National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

Jitter and Shimmer in Natural English Vowels

by

Janice L. Adlington

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF Master of Science
IN
Speech Production and Perception

Linguistics

EDMONTON, ALBERTA
Fall 1988

ISBN 0-315-45614-0

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Janice L. Adlington

TITLE OF THESIS Jitter and Shimmer in Natural English Vowels

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED Fall 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) ..Janice Adlington..

PERMANENT ADDRESS:

...787..Attlee St.,......
...SUDBURY,..Ontario...
...P3A...3H5...........

DATED ..July..28...... 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Jitter and Shimmer in Natural English Vowels submitted by Janice L. Adlington in partial fulfilment of the requirements for the degree of Master of Science in Speech Production and Perception.

A Rozsypal
....
Supervisor

John T. Hogan
....

....

....

Date 28 July 1988

Abstract

To establish normative values for natural production, jitter (glottal period perturbation) and shimmer (amplitude perturbation) were measured for nine Canadian English vowels, produced by eight male and eight female speakers in the sentence frame "Please say /hVd/ not /hVd/." The speech signals were digitized at a 20 kHz sampling rate. Following extraction of the vowel, the duration and peak amplitude of each period were measured using a semi-automatic peak-picking procedure with quadratic interpolation. Jitter and shimmer were determined as distance from a two-point linear trend line centered around the current period. Period measures were normalized by dividing this distance by the local mean period duration averaged across three periods; a similar measure was employed for shimmer.

Analysis of variance was used to examine the effects of vowel quality, speaker sex, and intonation (or sentence position) on the jitter and shimmer magnitudes. For both types of perturbation, unexpectedly large speaker differences were found. When the speakers were clustered into relatively homogeneous subgroups, significant main effects for vowel appeared, with more jitter [illegible] /ɩ/ than for /ɚ/, and more shimmer for /æ/ and /ʌ/ than for /u/ and /i/. Significant vowel by speaker interactions indicated that these effects would not necessarily hold for any given speaker. For shimmer, a significant main effect for

position emerged, with more shimmer in sentence-final words.

The relation between jitter and shimmer within the vowel was investigated by cross-correlating the signed jitter and shimmer perturbations of individual vowel periods. Significant correlations appeared for less than one quarter of the vowel tokens. The signed jitter and shimmer values were also autocorrelated for lags ranging from one to twelve periods, to test for regularities within the perturbations. No consistent long-term cycles in the perturbations were found.

Acknowledgments

I gratefully acknowledge the guidance and untiring supervision provided by Dr. Anton Rozsypal. This project would not have been possible without the attention and time he generously devoted to it. Many thanks are due. I would also like to express my appreciation to Dr. Baker for his guidance in matters statistical; and to Drs. Wilson and Hogan for their time and effort spent in reviewing the thesis manuscript.

Additional thanks go to the students of 4-23 Assiniboia (particularly Martha Smith, Katherine Tiede, and Jean Andruski) for help practical and otherwise, and to the assorted friends, relatives, and Pembina people who volunteered to act as speakers.

Financial support from the Social Sciences and Humanities Research Council of Canada, the Province of Alberta, and the Department of Linguistics was greatly appreciated.

Table of Contents

List of Tables

List of Figures

I. Introduction

It has long been recognized (Scripture, 1906; Simon, 1927; Moore & von Leden, 1958) that small perturbations or irregularities of glottal vibration are present in normal phonation. Within the last twenty-five years, the cycle-to-cycle variations of the glottal period, called jitter, and of the peak amplitude, called shimmer, have been quantitatively examined by researchers from various fields. Proponents of voice "psychological stress evaluators" have claimed a reduction in perturbation permits the reliable detection of lies (Disner, 982); singers and voice teachers noted the correlation between excessive perturbation and rough or harsh voice qualities (Murry & Large, 1979; Murry, 1980); and phoneticians, concerned with the production of high-quality synthetic speech, associated minute amounts of random perturbation with increased naturalness (Kersta, Bricker, & David, 1960; Askenfelt & Hammarberg, 1986). However, the greatest interest has perhaps been shown by laryngologists or clinicians, who have attempted to apply jitter and shimmer measures to the early detection of laryngeal pathologies. Many measurement studies have been conducted with this aim.

The diagnostic use of jitter and shimmer assumes that laryngeal neoplasms or lesions disrupt the normal functioning of the larynx, and so raise perturbation magnitudes above the levels expected for healthy speakers. In establishing these levels, the clinician attempts to eliminate all factors which do not relate directly to the physical-structural condition of the voice source. To control for the potential influence of phonetic context, stress, or intonation, measurements are commonly made from the central portion of sustained vowels. The speakers may be instructed to phonate as steadily as possible, and the trial, of several, with the lowest jitter or shimmer may be selected for analysis. Fundamentally different questions may thus be asked by the clinician and the phonetician: the former concentrates on the capabilities of the speaker, exploring the limits of his performance, while the latter may examine his natural performance under normal (laboratory) conditions.

This study measures the jitter and shimmer produced by healthy adults in natural English vowels, extracted from a sentential context. It is, in part, motivated by an experiment conducted by Rozsypal and Millar (1979), in which subjects rated the naturalness of sustained synthetic vowels containing controlled amounts of jitter and shimmer. They found the optimal amount of jitter to depend on the vowel sound, while the presence of shimmer always decreased naturalness. The data from natural, connected speech, which might explain these preferences, are not available: shimmer, in particular, has not been examined in this environment. The present study remedies this lack: its purpose is to develop a procedure, with appropriate software, to measure perturbations in natural vowels, and so to provide information on natural production. Although perceptual effects are not tested, this work additionally generates data which may be applicable to future studies of the perceived smoothness or roughness of the human voice.

A number of specific questions are addressed. The measurement sections aim to establish the magnitudes of jitter and shimmer in natural vowels, and the effect which the factors of vowel quality, speaker sex, and intonation (or sentence position) may have on these magnitudes. Correlations between the mean jitter and shimmer values of the vowel tokens, and cross-correlations between the signed jitter and shimmer values of individual vowel periods, explore the relation between the two types of perturbations. Time series analysis is used to determine whether the perturbations are random, as is often suggested, or whether they display some degree of regularity.

The following chapter reviews previous experimental evidence related to these questions. Many studies are drawn from the clinical literature, and investigate jitter and shimmer in sustained phonation; results are reported for comparative purposes. The presence of inconsistent or contradictory results particularly is noted. Chapter Three provides a detailed description of the jitter and shimmer extraction procedures followed in this study. In Chapter Four, the obtained measurements and correlations are presented. The final chapter discusses these results and offers an explanation, based on individual differences among "normal" speakers, for the inconsistencies of the literature.

II. Literature Review

In 1963, Philip Lieberman published a paper, entitled "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges". In it, he suggested that measurements of duration variations in successive periods could be applied to the detection of laryngeal pathologies. This idea, offering an objective, non-invasive way to assess the health of the larynx, inspired further measurement studies, initially of frequency perturbations alone, and later, also of amplitude perturbations. The aim of many of these studies was to develop a technique which could reliably discriminate between normal, healthy speakers and speakers with pathologies. Although normative data were generated, the degree of comparability among such studies was often limited by technical innovations, implemented to improve discrimination. The first portion of this chapter describes methodological factors which must be considered when examining these results.

Studies of jitter in connected speech may be within or outside of the clinical tradition. Early papers by Lieberman linked jitter to the expression of emotional content (Lieberman, 1961; 1962). A small number of clinical studies examined the absolute magnitudes of the jitter produced by normal and pathologic speakers in sentences or phrases. A discussion of these papers introduces the second major section, which presents results from previous measurement studies. In the following subsections, the effects of frequency, speaker sex, and vowel quality on jitter and shimmer are described, for the sustained vowels of healthy or control speakers. It is noted that the reported statistical effects do not appear to remain stable from study to study. The section concludes with a brief consideration of certain additional factors which have been shown to influence perturbation magnitudes.

The final sections of this chapter present speculations on the origins of the perturbations, and report on previous correlational and autocorrelational analyses.

A. Methodological Considerations

Information on factors which influence jitter and shimmer is obtained from studies which employ a variety of measurement techniques. This section describes the range of procedures available, and the manner in which the options selected may affect results. The features discussed must be determined for each study before results can be evaluated or compared.

Signal Transducers

Although perturbation studies commonly use high-quality conventional microphones to transduce the airborne speech signal, three other devices, contact microphones, electroglottographs (EGGs), and photoglottographs may be employed to more directly investigate variations in the glottal source. Contact microphones (including accelerometers) are attached externally to the skin in the pretrachial area of the throat; while passage through the tissues between glottis and surface filters out high frequencies, the fundamental is not affected. EGGs utilize two electrodes, symmetrically attached over the thyroid cartilege, to record impedance changes across the larynx; rapid decreases in impedance mark vocal fold closures. Photoglottographs illuminate the glottis with a light source; a photodetector then records the changes in luminous intensity which occur as vocal fold movements modulate the light. Each of these devices produces a simple waveform for analysis.

The choice of instruments does not appear to greatly influence jitter results. Comparability with standard microphones has been demonstrated by Horii (1982), who reported similar jitter magnitudes from a miniature accelerometer and a conventional microphone, and by Horiguchi, Haji, Baer, and Gould (1987), who found a significant correlation between EGG-jitter and airborne jitter. Photoglottography, though somewhat invasive, has been shown to provide reliable determination of the fundamental periodicity (Baken, 1987). Since jitter represents variations in the glottal period, instruments which are sensitive to the source vibrations, but not to the acoustic resonances of the vocal tract, may indeed improve measurement accuracy by simplifying pitch extraction. Valid results are thus produced by all four types of device.

The case is different for shimmer. Horii (1982) found accelerometer signals to have approximately half the shimmer of airborne ones, for the same utterances. This was attributed to phase shifts introduced by the resonances of the vocal tract, which affected the waveshape and peak amplitudes of the airborne wave, but which did not influence the neck-wall vibrations. The glottal area function provided by the photoglottograph may similarly display less shimmer than the airborne signal. For perceptual studies, the airborne signal should thus be analyzed. With EGGs, interpretation of the amplitude perturbations is problematic. The impedance sensed by the EGG represents the surface area of contact between the vocal folds; shimmer in the EGG trace reflects irregularities in the mode of contact. However, EGG-shimmer and airborne shimmer do not appear to be correlated (Horiguchi et al., 1987). In addition, the EGG's sensitivity to anatomical differences calls subject or sex effects into question: as Haji, Horiguchi, Baer, and Gould (1986) state, the greater EGG-shimmer they reported for females could be entirely due to a thicker fatty layer over the female throat and the attendant difficulty in obtaining a clear trace. For these reasons, shimmer results from voice microphones, accelerometers, photoglottographs, and EGGs should not be compared directly; shimmer determined from an EGG trace must, in particular, be treated as a distinct phenomenon.

Pitch Extraction

Literally hundreds of algorithms have been proposed to determine pitch in speech signals (where "pitch" refers to fundamental frequency and/or period duration, rather than their perceptual correlates); however, only a limited number of methods are appropriate for perturbation studies. The algorithms or devices used must, most obviously, be capable of measuring the signal on a period by period basis. This excludes the "short-term analysis" algorithms, based on frequency-domain analyses or correlations (Hess, 1982), since these estimate the average pitch from signal frames containing two or more periods.

The method most commonly employed is "peak-picking", with the maximum value of the major peak selected as the boundary point for each period. For jitter, the acoustic signal is often low-pass filtered prior to processing, to reduce the complexity of the input wave; this can produce a signal equivalent, in effect, to that from a contact microphone. The definition of shimmer requires that some form of peak-picking be used. Any pre-filtering in a shimmer study must be noted, however, as low frequency cut-offs, particularly those below 2 kHz, will decrease magnitudes (Titze, Horii, & Scherer, 1987). Peak-picking is comparatively simple to implement; its major disadvantage lies in a sensitivity to waveshape changes during segment transitions.

Two other procedures, zero-crossing detection and inverse filtering, may be used in jitter studies. Zero-crossing detection is generally preceded by extensive low-pass filtering. Measurements so obtained may be influenced by the formant structure within the pitch period, and by any noise or dc level in the signal (Rabiner, Cheng, Rosenberg, & McGonegal, 1976). Residue inverse filtering mathematically cancels the vocal tract transfer function and the glottal shaping function to estimate the timing of the source. Although conceptually simple, it is complex in realization. Amplitude perturbations may, incidentally, be calculated from the residue signal (Davis, 1976, 1979): however, their perceptual significance remains unclear.

Because of a perturbation study's need for extreme accuracy, the procedure used is rarely fully automated. Tests by Rabiner et al. (1976) of several major types of pitch extractors found none capable of matching the performance of a human working interactively with a sophisticated display of the speech waveform. Although pre-processing of the signal, or extraction exclusively from sustained vowels, may improve performance, the majority of studies maintain some degree of active human control.

Temporal Resolution

Details of the period measurement procedure are largely determined by the device used to record the speech waveform: continuous high-speed filming of an oscilloscope display, or galvanic recording on light-sensitive paper, are accompanied by painstaking hand measurements, while the digital sampling and storage of signals is normally followed by computer-assisted processing. However, such differences in technique become critical only when they affect the temporal resolution of measurement; given the minute nature of the jitter perturbations, the importance of this factor cannot be over-emphasized. Reported time resolutions, or noise floors, range from 500 µs (Moore & Thompson, 1965), with hand measurements from a phonelograph trace, to 2 µs (Ludlow, Bassich, Connor, Coulter, & Lee, 1987), with special-purpose circuits to directly process the signal. Resolutions in the centre of this range, of 50 or 25 µs, are most frequently chosen.

If, as in most studies conducted within the last decade, a computer is employed to store and analyze the speech signal, the temporal resolution will be determined by the sampling rate. A high sampling rate must, in particular, be used when sex or frequency effects are investigated: since fewer samples per period are taken for shorter periods, the uncertainty of measurement becomes proportionally greater. Rates up to 100 kHz have been implemented (for example, by Zyski, Bull, McDonald, and Johns, 1984). The storage requirements at this rate severely constrain the length of signal which may be recorded, however, and some compromise rate is usually accepted. A rate of 40 kHz, giving a resolution of 25 μs, has been judged adequate for the study of jitter in normal, male and female voices (Horii, 1979). While interpolation between sampling points could reduce the sampling rate required (Titze et al., 1987), this option has not found wide acceptance.

An inadequate measurement resolution calls results into question, or limits the conclusions which may legitimately be drawn. Moore and Thompson (1965) reported jitter values of .30 ms for a severely hoarse voice, and .06 ms for a moderately hoarse one; as Heiberger and Horii (1982) note, the high noise floor of their analysis system makes these results uninterpretable. Kasprzyk and Gilbert (1975) attempted to determine whether jitter varied as a function of tongue height in sustained vowels; with a Visicorder record providing a resolution of 250 μs, they could show only that differences above this magnitude did not occur among the tested vowels. Later studies, with finer resolutions, would indeed find such effects. Information on temporal resolution is thus required before results can be evaluated.

Amplitude Resolution

The analogous factor for shimmer is the amplitude resolution of the record-reproduce system. A high signal-to-noise ratio must be maintained during recording. If the signal is digitized, the A/D converter must use a sufficient number of bits, or quantization levels, to ensure that the quantization error does not obliterate the measured effect. As Titze et al. (1987) observe, though, amplitude resolution in general presents less of a problem than temporal resolution, since for most digitized speech waveforms, the number of bits per amplitude exceeds the number of samples per period; the measurement error is correspondingly smaller. Nine or more bits are considered sufficient to extract normal shimmer.

Definitions and Formulae

The greatest amount of variation occurs in the measures used to quantify perturbations: many indices have been proposed, in the belief that some reformulation may improve the discrimination between normal and pathologic speakers. These measures may be categorized into three general types, though: absolute measures, which ignore the overall fundamental frequency or amplitude of the utterance; relative measures, which normalize the perturbation magnitudes by either the mean frequency or the mean amplitude; and short-window measures, which determine perturbations as distance from the average over a small number of periods. Each type of measure incorporates the preceding type. In each of these measures, jitter may be defined in terms of either period duration, or its reciprocal, instantaneous frequency. Amplitude may refer to either the peak-to-baseline, or the peak-to-trough distance.

One of the first absolute jitter measures was Lieberman's Perturbation Factor (PF). Lieberman (1961) plotted the distribution of the differences in duration in adjacent periods P_i, that is, the distribution of

$$\Delta P = |P_i - P_{i-1}| \qquad 2.1$$

in sentences produced by healthy male speakers. A following study (Lieberman, 1963), which included speakers with pathologic larynges, showed that subjects with growths on the vocal folds produce a greater proportion of large differences than do healthy speakers. The PF, defined as the percentage of all perturbations equal to or greater than .5 ms, was then suggested as a suitable index with which to optimally separate the two groups. Although designed for this purpose, Lieberman's PF has also been applied to the investigation of vowel differences (Kasprzyk & Gilbert, 1975), and sex and age effects (Benjamin, 1981). A relatively gross measure, it provides little information on the productions of normal, healthy speakers.

An alternate approach is to calculate the mean of the durational differences $\Delta\overline{P}$ from individual periods P_i:

$$\Delta \overline{P} = \frac{1}{N-1}\left(\sum_{i=1}^{N-1} |P_i - P_{i+1}|\right) . \qquad 2.2$$

This may be useful when comparing the same subjects under different experimental conditions, as, for instance, in Sorensen, Horii, and Leonard (1980), who showed jitter to increase following administration of a laryngeal anesthetic. "Average Pitch Perturbation" (Zyski, et al., 1984) and "Mean Frequency Perturbation" (Ludlow, Coulter, & Genges, 1983a) have been used to describe this type of measure; its units may be milliseconds, microseconds, or Hertz. In practice, mean jitter so calculated is often presented in parallel with relative jitter values.

To quantify shimmer, Zyski et al. formulated the "Average Amplitude Perturbation" measure, by direct analogy to Average Pitch Perturbation. Although shown capable of separating normal and pathologic speaker groups, it has not been tested outside of their study. It remains the sole example of an absolute shimmer measure.

The appearance of relative jitter measures followed the discovery of a positive correlation between absolute jitter and period duration: larger differences in adjacent cycles tend to be associated with longer fundamental periods (Lieberman, 1963; Koike, 1973). This presented a problem to the use of absolute magnitudes for screening purposes, since pathologic speakers with higher fundamental frequencies were more likely to produce perturbations falling within the "normal" range. To compensate for this effect, Hollien, Michel, and Doherty (1973) suggested that mean absolute jitter be related to the mean frequency of phonation, in a measure called the "Jitter Factor":

$$\text{Jitter Factor} = \frac{\text{mean jitter [Hz]}}{\text{mean } f_0 \text{ [Hz]}} \times 100[\%]$$

$$= \frac{\frac{1}{N-1}\left(\sum_{i=1}^{N-1} |f_i - f_{i+1}|\right)}{\left(\frac{1}{N}\right)\sum_{i=1}^{N} f_i} \times 100[\%] , \qquad 2.3$$

where f_i is the instantaneous frequency of the i-th period, defined as the reciprocal of the period duration, P_i. The duration-based equivalent may be labelled "Percent Jitter" (Ramig & Ringel, 1983), "Average Percentage Pitch Perturbation" (Zyski, et al., 1984), or "Jitter Ratio" (Wilcox & Horii, 1980), although the last item can also refer to a measure scaled by a factor of 1000 rather than 100 (Sorensen et al., 1980). The relative jitter measures are the type most widely employed, both in normative studies and in research involving pathologic speakers; they remove a great deal of predictable variation from the analysis.

In a very small number of studies (Kitajima, Tanabe, & Isshiki, 1975; Haji et al., 1986; Horiguchi et al., 1987), a logarithmic frequency scale, the semitone scale, is used to compensate for variations in fundamental frequency. If cross-study comparisons are desired, the semitone values can be converted to percentages, by a procedure similar to that relating decibel and percentage shimmer values.

Amplitude is normally quantified on a relative scale: it therefore appears natural that differences in peak amplitude should be expresed in decibels, a unit based on amplitude ratios. Shimmer thus can be defined:

$$\text{Shimmer} = \frac{\left(20 \sum_{i=1}^{N-1} \left| \log \left(\frac{A_i}{A_{i+1}} \right) \right| \right)}{N-1} , \qquad 2.4$$

where A_i is the amplitude of the major peak of the i-th period. This measure, while frequently chosen, requires a logarithmic transformation in units if jitter and shimmer magnitudes are to be compared. Studies with this concern may instead calculate shimmer directly as a percentage; as in the "Average Percentage Amplitude Perturbation" of Zyski et al., and the "Shimmer Factor" of Klingholz and Martin (1985), peak amplitude may be simply substituted for period duration, or frequency, in corresponding jitter equations. A less common choice is to calculate shimmer from differences in rms rather than peak intensity. For sustained vowels, Hillenbrand (1987) found these two methods to produce similar results. However formulated, relative shimmer measures have the distinct advantage of controlling for variations in the overall level of signal transmission or recording.

The short-window measures form a sub-group of the relative measures. Koike (1973) noted that sustained vowels may exhibit slow and relatively smooth changes in period duration, as is evident in vibrato. To exclude the effect of these slow shifts, he proposed to determine jitter as distance from a smoothed trend line. A three-point moving average technique was used for this purpose, in the "Relative Average Perturbation" (RAP) measure:

$$\text{Relative Average Perturbation} = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{P_{i-1}+P_i+P_{i+1}}{3} - P_i\right|}{\frac{1}{N}\sum_{i=1}^{N} P_i} \qquad 2.5$$

The numerator here expresses the average absolute distance from the three-point average; the denominator normalizes this value by the average period duration across the entire segment. The corresponding frequency-based measure has been labelled the "Frequency Perturbation Quotient" (Takahashi & Koike, 1976). Hartmann and von Cramon (1984) weighted the elements of the three-point filter in a manner different from the RAP. They also normalized by a local average of period durations, although the isolated vowels they examined were apparently sustained at a constant pitch. Their measure, the "Fundamental Period Perturbation", is the most similar to that employed in the current study.

$$\text{Fundamental Period Perturbation} = \frac{1}{N-2}\sum_{i-2}^{N-1}\left(\frac{\left|\left(P_i - \frac{P_{i-1}+2P_i+P_{i+1}}{4}\right)\right|}{\frac{P_{i-1}+2P_i+P_{i+1}}{4}}\right) \times 100[\%] \quad . \tag{2.6}$$

A three-point window is not the only possibility. Kitajima et al. (1975) measured deviations from a five-point moving average, with weighting coefficients calculated by the least squares fitting method. Davis (1976) systematically investigated the effect of varying the window size, and concluded that, for jitter, a five-point window produced the best measure for normal-pathological discrimination. However, in spite of their utility or promise, these measures appear in comparatively few jitter studies.

The shimmer measures differ from each other primarily in the length of the averaging windows used. The "Relative Average Amplitude Perturbation" of Zyski et al., obtained by substituting peak amplitude, A_i, for P_i in the RAP equation (Eq. 2.5), calculates shimmer from a three-point average. In contrast, the "Amplitude Perturbation Quotient" of Takahashi and Koike (1976) defines the trend with an 11-point moving average. This measure was designed to control for long-term amplitude drifts, and, when used in conjunction with the Frequency Perturbation Quotient, to provide information on phenomena peculiar to the amplitude data. Davis (1976) found that, for shimmer as for jitter, a five-point window optimalized this function for diagnostic purposes. A five-point least squares trend line was employed by Kitajima and Gould (1976). It should be noted, however, that these studies, with the exception of that by Kitajima and Gould, use either contact microphones or inverse filtering: data are generally lacking on shimmer measured from trends in the acoustic wave.

The terms "jitter" and "shimmer" refer to sequential perturbation phenomena. They may be used in connection with perturbations in individual periods, or, as abbreviations for "mean jitter" and "mean shimmer", for the perturbations in entire vowel segments.

B. Jitter and Shimmer Magnitudes

Connected Speech

A small number of studies have measured jitter perturbations in connected speech. Two early papers by Lieberman (Lieberman, 1961; Lieberman & Michaels, 1962) were concerned with identifying factors related to speech quality, with the ultimate aim of enhancing the naturalness of synthetic speech production. Later research focussed almost exclusively on diagnostic applications: while sustained phonations were defended as the phonatory task most appropriate for screening purposes, phrases or sentences were also occasionally tested. Most of these studies employed absolute jitter measures, and so produced results of questionable generality.

