Emanuel, 1970; Yanagihara, 1967; Yumoto, Gould, & Baer, 1982; Yumoto, et al., 1984). Yanagihara (1967) reported acoustic analysis of narrow band spectrograms suggesting that hoarseness is related primarily to three acoustic properties: (1) noise components in the second and third formant of each vowel, (2) noise components above 3000 Hertz (Hz), and (3) loss of high frequency harmonic components. The results of this study suggest that noise components and changes of harmonic structure are significant factors related to the degree of perceived hoarseness. Trained listeners rated the degree of hoarseness in vowels produced by 167 hoarse patients. Spectrographic analysis at midpoint sections of the vowels
/u/, /ɛ/, and /o/ revealed a positive correlation (r = 0.65 at 0.01 level of significance) between perceived hoarseness and hoarseness ratings derived from visual inspection of a narrow-band spectrogram. Yanagihara concluded that two major acoustic factors relate to hoarseness (noise components and loss of harmonic components) and that this study revealed that noise components appear in formant ranges (especially 2nd and 3rd) at initial stages of hoarseness. With increasing levels of hoarseness, noise components predominate in the 2nd and 3rd formant ranges and then become evident above the 3000 Hz range. Synthesized signals with additive noise were investigated by the same subjective listening and visual examination paradigm. Yanagihara concluded that the classification system developed was clinically useful in two respects: (1) the degree of hoarseness can be classified into one of four numerical categories and (2) the degree of hoarseness based on this method closely agrees with the subjective, perceived degree of hoarseness. Yanagihara proposed that four types of hoarseness may be differentiated on the basis of level and frequency location of noise in vowel sonagrams.

In a comparison of normal and simulated rough vowels /i/, /ʌ/, /æ/, and /ɑ/, Emanuel and Sansone (1969) found increased noise levels across the frequencies studied (100 to 8000 Hz). A 3 Hz constant bandwidth wave analyzer reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
revealed that noise levels were greatest in the lower frequencies for these simulated rough vowels produced by subjects with normal voices. Greatest differences between normal and rough vowel productions were found for the vowel /i/ and smallest differences for the vowel /ae/. Visual inspection of the spectra suggested that harmonics tended to decrease in amplitude with increases in spectral noise, confirming an earlier finding by Nessel (1960; cited in Emanuel & Sansone, 1969). Sansone and Emanuel (1969) also found a high correlation between perceived vowel roughness and spectral noise levels by noting the lower observable level-recorder stylus marking. This measure provided a quantitative index of spectral noise levels in dB SPL.

These studies established that the degree of hoarseness can be evaluated by judging the extent to which noise replaces the harmonic structure of a sustained vowel. Some of these methods (visual inspection) are subjective, however, and therefore difficult to standardize. Yumoto et al. (1982) wanted to develop a harmonic-to-noise (H/N) ratio as an objective index of the degree of hoarseness. The steadiest 600 ms of a sustained /a/ vowel was recorded from 42 subjects without laryngeal or pulmonary complaints and from 20 subjects with laryngeal disorders. The H/N ratio is based on the assumption that the acoustic wave of a sustained vowel consists of two components: a periodic component that is very similar from cycle-to-cycle, and an
additive noise component that varies from cycle-to-cycle and has a zero-mean amplitude distribution. The H/N ratio was determined by subjecting the sampled waveforms to a 500Hz low-pass filter for pitch and noise component extraction. Pitch periods were marked on the filtered wave using a manual or semiautomated method based on zero crossings. The noise component for any pitch period was calculated by subtracting the average waveform from individual pitch pulses of the original waveform. Fifty consecutive noise components were concatenated to form a noise waveform which was theoretically equivalent to the original phonation. The H/N ratio was calculated from the 50 consecutive pitch periods and converted to a decibel scale. Two experts visually rated the spectrograms for the amount of noise relative to that of the harmonic components of vowels. Results revealed a highly significant agreement (rank correlation = 0.85) between the calculated H/N values and the subjective spectrogram evaluations. Yumoto et al. (1982) concluded that this procedure provides an objective method for detecting vocal pathology and for assessing the results of therapy for hoarseness.

Hillenbrand (1988) found that the spectral slope of the periodic component did not affect the magnitude of perceived dysphonia as was reported by Yanagihara (1967). Yanagihara found that high-frequency energy in the periodic component of a non-speech signal resulted in an increase in
perceived dysphonia. Contrary to this, however, Hillenbrand found that there was a very strong relationship between S/N ratio and listeners' perception of breathiness, independent of the spectral slope of the harmonic component. When signals were matched for signal-to-noise ratio without regard for spectral differences, Hillenbrand found that listeners tended to judge the signals as similar in amount of breathiness.

**Intercorrelations Among Parameters**

Studies by Wendahl (1966a,b) and Heiberger and Horii (1982) reported strong relationships between amplitude perturbation and perceived roughness using synthesized nonspeech signals. Wendahl reported that the roughness percept resulting from the introduction of amplitude perturbation in sawtooth waveforms was very similar to that associated with pitch perturbation. For example, Wendahl (1966a) found a signal with 10% jitter was perceived to be equally rough as a signal with 6 dB shimmer. Concluding a study with sawtooth wave nonspeech signals, Wendahl (1966b) states, "it is interesting to note that the roughness generated by these different procedures [jitter and shimmer] results in such similar auditory experiences" (p. 106). Heiberger and Horii investigated the perceptual relations between pitch and amplitude perturbation from synthesized triangular waves varying in jitter, shimmer, or
both. They also reported that the perception of jitter and shimmer can be equivalent.