Lieberman (1961) investigated the effect of simulated emotions on jitter magnitudes. Six male speakers were asked to read the sentence, "They have bought a new car", in eight emotional modes. The recorded waveforms were filmed from an oscilloscope display, and viewed on a microfilm reader. Individual periods were then measured, to the nearest .2 ms, from the leading edge of the major amplitude peak. The results were presented in terms of the relative frequency of occurrence of specific durational differences. Overall, Lieberman found the magnitude of the differences between adjacent periods to be greater than .6 ms 20% of the time, and greater than 1.0 ms 15% of the time. The most extreme changes occurred at the onset and end of voicing, and during sudden spectral shifts. When the distribution for each emotional mode was examined, smaller perturbations were seen to accompany doubt, fear, and happiness; a greater proportion of large perturbations occurred with boredom, confidentiality, and pomposity, and objective statements and questions. Although Lieberman noted that the magnitude of the differences increased with the period durations, and that doubt, fear, happiness, and confidentiality as a group contained shorter periods, he did not relate these observations, but suggested that the modes with reduced jitter require more conscious vocal control for their production. It would be interesting to isolate the contribution of this factor, with a measure which would compensate for the overall frequency ranges of the sentences.

A complementary study (Lieberman & Michaels, 1962) determined that pitch perturbations were pertinent to the transmission of emotional content. Three male speakers read eight sentences in eight emotional modes. The pitch pulses from these utterances were used to drive a fixed-formant synthesizer, which maintained formant values at 750, 1100, and 2450 Hz. The fundamental frequency information was systematically manipulated in the synthesized waveforms. The material was then presented to groups of naive listeners. For unprocessed speech, the subjects correctly categorized the emotional mode 85% of the time. When only pitch and amplitude information were included, in the first of the test tapes; the categorization rate fell to 47%. Smoothing pitch by a 40 ms time constant caused a further reduction to 38%; 100 ms smoothing produced 25% correct categorizations. When the results were examined in detail, Lieberman and Michaels noted that perturbation magnitudes appeared to be conditioned by the speech habits of the individual: different speakers favoured different acoustic parameters for expression of the same emotional modes. However, they concluded that the pitch perturbations, although less important than phonetic content or gross pitch changes, did contribute to the overall communication of emotion, and that it would be useful for speech transmission systems to preserve them.

The relation between jitter and emotion has not been pursued experimentally. Scherer (1986), as part of a model of affect expression, predicted the effect specific emotions would have on jitter magnitudes. He speculated that emotional arousal would produce tension changes in the striated musculature, which would in turn alter the size of the frequency perturbations. As an example, sadness would be accompanied by muscular hypotension, generating a "lax" voice with increased jitter. However, his predictions await empirical confirmation: since simulated emotions would not fulfill the conditions of the model, testing may prove difficult.

Research on jitter took a different direction following the publication of a paper by Lieberman (1963), in which he suggested that pitch perturbations might be used to detect laryngeal pathologies. Unlike the earlier studies, the focus was not on the productions of normal speakers, but rather on the degree to which healthy and pathologic speaker groups might be differentiated. Studies with this aim do not always report the jitter ranges of their subjects.

Lieberman (1963) recorded the voices of twenty-three speakers with pathologic larynges, and nine normal controls; he included an unspecified number of female subjects. Jitter was measured from the voiced sections of the sentences, "Joe took father's shoe bench out", and "They have bought a new car". The analysis procedure resembled that of the 1961 study; however, a finer vernier provided a temporal resolution of 50 μs. Lieberman reported the relative frequency of occurrence of the perturbations, for different ranges of the fundamental period, and observed, again, that larger perturbations tended to accompany longer periods. He also defined the Perturbation Factor, the percentage of the total number of perturbations greater than or equal to .5 ms, and plotted this measure against the median pitch period for each speaker. With normal speakers, perturbations of this magnitude generally occurred during rapid formant transitions, as, for example, near stops and voiced fricatives, and not during the steady-state portions of the vowels. This restriction did not hold for speakers with certain pathologies, who could produce large perturbations in all phonetic contexts.

Lieberman's Perturbation Factor was used in several later studies of jitter in connected speech (Smith & Lieberman, 1969; Hecker & Kreul, 1971; Heiberger & Horii, 1982). However, statements of "typical" values, such as 11.2% for a male or 3.7% for a female (from Heiberger & Horii), are not particularly useful, in the absence of details relating the perturbations to specific phonetic segments. Lieberman's results suggested that, although voicing transitions associated with vowels could generate large perturbations, the PF was not sensitive to jitter within the vowels of healthy speakers. Measurements from sustained phonations on the whole supported this observation. Iwata and von Leden (1970) found a total PF of 1.8% for 30 normal male and female speakers sustaining /a/. Zyski et al. (1984) reported that none of their 20 healthy speakers produced perturbations exceeding .5 ms in the central portion of this vowel. The magnitude of the PF for healthy speakers may thus be related to the number and types of transitions in the test passage; it appears too gross a measure for normal vocalic material.

As an absolute measure, the Perturbation Factor varies most directly with the fundamental frequency of the speaker; without data on this variable, comparisons among PF values are not meaningful. Relative measures, in contrast, provide more general information on jitter magnitudes by controlling for the overall f_0 levels. It is unfortunate that this type of measure has rarely been applied to the investigation of jitter in connected speech. Kitajima et al. (1975) argued that measurements from natural speech were potentially more sensitive to the disruptions of laryngeal function caused by pathologies. They measured the jitter produced by normal and pathologic speakers in both a sustained /a/, and in the Japanese all-continuant phrase, /aou umi/. Jitter in the vowel was determined as the mean of the difference between each pair of adjacent periods, in semitones (STs). To control for slowly-moving changes in frequency due to intonation, jitter in the phrase was calculated from a five-point least squares trend line, and expressed, again, in STs. If the trend indeed provided an appropriate correction for intonation, and if the measures thus were comparable, jitter appeared to be of the same approximate magnitude for the two types of material: normal male speakers produced perturbations ranging from .11 to .21 STs for the sustained vowel, and from .08 to .17 STs for the phrase, while normal females ranged from .15 to .24 STs for the vowel, and from .11 to .19 STs for the phrase. However, a sampling rate of 12,315 Hz gave a temporal resolution of only 81 μs. Tests on a saw-tooth signal of 100 Hz produced jitter of up to .13 STs; at 200 Hz, the jitter attributable to system error rose to .24 STs. The question of the comparative size of jitter in sustained or natural vowels thus cannot be resolved from this study; the magnitudes in all cases appear near to or below the system noise.

The measurement of jitter from sustained phonations is often defended with the statement that, in connected speech, the perturbations of interest for diagnostic purposes would be confounded with systematic perturbations due to phonetic context, stress, and intonation (Horii, 1979). It is clear, from Lieberman's work, that phonetic context may have a considerable influence on perturbation magnitudes. Stress and intonation are accompanied by changes in frequency levels which the absolute jitter measures will reflect. The effect of intonation on relative jitter has not been tested, however, and it is not immediately obvious that a relation exists. The discussion of frequency effects in the next section bears on this question, although frequency may not be the only relevant factor.

Research on perturbations in connected speech appears to be restricted to jitter; shimmer has not been investigated.

Frequency Effects

Frequency normalization controls for the linear effect of the overall f_0 on the jitter perturbations. A small number of studies have asked whether the relation is, in fact, linear, and whether a secondary effect for frequency might not remain following normalization. In attempts to resolve this question, jitter has been plotted and correlated with the fundamental, and measured from phonations sustained at specified frequencies. A consistent effect would have implications for jitter variations with intonation.

Three pitch-matching studies, Beckett (1969), Hollien et al. (1973), and Horii (1979), directly examined the relation between normalized jitter magnitudes and f_0 levels. Their results initially appear contradictory: Beckett reported his Perturbation Quotient to decrease with pitch, while the Jitter Factor of Hollien et al. increased, and the Jitter Ratio of Horii remained approximately constant for frequencies up to 210 Hz. Important methodological differences exist among these studies, however; attention must, in particular, be given to the measures used.

Beckett asked one male subject to sustain /a/ at four pitch levels, ranging from 130 Hz (C3) to 262 Hz (C4). The pitch periods were measured from a Visicorder Oscillograph trace, with an estimated temporal resolution of about 50 μs. The Perturbation Quotient, a measure unique to Beckett's study, was used to quantify jitter: it was defined as the sum of the perturbations, in .1 mm units, divided by the average frequency in Hz. Such a measure may be expected to decrease as frequency rises, if the absolute jitter stays constant or decreases, or if jitter increases at a slower rate than frequency. Unfortunately, absolute jitter values were not given, and it is not clear which situation was represented.

In the study conducted by Hollien et al., four young male adults sustained the vowel /a/, at frequencies of 100, 141, 200, and 282 Hz. The test material was recorded simultaneously with a reference signal on a two-channel tape recorder. The signals were then fed into separate channels of an oscilloscope, and filmed with a high-speed camera. Measurements were made by hand, with the reference signal as a guide; period durations were determined from the geometric centre of the major peak. The system error of this procedure was declared to be about .2%. Absolute jitter was given in Hz, representing the difference between the instantaneous frequencies of each successive pair of periods. Jitter so defined may be expected to increase with frequency, if jitter in ms is constant; it can increase even if jitter in ms decreases. This was indeed what was found: the absolute jitter, averaged over the four speakers, was .48 at 100, .76 at 141, .85 at 200, and 2.76 Hz at 282 Hz. When the absolute jitter was divided by the mean frequency, in the Jitter Factor, values of .47, .53, .43, and .97% respectively were produced for the four frequencies. An increase of relative jitter with frequency can be observed only at the highest f_0 level.

A similar effect was found in Horii's study. Horii had six male adults sustain the vowel /i/ at eleven f_0 levels, ranging from 98 to 298 Hz. The habitual f_0 levels of the speakers, as estimated from a short reading passage, ranged from 94 to 135 Hz. The speakers were instructed to phonate as steadily as possible. Their voice samples were recorded, then digitized via a 16-bit analog-to-digital converter at an effective sampling rate of 40 kHz; this gave a temporal resolution of 25 μs. A peak-picking program measured the period durations from the central portion of the vowels. Horii reported that the mean absolute jitter for the six speakers decreased from 51 μs, at the second f_0 level, to 24 μs, at the eleventh. Although the decreases, for individual speakers on individual trials, were neither monotonic nor invariable, the overall effect appeared consistent; a correlation of -.95 was noted between the eleven absolute jitter values and the frequency levels.

When the Jitter Ratio, a measure formed by dividing the mean absolute jitter by the mean period, was correlated with frequency, a positive relation was found (r=.78). ~~Horii observed,~~ however, that the increase of relative jitter with frequency was most evident above the eighth level. He suggested that, at the higher f_0's, the measurement resolution may have inflated jitter values, and concluded that frequency did not systematically influence relative jitter magnitudes. Hollien et al.'s data may be interpreted as supporting this idea; Beckett's results may or may not be in agreement.

Koike, Takahashi, and Calcaterra (1977) introduced an additional factor by suggesting that vocal effort be considered for speakers phonating over a range of frequencies. They instructed nine healthy adults to sustain /a/, first at a comfortable pitch, then at a higher and lower pitch level; they did not require the production of predetermined frequencies. The vowels were recorded with a contact microphone, and digitized at 20 kHz. Semi-automatic peak-picking was used to find the periods. Jitter was quantified with the Frequency Perturbation Quotient (see Eq. 2.5), a relative measure which determines perturbations from a three-point moving average. Plots of the FPQ against frequency, for the three tokens per speaker, showed no simple relation to exist between the two variables. However, the authors did note that the FPQ often assumed its lowest value for the central, comfortable pitch of the subject, "uncomfortable" or forced pitch levels increased perturbations. This effect could provide an alternate explanation for the results of Horii and Hollien et al., since some of the frequency levels they tested would be unusually high for male speakers. Intonation effects associated with frequency may then be predicted to occur only if the speaker utilizes the extremes of his range.

Ludlow et al. (1987) investigated the possibility that shimmer might vary with frequency. They correlated the mean shimmer values with the mean f_0's for 38 males and 61 females who sustained /a/ at a comfortable pitch. The vowels were recorded, then digitized with a 12 bit ADC. Shimmer was normalized by the mean peak level and expressed as a percentage. For the male speakers, they reported a Pearson r of -.41 ($p<.01$), indicating a tendency for shimmer to increase at lower frequencies; although significant, the coefficient does not suggest a strong relationship. For females, no significant correlation was found (r=-.09). Since each speaker contributed only one data point, explanations or speculations on the male correlation cannot be based on within-subjects effects, such as vocal effort. The relation between shimmer and frequency must be more precisely defined, in terms of frequency ranges or sex differences, before further statements may be made. Within-subjects effects, with the same speaker phonating at different frequencies, must also be examined.

Sex Effects

Although a number of studies have measured the jitter and shimmer produced by male and female speakers, results from both sexes are often pooled for analysis, with contrasts made only between healthy and pathologic speaker groups. The few studies which have directly examined sex differences, using a relative jitter or shimmer measure, provide a limited amount of data and no firm conclusions.

Sorensen and Horii (1983) reported the mean jitter and shimmer magnitudes of twenty adult females who produced /a/, /u/, and /i/. This study was one of a series conducted by Horii and his associates, and followed the same procedure as in Horii (1979). The speakers were instructed to sustain the vowels as steadily as possible for about five seconds. The items were recorded on tape and later digitized with a 12 bit ADC at a sampling rate of 40 kHz. A peak-picking program found the peak amplitudes and period durations from a 3-second section in the middle of the vowel. Each subject produced three repetitions of each vowel; only the trial with the lowest jitter, or the lowest shimmer, was included in the analysis. Jitter was normalized for frequency and expressed in percent, while shimmer was in decibels. The results were compared with those of Horii (1980), who had similarly measured jitter and shimmer for thirty-one adult males.

For jitter, Sorensen and Horii noted both a main effect for sex, with larger overall magnitudes for females, and a sex by vowel interaction. The values reported were .71% for /a/, .86% for /u/, and .96% for /i/, while Horii's male speakers had average magnitudes of .61, .60, and .72% for these vowels. T-tests showed a significant difference between the males and females for /i/ and /u/ (p<.05) but not for /a/. As the next section will show, though, few vowel effects appear to remain stable from group to group. It might be instructive, not only to compare Sorensen and Horii's females with other groups of male speakers, but to compare the male groups from Horii's various studies. Without such analyses, it would be premature to emphasize the sex by vowel interaction effect. However, the male speakers of Sorensen and Horii (1984), Wilcox and Horii (1980), and Horii (1982) did appear to produce less jitter overall than did the females: averaging over the three vowels, the magnitudes from these studies were .49, .55, and .66% respectively, as compared to the .84% of Sorensen and Horii's females.

Two other studies, Hartmann and von Cramon (1984) and Haji et al. (1986), tested for a sex effect on jitter, with differing results. Hartmann and von Cramon asked ten male and seven female normal speakers to sustain the vowels /a, e, i, o, u/. They digitized the voice signal at 20 kHz, then used the autocorrelation method to measure the durations of individual periods. When the five vowel values of each speaker were averaged together, females were seen to produce significantly more jitter than males (p<.005).

Vowel effects were not examined. In contrast, Haji et al., who used an EGG to record eighteen males and twelve females sustaining /a/, did not report a sex effect. The EGG signals were digitized at 20 kHz. Periods were defined from base-line crossings, and jitter was given in semitones. The source of the disparity in results is not known, but may perhaps depend upon the vowels measured, if sex by vowel interactions indeed replicate, or upon some detail of the measurement technique. These studies clearly do not resolve the question of sex effects for jitter.

For shimmer, Sorensen and Horii (1983) did not find a main effect for sex, but did note a sex by vowel interaction; their female speakers were again compared with the males from Horii (1980). Shimmer magnitudes for the females were 3.87% for /a/, 2.21% for /u/, and 2.68% for /i/, while Horii's males gave average values of 5.56, 3.87 and 4.35% for the three vowels. T-tests showed the females to produce significantly less shimmer for /i/ and /u/ ($p<.05$), but not to differ from the males for /a/. For consistency of presentation, the values have here been converted from decibels to percentages. As with jitter, though, there is some question whether the effects should in fact be attributed to the speaker's sex. Sorensen and Horii (1984) found magnitudes of 2.80% for /a/, 2.80% for /u/, and 2.56% for /i/, for twenty male speakers, values which are, for /a/ and /i/, lower than the averages for the female group. Had the tests compared the males of Sorensen and Horii (1984) with the females of Sorensen and Horii (1983), very different results would have been reported. The first task should thus be to identify factors which cause vowels to vary, within each sex or across the sexes; without this information, conclusions based on vowel interactions may not be extended beyond specific groups of male and female speakers.

Like Sorensen and Horii, Ludlow et al. (1987) failed to find a significant main effect for sex. Their thirty-eight male speakers produced an average shimmer of 5.1% for /a/, sustained for the subject's maximum phonation time, while their sixty-one female speakers gave a shimmer value of 5.3%. Unlike Sorensen and Horii, who examined only relatively young adults, Ludlow et al. included a wide range of ages in their groups.

Given that females as a group have higher speaking fundamentals than males, any between-subjects frequency effect on the relative jitter or shimmer measures would likely be reflected as a sex effect. The tendency, noted by Ludlow et al., for the shimmer of male speakers to increase as frequency decreases, does not appear to create a sex difference: neither of the two studies which examined this question reported a main effect for sex. With relative jitter, it is assumed that frequency does not affect the measures, if the speakers phonate at a comfortable pitch level, and if the temporal resolution is adequate. Comparing normal females with normal males, two of three studies found females to produce more jitter than males. Attempts should be made to replicate this effect, with particular attention given to the resolution of the measurement procedure. It may be trivially observed that a sex difference will be found if jitter is determined in absolute terms, with males producing more jitter than females (Iwata & von Leden, 1970; Benjamin, 1981); this is predictable from the relation between absolute jitter and period length.

Vowel Effects

A series of papers by Horii and his colleagues investigated the question of vowel differences for jitter and shimmer. Since these studies followed the procedure previously described for Sorensen and Horii (1983), they have the advantage of producing clearly comparable results. Perturbations were always measured from a central, three-second portion of the vowels /a/, /u/, and /i/, which the speakers sustained for five seconds. Both the reported statistical effects, and the relative magnitudes for the three vowels, will be examined.

Four of the five papers in this series reported significant vowel differences for jitter; the fifth paper, Sorensen and Horii (1984), did not test for effects. Horii (1980), recording thirty-one young male adults, and Sorensen and Horii (1983), with twenty young females, found /i/ to have significantly more jitter than /u/ or /a/.

Ramig and Ringel (1983), who measured the productions of forty-eight males; observed /i/ and /u/ to have significantly more jitter than /a/. Their speakers were classed into six groups, representing three chronological age ranges and two levels of physical condition; a speaker's condition was determined from measures of his resting heart rate, systolic and diastolic blood pressure, percentage of body fat, and forced vital capacity. Interactions with age or condition were not examined. Wilcox and Horii (1980), with twenty young males (mean age 23 years) and twenty old males (mean age 70 years), found significantly less jitter for /u/. However, the group means show the older speakers to be largely responsible for the effect: the young males gave average magnitudes of .53% for /a/, .51% for /u/, and .61% for /i/, while the values for the older speakers were .84, .58, and .76% for these vowels. The pooled data do not well describe the performance of the younger speakers.

When significant differences are considered, only two of the studies agree completely; however, three of the four show a difference between /i/ and /a/. To examine the relation between /i/, /a/, and /u/ in greater detail, the vowels from each group of speakers were ordered, from most to least jitter, irrespective of the significance of the differences. These orderings are presented in Table 1, below. Results from Ramig and Ringel's six groups are given separately, to allow comparisons based on age. From the table, it may be seen that ten of the eleven groups produced more mean jitter for /i/ than for /a/, with Wilcox and Horii's group of old male speakers as the sole exception. Ramig and Ringel's old males, with a similar mean age, gave the "normal" order. For /u/, no consistent relation appears: the comparative amount of jitter for this vowel varies with the group.

The signals in these studies were recorded with conventional microphones. Two other studies, Horii (1982) and Koike et al. (1977), examined vowel differences in the signal at the glottis. Horii used a miniature accelerometer to record eight English vowels sustained by twenty young male speakers. The procedure was otherwise identical to that of the above papers. He reported an order, from most to least jitter, of / ə, ɷ, o, æ, ɩ, i, u, a/, but noted no significant differences among the vowels. Koike et al., with a contact microphone, recorded twenty-one male and ten female speakers sustaining five English vowels. They did not statistically test for vowel effects, but plotted the Frequency Perturbation Quotient, a three-point short-window measure, for the vowels of individual speakers. No obvious trends could be attributed to vowel differences.

Shimmer was measured in four of the papers in Horii's series. Of these four, Horii (1980), with young males, and Sorensen and Horii (1983), with young females, noted significantly more shimmer for /a/ than for /i/ or /u/, while Ramig and Ringel, averaging over their male speakers

Sorensen & Horii, 1983 (young females):	/i, u, a/
Horii (young males):	/i, a, u/
Wilcox & Horii (young males):	/i, a, u/
Sorensen & Horii, 1984 (young males):	/u, i, a/
Ramig & Ringel (young males, good condition):	/u, i, a/
Ramig & Ringel (young males, poor condition):	/u, i, a/
Ramig & Ringel (middle aged males, good condition):	/i, u, a/
Ramig & Ringel (middle aged males, poor condition):	/i, a, u/
Ramig & Ringel (old males, good condition):	/i, u, a/
Ramig & Ringel (old males, poor condition):	/i, u, a/
Wilcox & Horii (old males):	/a, i, u/

Table 1. Reported vowel orders, from most to least jitter.

Sorensen & Horii, 1983 (young females):	/a, i, u/
Horii (young males):	/a, i, u/
Sorensen & Horii, 1984 (young males):	/ a/u, i /
Ramig & Ringel (young males, good condition):	/u, i, a/
Ramig & Ringel (young males, poor condition):	/u, a, i/
Ramig & Ringel (middle aged males, good condition):	/a, i, u/
Ramig & Ringel (middle aged males, poor condition):	/a, i, u/
Ramig & Ringel (old males, good condition):	/a, i, u/
Ramig & Ringel (old males, poor condition):	/a, i, u/

Table 2. Reported vowel orders, from most to least shimmer.

groups, found no significant differences among the 3-second vowels. As with jitter, Sorensen and Horii (1984) reported only the mean values for each vowel. When the relative magnitudes were examined, it was seen that eight of the nine tested groups produced more shimmer for /a/ than for /i/, with six groups giving the order /a, i, u/. The vowel order for each group is listed in Table 2.

Although vowel differences may or may not be statistically significant for a given group, there appears to be a strong tendency across groups for /a/ to be associated with both less jitter and more shimmer than /i/. Speculations on the possible origins of these effects will be given in the Discussion chapter; it may be peripherally noted that those studies which have found significant effects have offered no explanations. However, the presence or absence of significant differences does not seem to depend upon the speaker characteristics considered, since groups of male speakers may give results which are similar to those from females or older males, but which differ from comparable groups of young males. The speakers in these studies attempted to phonate as steadily as possible; it is not known whether results so obtained will generalize to more natural productions.

Other Factors

The preceding discussion focussed on factors which the current study examines experimentally. However, a number of additional factors, related to the task or speaker, have been shown to influence jitter and shimmer magnitudes. This section briefly reviews the effects which may be produced by variations in signal intensity, duration, and initiation, and speaker age, physical condition, and vocal training. Again, only the performance of normal, healthy speakers is considered, with all perturbations measured from sustained phonations.

While the relationship between voice frequency and absolute jitter encouraged the investigation of frequency effects, the consistent use of normalized shimmer measures, which control for variations in overall amplitude, did not motivate a similar examination of intensity. Incidental or casual observations appear to indicate, though, that intensity has a weak, between-subjects effect on jitter and shimmer magnitudes. Ludlow et al. (1987) correlated absolute jitter, in µs, and relative shimmer, in percent, with intensity for maximum duration /a/'s, which the speakers sustained at a comfortable loudness. They reported weak, but significant negative correlations for the jitter of their female speakers (r=-.37), and for the shimmer of their males (r=-.44) and females (r=-.44); as intensity decreased, the perturbations tended to increase. In contrast, Koike et al. (1977), who did not systematically pursue this question, noted no invariant effects for loudness on relative jitter or shimmer. They did, however, state that the variability attributable to loudness differences appeared to be smaller than that due to other phonatory factors, such as frequency. It must be emphasized that these results refer only to between-speakers effects. Since normative studies as a rule require speakers to either phonate at a comfortable loudness, or maintain a predetermined output level, information on within-speakers effects is lacking.

Beckett (1969) suggested that a reduction of air flow through the glottis, such as accompanies lower intensities, might affect the regularity of vocal fold vibration. In his study, he asked speakers to produce /a/ with high, medium, and low degrees of vocal constriction. The measured jitter was then seen to increase with the strength of the constriction. However, the speakers were trained to link high constriction with "increased vocal effort, a sense of tightness within the vocal tract, and a feeling of stress or strain," and it is not clear whether the increases should be attributed primarily to the decreased air flow, or to the abnormal effort of phonation.

Ramig and Ringel (1983), who considered the effect of vocal effort, noted that physically taxing tasks may cause performance breakdowns which are not present on habitual tasks. They measured jitter and shimmer from comfortable and maximum duration vowels, produced by speakers who represented three chronological age ranges, and good or poor levels of physical condition. While an analysis of variance revealed no significant effects for the comfortable duration vowels, differences did emerge for the maximum duration phonations, with more jitter and shimmer produced by speakers in poor condition, and more shimmer given by the oldest group than the youngest. The task, which required maximum subject effort, may here have accentuated existing speaker differences, or may have caused their unique manifestation. The perturbation magnitudes for the comfortable and extended durations were not compared.

Lieberman (1961, 1963) reported that greater jitter accompanied phonetic transitions, with the most extreme changes occurring at the onset and end of voicing. Although most studies have attempted to eliminate these transitions, both Hartmann and von Cramon (1984) and Koike (1973) confirmed Lieberman's observation for voice onset. Koike's study is particularly interesting, in that he further demonstrated the effect of different types of initiations. His thirty normal speakers produced the vowel /a/ with a simultaneous and a breathy onset. The RAP (Eq. 2.5), a three-point short-window measure, was then used to determine the jitter in the first seventeen vowel periods, which defined the onset portion, and in the following thirty-two periods, which represented the steady state. Koike found a significant difference between the two types of onsets, with an average jitter of 2.76% for the abrupt initiation as compared to 1.23% for the initiation portion preceded aspiration. A jitter of .46% was measured from the steady state. As might be expected, jitter appears to be sensitive to the different mechanisms involved in different transitions; these effects have not been further researched. The present study uses a breathy initiation for all vowel samples.

Of the speaker properties, the effect of age has received the greatest attention. The normal aging process may produce degenerative changes in the laryngeal tissues, or in the fine neurological control of the laryngeal mechanisms (Ramig & Ringel, 1983; Ryan & Burk, 1974). The perturbation increases which would result from such changes must be distinguished from those caused by pathologies or neoplasms. One goal of these investigations is thus to determine whether separate normative levels are needed for different age ranges.

The results are equivocal, with little agreement even among studies which utilize similar analysis procedures. Three papers in Horii's series, i.e. Ramig and Ringel (1983), Heiberger and Horii (1982), and Wilcox and Horii (1980), examined the perturbations of aged speakers. Ramig and Ringel reported no significant differences among their three age groups (mean ages 31, 53, and 68 years), for jitter and shimmer in comfortable duration vowels. An effect appeared only for the maximum duration task, with the oldest group giving more shimmer than the youngest. In contrast, Heiberger and Horii found that a group of older males, ranging in age from 60 to 80 years, produced both significantly more jitter, and more shimmer, than did a group of younger males, for perturbations measured from a central 1.2 second section of the vowels /a/, /u/, and /i/. Wilcox and Horii, who measured jitter from 3-second vowel segments, observed significantly more jitter for their older speakers (mean age 70) than for their younger ones (mean age 23). However, they also noted large individual differences, and a substantial degree of overlap in the distributions of the two groups: while the jitter ratios of their young males ranged from .32 to 1.28%, their older speakers gave values ranging from .31 to 1.89%. The only conclusions to be drawn are that some young and some old speakers differ, as do some groups of young and old speakers. Chronological age itself is not a good predictor of performance. Additional evidence for this last statement comes from Ludlow et al. (1987), who correlated the ages of male and female speakers with the jitter and shimmer magnitudes they produced. Speaker ages ranged from eighteen to over sixty. Perturbations were measured from maximum duration /a/'s. No significant correlations were found.

Ramig and Ringel (1983) suggested that the large intersubject variability within chronological age groups could be at least partially attributed to physiological differences among the speakers. They used such measures as the resting heart rate, the systolic and diastolic blood pressure, and the percentage of body fat to divide their speakers into two groups, whose members were in "good" or "poor" physical condition. Thirty males, in each of three age ranges, were tested; for each age range, the eight speakers with the best performance on the physical condition measures, and the eight with the worst, were selected to represent the two levels of physical condition. These speakers produced both comfortable and maximum duration vowels. As with age, an effect for physical condition was found only for the maximally prolonged phonations, with speakers in poor condition giving significantly more jitter and shimmer. Although the age-by-physical condition interactions were not statistically significant, the differences were most apparent among the elderly speakers. The authors concluded that, in addition to chronological age, measures of the age-related changes in body physiology, or physiological age, should be employed to classify normative data.

Murry and Large (1979) and Murry (1980) reported that voice training could reduce the amount of jitter produced in speech. While this claim could have interesting implications for the origins of the perturbations, it is unfortunately not clear that the between-groups differences observed in these studies were indeed due to this factor. In the earlier study, three female singers and five non-singers sustained the vowel /a/ in a comfortable conversational voice. No information was given on the age of the subjects or the sex of the non-singers. The utterances were tape recorded, then filmed from an oscilloscope display. Fifty periods in the steady-state portion of the vowel were manually measured from the projected film. Jitter was quantified with Hollien's "Jitter Factor" (Eq. 3). The singers were found to have an average jitter of .40%, while the non-singers gave an average of .88%. The second study followed the same procedure, but compared the jitter of four female singers, aged 26 to 37, with that of five male non-singers, aged 55 to 71. The average jitter of the singers, at .40%, was significantly smaller than the .99% of the non-singers. Murry and Large speculated that training may increase a singer's control over the expiratory airstream, allowing him to decrease perturbation magnitudes. While this is not an implausible hypothesis, the studies cannot be said to have demonstrated this effect, given the differences, particularly in age, between the groups.

C. Origins

On the most superficial level, jitter and shimmer derive primarily from the asymmetric and irregular vibration of the vocal folds. It then remains to be determined what factors, related to system properties or control, underlie this behavior. The following discussion describes some of the neurological, biomechanic, aerodynamic, acoustic, and psychosomatic sources which have been suggested in the literature. Many of the statements must be regarded as speculatory, since at present, few empirical investigations have attempted to connect specific physiological mechanisms with perturbations. In the absence of these data, no comprehensive model of the perturbation origins can be developed: the relative contribution of each source is not known.

One of the few experimental studies of the origins was conducted by Baer (1980), who attempted to relate jitter to laryngeal muscle control and to the inherent sloppiness of muscle excitation. His hypothesis relied on well-established muscle physiology. A muscle is composed of contractile fibers, organized into functional groups called motor units. The summed contractions of all motor units produce the contraction of the muscle as a whole. Since, at slow firing rates, the firings of individual units are unsynchronized, the outputs cannot sum to a perfectly constant result. Baer suggested that single-motor-unit twitches could cause perturbations in the outputs of individual laryngeal muscles, which would in turn be reflected in the vocal-fold tension and in the resulting voice periodicity. To test this theory, he made simultaneous recordings of the voice signal and of the electromyographic (EMG) activity from the cricothyroid muscle, for one male speaker sustaining a steady tone. The two signals were digitally processed. The instantaneous frequencies of the glottal periods were calculated from the voice waveform, and isolated single-motor-unit firings were identified in the EMG waveform. Comparison of the f_0 and

EMG traces showed corresponding changes in f_0, on the order of 1 to 2 Hz, to occur 70 to 80 ms after the single-motor-unit firings. Baer claimed that this indicated a relatively large neuromuscular contribution to the perturbations. He further speculated that these effects, though inherent, might be regulated, perhaps by vocal training, which might produce systematic differences in the motor control strategies of individual speakers.

While Baer asked why perturbations should be present, Sorensen et al. (1980) explored the mechanisms which restrict perturbation magnitudes. Ongoing adjustments in vocal fold tension are thought to be regulated by three brainstem reflexogenic systems, operating from sensitive mechano-receptor nerve endings embedded in the laryngeal tissues (Wyke, 1969; 1983). Sorensen et al. investigated the contribution of these receptors to frequency control by comparing normal jitter with that produced following application of a laryngeal topical anesthetic. In their study, five adult males sustained the vowel /i/ at eleven different frequencies, ranging from 98 to 298 Hz, under the two experimental conditions. Horii's recording and analysis procedure was used. Jitter was calculated both in absolute terms, in microseconds, and as Jitter Ratios, defined as the mean jitter divided by the mean period duration multiplied by 1000 (see Eq. 2.3). Overall, significantly more relative jitter was found under the anesthesia condition, at 1.20%, as compared to a jitter of .62% in the control recordings. The differences were particularly marked at the higher fundamental frequencies. While noting that mucous secretion or physical changes in the tissues could have caused the jitter increases, these researchers took the results to indicate the importance of tactile and proprioceptive feedback for the maintenance of the appropriate laryngeal tension during phonation.

Measurement studies with hearing-impaired speakers reveal that auditory feedback is particularly important for controlling the timing of the glottal source. Both Monsen, Engebretson, and Vemula (1979) and Metz, Whitehead, and Whitehead (1984) found abnormal cycle-to-cycle changes in frequency and intensity in the utterances of deaf speakers. Monsen et al., who examined the productions of twenty hearing-impaired adolescents, concluded that deafness may prevent a speaker from learning the phonatory consequences of the gestures which alter and maintain vocal-fold tension. Metz et al., who studied four hearing-impaired adults, observed that such speakers appear to lack control over the intrinsic laryngeal muscles responsible for maintaining vocal fold stiffness, resulting in tension imbalances between the two folds. Normally-hearing speakers automonitor their vocal output, readjusting the laryngeal mechanism on the basis of this information; long-term deprivation of such data increases perturbations.

While the same factors are commonly thought to produce both jitter and shimmer, Ludlow, Connor, and Coulter (1984) attempted to make a distinction for the pathological population, claiming that jitter is sensitive to changes in laryngeal morphology while shimmer is sensitive to changes in neurological control. This idea derived in part from the results of Ludlow, Coulter, and Gentges (1983b), who, for jitter, compared patients with vocal fold nodules, a neoplastic disorder which mechanically disturbs vibration, and patients with Parkinson's disease, a neuromotor disorder which may affect the intrinsic laryngeal musculature, with groups of age and sex-matched control speakers. They found that only the patients with nodules differed significantly from their normal controls. However, the results of Ludlow et al. (1987) forced these researchers to reconsider the distinction, as this later study found similar percentages of speakers with morphological and neurophysiological pathologies to be abnormal for each type of perturbation; the idea of a differential sensitivity for jitter and shimmer was then abandoned.

In addition to neuromuscular variations, physical-structural variations contribute to the perturbations. Laryngoscopic motion pictures reveal that slight asymmetries often exist in the configuration of the vocal folds (Heiberger & Horii, 1982); the tension, mass, or length of the two folds may differ. During phonation, the folds' mechanical properties may be modified by changes in the amount and location of mucous secretion (Hirano, 1979). These sources of random perturbations are not controversial, however, and are not the focus of experimental studies.

It is possible to view the vocal organs, from lungs to lips, as an aerodynamic system; the airstream mechanisms involved in phonation may then be seen to introduce perturbations into the signal. The airflow emerging from the glottis may become unstable or turbulent (Titze et al., 1987). Additive noise, which will be reflected in the jitter and shimmer measurements, may be generated by the leakage of air during the closed phase of the glottal cycle (Hillenbrand, 1987). Very large perturbations, exceeding .5 ms for jitter, may be produced by transient pressure changes across the glottis, caused by changing vocal tract configurations (Lieberman, 1963). Less obviously, the presence of either jitter or shimmer at the glottal source may create acoustic interactions in the vocal tract; source jitter may then produce shimmer in the acoustic signal, and vice versa. This last phenomenon is examined in the following section, in connection with Hillenbrand and Milenkovic's studies with synthetic speech.

The degree of control which an individual speaker may exert over the perturbations has yet to be determin[illegible] Lieberman (1961) suggested that some voluntary co[illegible] mechanism exists, since he found that the si[illegible]ed emotions which require greater conscious control [illegible] th[illegible] voice were accompanied by less absolute jit[illegible] As previously noted, it is thought that the vocal habits of a singer may, over time, be modified by training, allowing the singer to reduce perturbations magnitudes. Although these hypotheses are plausible, the influence of higher-order mechanisms has not yet been convincingly demonstrated.

D. Jitter and Shimmer Correlations

Correlations between jitter and shimmer bear indirectly upon the origins of the perturbations, by determining the degree of association between the two variables. However, caution must be exercised when interpreting these data. If it is assumed that jitter and shimmer may be measured independently of one another, a significant correlation may suggest that similar factors, related to the glottal source, underlie the regulation of individual period durations and amplitudes (Horii, 1980). Recent evidence from synthetic speech appears to indicate, though, that the acoustics of voice production cause interactions between the measured variables (Hillenbrand, 1987; Milenkovic, 1987). The measures may not then be directly interpreted as reflecting glottal events. Although the correlations cannot resolve this issue, a model of the origins must account for these relations.

The reported correlations vary considerably in strength. Horii (1980), Heiberger and Horii (1982), and Horiguchi et al. (1987) correlated the mean jitter and shimmer values for sustained vowels produced by healthy speakers. All perturbations were measured from the acoustic wave. Horii found a correlation of .47 (p<.001) for perturbations in three vowels sustained by thirty-one males. While noting that the predictive value was low, he took the correlation to support the idea that similar sets of physical forces, such as vocal fold tension, mass, length, and subglottic pressure, underlie both types of perturbations. Heiberger and Horii observed a correlation of .77 (p<.01) for three vowels produced by twenty elderly males. In contrast, Horiguchi et al. found no significant correlation between jitter and shimmer for the /a/ tokens of eighteen male and ten female speakers. It is not clear what factors might have caused these differences.

Hillenbrand (1987) suggested that significant correlations may be expected, due to the acoustic effects of vocal tract transmission. In a very interesting study, he manipulated synthetic vowels to demonstrate that jitter, introduced at the source, may produce shimmer in the acoustic signal, and that source shimmer may produce acoustic jitter. The stimuli were five-formant /a/ vowels, generated with a Klatt formant synthesizer. For the first test, he created a sequence of 22 vowels, in which mean source jitter ranged from 0 to 6.4%. The vowels were each 200 periods long, with a mean fundamental frequency of 130 Hz and a constant source amplitude; only the standard deviation of the fundamental frequency distribution was altered. Shimmer, calculated as the mean cycle-to-cycle difference in rms, was then measured from the synthesized waveforms. Hillenbrand found changes in jitter to have strong effects on these measurements, with shimmer increasing to .80 dB (9.64%) for the token with the greatest jitter. The second test reversed the conditions. Mean shimmer at the source was varied from 0 to 2.6 dB (34.89%), while the source frequency was kept constant. Jitter was then measured from waveform zero-crossings. The effect of shimmer on jitter was observed to be relatively weak, with jitter remaining around .01% for shimmer values below .5 dB (5.92%). A jitter value of .86% was obtained for the token with the greatest shimmer.

Hillenbrand discussed two ways in which pitch perturbations could influence peak amplitudes. The intensity of a pulse in the acoustic waveform is, first, partially determined by the relationship between the harmonics of the glottal source and the location of the vocal-tract resonances. These relations become more variable as jitter increases. The second, more important effect is the energy overlap between adjacent pulses. A glottal pulse will, in general, be generated before the previous pulse is fully damped; the current pulse is then superimposed on the tail of the preceding pulse. With increasing jitter, the degree of overlap becomes more variable. Since the overlap is greater for shorter fundamental periods, jitter may be predicted to have stronger effects on shimmer at higher fundamental frequencies. Tests by both Hillenbrand (1987) and Milenkovic (1987) indicate that this indeed occurs. Milenkovic additionally showed the effect of jitter on shimmer to vary for the synthetic vowels /a/, /u/, and /i/: the different vocal tract impulse responses of the vowels also influence the jitter/shimmer interaction.

As Hillenbrand noted, it would not be surprising to find that jitter and shimmer are actually correlated at the glottal source. Correlations between perturbations in the acoustic waveform cannot be directly interpreted as indicating a common source origin, though: they may reflect source effects, acoustic effects, or some combination of the two.

E. Time Series Analysis

While jitter and shimmer are often characterized as random perturbations (Hollien et al., 1973; Baken, 1987), time series analysis has revealed that the perturbations may vary in a regular manner. This type of analysis correlates a series of measurements with time-delayed copies of itself, for successively greater lags, expressed as numbers of voice periods: a significant correlation indicates the presence of a repeated pattern, with a period equal to the lag. Serial correlation coefficients have been calculated for measurements of peak amplitude (Koike, 1969; Von Leden & Koike, 1970), period duration (Hiki, Sugawara, & Oizumi, 1966), and jitter (Iwata, 1972).

Koike (1969) and Von Leden and Koike (1970) examined amplitude modulations in the sustained vowels of normal speakers and speakers with laryngeal pathologies. Koike included 20 normal speakers, while Von Leden and Koike looked at 35 such speakers: only results for these subjects are presented. The studies followed identical procedures. A contact microphone was used to record sustained productions of /a/. The signals were recorded on tape, then played back into a Visicorder, which produced signal traces at a paper speed of 200 cm/s. The amplitude envelopes of 30 consecutive periods, from the steadiest portion of the utterances, were then manually measured at the dominant amplitude peaks. These measurements were autocorrelated, for lags of 1 to 15 periods. Koike found a significant positive correlation between the amplitudes of consecutive periods (lag=1) for 13 speakers, while high positive correlations appeared for 32 of Von Leden and Koike's subjects. Von Leden and Koike did not test the statistical significance of their results. Correlograms, which express the correlation coefficients as a function of the lag, also showed the presence of long-term periodicities: a negative peak in the vicinity of lag 12 was "quite common" in Koike's study, and was found for 21 of Von Leden and Koike's speakers. Short-term periodicites of between 2 and

10 lags appeared for 2 of Koike's subjects and 11 of Von Leden and Koike's subjects.

Iwata (1972) used Koike's basic procedure to investigate periodicities of pitch perturbation. Twenty normal subjects were recorded. Period durations were measured as the peak-to-peak distance in the Visicorder traces. He found typical correlograms to show major negative peaks at lags of one and eight periods, and major positive peaks at lags of seven and fourteen periods. This may be contrasted with the results from Hiki et al. (1966). These researchers used a contact microphone to record Japanese vowels sustained by 32 male and 30 female normal speakers. The utterances were digitized, with a temporal resolution of 50 μs, and period durations were measured. No apparent periodicity in the perturbations was observed.

These studies, together, show that speakers may produce non-random perturbations, at least in sustained vowels, although the proportion of speakers who display this behavior is unclear. The long-term periodicities, over twelve periods for amplitude and seven periods for frequency, have been compared to the singer's vibrato. The task of sustaining or intoning a vowel at a constant pitch may well encourage the introduction of regular modulations. It is interesting to note, though, that the amplitude and frequency modulations do not correspond in period, as might be expected if they were to reflect a common control mechanism. Data on amplitude and frequency variations should be obtained from the same speakers. Several tokens from an individual should also be examined, to establish whether the variability evident among speakers also occurs within the individual.

III. Methodology

This chapter describes the procedure by which jitter and shimmer measurements were obtained in this study. The procedure followed can be divided into five main steps: (1) recording context sentences to the hard disk files of a micro-computer; (2) editing the signal files to isolate the vowel segments of interest; (3) peak-picking throughout each vowel segment to find the period durations and peak amplitudes; (4) calculating mean jitter and shimmer from the saved duration and peak values; and (5) computing correlational measures, based on the jitter and shimmer values for individual periods.

A. Subjects

The subjects were young, male and female adults, with no known laryngeal pathologies or history of pathology. From a set of twenty-one randomly selected individuals, the productions of eight male and eight female speakers were chosen for analysis, based on the technical quality of the recordings. The male speakers ranged in age from 19 to 38 years, with a mean age of 28.25; the females ranged from 19 to 35, with a mean age of 25.37. Although smoking was not a factor controlled for, the smoking history of each subject was noted, as it was thought possible that it could affect the laryngeal tissues and hence the jitter and shimmer measures. Five male and five female speakers had never smoked. Information on the ages and smoking habits of the

Speakers

	Age	Smoking
Male Speakers		
M01	32	Infrequently, for 10 years
M02	37	Pack a week (Colts cigarillos), for 20 years
M03	38	Never
M04	32	Two cigarettes per week, for 2 months
M06	25	Never
M08	20	Never
M09	23	Never
M10	19	Never
Female Speakers		
F01	23	On and off, for 9 years
F02	29	Never
F03	35	Quit 10 years ago
F04	30	Never
F08	19	Never
F09	21	Never
F10	25	Quit 7 years ago
F11	21	Never

Table 3. Age and smoking background of the speakers.

subjects is included in Table 3. Additional data were gathered concerning the subjects' hearing status and their linguistic and musical backgrounds. No significant loss of hearing was reported. All subjects were native speakers of either Western Canadian or Ontario English dialects. Only one subject, F08, had received any formal vocal training.

B. Test Material

Jitter and shimmer were measured for nine Canadian English vowels: phonemically, these were /i, ι, ϵ, æ, ɚ, ʌ, u, ω, ɒ/. Since period durations were determined as peak-to-peak distance, and since this method of period extraction is sensitive to changes in formant structure (Rabiner et al., 1976), efforts were made to minimize transition and coarticulation effects: only vowels traditionally described as monophthongal, produced in the "neutral" context hVd, were examined. The /h/ initiation allowed presetting of the articulators to the position of the following vowel; the /d/ increased vowel duration without greatly altering formant values from those expected for isolated vowels (Stevens & House, 1963).

Subjects were given a list of sentences of the form "Please say hVd not hVd". They were not required to read phonetic symbols, but saw the items as the words "heed, hid, head, had, heard, Hud, who'd, hood," and "hod". Intonation effects were investigated by placing the words in two stressed positions in the sentence; there were two replications in each position. The order of the thirty-six test words (9 vowels × 2 positions × 2 replications) was randomized in two blocks of nine sentences, with each block containing each word in each position. No word could occur twice in the same sentence. The "RANDOM" program, given in Appendix A, generated an individual list for each subject.

During analysis, the tokens for a subject were uniquely identified by a four-letter sequence, consisting of a two-letter code for the vowel, the letter B (between words) or E (end of sentence) for the position, and the numbers 1 or 2 for the replication. The Latin-letter representations of the vowels were:

IY : i
IH : ι
EH : ϵ
AE : æ
ER : ɚ
AH : ʌ
UW : u
UH : ω
AW : ɒ

Thus, the first replication of the vowel /i/, between words, would appear as IYB1. This notation will occasionally be used when individual tokens are discussed.

C. Apparatus

The technical specifications of the instruments used in this study are given below.

1. Condenser Microphone: Sennheiser MKE 405
    Frequency response: 200-2000 Hz ± 2 dB
    50-20,000 Hz ± 5 dB
    Directionality: cardioid
    Windscreen

2. Microphone Power Supply: Sennheiser MZN 5-1

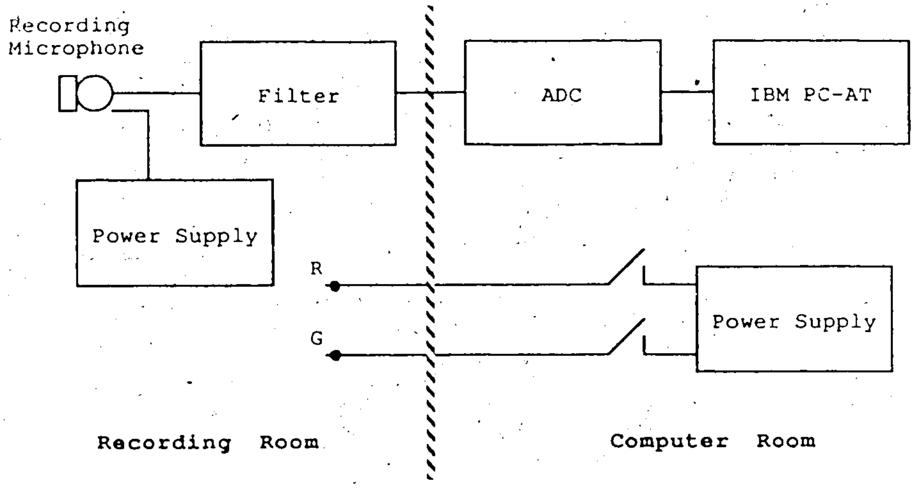
3. Audio-Frequency Filter: Rockland Wavetek 852 Dual Hi/Lo Filter
    Frequency range: 0.01 Hz to 111 KHz
    Cutoff frequency accurancy: ≃ 2%
    Attenuation slope: 48 dB/oct
    Filter characteristics: Butterworth
    Attenuation at cutoff: 3 dB

4. Analog-to-Digital Converter: Tecmar Lab Master
    Amplitude Resolution: 12 bits
    Input range: -10 to +10 V
    Sampling rate (max): 30 kHz

5. Microcomputer: IBM PC-AT, for speech sampling
    Memory: 640 KBytes RAM
    20 MByte hard disk
    1.2 MByte floppy disk
    EGA (Enhanced Graphics Adapter)
    Sampling rate (max): 20 kHz
    Operating system: DOS 3.0

6. Microcomputer: IBM PC-XT, for data analysis

a) Recording



b) Monitoring

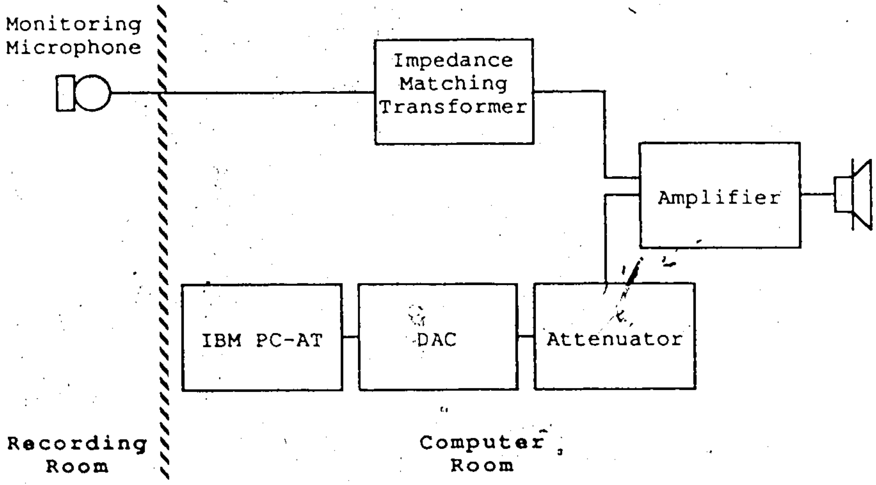


Figure 1. Block diagrams for recording and monitoring.