Several acoustic studies of disordered voices have reported significant intercorrelations among individual acoustic parameters such as jitter, shimmer, and signal-to-noise ratio (Davis, 1976; Deal & Emanuel, 1978; Heiberger & Horii, 1982; Horii, 1980; Kempster, 1984; Kempster & Kistler, 1984; Yumoto et al., 1984). Hillenbrand (1987) also suggests the acoustic analysis techniques that have been used to measure perturbation are not always able to discriminate among various sources of aperiodicity.

Horii (1980) found significant intercorrelations among shimmer, jitter, and fundamental frequency standard deviation during sustained vowel phonations on 31 adult males with normal voices. These correlations are thought to support the notion that similar sets of physical forces underlie the regulation of individual fundamental period and intensity of laryngeal sounds.

Hillenbrand (1987) offered alternative conclusions to those of Horii (1980). He hypothesized that strong measurement interactions among acoustic variables (e.g., jitter, shimmer, and signal-to-noise ratio) are responsible for the intercorrelations among these time domain variables. Hillenbrand found, for example, that adding increasing amounts of noise to an otherwise perfectly periodic voice signal resulted not only in decreases in
measures of harmonic-to-noise ratio values, but also substantial increases in measured values of pitch and amplitude perturbation. For these reasons, Hillenbrand (1987) concluded it may be very difficult to make precise determination of the source of aperiodicity in voice waveforms.

Methodological Issues

Direct comparison of perturbation values from the existing studies must be made with caution given the variation of measurement techniques and computational methods used by the different researchers. Heiberger and Horii (1982) discuss a variety of methodological issues that voice perturbation research must take into consideration. Many of the issues raised are concerned with instrumental and procedural issues that affect jitter and shimmer measures and problems common with high quality recording.

Temporal and amplitude resolution factors are important given the small size of the perturbations that are being measured and compared. In a review of thirteen studies on jitter in naturally produced voices, the time resolution factor varied from 0.5 msec to 0.02 msec (mean = 0.08 msec; mode = 0.05 msec). Given the finding that typical jitter values for sustained vowel phonations produced by young adults with normal voices is less than
1.0% (Hollien et al., 1973; Horii, 1979; Kempster, 1984), this clearly is an issue to be considered when evaluating existing results and designing new investigations.

A variety of mathematical methods have been used in the calculation of the pitch and amplitude perturbation. These different methods frequently do not result in the same measured values for the same data. Random single cycle measures and consecutive cycle-to-cycle measures have both been investigated. Consecutive measures are used most often, but even these can vary in the number of consecutive cycles taken into consideration. Calculated measures also vary significantly. For example, average deviation of vocal periods about a mean period (Deal & Emanuel, 1978), cycle-to-cycle or period-to-period deviations (Horii, 1980), and perturbation factors (Lieberman, 1961) have been used to mention just a few.

Differences in instrumentation between studies is another variable related to the kinds of results obtained. The quality of recording equipment (analog and digital) has had a significant impact on perturbation research. Horii (1980) investigated quantitized vocal shimmer using a 16-bit analog-to-digital converter but determined that the magnetic tape the signals were stored on only transferred 12-bits of information. Theoretically, therefore, it was determined that 0.04 dB of system noise was present (including temporal and amplitude quantity errors) when the
dynamic range of the A/D converter was optimally used (Horii, 1980).

Sampling rate likewise has a direct impact on resolution. It is obvious that a greater sampling rate will produce a recording that has greater resolution thus allowing for a more accurate measure.

Vowel-dependent jitter and shimmer magnitude differences have also been documented and discussed (Horii, 1980; Heiberger & Horii, 1982). Speaking conditions (e.g., intensity, fundamental frequency), and thus age and sex have likewise been shown to have a significant effect on the perception of vocal roughness (Kempster, 1984; Kempster, Kistler, & Hillenbrand, submitted; Kioke, 1973; and Wendahl, 1963).

Multidimensional Scaling

Multidimensional scaling (MDS) procedures represent stimuli judged to be subjectively similar to one another as points close to each other in a derived spatial map. Likewise, stimuli judged to be dissimilar are represented as points distant from one another. MDS derives a space with dimensions that are relevant to the subjects making the similarity judgments. A priori knowledge of the attributes of the stimuli to be scaled is not necessary. However, determination of the relevant attributes is necessary to interpret the MDS space. Four types of data
can be gathered and analyzed in a MDS experiment: (1) similarity or dissimilarity judgments among all pairs of stimuli, (2) ratings of stimuli on descriptors such as adjectives, (3) objective measures relating to the physical properties of the stimuli, and (4) information about the subjects. The spatial configuration is derived by the MDS algorithm from the similarity judgments among stimuli. The other types of data gathered are primarily used to interpret the derived MDS space.

Murry et al. (1977) demonstrated the application of a multidimensional technique to determine the perceptual attributes of a group of disordered voices. An interval scale was used to obtain judgments of similarity for 20 speakers sustaining the vowel /a/. Six physical and three psychophysical measures were also obtained and used to interpret the MDS results. A five dimensional solution was found to be the most appropriate solution for explaining the variance. These findings suggested that there exist distinct perceptual features of the disordered voice which have physical and psychophysical correlates.

Kempster (1984) also utilized MDS to investigate disordered voices. Thirty dysphonic female voices were studied in an attempt to identify perceptually relevant features common between them. Similarities were noted between a group of subjects that had bilateral vocal fold nodules and a group that was dysphonic but exhibited no
laryngeal pathology. A three dimensional solution was 
extracted from each comparison that accounted for ap­
proximately 60% of the variance in the listener judgments. 
The three dimensions were determined to be intensity, 
fundamental frequency, and perturbation.
CHAPTER III

DESIGN

The goal of this study was to investigate five acoustical parameters that have been implicated in the perception of dysphonic vocal quality. It is evident from the literature review that no one acoustic dimension can be used to identify and/or quantify a voice disorder.

The primary purpose of the present study was to investigate the relationships between variation in each of the acoustic dimensions and the perception of dysphonic voice quality. The specific interest was to determine the perceptual similarities among five acoustic parameters that were varied in a systematic manner. For this reason, synthetically generated vowel signals were investigated. These signals varied in discrete amounts in only one dimension, thus allowing a unique, systematic comparison. Listeners first performed a magnitude estimation task on all signals to rate the signals for the degree of perceived dysphonia. From these results, three levels of overall severity were selected. Listeners then rated the dissimilarity of quality between pairs of voice samples from the same overall severity level. By comparing signals from the same quantitative level, dissimilarity judgments were
controlled by requiring them to be based solely on qualitative characteristics of the signal pairs.

This investigation combined experimental and statistical procedures to determine perceptual effects of perturbing five acoustic parameters of the glottal wave. Multidimensional scaling (MDS) was utilized in this study to investigate the perceptual dimensions underlying the listeners' judgments of the dissimilarity of synthesized voice signals with varying degrees of glottal waveshape perturbation, duty cycle perturbation, fundamental frequency perturbation (jitter), additive noise, and amplitude perturbation (shimmer). MDS analysis allowed the signals to be compared such that the qualities that are associated with the five different parameters could be investigated. For example, do jitter and shimmer signals really sound similar as suggested by previous research (Wendahl, 1966b)? MDS also permitted investigation regarding the effect of the qualities of the different signal types on overall severity. Furthermore, the MDS procedure was selected as the experimental paradigm for this investigation because it has the advantage of being low in experimenter contamination and has been found to be useful in evaluating stimuli with multiple psychophysical dimensions (Kruskal & Wish, 1983; Schiffman, Reynolds, & Young, 1981).
CHAPTER IV

METHODS

Stimuli

Stimuli for all the experiments were synthesized with a modified version of Klatt’s (1980) digital formant synthesizer. Modification of the Klatt formant synthesizer allowed the formant resonators to be driven by an externally generated glottal waveform. All stimuli were: (a) 1 sec in duration, (b) synthesized at 20 kHz with 12 bits of amplitude resolution, (c) equated for overall RMS energy, and (d) synthesized with a formant frequency pattern appropriate for the vowel [a] (F1 = 720 Hz, F2 = 1240 Hz, F3 = 2400 Hz, F4 = 3300 Hz, F5 = 3700 Hz). With the exception of the jitter continuum, all signals were generated with a constant 130 Hz fundamental frequency. The jitter continuum was synthesized by introducing random perturbations around an average fundamental frequency of 130 Hz.

The formula used to generate the glottal waveforms (except the waveshape perturbation continuum) was from Rosenberg (1971, see Figure 1). Rosenberg concluded that listeners preferred glottal waveforms generated with this formula.
Figure 1. Rosenberg's Waveform.
**Fundamental Frequency Perturbation**

Jitter was controlled by introducing a specific amount of variability in the fundamental frequency of the glottal waveform produced with the Rosenberg (1971) formula. A simple white-noise random number generator was used to produce the sequence of jittered frequencies, which centered around 130 Hz.

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>Absolute Jitter</th>
<th>Percent Jitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.12</td>
<td>1.50</td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
<td>2.00</td>
</tr>
<tr>
<td>9</td>
<td>0.19</td>
<td>2.50</td>
</tr>
<tr>
<td>10</td>
<td>0.23</td>
<td>3.00</td>
</tr>
<tr>
<td>11</td>
<td>0.31</td>
<td>4.00</td>
</tr>
<tr>
<td>12</td>
<td>0.38</td>
<td>5.00</td>
</tr>
<tr>
<td>13</td>
<td>0.46</td>
<td>6.00</td>
</tr>
</tbody>
</table>

The distribution of fundamental frequency was set high for large jitter values and low for small jitter values. A 13-step continuum was created that varied from 0 to 462 μsec in absolute terms or 0.0-6.0% in relative terms.
Amplitude Perturbation

The same technique was utilized to synthesize stimuli varying in shimmer, except that the random number generator output was used to control the sequence of pitch-pulse amplitudes. A 13-step continuum was created that varied from 0.0 to 2.4 dB.

Table 2
Stimulus Values for 13-Step Shimmer Continuum

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>Absolute Shimmer (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>7</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>1.6</td>
</tr>
<tr>
<td>10</td>
<td>1.8</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
</tr>
<tr>
<td>12</td>
<td>2.2</td>
</tr>
<tr>
<td>13</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Additive Noise

An additive noise continuum was created by mixing perfectly periodic vowels with varying amounts of synthesized noise. The noise was synthesized by passing
white noise through formant resonators set appropriate for [a]. The periodic components were generated with the Klatt synthesizer using the parameter settings mentioned previously. The periodic and aperiodic signals were then mixed such that a 13-step continuum was produced varying in signal-to-noise ratio from 26 dB to -10 dB.