D. Recording

The microphone signal was digitized through a 12-bit analog-to-digital converter at a sampling rate of 20 kHz. The signal was low-pass filtered at 7800 Hz to prevent aliasing, and high-pass filtered at 50 Hz to eliminate the dc effect of breath expiration. The low cut-off for the high-pass filter was selected to capture the frequency sidebands caused by the FM and AM modulation, due to jitter and shimmer, of the lowest voice components. Rather than filtering out any 60 Hz power-line hum, rigorous attention was given to the proper grounding and shielding of all equipment and connecting cables.

Subjects were individually recorded in a sound-treated and sound-insulated recording room. Each subject was given a list of twenty-one sentences to read. The first two sentences were identical for all speakers, and contained combinations of the words "hayed", "hoad", and "hide"; these were not stored, but were used to test the amplitude level and accustom the subjects to the task. A final sentence containing these words was also included to prevent any list-final intonation effects. A typical list may be seen in Appendix B.

The amplitude of the recordings could be adjusted only by placing the subjects nearer to or further from the microphone suspended in front of them. No amplifier was used, in an attempt to keep instrumentally-induced noise and hum at a minimum. As a consequence, while the full range of the ADC allowed for 4096 quantization levels, two bits were commonly lost. Utilizing approximately 10 bits (1024 levels), a signal-to-quantization noise ratio of about 60 dB was achieved (Rozsypal, 1976); this was felt to be sufficient.

Once placed, subjects were instructed to maintain a constant distance from the microphone and to speak at a volume comfortable to them. Their utterances were monitored during recording using a separate microphone, and each sampled sentence could be immediately played back to verify the correctness of the recording. Communication with the subject was effected by means of two small red and green signal lights. The green light told the subject to proceed to the next sentence; a red light indicated that he was to repeat the current one. Repetitions could be required for coughs or stutters, mispronounced vowels, or mistimed triggerings of the sampling procedure by the operator.

Sampling was controlled by the "A/D Sampling" module of the Alligator[1] (Eagles, Morrow, & Sannino, 1986), which was programmed to digitize 3 seconds of speech at the 20 kHz sampling rate. A 15-30 second pause between sentences was required for each utterance to be saved and displayed. Recording took approximately 10 minutes per subject.

---

[1] The Alligator is a multi-purpose signal editing program written in Turbo Pascal for use on IBM PC's.

E. Vowel Gating

The signal editor of the Alligator was used to isolate the vowel segment in each test word. If the amplitude at vowel onset increased sharply over a small number of periods before achieving a roughly stable level, all "onset" periods were eliminated; however, if no stable level existed, the vowel segment was defined as beginning at the first period having an amplitude at least half that of the greatest point in the signal. A vowel was regarded as ending with the disappearance of the waveshape variations which indicate a formant structure. The vowel/d boundary was, though, somewhat arbitrarily established at a point preceding this, where waveshape changes suggested the start of the transition between the two sounds, or where amplitude suddenly decreased. In spite of these restrictions, there was no attempt to capture the vowel "nucleus" or the "steady-state" portion of the vowel as such, since this would often, particularly with male speakers, have produced signals containing very few periods. Playback of selected signal segments helped verify the appropriateness of these criteria, or motivated further gating; the procedure continued until a vowel percept was produced.

Those utterances where the pitch fell into creak or fry presented special difficulties. Often, when this occurred, only a few periods in what was clearly the vowel/d transition were affected· these periods were removed. However, in those rare instances when the fry both appeared through a substantial portion of the signal, and contained pulses of the same general shape as the vowel, it was decided that it should be retained, as a natural phenomenon, with the full knowledge that its inclusion would greatly increase the jitter and shimmer values for these tokens. The effect of this decision will be discussed in the next chapter.

An IBM PC-XT microcomputer was used for gating and for all subsequent analysis of the vowels.

F. Period and Amplitude Measurement

A semi-automatic peak-picking program, "JSPEAKS", was used to establish the duration and peak amplitude of each period in the vowel segments. The program searched for the greatest point under a horizontal bar, spanning forty sampling points, roughly placed on a major signal peak. After the operator had positioned and confirmed the bar for the first two peaks, the program took the number of sampling points between these peaks as the interval by which to advance the signal display for every following period; the signal then appeared to shift forward by approximately a period while the bar remained stationary. The operator would, if necessary, correct the bar's position, or would simply confirm the appropriateness of its placement. In this first pass though the signals, the length of the bar relative to the period lengths made few adjustments necessary.

Once the greatest point in a period had been located, JSPEAKS calculated the maximum of a parabola passing through this point and the ones immediately preceding and following. The y value of the parabola maximum was stored as the peak amplitude; the corresponding x coordinate allowed the period duration, in fractions of sampling intervals, to be computed, a figure which was converted to milliseconds before storage.

This interpolation was necessary to improve the temporal resolution, critical to a jitter study. At the 20 kHz sampling rate, one sample was taken every 50 µs. Since the resolution without interpolation is equal to the duration [illegible] the sampling interval, the accuracy of measurement [illegible]uld be no better than 50 µs. This is of the same [illegible] as reported mean jitter, at least for that measured from the steady-state portion of sustained vowels (Horii, 1979; Wilcox & Horii, 1980; and others). Titze et al. (1987) have demonstrated, though, that parabolic interpolation with peak-picking can greatly reduce the number of sampling points per cycle required to resolve jitter. They suggest that a 10 kHz sampling rate with interpolation would be acceptable for fundamental frequencies up to 200 Hz. In this study, the efficiency of the parabolic interpolation was tested on a sinusoid of 1001 Hz; this frequency was selected since the peaks found by the peak-picking procedure are strongly influenced by the second vowel formant. A mean period duration of .999005 ms was found over 100 periods. The standard deviation of the measurements was .388 µs; values deviated from the mean by a maximum of .940 µs.

For amplitude, the correction provided by the interpolation could exceed the size of the quantization interval. Since the first two formants are presumed to exert the greatest influence on airborne shimmer, tests were made on sinusoids in the 1000 Hz range, having peak amplitudes of 500 quantization levels. Corrections of over five quantization intervals, from a peak sample point of 494, to an interpolated peak of 499.5, were observed.

The assumptions about the nature of the original signal, needed for parabolic interpolation, were believed to be approximately valid within the restricted environment of a vowel peak. However, this type of interpolation could not be used where three or more successive points at the peak had the same value. Instead, a sepa[illegible]te procedure within JSPEAKS dealt with flat peaks. These were not due to peak clipping, but occurred naturally with broadly-peaked vowels, most notably with /u/, for which the wave could take on a smooth, almost sinusoidal shape. In these cases, the centre of the flat portion was found and used to calculate period duration; the peak amplitude was assumed to be that of the peak points.

JSPEAKS existed in two versions, allowing the period durations to be extracted by two methods. In the first, called "maximum peak-picking", the greatest point in the period was always taken as the reference for period calculation. Although this may seem implied by the term "peak-picking", it produced some unusually variable durations, when a secondary peak suddenly became prominent in a period, or when two peaks alternated in prominence in a series of periods. This last appeared only with some female speakers' productions of /ɒ/ and /ɚ/, vowels in which the first two formants were narrowly separated, and had more to do with the supraglottal resonances and damping of the vocal tract than with any variations in the glottal source. Comparison by listening failed to distinguish these tokens from ones without the effect.

The second version of the program was substantially similar to the first, but allowed the operator to more precisely control the length and location of the bar; he could then position the bar on a selected, characteristic peak, whether or not this peak was the greatest in the period. To be chosen, a peak had to be the most prominent at some place in the vowel. The operator then tracked this peak throughout the vowel, using it as the startpoint reference for all periods. In many instances, the two program versions produced identical results. However, for 255 of the 576 tokens there was some difference, involving at least one, and usually several, period markers. Table 4, taken from a vowel section with peak-switching, suggests the manner in which measurements generated by the two methods could deviate; Figure 2 schematically illustrates this situation. While jitter values calculated from the "maximum-peak" period durations will be presented in the following chapter, the analysis concentrates on "characteristic-peak" jitter, as more closely representing the phenomenon of interest. It should incidentally be noted that studies examining the steady-state section of sustained vowels need not contend with the period-to-period shape variations here encountered, as period shape in such cases essentially does not vary.

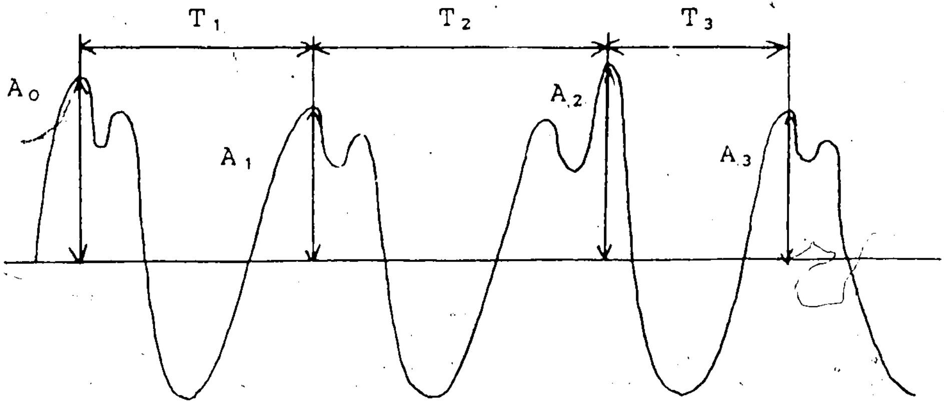
JSPEAKS permitted the inversion of signals, an option which was occasionally chosen for characteristic-peak extraction, when the negative-going peaks were more clearly defined or consistent than the positive ones. Inversion was also considered for maximum-peak extraction within the 26 tokens for which the absolute value of the signal minimum exceeded the maximum. This was rejected, however, since large negative peaks appeared only in relatively short portions of the vowels, produced by different female speakers, and clearly did not result from any overall reversal of signal polarity in the apparatus.

In contrast to period durations, peak amplitudes were, by definition, always determined from the maximum point in the period, and were always obtained from positive peaks. The characteristic-peak version of the program did not

Period Number	Maximum Peak Picking	Characteristic Peak Picking
4	3.70	3.65
5	3.80	3.78
6	2.72	3.81
7	3.86	3.86
8	5.02	3.91
9	3.92	3.93
10	2.85	3.94
11	3.97	3.97

Table 4. Period length in ms, as determined by two methods of period extraction, for a section in which the first and second peak in the periods alternate in prominence. (From F11AWB2, values truncated).

a) Maximum peak-picking



b) Characteristic peak-picking

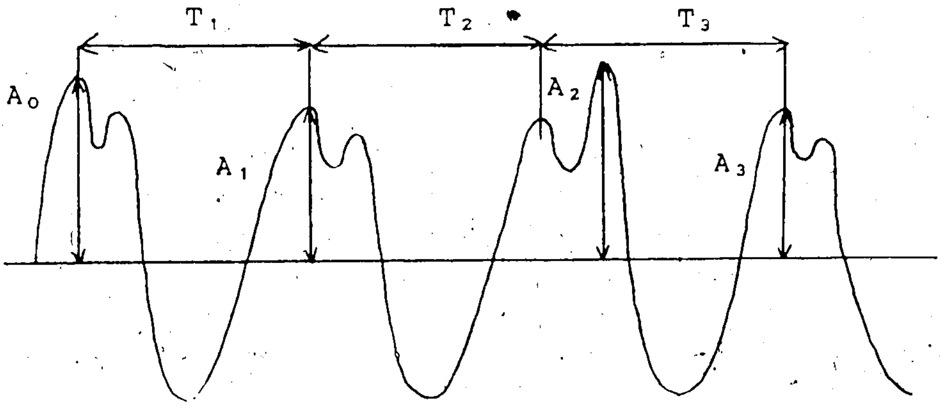


Figure 2. Two methods of period determination. Note that amplitude is always measured from the greatest peak in the period.

report amplitude measures. "Maximum-peak" amplitudes and "characteristic-peak" period durations could thus have a certain independence, in terms of the points from which they were calculated.

The output of the maximum-peak version of JSPEAKS consisted of the following data, for every period: the period number, the amplitude values of the three peak sample points, the array index of the maximum point, the interpolated period duration in milliseconds, and the interpolated peak amplitude. With the exception of amplitude, the output format was identical in the characteristic-peak run. A listing of the maximum-peak version of the JSPEAKS program, and a sample output listing, can be found in Appendix C.

G. Jitter and Shimmer Measurement

The period durations and peak amplitudes obtained from JSPEAKS served as input to a second program, "JSEXTR", which calculated mean jitter and shimmer for each vowel, and plotted the signed jitter and shimmer values for each period. An additional procedure in the program autocorrelated signed jitter and shimmer values for various lags, and cross-correlated jitter with shimmer for lags of -1, 0, and 1 period. A listing of JSEXTR is given in Appendix D.

Mean jitter was calculated according to the following formula:

$$\text{Mean Jitter} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left(\frac{\left| T_i - \frac{T_{i-1} + T_{i+1}}{2} \right|}{\frac{T_{i-1} + T_i + T_{i+1}}{3}} \right) \times 100[\%] \quad , \qquad 3.1$$

where T_i is the duration of the i-th period in ms and N is the number of periods in the vowel. The numerator measured the absolute difference between T_i and an arithmetic average of the preceding and following period durations. This ensured that duration variations resulting from steadily increasing or decreasing period trends would not be included in the measure as jitter. The denominator normalized this value for frequency, since jitter magnitudes without normalization increase with period length (Lieberman, 1963; Horii, 1979); a local average over three periods was used to control for the range of frequencies which intonation produces within a vowel. The overall mean was then obtained by dividing the sum total of the absolute, normalized, deviations from linear trend by the number of periods included in the analysis, minus two periods, one lost at the beginning and one at the end. This formulation of jitter closely resembles the "Relative Average Perturbation" (Eq. 2.5) measure of Koike 1973) and the "Fundamental Period Perturbation" (Eq. 2.6 of Hartmann and von Cramon (1984).

An exactly analogous formula was used for shimmer:

$$\text{Mean Shimmer} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left(\frac{\left| A_i - \frac{A_{i-1} + A_{i+1}}{2} \right|}{\frac{A_{i-1} + A_i + A_{i+1}}{3}} \right) \times 100[\%] \quad , \qquad 3.2$$

where A_i is the peak amplitude of the i-th period. Defining shimmer as a ratio of amplitude variations to local average amplitudes prevented differences in recording levels from directly affecting shimmer magnitudes. The formula is similar to the "Relative Average Amplitude Perturbation" measure described in Zyski et al. (1984).

The percentages produced by these formulae must be interpreted cautiously. For the artificial case of alternating small and large peaks, with the small peaks exactly half the size of the large, the formula would not, for example, measure shimmer at the 50% or 100% which might be expected, but would give a value over three periods of 75%. This is determined by the weighting of elements in the denominator: dividing by $(A_{i-1} + 2A_i + A_{i+1})/3$ would give 50% for this case, while $(A_{i-1} + A_{i+1})/2$ would produce 100%. There appears to be no compelling reason for choosing among these weightings, however.

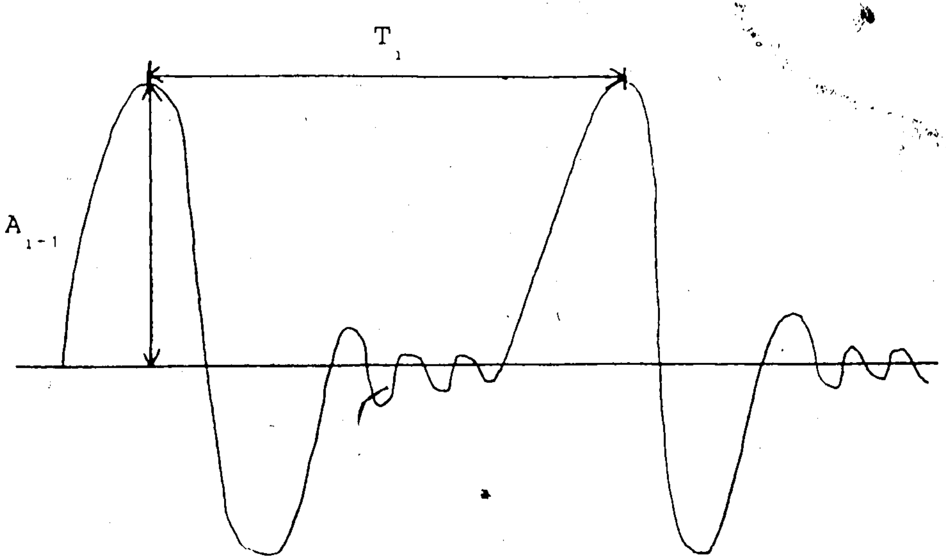
After calculating mean jitter and shimmer, and the mean fundamental frequency, JSEXTR displayed a plot of the positive or negative jitter and shimmer values for each period in the vowel. These plots allowed visual evaluation of the perturbation pattern, and greatly assisted interpretation of the correlation data. Copies of the plots were optionally printed, and examples can be found in the Results chapter (Figures 16 to 19). The vertical plot scales varied, to provide good resolution for each token.

The program's AutoCorr procedure investigated possible cycles within the signed jitter and shimmer values for the periods. In this procedure, the measures were autocorrelated for period lags ranging from one, to one-third the total number of periods, to a maximum lag of twelve periods. Significant correlations thus indicate the presence of at least three cycles of regular perturbation. The twelve period maximum lag was selected since the previous time series analyses of jitter or shimmer (Koike, 1969; Von Leden & Koike, 1970; Iwata, 1972) have shown maxima to occur within this range. However, the full twelve lags were rarely reached for male speakers, as their vowels often contained fewer than the thirty-six periods required.

The procedure also explored the period-to-period relations between signed jitter and shimmer by cross-correlating the two parameters at three period lags. At a lag of 0, the duration of a period was correlated with the height of the following peak; at a lag of 1, it was correlated with the peak within the same period. These relations are schematically illustrated in Figure 3, although it must be emphasized that the calculations were in fact carried out on the signed jitter and shimmer values rather than the actual period durations and peak amplitudes. As jitter and shimmer measure deviations from linear trend, however, the correlations can be interpreted as relating durations and heights with these linear trends removed. The lag of -1, which correlated period duration with the height of the peak of the second following period, was included solely to give flexibility to the program; interesting data were not produced.

JSEXTR produced three output files for each subject. One contained, for each vowel, the values for mean jitter and shimmer, and mean fundamental frequency; the second stored the autocorrelation coefficients for the different lags; and the third held the cross-correlation coefficients for the three lags.

a) Lag=1



b) Lag=0

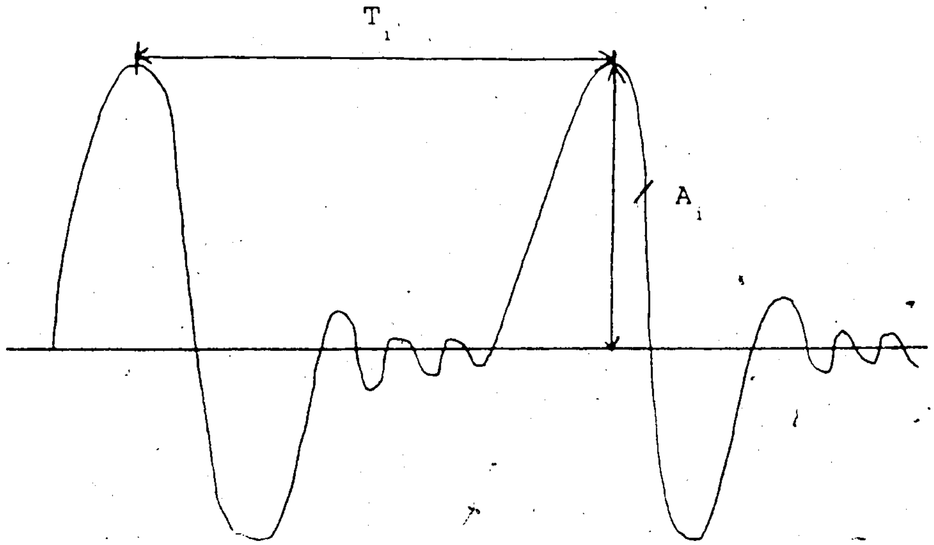


Figure 3. Cross-correlations at lags of zero and one.

IV. Results

The initial portion of this chapter reports the jitter and shimmer magnitudes measured, and describes a procedure for the exploration and adjustment of anomalous items. The effects of vowel, speaker sex, and intonation, as indicated by analyses of variance, are then detailed. The final section presents the jitter and shimmer correlations, based on the mean values for the tokens, and on the period-to-period deviations.

A. Maximum Peak Jitter

As explained in the "Period and Amplitude Measurement" section, period durations were determined in two ways, "automatically", from the maximum point in each period, and "intelligently", with the program user tracking a selected characteristic peak throughout the vowel. For purposes of comparison, Table 5 presents the jitter magnitudes calculated from the two period measurement sets, averaged across the tokens of each subject. With both methods, the jitter values of many speakers can be seen to exceed the .5 to 1% considered typical for jitter in sustained phonations (Heiberger & Horii, 1982). While it was thought possible that the stress and intonation of natural vowels might elevate overall jitter levels, notably large values and large speaker differences were expected only with "maximum-peak" jitter, reflecting this measure's added sensitivity to waveshape perturbations. However, while

Jitter

	Maximum Peak Picking [%]	Characteristic Peak Picking [%]
Male Speakers		
M01	.91	.54
M02	2.55	2.00
M03	5.56	4.81
M04	.75	.46
M06	1.67	1.21
M08	.88	.57
M09	.82	.54
M10	1.06	.70
Female Speakers		
F01	1.02	.43
F02	.88	.49
F03	.99	.79
F04	2.10	1.11
F08	1.72	1.30
F09	3.34	2.57
F10	1.76	.72
F11	1.31	.80

Table 5. Jitter values for each subject, averaged over thirty-six tokens, for the two methods of period determination. Similar speaker differences appear with both methods.

maximum-peak jitter magnitudes were consistently larger than "characteristic-peak" ones, the measurement technique itself did not contribute greatly to the inter-speaker variability.

An exploratory analysis of variance performed on the maximum-peak jitter data revealed no significant effects apart from those associated with speaker differences (subject, $p<.01$; subject × vowel, $p<.05$; subject × position, $p<.05$; subject × vowel × position, $p<.01$). In the absence of evidence relating period-to-period waveshape changes to perceived voice quality (as opposed to vowel quality), further examination of these data was not motivated, although an analysis parallel to that which will be presented for characteristic-peak jitter could be performed.

For the remainder of this study, "jitter" can thus be taken to refer exclusively to "characteristic-peak jitter".