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>Signal-to-Noise Ratio (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.0</td>
</tr>
<tr>
<td>2</td>
<td>23.0</td>
</tr>
<tr>
<td>3</td>
<td>20.0</td>
</tr>
<tr>
<td>4</td>
<td>17.0</td>
</tr>
<tr>
<td>5</td>
<td>14.0</td>
</tr>
<tr>
<td>6</td>
<td>11.0</td>
</tr>
<tr>
<td>7</td>
<td>8.0</td>
</tr>
<tr>
<td>8</td>
<td>5.0</td>
</tr>
<tr>
<td>9</td>
<td>2.0</td>
</tr>
<tr>
<td>10</td>
<td>-1.0</td>
</tr>
<tr>
<td>11</td>
<td>-4.0</td>
</tr>
<tr>
<td>12</td>
<td>-7.0</td>
</tr>
<tr>
<td>13</td>
<td>-10.0</td>
</tr>
</tbody>
</table>

**Duty Cycle Perturbation**

Duty cycle perturbation was created by introducing random cycle-to-cycle variations in the duration of the open phase of the glottal cycle, keeping all other parameters constant. The average duty cycle for all stimuli on the continuum was 5.87 msec (76% of 7.72 msec), with the
standard deviations of duty cycle ranging from 0.0 to 0.40144 msec (5.2%) across the thirteen signal continuum.

Table 4

Stimulus Values for 13-Step Duty Cycle Continuum

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>Duty Cycle Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.435</td>
</tr>
<tr>
<td>3</td>
<td>0.869</td>
</tr>
<tr>
<td>4</td>
<td>1.304</td>
</tr>
<tr>
<td>5</td>
<td>1.738</td>
</tr>
<tr>
<td>6</td>
<td>2.100</td>
</tr>
<tr>
<td>7</td>
<td>2.607</td>
</tr>
<tr>
<td>8</td>
<td>3.042</td>
</tr>
<tr>
<td>9</td>
<td>3.476</td>
</tr>
<tr>
<td>10</td>
<td>3.911</td>
</tr>
<tr>
<td>11</td>
<td>4.345</td>
</tr>
<tr>
<td>12</td>
<td>4.780</td>
</tr>
<tr>
<td>13</td>
<td>5.214</td>
</tr>
</tbody>
</table>

Glottal Waveshape Perturbation

Synthesizing glottal waveshape perturbation involved introducing random cycle-to-cycle variations in the gross shape of the open portion of the glottal pulse, keeping all other parameters constant. The procedure involved introducing random perturbations in waveform details such as the slope of the leading edge, the slope of the trailing edge, and whether the peak was rounded or sharp. Perturbation was measured by computing cycle-to-cycle spectral differences of each period (one FFT bin) of the signal and then
averaging those differences. The numbers in table 5 represent this averaged spectral difference increasing from 0.0 to 2.73 dB/FFT bin across the 13-step duty cycle continuum.

### Table 5

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>Spectral Difference (dB/FFT bin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.54</td>
</tr>
<tr>
<td>3</td>
<td>0.84</td>
</tr>
<tr>
<td>4</td>
<td>1.13</td>
</tr>
<tr>
<td>5</td>
<td>1.37</td>
</tr>
<tr>
<td>6</td>
<td>1.59</td>
</tr>
<tr>
<td>7</td>
<td>1.81</td>
</tr>
<tr>
<td>8</td>
<td>1.98</td>
</tr>
<tr>
<td>9</td>
<td>2.10</td>
</tr>
<tr>
<td>10</td>
<td>2.21</td>
</tr>
<tr>
<td>11</td>
<td>2.33</td>
</tr>
<tr>
<td>12</td>
<td>2.43</td>
</tr>
<tr>
<td>13</td>
<td>2.73</td>
</tr>
</tbody>
</table>

### Subjects

The listeners were nine graduate students (seven females and two males) studying Speech-Language Pathology at Western Michigan University, Kalamazoo. The listeners' mean age was 26 with an age range of 23 to 31 years. All listeners reported normal hearing sensitivity as indicated by pure tone testing at 500 Hz, 1000 Hz, and 2000 Hz presented at 20 dB HL or less in a sound-treated booth less
than one year prior to this study.

Instrumentation

Experiments took place in a sound treated room. Stimuli were stored in the disk memory of a PDP 11/73 computer. Signals were sent through a Shure Model M67 Preamplifier, an Alison AL-2ABR Variable Filter (high pass setting 8160 Hz), a Grason-Stadler M162K Booster Amplifier and then were presented over a single Boston Acoustics A60 speaker.

Procedures

All listeners were tested individually in a sound treated room. Signals were presented free field at an intensity of 75dBA at one meter.

Scaling of Overall Severity

The main purpose of this study was to investigate the degree of dissimilarity in quality between pairs of signals. The preliminary investigation was designed to determine which signals on each 13-step continuum were equal in magnitude of severity and thus assure that listeners' judgments in the MDS study were based on qualitative, not quantitative differences.