B. Jitter and Shimmer Magnitudes

Over the 576 tokens, jitter had a mean value of 1.19%, with a minimum of .11% and a maximum of 17.26%; shimmer had a mean of 3.99%, and ranged from .73% to 84.43%. With the given formulae for jitter and shimmer (Eqs. 3.1 and 3.2), large values do not directly express the depth of variation: however, a measure of 84.43% indicates that the amplitude for this token varied, on average, by more than one-half across the three-period sets. This degree of variation was not anticipated.

Histograms of the jitter and shimmer measures showed that, while the majority of items were distributed around the means, a number of unexpectedly large values had been produced. For jitter, these included five values above 15%, seven between 10 and 15%, and nine between 5 and 10%; for shimmer, three values exceeded 50%, four fell between 30 and 50%, and nine were between 15 and 30%. The production of outliers appeared to be characteristic of certain speakers: five of the sixteen speakers were responsible for all such items while multiple outliers occurred only with M02, M03, and F09. Table 6, which gives the mean jitter and shimmer magnitudes for each speaker, suggests the effect these values had in creating speaker differences.

Outliers were interpreted as representing the presence of some phenomenon additional to normal jitter and shimmer, the nature of which remained to be determined. If the effect appeared throughout the data, with the outliers as its extreme manifestation, some covariance measure could be devised; however, if it were unique to the outliers, a separate analysis of these tokens would be appropriate. The following section details one attempt to define the phenomenon.

	Jitter [%]	Shimmer [%]
Male Speakers		
M01	.54	3.34
M02	2.00	5.08
M03	4.81	14.27
M04	.46	2.24
M06	1.21	3.89
M08	.57	2.15
M09	.54	2.45
M10	.70	4.11
Female Speakers		
F01	.43	2.13
F02	.49	2.33
F03	.79	2.28
F04	1.11	3.14
F08	1.30	5.05
F09	2.57	6.22
F10	.72	2.64
F11	.80	2.45

Table 6. Jitter and shimmer magnitudes, averaged over thirty-six tokens for each speaker. Outlying tokens were produced by M02, M03, M06, F08, and F09.

C. Alternate Period Measures

In the waveform of certain tokens, alternate pulses were observed to correspond more closely, both in period duration and peak amplitude, than adjacent ones. This characteristic of the wave, called "double-periodicity" (Shoup & Pfeifer, 1976), could produce very large values for perturbation measures on successive pulses. To test the extent to which double periodicity existed in the data set, and the extent to which it could account for the outlier values, jitter and shimmer were recalculated from alternate periods. This measure, for jitter, had the form

$$\text{Alternate Period Jitter} = \frac{1}{N-4} \sum_{i=3}^{N-2} \left(\frac{\left| T_i - \frac{T_{i-2} + T_{i+2}}{2} \right|}{\frac{T_{i-2} + T_i + T_{i+2}}{3}} \right) \times 100[\%] \quad 4.1$$

A similar measure was used for shimmer. For tokens containing double periodicity, alternate period jitter and shimmer values would be smaller than those previously calculated; for tokens without it, values very similar to the original magnitudes were predicted.

The results, for both jitter and shimmer, indicated that this effect was not consistently present. For a number of speakers, a roughly proportional relation could be seen between the two types of measures, with the alternate period perturbation magnitudes approximately equalling the successive period ones. The plot of jitter with alternate period jitter for F11 (Fig. 4) provides an example of this type; with this speaker, there is no evidence for any phenomenon apart from "normal" jitter. In contrast, the plots for speakers F08, F04, and M03 (Figs. 5, 6, and 7) show the presence of tokens with double periodicity, but suggest various relations between the measures. Note that the horizontal and vertical scales of these plots vary. The plot for F08 reveals a roughly linear trend between jitter and alternate period jitter, with values for normal jitter greater than those for alternate period jitter; one extreme jitter outlier deviates markedly from the trend. This can be compared with the relation in F04, which would be best described by a quadratic, and with the scatter in M03, which is not suggestive of any regular relation. Such lack of consistency, both in presence and effect, did not permit the use of the double periodicity measures as linear covariates.

Alternate period values were predicted to be smaller than or equal to successive period values, but were not expected to exceed them. That this last possibility did, in fact, occur may be traced to the assumption, implicit in the measures, that systematic variations in period duration or amplitude will follow linear trends. With natural vowels, this is more likely to appear true over three periods than five; since the measurements are of deviations from such trends, the longer averaging window for alternate periods could increase the measured magnitudes. The specific values calculated may thus be used only as

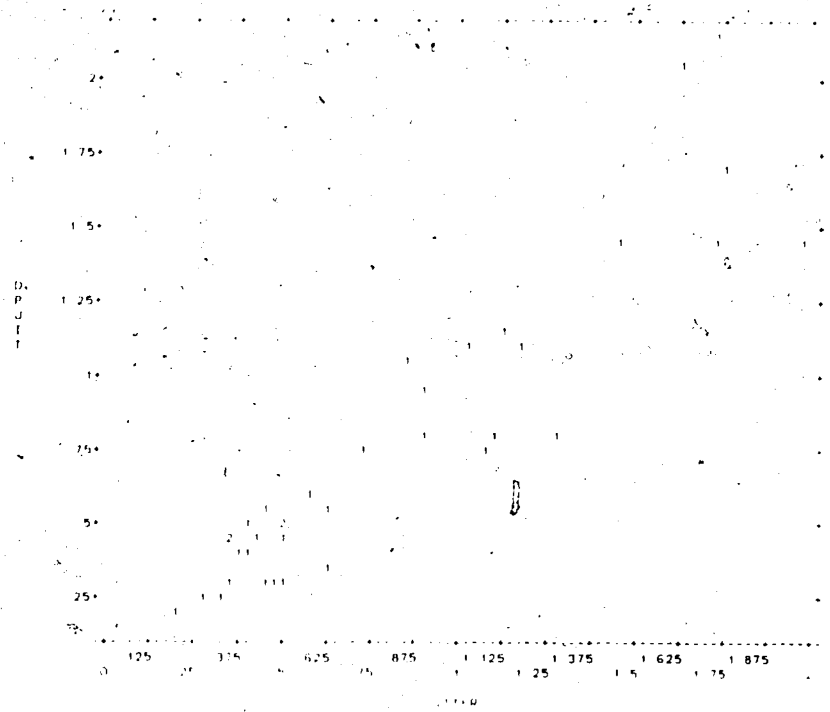


Figure 4. Plot of jitter with alternate period jitter for the thirty-six tokens of female speaker F11.

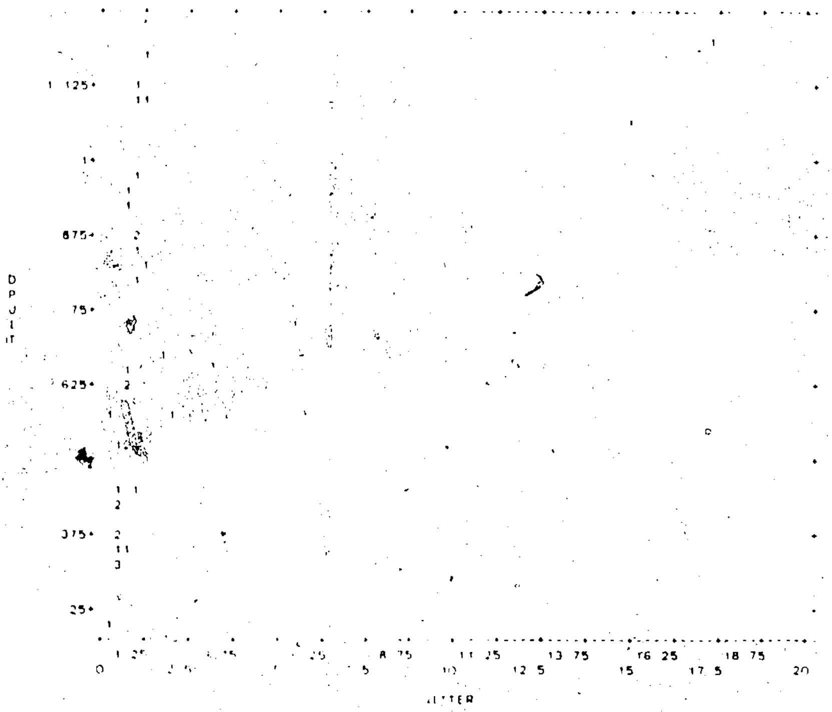


Figure 5. Plot of jitter with alternate period jitter for the tokens of female speaker F08. Note the position of an extreme jitter outlier.

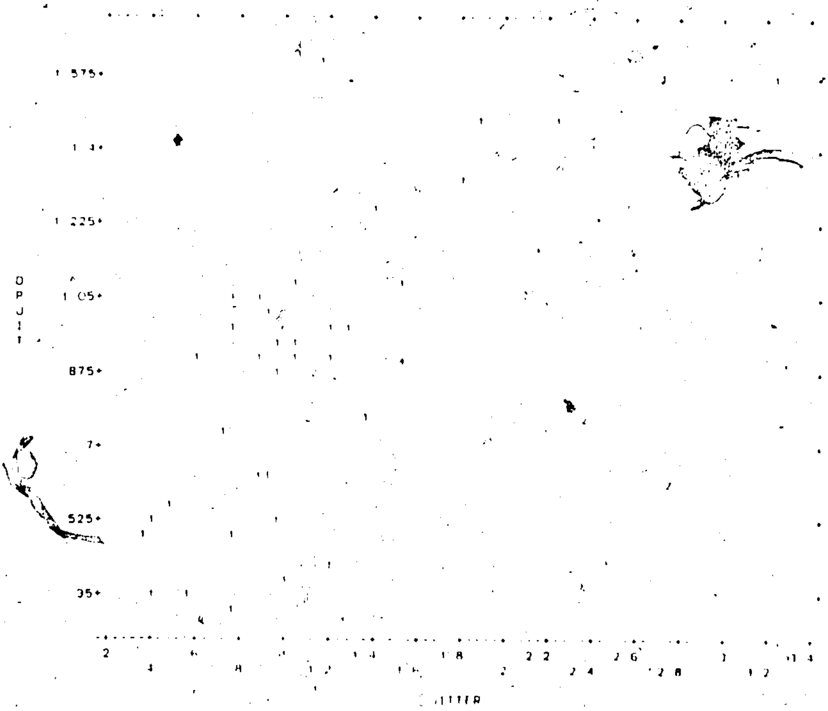


Figure 6. Plot of jitter with alternate period jitter for the tokens of female speaker F04.

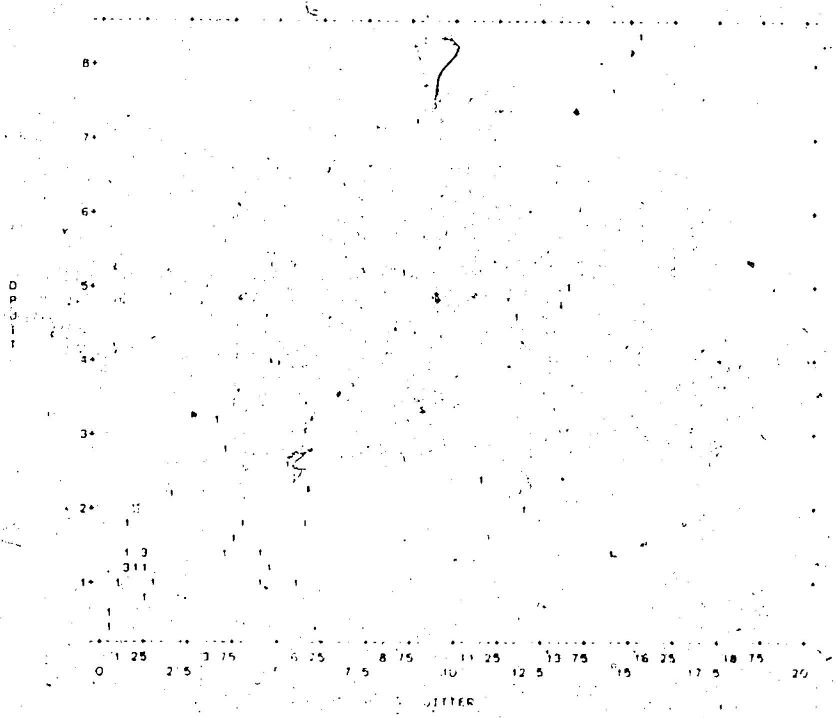


Figure 7. Plot of jitter with alternate period jitter for the tokens of male speaker M03.

approximate guides to the amount of double periodicity present.

With this caution, the alternate period values for the outlier tokens were examined, to determine the contribution of double periodicity to these items. A list of such tokens, with the successive and alternate period values, is given in Appendix E. In general, the jitter and shimmer values from alternate periods can be seen to be greatly reduced from the original magnitudes, indicating a notable degree of double periodicity. However, the range of values produced suggests that this characteristic alone was not responsible for all of the exceptional perturbations. Tokens such as ERE2, spoken by M06, and AWE2, by F09, contained sections of glottal fry, which may be described as double periodicity with added irregularities. Token UHE2, by M03, showed variable double periodicity, with two pulses of the same type occasionally adjacent. Other tokens differed in the consistency of the effect, having it throughout the vowel or only in certain portions. For these reasons, the outliers may not be regarded as a truly homogeneous group; double periodicity was not the sole phenomenon displayed.

D. Analysis of Variance: All Speakers

Although the alternate period analysis failed to explain all of the outlier values, it did support the idea that large values represented some phenomenon or phenomena

	Jitter [%] Outliers Adjusted	Shimmer [%] Outliers Adjusted
Male Speakers		
M01	.54	3.34
M02	1.12	4.73
M03	1.27	5.46
M04	.46	2.24
M06	.78	3.89
M08	.57	2.15
M09	.54	2.45
M10	.70	4.11
Female Speakers		
F01	.43	2.13
F02	.49	2.33
F03	.79	2.28
F04	1.11	3.14
F08	.85	2.80
F09	1.38	4.51
F10	.72	2.64
F11	.80	2.45

Table 7. Mean jitter and shimmer magnitudes for each speaker, following adjustment of the outlier token values. Note the change from Table 6 for speakers M02, M03, M06, F08, and F09.

extra to, and qualitatively different from, "normal" jitter and shimmer. If included in the general analysis, values produced by these effects, whether double periodicity, or fry, would conceal or distort the variations due to "normal" jitter and shimmer. Arbitrary upper limits on acceptable jitter, at 4%, and shimmer, at 15%, were therefore established, based on distributional criteria. Values above these levels were adjusted to 1%, for jitter, and 3.5%, for shimmer, the approximate overall means of the parameters. This step effectively eliminated the outliers from the analysis, and provided a distribution which better satisfied the normality requirements of the statistical tests.

Using the above criteria, 29 (5.0%) jitter and 16 (2.8%) shimmer values were classed as outliers, from the total of 576 tokens; with one exception, the outlier tokens for shimmer were also outliers for jitter. The list in Appendix E shows that these did not depend on either vowel, position, or replication, although their production was restricted to particular speakers. The effect their elimination had on the means for these speakers can be seen by comparing Tables 6 and 7. Their peculiarities will be further discussed in the "Correlational Analyses" section.

Following adjustment of the outliers, analyses of variance (ANOVAs) were performed on the jitter and shimmer data. The ANOVAs were generated by the ANOVAR user procedure, called from SPSSx. A mixed repeated measures design was used, with sex as a between-subjects factor, and with vowel, position, and replication as within-subjects factors. Sex, vowel, and position were treated as fixed factors. Subjects, implicitly random, were nested within sex. Repeated measures were made on vowel and position.

The ANOVA sums of squares, F ratios, and probabilities for jitter and shimmer are presented in Tables 8 and 9. For both perturbation measures, strong subject effects were found ($p<.01$). For jitter, the sex by vowel interaction ($p<.01$) and the vowel by position interaction ($p<.05$) were also significant; for shimmer, significant main effects for vowel ($p<.01$) and position ($p<.05$) emerged.

To examine the jitter interactions, tests on simple main effects (Winer, pp. 544-5) were performed. In this case, this involved constructing F ratios to test the effect of sex, and position, at each of the nine vowel levels. For sex, the results indicated a difference only for the vowel /ι/, with females having significantly more jitter than males [$F(1,34) = 7.497$, $p<.05$]. For position, /ɚ/ was observed to have significantly more jitter at the end of the sentence tha[illegible] between words [$F(1,126) = 11.866$, $p<.01$], while /ω/ had more jitter between words [$F(1,126) = 4.859$, $p<.05$]. Plots of these interactions may be seen in Figures 8 and 9.

For shimmer, a Tukey (a) test (Winer, p. 198) was conducted to determine which pairs of vowels differed. From

Source	SS	df	MS	F	p
Sex	.828	1	.828	.252	.624
Subjects (Sex)	46.008	14	3.286	16.851**	
Vowels	3.247	8	.406	1.615	.128
Sex x Vowels	6.133	8	.767	3.051**	.004
Vowels x Subject (Sex)	28.142	112	.251	1.287	
Position	.682	1	.682	3.036	.103
Sex x Position	.099	1	.099	.440	.518
Position x Subject (Sex)	3.142	14	.224	1.149	
Vowels x Position	3.699	8	.462	2.336*	.023
Sex x Vowels x Position	1.530	8	.191	.966	.466
Vowels x Position x Subj (Sex)	22.167	112	.198	1.015	
Replication	.040	1	.040	.202	.660

** $p < 0.01$
* $p < 0.05$

Table 8. Analysis of variance summary table for jitter, following adjustment of the outlier tokens.

Source	SS	df	MS	F	p
Sex	70.387	1	70.387	1.670	.217
Subjects (Sex)	590.129	14	42.152	14.135**	
Vowels	100.914	8	12.614	3.448**	.001
Sex x Vowels	46.436	8	5.804	1.587	.137
Vowels x Subject (Sex)	409.730	112	3.658	1.227	
Position	18.471	1	18.471	4.709*	.048
Sex x Position	6.234	1	6.234	1.589	.228
Position x Subject (Sex)	54.918	14	3.923	1.316	
Vowels x Position	13.777	8	1.722	.702	.689
Sex x Vowels x Position	33.598	8	4.200	1.712	.103
Vowels x Position x Subj (Sex)	274.793	112	2.454	.823	
Replication	.521	1	.521	.081	.780

** $p < 0.01$
* $p < 0.05$

Table 9. Analysis of variance summary table for shimmer, following adjustment of the outlier tokens.

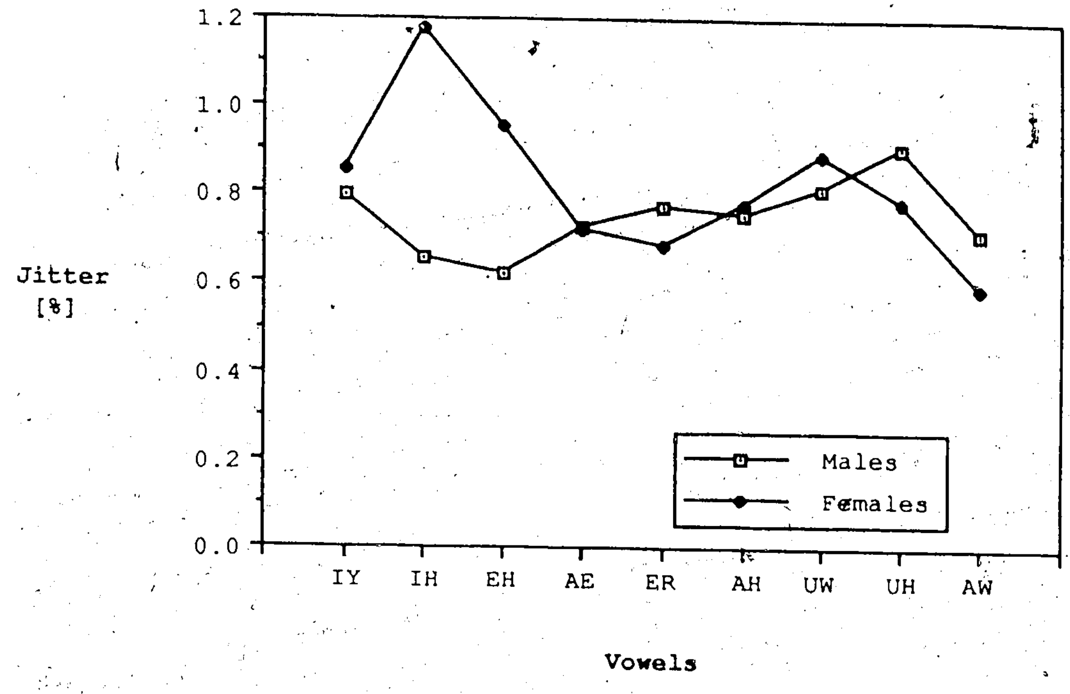


Figure 8. Sex by vowel interaction for jitter.

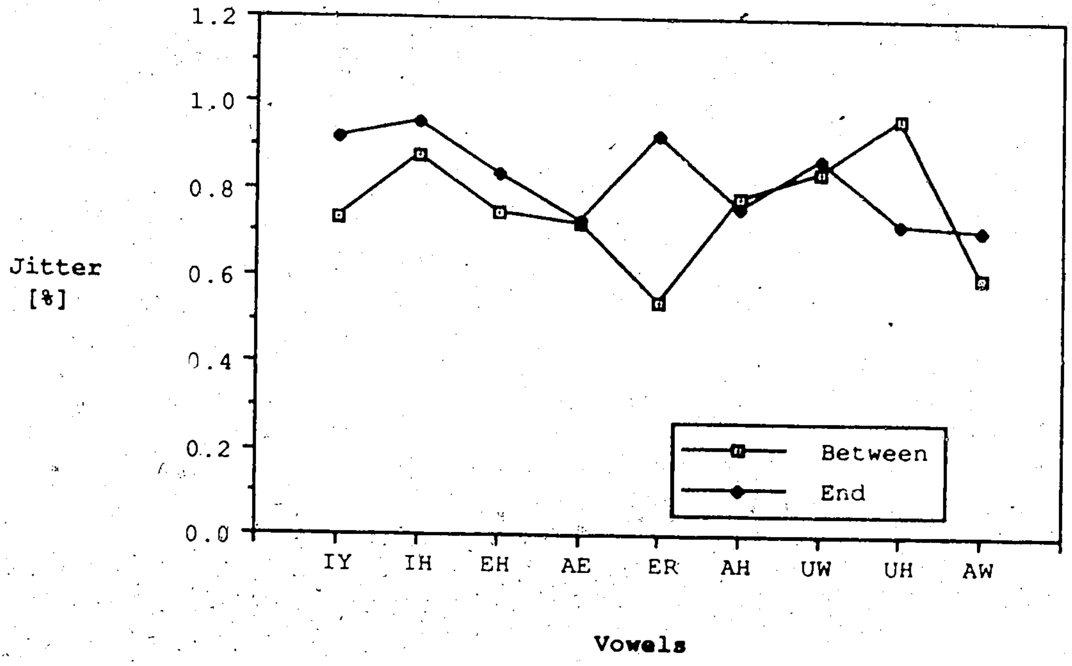


Figure 9. Vowel by position interaction for jitter.

most to least shimmer, the vowels were ordered /æ, ɑ, ʌ, ɛ, ɒ, ɪ, ɚ, u, i/. At the .01 level, /æ/ and /ɑ/ displayed more shimmer than /i/. For position, more shimmer was found at the end of the sentence than within it (mean shimmer, sentence-final = 3.38%; between words = 3.02%).

The large subject effects, apparent with both jitter and shimmer, suggested that the speakers had been drawn from more than one subject population. To explore the speaker differences, and their effect on this analysis, cluster analyses of the subjects were then generated.

E. Cluster Analysis

Cluster analysis is a technique which groups entities into homogeneous subgroups on the basis of the similarity of their response profiles (Lorr, 1983). The analysis takes as input a matrix of the raw data from which distance measures between each pair of entities (in this case, subjects) are computed. The agglomerative hierarchical method of clustering sequentially combines the set of subjects, taking at each stage the two subjects or clusters which are closest, to produce a tree-structure representation of the distances. Height on the resulting tree may thus be equated with distance. Subjects with similar profiles will be joined at a low level in the hierarchy. Groups of similarly responding subjects will be combined at a low level among themselves, but will only merge with other groups at a relatively high level.

Subjects grouped at a high level may have little in common; since the procedure continues until all entities are included, spurious groupings may be formed.

Using SPSSx, two subject cluster analyses were produced, one with the jitter and the other with the shimmer measures as the profile variables; the adjusted values were used for the outliers. Distances among the subjects were determined by the squared Euclidean dissimilarity metric, and clustering proceeded by Ward's minimum variance method (Lorr, p. 90).

For jitter (Fig. 10), the subjects can be seen to cluster into two groups of six subjects each. Group membership was not determined by sex: the first group, the most homogeneous, contained two female and four male speakers, while the second, connected at higher levels, held four females and two males. The height at which the four remaining speakers were added indicated that these four, though linked in the diagram, were best treated as individuals.

For shimmer (Fig. 11), one group of nine relatively similar speakers (six female, three male) was formed. A second group, consisting of F04, M10, and M06, was more loosely defined, and four speakers again were best regarded as ungrouped. Three of the four ungrouped speakers, for both jitter and shimmer, were M02, M03, and F09, the speakers responsible for most of the outliers: even following removal of the extreme values, their measures

```
Dendrogram using Ward Method

                         Rescaled Distance Cluster Combine

   C A S E     0         5        10        15        20        25
  Label  Seq   +---------+---------+---------+---------+---------+

  MO4     12    -+-+
  MO8     14    -+ |
  MO1      9    ---+-+
  FO1      1    -+-+ +---------+
  FO2      2    -+   |         |
  MO9     15    -----+         |
  FO3      3    -+---+         +---------------------------------+
  F10      7    -+   +---+     |                                 |
  M10     16    -----+   +---+ |                                 |
  FO8      5    -------+-+   +-+                                 |
  MO6     13    -------+     |                                   |
  F11      8    -------------+                                   |
  FO4      4    ---------------------+-----------+               |
  MO3     11    ---------------------+           +---------------+
  FO9      6    -------------------------+-------+
  MO2     10    -------------------------+
```

Figure 10. Subject clusters for jitter, following adjustment of the jitter outlier values.

```
Dendrogram using Ward Method

                         Rescaled Distance Cluster Combine

   C A S E     0         5        10        15        20        25
  Label  Seq   +---------+---------+---------+---------+---------+

  FO1      1    -+
  MO8     14    -+
  FO3      3    -+-+
  F11      8    -+ +---+
  MO4     12    -+ |   +-+
  MO9     15    ---+   | +---------+
  FO2      2    -------+ |         |
  FO8      5    ---------+         +-----------------------------+
  F10      7    ---+               |                             |
  FO4      4    -----+-----+       |                             |
  M10     16    -----+     +-------+                             |
  MO6     13    -----------+                                     |
  MO1      9    -------------+-----------+                       |
  MO2     10    -------------+           +---------+             |
  FO9      6    -------------------------+         +-------------+
  MO3     11    -----------------------------------+
```

Figure 11. Subject clusters for shimmer, following adjustment of the shimmer outlier values.

appeared abnormal. The other ungrouped speaker for jitter, F04, was included in the second shimmer group; the fourth speaker with abnormal shimmer, M01, was in the first group for jitter. As a general rule, similarity for jitter did not imply similarity for shimmer.

Smoking did not appear to affect group membership, or non-membership. The two jitter groups and the largest shimmer group each contained a mixture of smokers and non-smokers; of the consistently ungrouped speakers, only M02 had a history of smoking.

F. Analysis of Variance: Grouped Subjects

With the subjects nested within relatively homogeneous subgroups, much of the variance attributable to speaker differences could be eliminated. This would provide added sensitivity to tests on vowel or position effects. The subject groups defined by the cluster analyses were therefore used in a second set of ANOVAs, which explored the jitter and shimmer variations within the groups.

The ANOVA for jitter followed the same basic design as the previous analysis, with the two groups replacing the two sexes as the between-subjects factor, and with the same three within-subject factors. Only the twelve grouped subjects were included. Group, vowel, and position were declared as fixed; subjects, nested within groups, and replication were random. The results, presented in Table 10, showed a significant main effect for vowel ($p<.05$) and a significant subject by vowel interaction (p<.01). As expected, groups were also significant (p<.001).

From most to least jitter, the vowels were ordered /ι, u, ω, i, ϵ, æ, ʌ, ɒ, ɚ/, although the subject by vowel interaction indicated this would not necessarily hold for any given speaker. A Tukey test showed only that, overall, the extreme cases, /ι/ and /ɚ/, differed (p<.05). This may be related back to the sex by vowel and vowel by position effects of the first ANOVA, which involved these vowels. Plots of the subject by vowel interactions for the two groups (Figs. 12 and 13), generated to test whether any subset of speakers behaved consistently, suggested that the production of jitter for different vowels should be examined on an individual basis. Any vowel effects may thus be regarded only as tendencies.

For shimmer, only the nine subjects clustered into the first group were analyzed; there were thus no between-subjects factors. The three within-subject factors were again vowel, position, and replication. As with jitter, the results showed a vowel effect (p<.01) and a subject by vowel interaction (p<.01); a significant position effect also appeared (p<.05). This analysis is summarized in Table 11.

From most to least shimmer, the vowels were ordered /æ, ʌ, ω, ɒ, ϵ, ι, ɚ, i, u/. A Tukey test indicated that /æ/ and /ʌ/ differed significantly from /u/ (p<.01) and from /i/ (p<.05), though these effects again did not hold

Source	SS	df	MS	F	p
Groups	7.726	1	7.726	76.895**	.001
Subjects (Groups)	1.005	10	.100	1.127	
Vowels	3.889	8	.486	2.485*	.018
Groups x Vowels	.477	8	.060	.305	.962
Vowels x Subject (Groups)	15.653	80	.196	2.210**	
Position	.452	1	.452	4.513	.060
Groups x Position	.004	1	.004	.040	.845
Position x Subject (Groups)	1.002	10	.100	1.127	
Vowels x Position	.767	8	.096	1.012	.434
Groups x Vowels x Position	.754	8	.094	.995	.447
Vowels x Pos'n x Subj (Groups)	7.576	80	.095	1.071	
Replication	.036	1	.036	.231	.641

** $p < 0.01$
* $p < 0.05$

Table 10. Analysis of variance summary table for jitter, with 12 speakers in two groups of six.

Source	SS	df	MS	F	p
Subjects	14.623	8	1.828	1.936	
Vowel	62.046	8	7.756	4.621**	.001
Vowel x Subject	107.416	64	1.678	1.778**	
Position	9.283	1	9.283	6.140*	.038
Position x Subject	12.094	8	1.512	1.602	
Vowel x Position	.710	8	.089	.120	.998
Vowel x Pos'n x Subj	47.457	64	.742	.786	
Replication	.066	1	.066	.030	.867

** $p < 0.01$
* $p < 0.05$

Table 11. Analysis of variance summary table for shimmer, for one group of nine speakers.

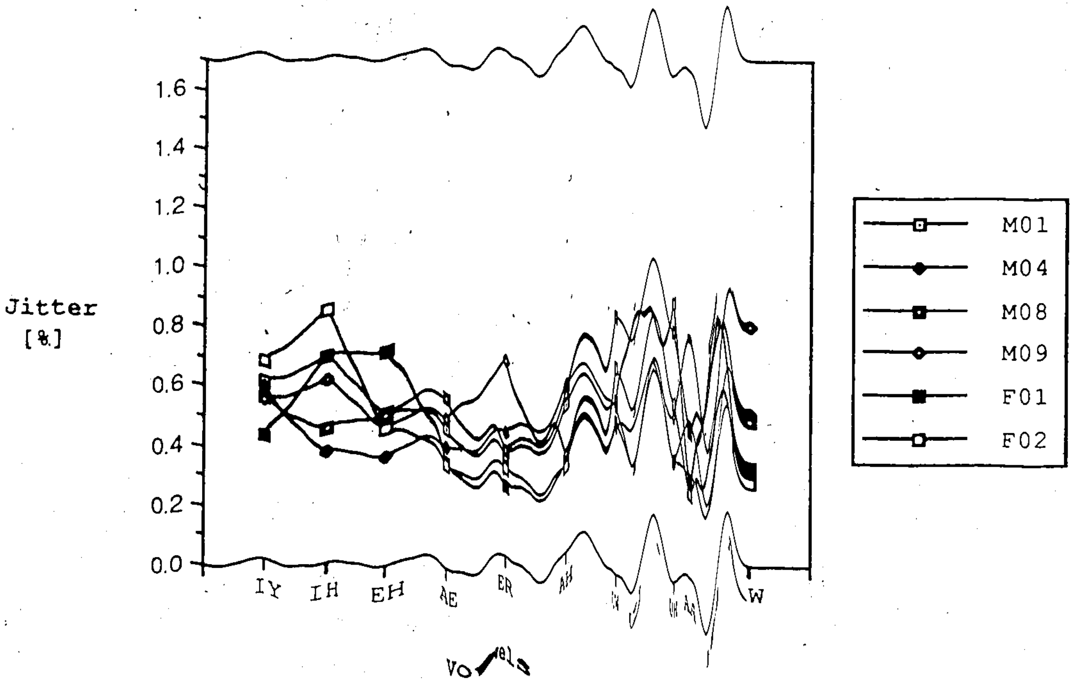


Figure 12. Subject by vowel interactions for the six speakers of the first jitter group. The vowel values are averaged over four tokens.

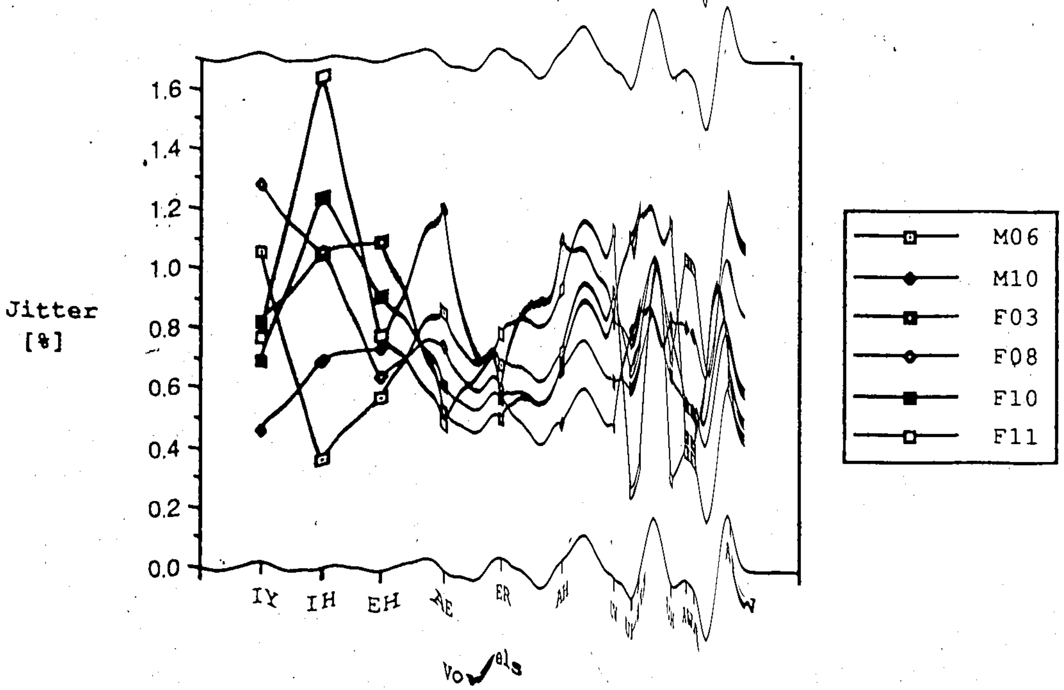


Figure 13. Subject by vowel interactions for the six speakers of the second jitter group.

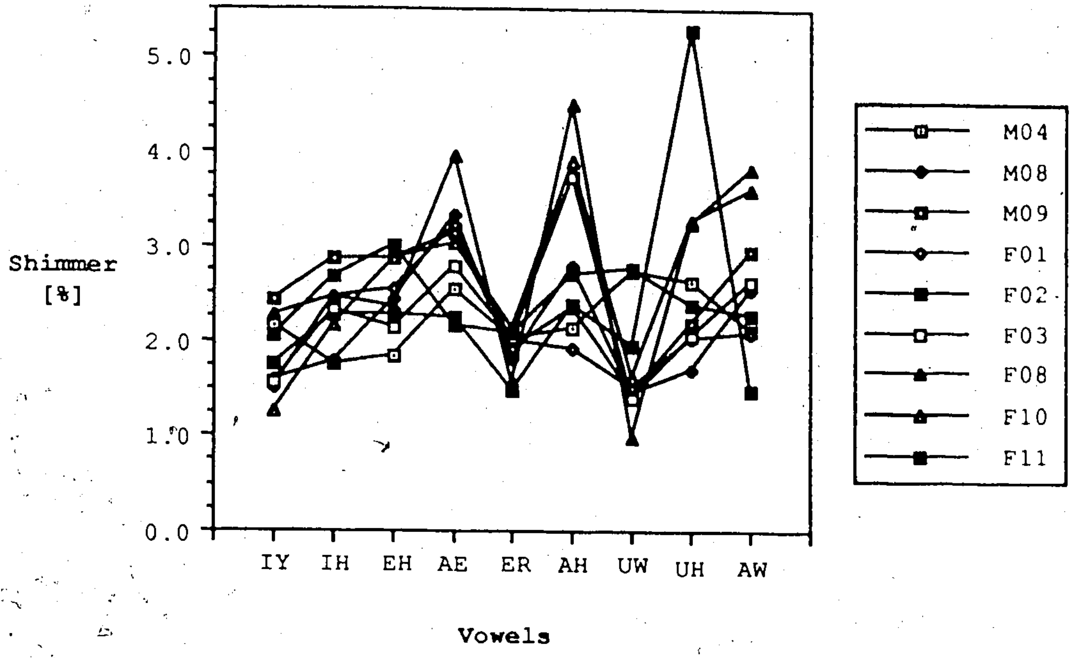


Figure 14. Subject by vowel interactions for the nine speakers of the most homogeneous shimmer group.

for all speakers. A plot of the subject by vowel interaction, given in Figure 14, shows the lack of consistency within the group. More shimmer appeared in sentence-final tokens (mean shimmer = 2.56%) than in those occurring between words (shimmer = 2.22%).

G. Correlational Analyses

The preceding analyses examined the effect of certain factors on jitter and shimmer magnitudes, without attempting to relate the two measurement parameters. The correlational analyses explicitly investigate this relationship, both in the mean magnitudes for the tokens and in individual periods.

Token Magnitudes

To test the degree of association between mean jitter and shimmer, the Pearson product-moment correlation coefficient was calculated from the 576 tokens. With the 30 jitter or shimmer outliers included, the Pearson r was found to equal .846 (p=.000): the linear correlation accounted for approximately 72% of the variance in the parameters. However, when the outlier tokens were excluded, the coefficient was reduced to .491 (p=.000), indicating a shared variance of only 24%. The outliers, which as a group were exceptional for both jitter and shimmer, produced an impression of linear dependence between the measures which the normal tokens did not strongly support.

	Number of Tokens	r	p
Male Speakers			
M01	36	.0800	.321
M02	33	.1972	.136
M03	19	.7841	.000 **
M04	36	.1040	.273
M06	35	.3331	.025 *
M08	36	-.1402	.207
M09	36	.7380	.000 **
M10	36	.5207	.001 **
Female Speakers			
F01	36	.1686	.163
F02	36	.3008	.037 *
F03	36	-.0523	.381
F04	36	.5288	.000 **
F08	35	.4354	.004 **
F09	28	.7279	.000 **
F10	36	.0616	.361
F11	36	.6241	.000 **

Table 12. Correlations between jitter and shimmer for the normal tokens of each speaker.

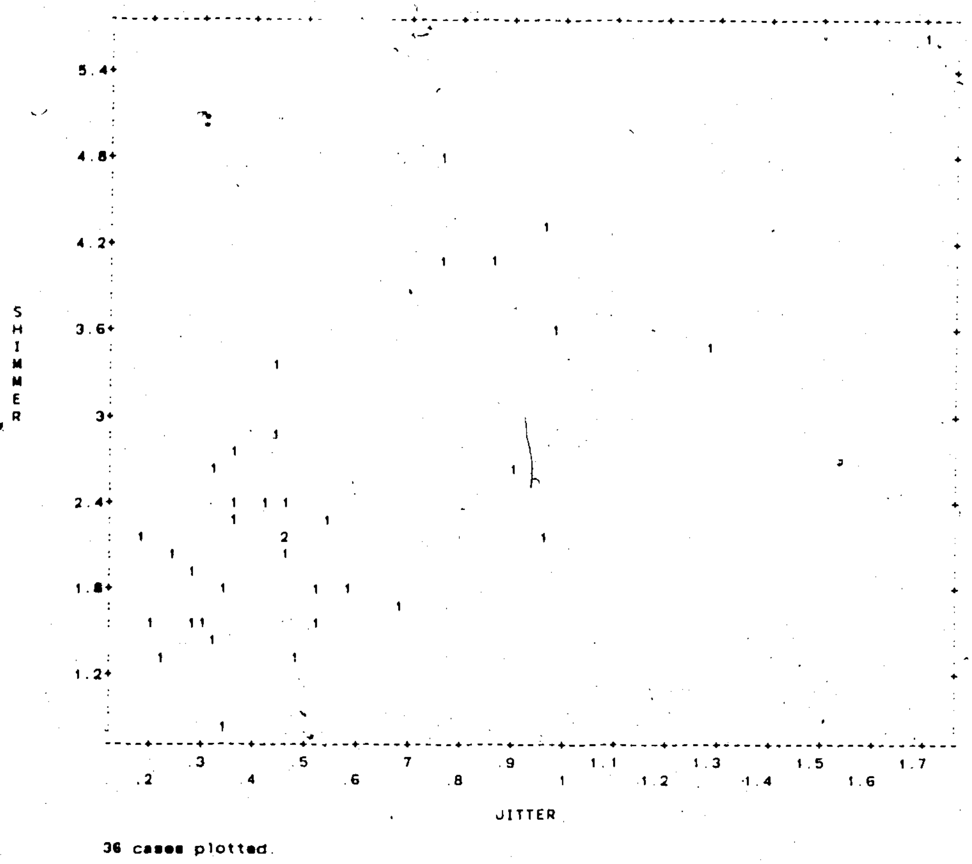


Figure 15. Plot of jitter with shimmer for the thirty-six tokens of male speaker M09. This speaker produced no outliers. r=.738, p=.000.

To obtain a more detailed view, correlation coefficients for the normal tokens of individual subjects were calculated. The results, presented in Table 12, showed jitter and shimmer to be significantly correlated at the .01 level for seven of the sixteen speakers. This should not be immediately attributed to speaker differences, however, as much of the apparent linear dependence could be seen to derive from a small number of tokens. The plot for M09 (r=.738), the subject having the highest correlation with 36 tokens, provides an example of this type: through the greater portion of this plot, a substantial degree of scatter is displayed (Fig. 15). Claims for a predictive relation between the mean jitter and mean shimmer magnitudes were thus not appropriate.

Cross-Correlations

The cross-correlations investigated the association between the signed jitter and shimmer values for individual periods within a token. At a time lag of zero, the jitter for a period, defined as the signed deviation from the linear trend in period duration, was correlated with the shimmer from the following peak; at a lag of one, it was correlated with the shimmer from the peak within the same period. These relations were illustrated in Figure 3. The cross-correlations thus relate the perturbations in period duration with the perturbations in peak amplitude, on a period-to-period basis. Significant correlations would support the hypothesis that the same causative mechanism underlies both perturbations, although the correlation itself could not, logically, provide insight into the nature of this mechanism.

The results showed the signed jitter and shimmer to be significantly correlated, at the .01 level, for 140 of the 576 tokens (24.3%). Significant negative correlations at a lag of zero, or positive correlations at a lag of one, accounted for 115 of these cases (19.9%). In these tokens, long periods followed large peaks, and short periods followed small peaks. By contrast, positive correlations at lag zero, or negative correlations at lag one, appeared in only 25 cases (4.3%), representing the reverse situation.

When the outliers were examined alone, 26 of these 30 tokens were found to be included among the 115 "large peak following long period" items; none demonstrated the opposite possibility. As a group, they were remarkable for the strength of their correlations: twelve of the twenty-six correlations were above .90, and all but three exceeded .70. The results also tended to be consonant with the alternate period measures, although the plot (Fig. 16) of the signed jitter and shimmer in ERE1, by M03 (lag=1, r=-.25, not significant), shows one way in which the analyses could deviate: although the alternate period jitter and shimmer measurements were both smaller than the successive period values, no consistent relation between jitter and shimmer was maintained. However, for most of these tokens, the correlations supported the idea that double periodicity, involving consistent alternations in both peak height and period duration, was present.

Without the outliers, 89 of the 546 normal tokens (16.3%) fell into the "long period following large peak" group; the number of "short period following large peak" tokens remained the same. The majority of these correlations were comparatively weak, however: of the 114 significant coefficients, only 4 were greater than .90, while 39 were between .70 and .89, and 71 fell between .40 and .69. The plots for tokens AEE1 (lag=1, r=.97) and AEE2 (lag=1, r=.55), produced by F08, reveal the qualitative difference between strong and weak correlations (Figs. 17 and 18). These plots link the jitter of a period with the shimmer from the following peak; this corresponds to a lag of 0. Token AEE1, an extreme outlier (jitter = 17.20%, shimmer = 84.43%), displays the exaggerated, regular pattern characteristic of double periodicity. In contrast, AEE2 (jitter = .55%, shimmer = 5.01%) shows what could be described as an inconsistent regularity: the relation between the parameters does not account for a great deal of the variance. As the autocorrelation analysis will show, the correlations for most of these tokens should not be associated with double periodicity.

The plot for a third /æ/ by F08, AEB2 (Fig. 19), provides an example of a token with no significant correlations (lag=0, r=-.16; lag=1, r=.09). This was the

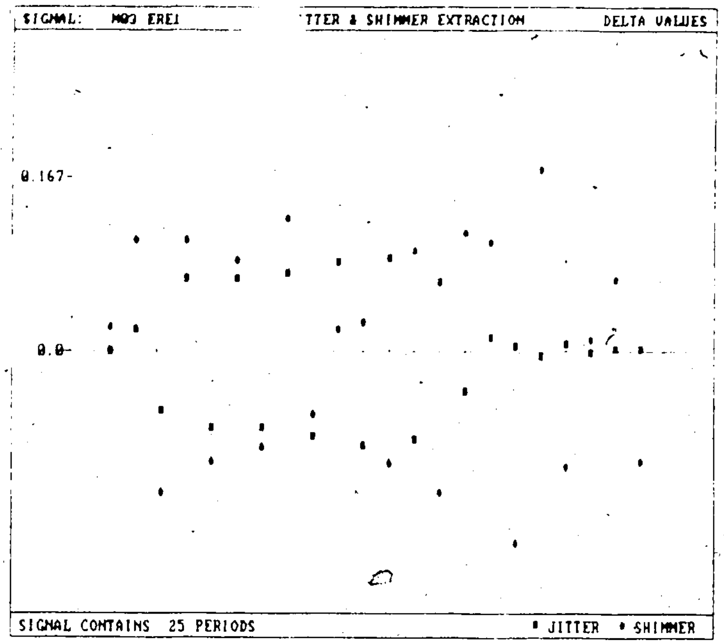


Figure 16. Signed jitter and shimmer for each period in token ERE1, produced by male speaker M03. The points represent relative perturbation values prior to the conversion to percentage. Jitter=4.59%, shimmer=9.54%, r=-.25 at lag=1.

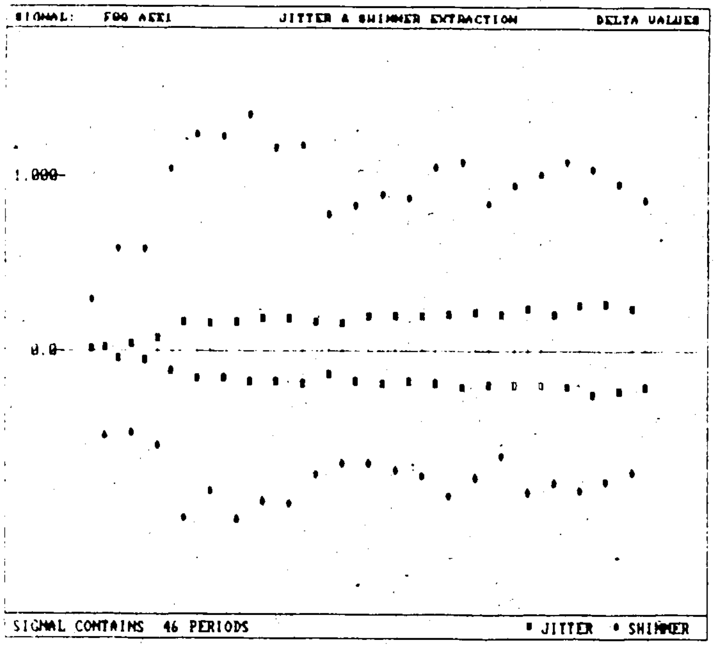


Figure 17. Signed jitter and shimmer for each period in the outlier token AEE1, produced by female speaker F08. The regular alternations are characteristic of double periodicity. Jitter=17.20%, shimmer=84.43%, r=.97 at lag=1.