Each listener performed magnitude of dysphonia ratings on a pool of signals comprised of the five 13-step continua.
(65 signals). Each block of 65 signals was presented in random order. Four blocks of signals were presented during each of ten listening sessions. Signals were presented one at a time and each listener was instructed to rate the amount of dysphonia heard in each of the signals. A self-determined scale was employed such that larger numbers represented a more dysphonic voice. The very first block of 65 stimuli was disregarded to allow each subject an opportunity to become familiar with the task. The direct magnitude estimates provided by the listeners were later rescaled so that all estimates ranged from 10 to 90.

Dissimilarity Ratings

Listeners made dissimilarity judgments between pairs of synthesized stimuli varying in each of the five parameters at three different levels of overall severity. Results of the quantity experiment were used to determine the value of the signals for the five acoustic parameters at each of three levels—high, medium, and low (see Table 6). A signal was compared to itself and to the other four signals from within that severity level (5x5=25 signal pairs per level). Signal pairs from all three levels were then pooled (3x25=75 pairs). Each block of 75 signal pairs was randomized and presented twice during each listening session ([25 signal pairs/level x 3 levels] x 2 presentations = 150 signal pairs per listening session). The
interstimulus interval was 500 msec. The intertrial interval was 1 sec. The very first presentation of signal pairs (75 signal pairs) was not included in the analyses to allow each subject time to become familiar with the task. Subjects were allowed to hear any pair of stimuli repeated as often as they wished before entering a dissimilarity rating. Each subject was instructed as follows:

In this experiment you will hear synthetically generated voice signals presented in pairs over a loudspeaker. Your job will be to judge how different the two voice qualities are by entering a number on the computer terminal. If the two voices are very different, enter a large number, and if the two voices are very similar, enter a small number. You can use any number equal to or greater than zero. It makes no difference what range of numbers you use; that is, you can scale the pairs from 0 to 10, from 10 to 90, from 100 to 10,000, or whatever range of numbers you choose. If you wish to hear a pair presented again before making your judgement, enter a '-1' for the trial. You can ask for a pair to be repeated as often as you wish. It is important to emphasize that your task is to judge the dissimilarity of each pair in voice quality. The term voice quality here refers to whether a given voice sounds "rough," "breathy," "harsh," "strident," "clear," "weird," "bright," "thin," "metallic," or whatever other subjective term comes to mind. All of the stimuli in a given set were intended to be similar in the overall severity of dysphonia. However, individual listeners may differ from one another in their judgment of overall severity. We would like you to ignore as much as possible any differences that you may hear in the overall severity of dysphonia and base your judgement on the difference in quality between the signals.

A total of seven subjects took part in the MDS study. One subject from the first experiment was dropped from the
final experiment because of unavailability. An additional subject was omitted because of unresolvable problems analyzing the dissimilarity data generated.

Table 6

<table>
<thead>
<tr>
<th>Stimuli Type</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waveshape</td>
<td>13</td>
<td>06</td>
<td>04</td>
</tr>
<tr>
<td>Duty Cycle</td>
<td>07</td>
<td>03</td>
<td>02</td>
</tr>
<tr>
<td>Jitter</td>
<td>11</td>
<td>07</td>
<td>05</td>
</tr>
<tr>
<td>Additive Noise</td>
<td>06</td>
<td>04</td>
<td>02</td>
</tr>
<tr>
<td>Shimmer</td>
<td>09</td>
<td>05</td>
<td>03</td>
</tr>
</tbody>
</table>

Data Analysis

The MDS analysis was accomplished utilizing ALSCAL of the SPSS statistical analysis package. Intrasubject and intersubject differences were investigated by specifying the INDSCAL model subcommand. Intrasubject matrix weights were arrived at by stacking a subject's response matrices and analyzing them with ALSCAL as stated above. Intersubject response weights resulted from averaging each individual subjects' response matrices, stacking the eight subjects' averaged matrices, and analyzing the data file with ALSCAL as above. For all analyses, the order of the
stacked files was varied in at least three different sequences to investigate the reliability of the original analysis. Results for every analysis were found to be very reliable.
CHAPTER V

RESULTS

Overall Severity Estimation

The results are shown in Figure 2 which displays the degree of perceived dysphonia as a function of stimulus number along the various 13-step continua. It can be seen that the magnitude of perceived dysphonia increases systematically with increasing amounts of aperiodicity across all five continua. The rate of change in magnitude estimates was similar, but not identical for the five continua. Results also demonstrated a flattening of each continuum for larger amounts of perturbation. This is particularly true for the noise, duty cycle, and wave shape continua.

Dissimilarity Ratings

Intersubject Analysis Reliability

Multidimensional scaling solutions were obtained using the individual differences model (INDSCAL) of the SPSS ALSCAL program. Separate one-, two-, three-, and four-dimensional space solutions were derived. Table 7 shows the goodness-of-fit measures for group data subjected to
Figure 2. Quantity Magnitude Estimation as a Function of Stimulus Number.
each of the four solutions. These data consisted of each listener's averaged quality magnitude estimation response files, i.e., the average of each listener's ten response files. Three additional data sets that consisted of the same ten averaged data matrices

TABLE 7
Comparison of Goodness-of-Fit Values* for One to Four Dimensional Solutions for Group Data

<table>
<thead>
<tr>
<th>Number of Dimensions</th>
<th>Goodness of Fit (RSQ*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.878</td>
</tr>
<tr>
<td>2</td>
<td>.879</td>
</tr>
<tr>
<td>3</td>
<td>.932</td>
</tr>
<tr>
<td>4</td>
<td>.938</td>
</tr>
</tbody>
</table>

*Values are squared correlations

stacked in varying order revealed results identical to Table 7. The goodness-of-fit values are squared correlations (RSQ) which reflect the amount of variance in the listeners' perceptions accounted for by each solution.

Based on these results, the 3-dimensional solution was selected as the most appropriate solution. This solution accounted for more of the variance than did the 2-dimensional solution. The 4-dimensional solution, alternatively, was not chosen as the improvement in variance accounted for beyond the 3-dimensional solution was relatively small.
3-Dimensional Analysis

The results of 3-dimensional analysis of the listener's perception of the dissimilarity in quality between signal pairs is displayed in Figure 3. It can be seen that signals were perceived to be overall most different from one another at high levels of perturbation. Duty cycle (D) and jitter (J), however, were consistently judged to be very similar by all listeners. Additive noise (N) was perceived to be significantly different from the other four experimental parameters by all listeners. Waveshape (H) and shimmer (S) were judged to be relatively similar to one another and distinct from both additive noise and duty cycle/jitter.

At medium levels of perturbation, no consistent pattern of interassociation among the five parameters was noted except that the five parameters were clustered relatively more closely together than at high levels of perturbation. All five signals at low levels of perturbation were consistently perceived to be very similar.

Intrasubject Analysis

The importance associated with each dimension in the solution for each subject is shown in Table 8. The results demonstrated that more than half of the weight given for
Figure 3. MDS Analysis Results.
TABLE 8

Individual Listener Weights and Goodness-of-Fit Values* for 3-Dimensional Solution

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>FIRST</th>
<th>SECOND</th>
<th>THIRD</th>
<th>RSQ*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB</td>
<td>.6642</td>
<td>.1390</td>
<td>.0956</td>
<td>.899</td>
</tr>
<tr>
<td>RC</td>
<td>.6684</td>
<td>.1774</td>
<td>.0976</td>
<td>.943</td>
</tr>
<tr>
<td>LG</td>
<td>.5282</td>
<td>.2096</td>
<td>.1960</td>
<td>.934</td>
</tr>
<tr>
<td>RN</td>
<td>.4772</td>
<td>.2550</td>
<td>.1905</td>
<td>.923</td>
</tr>
<tr>
<td>CO</td>
<td>.3640</td>
<td>.3323</td>
<td>.1985</td>
<td>.895</td>
</tr>
<tr>
<td>JS</td>
<td>.6628</td>
<td>.2044</td>
<td>.0876</td>
<td>.954</td>
</tr>
<tr>
<td>JT</td>
<td>.5491</td>
<td>.3193</td>
<td>.0656</td>
<td>.934</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>.5591</td>
<td>.2339</td>
<td>.1331</td>
<td>.926</td>
</tr>
</tbody>
</table>

* Values are squared correlations.

Each subject's solution was accounted for in the solution of the first dimension for the majority of listeners. Less importance was always given to the second and third dimensions respectively.
CHAPTER VI

DISCUSSION

Scaling of Overall Severity

The primary conclusion of the severity scaling study was that all five parameters were correlated with the perception of dysphonia. The data for all five acoustic parameters clearly indicated that as the magnitude of aperiodicity increased (represented by increasing stimulus number in Figure 2), so did the corresponding level of dysphonia assigned to the stimuli. The finding that the quantity of dysphonia tended to plateau at high levels of perturbation confirmed earlier results by Heiberger and Horii (1982), who concluded that, "beyond a certain point, relatively large increases in [jitter] did not result in similarly large increases in roughness level" (p. 321). In this present study, all five acoustic measures were found to follow this pattern of lesser increases in the amount of perceived dysphonia with continued increases in the amount of acoustic perturbation at some point in each continuum (See Table 9).
Table 9
Values of Perturbation at Which Increases in Perception of Dysphonia of Each Acoustic Parameter Lessened with Continued Increases in Perturbation

<table>
<thead>
<tr>
<th>Acoustic Parameter</th>
<th>Amount Perturbation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jitter</td>
<td>4%</td>
</tr>
<tr>
<td>Shimmer</td>
<td>1.0 dB</td>
</tr>
<tr>
<td>Duty Cycle</td>
<td>3.6%</td>
</tr>
<tr>
<td>Waveshape</td>
<td>2.10 dB/FFT Bin</td>
</tr>
<tr>
<td>Additive Noise</td>
<td>2.0 dB</td>
</tr>
</tbody>
</table>

Findings for the jitter percept and shimmer percept can be directly compared to the results of Heiberger and Horii (1982). In this earlier study, the point where the perception of increasing roughness lessened with increasing levels of jitter was found to be between 5% and 10%. Hillenbrand (1988) determined this similar point to be at 2% jitter. In the present study, the point of changing slope was found to occur at 4% jitter.

A similar comparison can be made for the shimmer percept. Heiberger and Horii (1982) determined that the point of changing slope occurred at 4.0 dB shimmer. Hillenbrand (1988) concluded this point occurred at 1.2 dB shimmer, while the results of this study found the slope to vary at 1.0 dB.