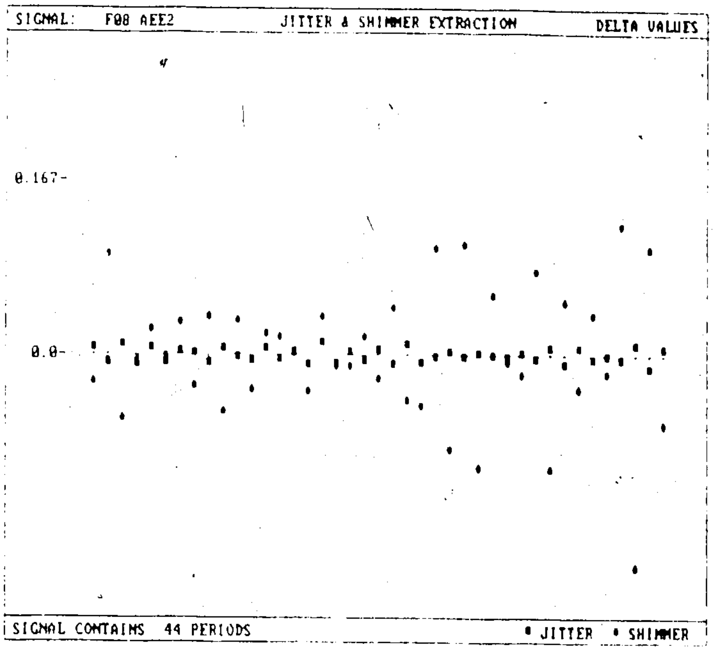


Figure 18. Signed jitter and shimmer for each period in AEE2, a normal token with a significant jitter/shimmer correlation, produced by speaker F08. Jitter=.55%, shimmer=5.01%, r=.55 at lag=1.

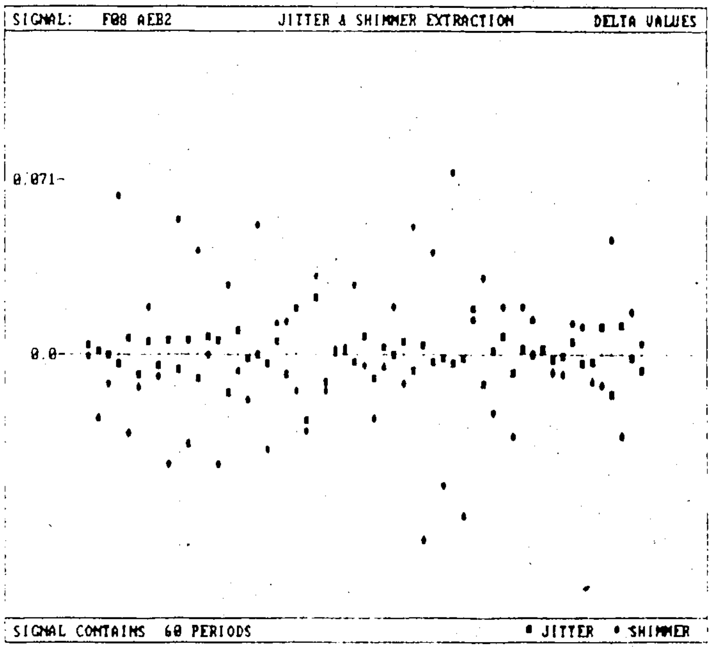


Figure 19. Signed jitter and shimmer for each period in AEB2, a normal token with no significant jitter/shimmer correlation, produced by speaker F08. Jitter=.63%, shimmer=2.54%, r=.09 at lag=1.

case for approximately four out of five of the normal tokens. Such a degree of independence suggests that separate causative mechanisms for jitter and shimmer should be considered.

H. Autocorrelations

The autocorrelation analysis tested for cycles within the signed jitter or shimmer deviations of each token. The array of period values for jitter, or for shimmer, was correlated with versions of the same series, delayed by integer numbers of periods; significant correlations indicated the presence of regularly repeated perturbations, with periods equal to the time delays. The delays, or lags, ranged from one period, to one-third the total number of periods, to a maximum lag of twelve periods: to be valid, at least three cycles had to be represented within a token. Since the number of lags varied for different tokens, results are reported both by the number of tokens significant at a given lag, and by percentages, with the number of significant tokens divided by the number of items for which the lag was calculated.

Jitter and shimmer values were based on three successive periods. Thus, the correlated measures were not independent at lags of one or two periods, lags for which the trend line windows overlapped. This tended to produce negative correlations at a lag of one period, and positive correlations at lag two, results which cannot be directly interpreted in terms of perturbation periodicities. For this reason, only results for lags three to twelve are presented.

The normal and outlier tokens were examined separately. In addition, a distinction was made between "sequential" and "non-sequential" correlations. With the former, significant correlations appeared for a consecutive number of lags, starting at lag one; with the latter, correlations at longer time lags were significant, while those at shorter lags were not. Sequential correlations were believed to arise from a relation between alternate periods. The coefficients for these tokens were negative at odd lags and positive at even ones; they decreased in value as the lag increased. The non-sequential correlations, reflecting long-term cycles in the underlying signal, represented the more interesting case.

Results for the normal tokens are given in Tables 13 and 14. Overall, very few correlations were significant at $p<.01$. For the non-sequential items, the numbers appear at, or near to, chance levels. To determine the number of correlations which might be expected due to chance alone, the arrays for each token could be randomized, and again autocorrelated. However, the minimal effects observed for both jitter and shimmer did not motivate this procedure: consistent non-sequential correlations did not occur at any lag. The sequential correlations suggested that periodic relations across two periods existed within a small number of normal tokens. For jitter, this effect involved less than 6% of the tokens (negative correlations at lag three = 5.87%); for shimmer, it was present for less than 3.5% (lag three = 3.41%). Tokens with sequential correlations were produced by fourteen of the sixteen speakers, but were not frequently generated by any individual.

The thirty tokens for which jitter or shimmer was an outlier showed a high proportion of sequential correlations. This could be expected from the regular alternations visible in many of the token perturbation plots. However, as with the normal group, no consistent longer cycles were apparent. Results for these tokens are presented in Tables 15 and 16.

The relation between jitter and shimmer differed for normal and outlier tokens. Sequential correlations affecting both parameters occurred in only six normal tokens; the remaining correlations showed that cyclical variations in duration, or in amplitude, could exist independently of each other. In contrast, parallel correlations for jitter and shimmer appeared in thirteen of the thirty outliers. This effect, associated with consistent double periodicity, should be distinguished from the independent alternations.

The token plots previously presented demonstrate some of the possible relations between the parameters. For token ERE1, by M03 (Fig. 16), significant correlations appeared at lags of one to five periods for jitter, but only at lag one for shimmer; in the plot, jitter can be seen to vary regularly throughout the token, while shimmer fails to maintain consistent alternations. The correlations for AEE1 (Fig. 17), an outlier with marked double periodicity, were highly significant at lags one through twelve for both jitter and shimmer. AEE2 and AEB2 (Figs. 18 and 19) provide examples of normal tokens with no correlations at lags greater than two; this was the case for the vast majority of tokens.

Lag	Number of Tokens	Non-Sequential				Sequential			
		Positive Correlations		Negative Correlations		Positive Correlations		Negative Correlations	
		#	%	#	%	#	%	#	%
3	528	6	1.14	1	.19			31	5.87
4	499	7	1.40	10	2.00	21	4.21		
5	464	8	1.72	3	.65		—	16	3.45
6	405	7	1.73	9	2.22	8	1.98		
7	350	9	2.57	8	2.28			7	2.00
8	301	6	1.99	9	2.99	4	1.33		
9	254	6	2.36	7	2.75			3	1.18
10	207	4	1.93	3	1.45	1	.48		
11	166	2	1.20	3	1.81			1	.60
12	124	1	.81	1	.81	1	.81		

Table 13. Number of normal tokens with significant autocorrelations for signed jitter at lags ranging from 3 to 12 periods ($p<.01$).

Lag	Number of Tokens	Non-Sequential				Sequential			
		Positive Correlations		Negative Correlations		Positive Correlations		Negative Correlations	
		#	%	#	%	#	%	#	%
3	528	11	2.08	7	1.32			18	3.41
4	499	1	.20	8	1.60	11	2.20		
5	464	6	1.29	5	1.08			7	1.51
6	405	5	1.23	4	.99	5	1.23		
7	350	3	.86	4	1.14			3	.86
8	301	1	.33	4	1.33	2	.66		
9	254	0		3	1.18			2	.79
10	207	1	.48	3	1.45	2	.97		
11	166	2	1.20	3	1.81			1	.60
12	124	2	1.61	6	4.84	1	.81		

Table 14. Number of normal tokens with significant autocorrelations for signed shimmer at lags ranging from 3 to 12 periods ($p<.01$).

Lag	Number of Tokens	Non-Sequential				Sequential			
		Positive Correlations		Negative Correlations		Positive Correlations		Negative Correlations	
		#	%	#	%	#	%	#	%
3	30	1	3.33	0				18	60.00
4	29	0		0		12	41.38		
5	27	1	3.70	0				9	33.33
6	24	0		1	4.17	8	33.33		
7	19	1	5.26	0				6	31.58
8	14	0		0		6	42.86		
9	10	0		0				3	30.00
10	8	1	12.50	0		3	37.50		
11	8	0		1	12.50			3	37.50
12	7	1	14.28	0		3	42.86		

Table 15. Number of tokens with significant autocorrelations for signed jitter at lags ranging from 3 to 12 periods, from the 30 tokens for which jitter or shimmer is an outlier ($p<.01$).

Lag	Number of Tokens	Non-Sequential				Sequential			
		Positive Correlations		Negative Correlations		Positive Correlations		Negative Correlations	
		#	%	#	%	#	%	#	%
3	30	0		0				15	50.00
4	29	1	3.45	0		12	41.38		
5	27	1	3.70	1	3.70			9	33.33
6	24	1	4.17	1	4.17	6	25.00		
7	19	0		0				3	15.79
8	14	0		0		2	14.28		
9	10	0		0				2	20.00
10	8	0		0		2	25.00		
11	8	0		0				2	25.00
12	7	1	14.28	0		2	28.57		

Table 16. Number of tokens with significant autocorrelations for signed shimmer at lags ranging from 3 to 12 periods, from the 30 tokens for which jitter or shimmer is an outlier ($p<.01$).

V. Discussion

As in the Results chapter, the first sections of this chapter briefly discuss the method of pitch extraction selected and the magnitudes measured in this study. The statistical effects are then reviewed, with comparisons made to results previously reported in the literature. The effect which individual differences may have on the statistical analyses, and the implications for the development of a screening procedure, are noted. Beyond these differences, possible causative factors related to the sex, vowel, and intonation effects are suggested, although much of this material must be regarded as speculative. With the correlations, particular attention is given to the problem of acoustic interactions between jitter and shimmer. The final section summarizes the findings of this thesis and offers some suggestions for further research.

A. Jitter and Shimmer Magnitudes

Jitter Measurement

Jitter determined from peaks in the acoustic waveform is not only influenced by timing perturbations in the glottal period: the generation of additive noise, or changes in the vocal tract configuration, may also contribute to the measured perturbations. The characteristic-peak method of period extraction attempts to minimize the effect of the last factor, by providing a consistent reference for the measurement of all vowel periods. In this approach, it is assumed that the perception of periodicity does not depend upon the location of the greatest amplitude point in the period, and that cycle-to-cycle waveshape variations in the acoustic wave do not directly contribute to the perception of roughness. These assumptions should be tested empirically.

Informal listening suggested that glottal perturbations and waveshape variations do not produce similar perceptual effects. Tokens with peak-switching, in which waveshape changes brought different peaks into prominence in successive periods, could have large jitter values when period durations were found by the "maximum-peak" method, but comparatively little characteristic-peak jitter. The token AEE2, by F10, is an example of this type, with a maximum-peak jitter of 14.31% and a characteristic-peak jitter of .97%. Unlike the characteristic-peak outliers, which were perceived as notably rough, this token was heard as a normal item. The listening conditions were not those which would be used in a perceptual experiment, however, in that playback occurred over a loudspeaker, in the presence of equipment noise. It thus cannot be claimed that waveshape changes have no effect on the quality of the signal, only that the effect is not obvious. It does not, though, appear useful, for perceptual purposes, to combine glottal perturbations with waveshape variations in a single measure.

Signals from electroglottographs or contact microphones, or low-pass filtered acoustic signals, provide a simple representation of the periodicity at the glottis. For the perception of roughness or naturalness, jitter should initially be defined in terms of these signals, with the contributions of other features of the acoustic wave, such as shimmer or waveshape changes, investigated separately. Differences between the maximum-peak and characteristic-peak jitter roughly indicate the presence of waveshape variations, and could facilitate the examination of this phenomenon.

Magnitudes in Natural Vowels

When the outlier tokens are removed from the analysis, the average magnitudes of the characteristic-peak jitter, at .79%, and of the shimmer, at 3.20%, appear similar to values reported for sustained vowels. Any detailed comparison of magnitudes would assume, however, that the measures used in this study, with perturbations normalized by a local average over three periods, are equivalent in effect to measures which normalize by the average pitch or amplitude across a sustained utterance. The degree of validity in this assumption could only be determined by using the "local-average" measures on sets of natural and sustained vowels, produced by the same speakers. In the absence of these data, it is not possible to distinguish between the effects of the different measures, and the different phonatory tasks. Such an analysis could, incidentally, reveal the extent to which speakers can consciously control or reduce perturbation magnitudes, if the speakers were instructed to sustain the prolonged vowels as steadily as possible. The relation between the capabilities of normal speakers, and their performance under more natural conditions, could thus be explored.

There is, in general, some question as to whether it is meaningful to compare magnitude levels across studies, given that the magnitudes are often reported only as group means. The variation among individuals, noted in this study, and the relatively small number of speakers often included in a subject group, suggest that misleading conclusions may be drawn from pooled data. The next section focusses upon this problem.

Individual Differences

The most surprising result from this study was the degree to which the sixteen, "normal" speakers varied in their production of jitter and shimmer. This variation was observed at every stage of the analysis, from the initial examination of the full 576 tokens, with the outliers included, to the subject by vowel interactions within the groups of "homogeneous" speakers. With a few exceptions, the presence of differences among healthy speakers is only hinted at in the clinical literature. However, if such differences are indeed common, they could explain both the current lack of success in devising a feasible screening procedure, and the inconsistent effects reported for such factors as vowel, sex, and age.

As a rule, clinicians are less concerned with exploring differences within groups of healthy speakers than with using these speakers to define the norm. This may be done in several ways. The simplest approach, which gives what Koike et al. (1977) call the "naive" norm, determines a borderline, between the normal and the abnormal, from a range of values measured from selected healthy speakers. An example is provided by Horii (1980), who calculated a critical value of .98 dB (11.94%) from the shimmer produced by thirty-one healthy males: with 95% confidence, shimmer above this level can be judged abnormal for this type of speaker. Abnormality is not, here, explicitly identified with pathology. A more sophisticated procedure is to "randomly" sample from both the normal and the pathologic populations. Since a screening procedure must decide the status of individuals, statistical differences are not useful; the range of values instead is examined. If the upper 90 percent confidence limit of the normal speakers is computed, the discrimination provided by the measures can be evaluated by using this limit to assign speakers to normal or pathologic groups. The accuracy of the jitter and shimmer measures, when so tested, is far from ideal.

It is possible that the discrimination between normal and pathologic speakers may be improved if norms are found for subgroups of the healthy population. This idea is similar to that used by the control study, which matches selected pathologic and normal speakers on factors, such as sex or age, which the researcher suspects may influence results. Thus, Horii would apply the .98 dB shimmer critical value only to young males, while suggesting a value of .48 dB for young females (Sorensen & Horii, 1983). Ludlow et al. (1987) similarly found separate upper limits for their normal male and female speakers, although they did not choose to control for age differences. Their study also attempted a second type of control, based on a linear multiple regression model. The obtained jitter and shimmer values were first correlated with various characteristics of the signals and speakers. Factors which showed significant correlations were then used in regression equations, from which the normal standards for males and females were derived. However, none of the speaker characteristics considered, i.e. age, smoking, or drinking, correlated significantly with the perturbation magnitudes, and their two analyses did not appreciably differ. As in previous tests (Zyski et al., 1984; Horiguchi et al., 1987), approximately half of the pathologic cases were undetected, while roughly 15% of the healthy speakers were classed as pathologic.

Th present study initially assumed that randomly selected healthy young adults would form a homogeneous subject group. The age range of the speakers was restricted, as the inclusion of older speakers has been shown to increase the inter-subject variability (Wilcox & Horii, 1980). When the analyses of variance revealed large subject differences, an exploratory type of analysis, hierarchical clustering, was chosen to group speakers on the basis of their similarities in performance; it was hoped that the clustering results might correlate with identifiable speaker characteristics. The results were negative. Sex did not play a determining role in clustering, with both sexes represented in every group. Smoking also did not appear to influence group membership: the first and most homogeneous jitter group contained three non-smokers and three current, though light smokers, while the shimmer group held five non-smokers and four smokers. The factors which might cause speakers to be similar or different in their production of perturbations cannot be deduced from the available information; it is not, in fact, fully clear what kinds of speaker characteristics need to be considered.

However, models of the perturbation origins suggest potentially relevant factors. Baer (1980) speculated that systematic differences in motor control strategies could exist between speakers. Physical-structural variations, including slight asymmetries in the conformation of the vocal folds, may be present. An individual's learned speech habits may partially condition the perturbation magnitudes (Lieberman & Michaels, 1962). The role of these and other factors must be investigated, before any meaningful "norms" can be developed. The problem is further complicated when pathologic speakers are considered, as different compensatory mechanisms can presumably be engaged, depending on the extent and involvement of the pathology. Some speakers with laryngeal pathologies are thus capable of producing relatively small amounts of jitter and shimmer, while some healthy speakers may, for whatever reasons, produce larger magnitudes. This current lack of information prevents the synonymous use of the terms "abnormal" and "pathologic", desired for screening purposes.

Changes in perturbation magnitudes have been associated with psychological stress (Inbar & Eden, 1976; Eden & Inbar, 1978); attention must therefore be given, not only to a speaker's characteristics, but to his reaction to the experimental task. It is possible that the recording conditions in this study may have induced stress in certain speakers. The subjects were isolated in a small recording booth, with no visual contact with the experimenter. They were required to watch for the flashing of a signal light, which instructed them to read a sentence. This environment and type of task were familiar to some of the speakers, but not to others. It should be noted, though, that stress and tension are expected to reduce perturbation magnitudes below the levels which a given speaker would ordinarily produce; they do not explain the presence of large perturbations.

Three subjects in this study, M02, M03, and F09, produced unusually large magnitudes for both jitter and shimmer. Each gave multiple outlier tokens; even following the elimination of the outliers, the cluster analyses showed their values to be abnormally high. No explanation for their behavior can be offered. However, the tendency to switch phonatory modes, from a quasi-periodic to a double-periodic type of vibration, suggests abnormal laryngeal adjustments on the part of these subjects. Although the production of a double periodic, or diplophonic voice is often regarded as a symptom of laryngeal pathology, reports in the clinical literature (e.g. Ward, Sanders, Goldman, & Moore, 1969; Moore, 1976), and in the pitch extraction literature (Rabiner et al., 1976; Hess, 1983, p. 50), indicate that such a voice may also be generated by healthy speakers. It appears to be caused by the independent vibration of the two vocal folds, which may, in a healthy individual, result from differential contractions in the vocal fold muscles (Ward et al., 1969). Transient diplophonic effects may be caused by the differential loading of the vocal folds with strands or chunks of mucous. Phonation in this mode is clearly habitual for speaker M03: from thirty-six tokens, he gave seventeen outliers, each showing some degree of double periodicity (Appendix E). Such tokens were less commonly produced by M02 and F09. Particularly irregular double periodic alternations appeared in F09's tokens. These speakers could switch modes both between and within tokens: the mechanisms involved in such switches again are not known.

Unusual features of the voices were noted, in a general way, during the vowel gating procedure. The voice qualities of these speakers appear to reflect their abnormalities. Speaker M02, who had the highest mean fundamental frequency among the male subjects, had a strained, slightly hoarse voice. Speaker M03 had a voice which could only be described as peculiar: it was characterized by a moderate degree of roughness and a very distinctive timbre. F09's voice was perceived as rough. There is a tendency, in discussion, to group these three speakers. It must be remembered that, as well as differing from the clustered speakers, they also differ from each other.

The inclusion of anomalous speakers can have large effects on the statistical analyses. In this study, the vowel by sentence position interactions from the first jitter ANOVA (Table 8), which followed the adjustment of the outliers but preceded the cluster analysis, can be largely attributed to the four speakers not clustered for jitter, M02, M03, F04, and F09. This ANOVA showed /ɚ/ to have significantly more jitter at the end of the sentence than within it, while /ɷ/ had significantly less jitter in sentence-final position. However, when the mean values for the grouped and ungrouped speakers were examined separately, as in Table 17, the positional differences for the grouped speakers were seen to be relatively small. The second ANOVA (Table 10), which included only these subjects, did not find the differences significant. A few subjects, each producing large differences, have here created effects which do not consistently hold for the majority of speakers.

A lack of homogeneity in the healthy population could explain some of the differing results reported in the literature, as the statistical effects found may then depend on the particular speakers tested. The false alarm rate from the clinical studies, with normal speakers classed as pathologic, suggests that large speaker differences may occur with sustained as well as natural vowels. A better understanding of the origins of the perturbations in healthy speakers might allow the researcher to control for some of this variability. This is particularly important when perturbations are to be applied in a screening procedure, or when the effect of specific factors on jitter and shimmer magnitudes is at issue. If a study's aim is to obtain data to improve the natural quality of synthetic speech, it might be appropriate to select speakers on the basis of voice quality, rather than

	/ɚ/ Within-Sentence	/ɚ/ Sentence-Final
Four Ungrouped Subjects	.90 (2 outliers)	2.04 (1 outlier)
Twelve Grouped Subjects	.35	.41 (1 outlier)
All Subjects	.53	.92

	/ɑ/ Within-Sentence	/ɑ/ Sentence-Final
Four Ungrouped Subjects	1.60	.91 (1 outlier)
Twelve Grouped Subjects	.66	.57
All Subjects	.96	.72

Table 17. Contribution of the ungrouped speakers M02, M03, F04, and F09 to the vowel by position interactions for jitter. The outlier tokens have been adjusted. All values are in percentage.

randomly sampling from the population. Perturbations may then be measured only from voices judged to be pleasant, or, at least, from voices with no blatant perceptual abnormalities. It is further possible to find the differences themselves of interest, as these suggest a potential use for jitter and shimmer in the identification of individual speakers. This avenue of research should be pursued.

In view of the variability displayed, the number of subjects examined in this study may appear small: the generality of the results then must be considered. It is assumed that the clustered speakers represent a single population, although the defining features of this population remain unspecified. However, a substantial proportion of young, healthy speakers will presumably behave in a manner similar to these subjects.

With the outlier tokens, the variability within subjects can be as great as that between subjects. Several phonatory samples are therefore required before a speaker may be classified.

Sex Effects

The first set of ANOVAs (Tables 8 and 9), which included all sixteen speakers, tested the effect of speaker sex on jitter and shimmer magnitudes. No main effects were found. However, a sex by vowel interaction for jitter did appear, with females giving significantly larger values for the vowel /ɩ/. To determine the influence of the four ungrouped speakers on this effect, the average magnitudes for the sixteen speakers were listed. These magnitudes, presented in Table 18, show a consistent tendency for females, whether grouped or ungrouped, to produce more jitter for this vowel; unlike the vowel by position interaction, the effect is not due to the presence of the anomalous subjects. Given the observed speaker differences, and the general lack of agreement on effects involving sex and vowels, the question must then be whether a second, preferably larger, group of speakers would replicate the effect. The unique features of /ɩ/ which could cause an interaction with sex are not known.

As might be expected, the results here reported agree with certain previous studies, and disagree with others. Haji et al. (1986), measuring jitter in /a/, failed to find a sex difference; Ludlow et al. (1987), for shimmer in /a/, and Sorensen and Horii (1983), for shimmer in /a/, /u/, and /i/, also noted no main effects. However, sex differences for jitter were observed by Hartmann and von Cramon (1984) and Sorensen and Horii (1983), with females producing more jitter than males. Sorensen and Horii additionally reported sex by vowel interactions, with their female group giving more jitter and less shimmer than their males for the vowels /i/ and /u/. These studies are discussed in detail in the literature review. Jitter and shimmer have rarely

Male Speakers

M01	.68
M02	(.76)
M03	(1.27)
M04	.37
M06	.36
M08	.44
M09	.61
M10	.44

Female Speakers

F01	.68
F02	.84
F03	1.04
F04	(1.15)
F08	1.05
F09	(1.75)
F10	1.23
F11	1.63

Table 18. Jitter values for /ι/, averaged over four tokens for each speaker, following adjustment of the outliers. Values for the ungrouped speakers are given in parentheses.

been measured from /ι/, and possible sex effects have not been explored.