Hillenbrand (1988) compared his results with those from the Heiberger and Horii (1982) study. Hillenbrand
credited the lower values to be due to three factors: (1) differences in mean fundamental frequency (130 Hz versus 165 Hz), (2) differences in the perception of triangular waves versus "the more harmonically rich voice signals" (p. 2365), and (3) the fact that Heiberger and Horii presented the test stimuli over head phones as opposed to a loudspeaker as in the Hillenbrand study. Hillenbrand concludes, "these differences would be expected to make the Heiberger and Horii stimuli sound less rough than the stimuli used in the [Hillenbrand] study (Wendahl, 1963, 1966a,b; Coleman, 1969; Wilde et al., 1986), which might have the effect of moving the entire roughness-perturbation function to the right" (p. 2365).

Heiberger and Horii (1982) state that "it may be concluded that both jitter and shimmer magnitudes appropriate for human voices do not correlate well with the perceived roughness" (p. 323). The signals in this present study, however, included many signals that are found to exist in naturally produced voices and as is evident in Figure 2, listeners were able to perceive an apparently significant difference between signals that contained even the low degrees of perturbation. This perceived difference is revealed by the spread of data points across the vast majority of Figure 2. The lower magnitude signals, i.e., those thought to be within the range of naturally produced voices, have been perceived to
be relatively more similar as a group than the higher magnitude signals. When asked to judge these signals on perceived magnitude of dysphonia, however, all listeners agreed that there was a difference in magnitude between every signal within each continuum except for the first stimulus, i.e., the signal that contained no perturbation.

Results from the quantity magnitude experiment also revealed that the listeners judged 1.5% jitter, 0.8 dB shimmer, 17.0 dB additive noise, 1.2% duty cycle perturbation, and 1.59 dB/FFT bin waveshape perturbation to sound equally dysphonic. These findings support the results of Heiberger and Horii (1982) that the dysphonia produced by 1.0 dB shimmer is comparable to that produced by the 2.0% jitter stimulus.

Dissimilarity Ratings

Perception of Jitter and Shimmer

These results are not in agreement with the findings of Wendahl (1966b). As is evident particularly at the high level of perturbation factor in Figure 3, jitter and shimmer do not lie in close proximity to one another. Using sawtooth wave signals, Wendahl concluded that jitter and shimmer signals were virtually identical:

It is interesting to note that the roughness generated by these different procedures [synthesized jitter and shimmer stimuli] results in such similar auditory experiences. Some highly trained listeners were able to distinguish
between the two types of signals, but the writer, who has had years of listening experience with such stimuli, is able to discriminate between the program types only at the extremes of the continuum (Wendahl, 1966b, p. 106).

The differing results between this study and those of Wendahl (1966b) might be accounted for by the different methods of synthesis. Synthesis methods used in this study produced a signal that was a closer approximation of a voice signal than were Wendahl’s nonspeech sawtooth wave signals.

The findings of this study corroborate the results of Hillenbrand (1988) in that signals varying in pitch perturbation and those varying in amplitude perturbation were perceived by listeners as having different qualities of roughness. Hillenbrand asked subjects to classify stimuli into two categories: (1) from a jitter continuum or (2) from a shimmer continuum. Hillenbrand concluded that:

with the obvious exception of stimuli with zero perturbation values, subjects were generally able to determine whether the stimulus represented perturbations in pitch or amplitude ... the subjective qualities produced by jitter and shimmer in synthetic vowels are quite different, except at very low levels of aperiodicity (p. 2367).

In the present study, the difference in quality was particularly evident at the high severity level. Signals varying in pitch perturbation and those varying in amplitude perturbation at the medium level of perceived dyspho-
nic magnitude were judged to be more similar to each other than at the high level, but were still judged to be different by the majority of listeners. The only jitter and shimmer stimuli that were perceived to be relatively similar in the present study were perturbation signals initially judged to be low in magnitude. This was not remarkable, however, as all signals judged to be low in magnitude were perceived to be virtually identical in quality.

**Fundamental Frequency Perturbation and Duty Cycle Perturbation Perception**

Listeners consistently judged signals varying in fundamental frequency perturbation to be similar to those varying in duty cycle perturbation at virtually all levels of estimated dysphonia magnitude. These two time domain measures have not been associated like this before. It is extremely interesting that a glottal source measure is perceived to be so similar to an oral output source measure. From this finding, it could be proposed that cycle-to-cycle variations in duty cycle is a glottal event responsible for the output measure of jitter. That is, at least one origin of cycle-to-cycle variations in fundamental frequency may be from cycle-to-cycle variations in duty cycle.
Limitations of the Present Study

The greatest potential limitation of this investigation is that the MDS analysis procedure was used with univariate stimuli. MDS was designed to investigate perceptual attributes common to multiple stimuli, but the stimuli have always been multivariate, i.e., varying in more than one dimension at a time (see Danhauer & Singh, 1975; Kempster, 1984; Murry et al., 1977). The use of univariate stimuli is an unorthodox utilization of the procedure and must be taken into consideration until further research validates or invalidates this use of MDS.

The statistical analysis of intra- and intersubject response reliability would provide a better understanding of listener reliability in rating overall severity and dissimilarity of dysphonic voices. These additional analyses would broaden our current knowledge of listener ability to reliably classify disordered voices. Previous research has shown that classroom teachers can reliably identify children with voice disorders (DeGregorio & Polow, 1985; Wertz & Mead, 1975). The reliability of this skill for teachers is known to be most highly correlated with academic experience, age, years of professional experience, and accessibility to a speech-language pathologist (Phillips, 1976). The further statistical analysis of the data from the present study would provide similar information on
listener ability to quantify and to qualify disordered voices that would have important implications on further perceptual research.

An additional limitation of this project is the potential of unknown interactions between acoustic variables similar to those Hillenbrand (1987) has documented. It is uncertain at this time if it is possible to avoid these interactions in the production of synthesized speech stimuli. Research needs to establish if this interaction phenomenon is a natural process, a result of the synthesis technique, or a result of the measurement paradigm. In determining which is the case, valuable information about voice production, voice recognition, and voice synthesis would be gained.

Proposed Future Studies

The finding that the perception of the glottal measure duty cycle is similar to jitter is very interesting. Both of these measures are defined as glottal events. Jitter, however, can simply be quantitized from a voice sample. Duty cycle, on the other hand, is a glottal event that is relatively difficult to measure. The finding that these two parameters were perceived to be similar could have implications of common physiological origins for these two measures.

Any understanding of duty cycle, however, is limited
because of the paucity of information available. Studies on waveshape of the human voice are likewise few and far between. Further investigation of cycle-to-cycle variations of both these parameters are important to the complete understanding of voice intervention and perception.

As previous studies have been inconclusive, the existing knowledge base of acoustic correlates of vocal roughness requires further research. Investigating the additive effect of jitter and shimmer using perturbation values that are within the range of production of natural voices is a necessary and very important study. Attempts need to be made to determine if the interaction between acoustic variables documented by Hillenbrand (1987) is a result of the measurement process, the synthesis process, or is a natural process that we have yet to appreciate.

As synthesized voices are increasingly used to study both normal and abnormal voice production, voice scientists must strive to discover the limits of the natural voice production system. Early synthesis studies have failed to reach conclusive results with stimuli believed to be within the range of natural production. Voice scientists need to establish what levels of each of the acoustic parameters being studied are possible for the human vocal tract to produce at comfortable levels of effort given the physiological constraints of the voice production system.
Further perceptual studies are also required. In an attempt to gain a more thorough understanding of how listeners qualify dysphonic voices, listeners could be asked to qualitatively rate the same signals used in this present study given a closed set of descriptors (e.g., breathy, rough, natural, hoarse, un-natural, etc.). Attention would be given to alternative descriptors that the subjects devise while listening to the signals. A multidimensional analysis could then be performed on subjects' responses in an attempt to arrive at the subjective dimensions underlying these judgments.
CHAPTER VII

SUMMARY AND CONCLUSIONS

Numerous recent studies have attempted to identify acoustic dimensions that can be used to aid in the evaluation of voice disorders. No one acoustical parameter has been found to achieve this goal. This study investigated the perceptual effects of variations in five acoustical parameters.

Duty cycle perturbation, waveshape perturbation, fundamental frequency perturbation, amplitude perturbation, and additive noise were manipulated individually in discrete quantities through the use of synthesized speech signals. Listeners were first asked to rate the magnitude of perceived dysphonia of single stimuli. Results revealed that all five domains were correlated to the perception of magnitude of dysphonia. As the degree of perturbation was increased, evidence showed that the perception of dysphonia increased as well. The perceived dysphonia increased linearly to a point and then the slope of the line decreased slightly for all five parameters studied.

Based on these results, subjects were then presented pairs of stimuli from three levels of dysphonia severity and were asked to judge the degree of dissimilarity between
them. Judgments of quality of dysphonia between stimuli were analyzed using multidimensional scaling. Results revealed that signals varying in duty cycle and fundamental frequency perturbation were perceived to be very similar at all levels of severity ratings. Patterns for signals varying in amplitude and waveshape were not consistently found. The additive noise parameter was consistently judged to be unique in dysphonic quality at all three levels of estimated quantity.

The lack of significant correlations between the perception of jitter and shimmer counters the notion put forth by Horii (1980) that similar sets of physical factors (such as subglottic pressure, vocal fold elasticity) underlie the regulation of the individual fundamental period and intensity of laryngeal sounds.

Working within the confines of a steady pitch period, however, perturbing any one acoustic measure (e.g., fundamental frequency or amplitude) of a signal logically would also change other time domain measures of that signal. This co-varying phenomena of the acoustic signal could account for the results Hillenbrand (1987) found in his study. By adding perturbation and additive noise to synthetically generated voice signals that were otherwise perfectly periodic, Hillenbrand found very strong measurement interactions among the three variables jitter, shimmer, and additive noise. Lieberman (1963) concluded
pitch perturbations reflect variations in waveshape of the glottal area, as well as the periodicity of vocal cord vibration in both normal and pathological larynges. Given these results, it is apparent that there could be a fundamental parameter common to jitter, shimmer, and additive noise.

It is apparent from these findings, that further investigation into the perception of acoustic measures more closely aligned with values found to exist in naturally produced voices (jitter less than 1.5%, shimmer less than 1.0%) is necessary. As discussed previously, the additive dysphonic effect of bivariate (and multivariate) stimuli needs to be investigated as well. Only after the database of these basic voice science measures is expanded, can we expect to better understand how to objectively and accurately diagnose the disordered voice, to produce a more natural sounding synthesized voice, and begin to understand how we physiologically perceive the voice.
BIBLIOGRAPHY