From the literature it must be asked, first, why sex differences are not consistently present, and second, why the differences for jitter, when present, are always in the direction of more jitter for females than males. An answer to the first question may be found in the variability apparent among individual speakers. This study's cluster analyses showed some females to differ from some males. With random sampling, the differences may presumably coincide, in some cases, to create between-groups effects. The particular groups compared may then determine the results. This was previously suggested in the literature review, in connection with the effects reported by Sorensen and Horii (1983). Their study compared the jitter and shimmer of a group of female speakers with the values obtained in Horii (1980) for a group of males. It was noted, in particular, that identical sex by vowel interactions for shimmer would not have appeared had the male group from Sorensen and Horii (1984) instead been tested, as the average shimmer values produced by these speakers for /i/ and /a/ were smaller, not larger, than those of the females. The replicability of the sex by vowel interactions may similarly be questioned. However, the consistent direction of the jitter effects does indicate a tendency for females to give more relative jitter than males: while a female and male will not necessarily differ, when they do, the female will be more likely to produce larger magnitudes.

It may be possible to describe this tendency in terms of voice frequency rather than sex. In this study, the correlation between jitter and frequency for the 546 normal tokens was found to be .18 (p=.000), indicating a significant, but weak tendency for relative jitter to increase with frequency. Stronger correlations may be expected from studies with significant sex effects. However, for jitter, there appears to be no explanatory advantage to this approach. In contrast, the work of Hillenbrand (1987) suggests that frequency may be important for the interpretation of shimmer effects. Hillenbrand manipulated the source properties of synthetic vowels to show that shimmer in the acoustic signal is influenced by both the source jitter and the fundamental frequency. As frequency rises, absolute jitter magnitudes decrease, but the energy overlap between adjacent vowel periods increases: the jitter present at higher frequencies then has stronger effects in producing shimmer. Although the complexity of these relations makes predictions difficult, the observed increases in relative jitter could be expected to cause some increase in shimmer, which could in turn be reflected as a sex effect. That this does not occur requires explanation: shimmer must be considered both at the glottis and in the acoustic waveform. The associations between frequency, jitter, and shimmer will be further discussed in the Intonation Effects section.

The investigation of sex effects has been motivated primarily by practical concerns, centered on the need to establish normative levels for screening tests. Little attention has been given to the theoretical reasons why the sexes should, or should not, differ. From the cluster analyses of this study, speaker sex does not appear to play a major role in determining similarities or differences in performance. Some males may differ from some females; from the literature, some groups of males and females may differ. However, at present, sex differences for relative jitter or shimmer must be regarded as secondary to overall speaker differences: the inclusion of sex as a factor does not appreciably reduce the inter-subject variability.

Vowel Effects

The main effects for vowel again show the way in which results may vary depending on the speakers tested. The first ANOVA for jitter (Table 8), which included all subjects, found no significant differences among the vowels. However, for the twelve grouped subjects of the second ANOVA (Table 10), an effect emerged, with /ι/ having significantly more jitter than /ɚ/. The unusually high values given by female speakers for /ι/ (see Table 18) are largely responsible for this effect: although "group" has replaced "sex" as a factor in this analysis, the sex by vowel interaction from the first ANOVA implicitly remains. For shimmer, the details of the effects differ between the two analyses, with /æ/ and /ɷ/ having significantly more shimmer than /i/ with sixteen speakers, and /æ/ and /ʌ/ having more shimmer than /u/ and /i/ with the nine speakers of the most homogeneous group. Before attempts are made to interpret these results, though, it must be noted that subject by vowel interactions among the clustered, homogeneous speakers appeared for both jitter and shimmer. These interactions were plotted in Figures 12, 13, and 14 of the preceding chapter. Individual speakers can clearly deviate from the statistical trends.

Perturbations have most frequently been measured from the cardinal vowels /a/, /u/, and /i/. In the literature review, it was noted that, while the statistical significance of the results may vary, /i/ tends across studies to have more jitter and less shimmer than /a/. No consistent relation appears with /u/ for jitter; there may be a weak tendency for this vowel to have the least shimmer. The vowel orders found in previous studies may be seen in Tables 1 and 2. In this study, the vowel order for jitter, both for the clustered speakers and all speakers, was /u, i, ɒ/, with nine of the twelve clustered speakers and eleven of the sixteen speakers giving higher jitter values for /i/ than /ɒ/. For shimmer, the vowel orders were /ɒ, i, u/ for the nine clustered speakers, and /ɒ, u, i/ for all speakers. Seven of the nine speakers and twelve of the sixteen speakers gave more shimmer for /ɒ/ than /i/. Although the differences between these vowels are not significant in this study, the consistency with which these orderings appear requires explanation. Results are similar for both sustained and naturally-produced vowels.

Evidence from the literature suggests that vowel differences originate primarily in the vocal tract. Horii (1982), who used a miniature accelerometer to transduce the signal at the glottis, found no statistical differences for jitter or shimmer among eight English vowels. Koike et al. (1977), who employed a contact microphone to record five English vowels, noted that the jitter and shimmer variations for the vowels of individual speakers seemed "quite random". They did not statistically test for vowel effects. However, the most compelling data come from Milenkovic's 1987 study of synthetic speech. Milenkovic introduced different amounts of jitter and shimmer into the excitation signal for the vowels /a/, /u/, and /i/. He then used an autocorrelation procedure, with interpolation, to measure the perturbations present in the waveforms. In the first test, jitter was varied, while the source amplitude remained constant. He found a given amount of jitter to have the strongest effect in producing shimmer in /a/, a weaker effect in /i/, and the weakest effect in /u/ (graphs, p. 535). In the second test, shimmer was varied, while the fundamental frequency was kept constant. Shimmer was seen to produce the most jitter in /u/, less in /i/, and the least in /a/ (p. 536). These results correspond to the vowel orders often reported for natural production. Milenkovic suggested that the energy overlap between successive pulse responses may generate acoustic jitter and shimmer. He observed that the shimmer measured in the first test, and the jitter in the second, tended to increase for shorter fundamental periods, for which the overlaps were greater. However, the increases with frequency differed for different vowels. It may be that the formant frequencies of the vowels determine the amount of energy in the tail portion of the pulse responses, and so the strength of the effects. Milenkovic did not attempt to describe the acoustic interactions involved; he did, though, advocate the further study of vowel effects, both at the glottal source and in the acoustic wave.

While Heiberger and Horii (1982) stated that they could find no physiological reason to expect jitter differences among sustained vowels, a possible influence of tongue height on laryngeal tension, and of tension on jitter, should be considered. Lehiste (1970, p. 70) noted that, in the production of high vowels, the tongue tends to be pulled upwards, stretching the laryngeal muscles and increasing tension. Increased laryngeal tension has been associated with reductions in jitter (Klingholz & Martin, 1985). This does not explain why more jitter should appear for /i/, the vowel with the highest tongue position, than for /a/, but does identify a potentially relevant source factor. It is presumed that the vowel effects in this study derive, at least partially, from the glottal source: as the vowels involved share no unique features, the generation of effects through acoustic interactions seems unlikely.

The measurement data for the vowels /a/, /u/, and /i/ do not well agree with the results reported by Rozsypal and Millar (1979) for a naturalness-rating experiment with synthetic vowels. These researchers synthesized three-formant tokens of the vowels, introducing different amounts of jitter and shimmer at the source. Five levels of jitter were programmed, ranging from 0 to 3.20%. The five shimmer levels ranged from 0 to 15.19%. The vowel tokens were each 1200 ms long, with a mean fundamental frequency of 110 Hz and a stationary pitch contour. The seventy-five stimuli (5 jitter levels × 5 shimmer levels × 3 vowels) were presented once over a loudspeaker system to twenty-seven subjects, who rated the naturalness of the tokens on a seven point scale. Rozsypal and Millar found the optimal amounts of jitter to be .80% (level 2) for /u/ and 1.60% (level 3) for /a/, with the optimal amount for /i/ ranging between 0 and 1.60% (levels 1 and 3). The presence of shimmer always decreased naturalness. It may be noted, though, that this study did not consider the effect of acoustic interactions in the vocal tract. The mean perturbation magnitudes in the acoustic wave, particularly for shimmer, almost certainly exceeded their nominal values. More importantly, jitter and shimmer were not independent in the stimuli, making interpretation of the results difficult. It might be useful to repeat this experiment, with measurements made of the jitter and shimmer present in the acoustic signals. From the magnitudes observed in this thesis, smaller increments for source shimmer might also be appropriate.

Rozsypal and Millar had expected to find a "trading" relation between jitter and shimmer for perceived naturalness. They noted that the two types of perturbations can be spectrally similar, producing components which differ in phase but not in frequency. Since hearing is known to be relatively insensitive to phase differences within complex signals, jitter and shimmer could be predicted to have equivalent and additive effects on naturalness. Heiberger (1980) claimed to have found an additive effect for the perceived roughness of triangular waves containing more than 1.0% jitter and 1 dB shimmer. However, at lower perturbation levels (.5% jitter and .5 dB shimmer), the roughness of the combined stimuli were seen to be approximately equal to the roughness of either the jitter, or the shimmer, included. Tests should be made for a similar effect on naturalness, using small amounts of jitter and shimmer. It must be asked, first, whether the significant differences indicated by the production measures for vowels indeed correspond to differences in perceived naturalness, and second, whether small amounts of jitter are perceptually distinguishable from small amounts of shimmer. The relation between jitter and shimmer in the vowels /i/ and /a/ could be interesting in this respect.

This study was designed to provide information which may be used in a complementary study of voice naturalness. Subjects could be instructed to rate the naturalness of various tokens, to establish "good" levels of perturbation; further experiments with synthetic speech could then be undertaken. The subject by vowel interactions in this study suggest that a certain variability for perturbations in vowels may be acceptable to listeners.

Intonation Effects

To test whether relative jitter and shimmer magnitudes vary with intonation, perturbations were measured from words in two sentence positions. The jitter ANOVAs (Tables 8 and 10) showed no main effects for position: the jitter values of vowel segments from words within and at the end of the sentence did not significantly differ. A vowel by position interaction was found in the first ANOVA, which included all subjects. However, as previously noted, this could be attributed to the four speakers not grouped for jitter; the interaction did not appear for the clustered speakers. For shimmer, a position effect emerged in both ANOVAs (Tables 9 and 11), with significantly more shimmer produced in sentence-final segments. The factors which vary with position must then be considered.

It was initially assumed that differences with intonation, if such appeared, would primarily reflect a frequency effect on the relative jitter or shimmer measures. While frequency normalization controls for the linear effect of frequency on absolute jitter, the linearity of the relation has not been fully established: it is possible that a residual effect may remain following normalization. From the literature, frequency may also have a within-subjects influence on relative jitter for subjects phonating over a wide range of frequencies (Koike et al., 1977). For shimmer, Ludlow et al. (1987) reported a negative correlation with frequency for their male subjects, revealing a between-subjects tendency for shimmer to increase as frequency decreases. To examine the effect of frequency in this study, the jitter and shimmer measures were correlated with the mean frequency values for the 546 normal tokens. A correlation of .18 (p=.000) was found between jitter and frequency; the correlation for shimmer was -.15 (p=.000). Although significant, coefficients of such small magnitudes have little explanatory value.

The direction of the shimmer correlation was somewhat surprising. As previously observed in the Sex Effects section, the acoustic interaction effect of jitter on shimmer is stronger at higher frequencies. Since relative jitter showed a slight tendency to increase with frequency, a decrease in shimmer would not have been predicted. Some source effect instead seems indicated. It may be suggested that, for shimmer, intensity rather than frequency is the relevant factor. Although measurements were not made, intensity, as frequency, may be assumed to fall towards the end of an imperative sentence. A between-subjects effect of intensity on shimmer has been reported by Ludlow et al. (1987), who found shimmer to increase as intensity decreased (r=-.44). It thus seems possible that an intensity effect on the relative shimmer measures might underlie the significant differences for sentence position.

As this study tested only one sentence type, only preliminary conclusions may be drawn. It appears that intonational changes do not necessarily produce significant effects on relative jitter measures; however, such effects for shimmer may occur. The potential effects of intonation, or its physical correlates, must therefore be considered in the design of measurements studies with connected speech.

B. Correlations

Magnitude Correlations and Cross-Correlations

Magnitude correlations determine the degree of association between the mean jitter and shimmer values of vowel tokens. When perturbations in the acoustic signal are correlated, though, the interpretation of "association" can be problematic. Horii (1980) and Heiberger and Horii (1982), who reported significant correlations, took their results to indicate that common source factors may underlie both types of perturbations. In contrast, Hillenbrand (1987) and Milenkovic (1987) suggested that a dependent relation may hold, as their tests with synthetic vowels showed source jitter to have strong effects in producing acoustic shimmer; source shimmer additionally had a weak effect on acoustic jitter. The correlations themselves reveal the strength but not the cause of the association.

From the results of Hillenbrand and Milenkovic, significant correlations may be predicted to occur. In this study, the correlations betwen mean jitter and shimmer were found to be .846 (p=.000) when the outlier tokens were included, and .491 (p=.000), when these thirty tokens were eliminated. The outlier tokens, whether displaying double periodicity, glottal fry, or simple exaggerated irregularities, tended to be abnormal for both jitter and shimmer. However, the correlation for the normal tokens did not account for a great deal of the variability in the measures, and it appeared possible for jitter and shimmer to be uncorrelated for the tokens of individual speakers. These findings may be reconciled with those of Hillenbrand and Milenkovic if it is assumed that additional factors interact with the measures, and so obscure existing relationships. Both Hillenbrand and Milenkovic observed that the strength of the effect of jitter on shimmer, and of shimmer on jitter, varied with the mean fundamental frequency (for constant amounts of absolute jitter or shimmer introduced at the source); Milenkovic also noted what could be described as a jitter by shimmer by fundamental frequency by vowel interaction. The present study measured perturbations from nine vowels; with natural intonation, the frequency of phonation changed both between and within tokens. Direct correlations between jitter and shimmer, over all conditions, thus may not reveal a relationship.

The studies with synthetic speech examined only the dependencies produced through acoustic effects. As Hillenbrand observed, though, it would not be surprising to find that jitter and shimmer are additionally correlated at the glottal source. This study's cross-correlation analyses attempted to explore the source relations by correlating the signed jitter and shimmer magnitudes of individual vowel periods. It was assumed that acoustic interactions would not generate consistent positive or negative perturbations. The interactions are thought to derive primarily from the energy overlaps between adjacent periods. With source jitter and shimmer, the superimposed components vary from cycle to cycle. The acoustic effects then add random noise to the cross-correlation data. significant cross-correlations can be expected to be more significant at the source. The analyses found significant correlations for 24.3% of the tokens overall. Negative correlations at a lag of zero, or positive correlations a lag of one, appeared for 16.3% of the normal tokens and for 86.6% for the outliers, indicating some tendency for high peaks to be followed by long periods. In contrast, positive correlations at lag one, or negative correlations at lag zero, were seen for only 4.6% of the normal tokens and for none of the outliers; chance alone could have produced these effects. If the goal is to provide information which would allow a synthetic source to simulate the behavior of the glottis, these relations would be better examined directly at the glottal source; however, for perceptual purposes, it may be sufficient to note that, for some small percentage of tokens, the period to period jitter and shimmer deviations will not be independent of one another in the acoustic signal.

It may be observed that significant magnitude correlations do not imply the presence of significant cross-correlations, as the token magnitudes are calculated from the absolute values of the period deviations.

Autocorrelations

The autocorrelation analysis attempted to determine whether the jitter and shimmer deviations of periods within the tokens are random, or whether they vary in a cyclical fashion. Previous time series analyses with sustained vowels have reported long-term cycles, over seven periods for jitter (Iwata, 1972) and twelve periods for shimmer (Koike, 1969; Von Leden & Koike, 1970). These effects have been compared to the singer's vibrato: it is suggested that they emerge when speakers attempt to maintain a vowel at a steady pitch and loudness. Between successive cycles, negative correlations for jitter, and positive correlations for peak amplitude, were also observed. The autocorrelations in this study tested for effects in natural, dynamic vowels, produced in a sentential context. In contrast to the studies cited, the perturbations were measured from the acoustic, not the glottal, wave.

Since jitter and shimmer were determined from a three-point trend line, the correlated measures were not independent at lags of one or two periods. Correlations between alternate periods were therefore inferred from "sequential" lags, where the coefficients for each lag, from one up to (at least) three, were significant. For jitter, thirty-one (5.8%) of the normal tokens showed such correlations; for shimmer, the number was eighteen (3.4%). Such values could have been produced by chance.

Significant non-sequential correlations were similarly rare: no evidence was found for any consistent long-term cycles. These results suggest that a random modulating function may be appropriate for the generation of jitter or shimmer in non-stationary synthetic vowel stimuli.

C. Summary and Suggestions

This thesis attempted to answer the following very specific questions related to the natural production of jitter and shimmer by normal speakers:

1. Are there sex differences for jitter or shimmer?

2. Are there vowel differences?
3. Are there differences with intonation?
4. Are mean jitter and shimmer magnitudes correlated?
5. Are jitter and shimmer correlated on a period-to-period basis?
6. Are jitter and shimmer random perturbations?

It was designed to provide data which might, following complementary perceptual studies, be applied to enhance the naturalness of synthetic speech.

The results for the factors of sex and vowel appear secondary to the large individual differences found. While some males may differ from some females, the cluster analyses did not show sex to play a determining role in these differences. For vowels, both main effects and subject by vowel interactions emerged. The significance of these effects should be evaluated in perceptual terms, though: the preferences of listeners must be established. The individual differences themselves may be interesting, as they suggest a potential use for perturbations in speaker identification or recognition. However, the consistency of speakers over time has yet to be examined. A better understanding of the origins of these differences is also required if a screening procedure for laryngeal pathologies is to be based on perturbation magnitudes.

Intonation was considered to determine whether measurement studies with connected speech need to control for this factor. The variations with intonation should be described in physical terms, as changes in frequency or intensity. No effect appeared with relative jitter for the one sentence type tested; however, shimmer magnitudes did differ between two sentence positions.

The last three questions are concerned with signal features which might be emulated when introducing perturbations into synthetic stimuli. From the autocorrelation analyses, regular modulations in the perturbations do not extend over more than two vowel periods. Such relations tended to occur in tokens judged to be outliers. The magnitude correlations between the mean jitter and shimmer values for the tokens, although significant, accounted for little of the variance in the measures: no predictive relation could be claimed. However, the cross-correlations did indicate a relationship to exist between the jitter and shimmer of individual vowel periods for approximately one out of five of the tokens. These relations should be examined in the glottal signal. It must be asked whether correlations at the source, if obscured by acoustic interactions in the vocal tract, have any perceptual significance. On the basis of the production data, the modulating functions which introduce jitter and shimmer into synthetic stimuli may be random, but not mutually independent.

References

Askenfelt, A.G. and Hammarberg, B. (1986) Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research 29*, 50-64.

Baer, T. (1980) Vocal jitter: A neuromuscular explanation. In *Transcripts of the Eighth Symposium on the Care of the Professional Voice*. Eds. V. Lawrence & B. Weinberg. New York: The Voice Foundation, Vol. 1, 19-22.

Baken, R.J. (1987) *Clinical Measurement of Speech and Voice*. Boston: Little, Brown & Company.

Beckett, R.L. (1969) Pitch perturbation as a function of subjective vocal constriction. *Folia Phoniatrica 21*, 416-425.

Benjamin, B.J. (1981) Frequency variability in the aged voice. *Journal of Gerontology 36*, 722-726.

Davis, S.B. (1976) Computer evaluation of laryngeal pathology based on inverse filtering of speech. *SCRL Monograph No. 13*. Santa Barbara, California: Speech Communications Laboratory.

Davis, S.B. (1979) Acoustic characteristics of normal and pathological voices. In *Speech and Language: Advances in Basic Research and Practice*. Ed. N.J. Lass. New York: Academic Press, Vol. 1, 271-335.

Disner, S.F. (1982) Stress evaluation and voice lie detection: A review. In *UCLA Working Papers in Phonetics 54*, 78-92.

Eagles, I., Morrow, C. and Sannino, M. (1986) *Alligator 2.0*. University of Alberta, Department of Linguistics.

Eden, G. and Inbar, G.F. (1978) Physiological model analysis of involuntary human-voice tremor. *Biological Cybernetics 30*, 179-185.

Haji, T., Horiguchi, S., Baer, T. and Gould, W.J. (1986) Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. *Journal of the Acoustical Society of America 80*, 58-62.

Hartmann, E. and von Cramon, D. (1984) Acoustic measurement of voice quality in central dysphonia. *Journal of Communication Disorders 17*, 425-440.

Hecker, M.H.L. and Kreul, E.J. (1971) Descriptions of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency. *Journal of the Acoustical Society of America 49*, 1275-1282.

Heiberger, V.L. (1980) *Roughness levels of jittered/shimmered stimuli*. Unpublished Master's thesis, Purdue University.

Heiberger, V.L. and Horii, Y. (1982) Jitter and shimmer in sustained phonation. In *Speech and Language: Advances in Basic Research and Practice*. Ed. N.J. Lass. New York: Academic Press, Vol. 7, 299-332.

Hess, W.J. (1982) Algorithms and devices for pitch determination of speech signals. In *Automatic Speech Analysis and Recognition*. Ed. J.P. Haton. Dordrecht: D. Reidel Publishing Co., 49-67.

Hess, W.J. (1983) *Pitch Determination of Speech Signals: Algorithms and Devices*. New York: Springer Verlag.

Hiki, S., Sugawara, K. and Oizumi, J. (1966) On the rapid fluctuations of voice pitch. *Journal of the Acoustical Society of Japan 22*,

Hillenbrand, J. (1987) A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research 30*, 448-461.

Hirano, M. (1979) The sound and its origins: Panel discussion. In *Transcripts of the Seventh Symposium on the Care of the Professional Voice*. Ed. V. Lawrence. New York: The Voice Foundation, Vol. 1, 54.

Hollien, H., Michel, J. and Doherty, E.T. (1973) A method for analyzing vocal jitter in sustained phonation. *Journal of Phonetics 1*, 85-91.

Horiguchi, S., Haji, T., Baer, T., and Gould, W.J. (1987) Comparison of electroglottographic and acoustic waveform perturbation measures. In *Laryngeal Function in Phonation and Respiration*. Eds. T. Baer, C. Sasaki, & K. Harris. Boston: Little, Brown & Company, 509-518.

Horii, Y. (1979) Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research 22*, 5-19.

Horii, Y. (1980) Vocal shimmer in sustained phonation. *Journal of Speech and Hearing Research 23*, 202-209.

Horii, Y. (1982) Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research 25*, 12-14.

Inbar, G.F. and Eden, G. (1976) Psychological stress evaluators: EMG correlation with voice tremor. *Biological Cybernetics 24*, 165-167.

Iwata, S. (1972) Periodicities of pitch perturbations in normal and pathologic larynges. *Laryngoscope 82*, 87-96.

Iwata, S. and von Leden, H. (1970) Pitch perturbation in normal and pathologic voices. *Folia Phoniatrica 22*, 413-424.

Kasprzyk, P.L. and Gilbert, H.R. (1975) Vowel perturbation as a function of tongue height. *Journal of the Acoustical Society of America 57*, 1545-1546.

Kersta, L.G., Bricker, P.D. and David, E.E., Jr. (1960) Human of machine?--A study of voice naturalness. *Journal of the Acoustical Society of America 32*, 1502.

Kitajima, K. and Gould, W.J. (1976) Vocal shimmer in sustained phonation of normal and pathologic voice. *Annals of Otology, Rhinology, and Laryngology 85*, 377-381.

Kitajima, K., Tanabe, M. and Isshiki, N. (1975) Pitch perturbation in normal and pathologic voice. *Studia Phonologica IX*, 25-32.

Klingholz, F. and Martin F. (1985) Quantitative spectral evaluation of shimmer and jitter. *Journal of Speech and Hearing Research 28*, 169-174.

Koike, Y. (1969) Vowel amplitude modulations in patients with laryngeal diseases. *Journal of the Acoustical Society of America 45*, 839-844.

Koike, Y. (1973) Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica VII*, 17-23.

Koike, Y., Takahashi, H. and Calcaterra, T.C. (1977) Acoustic measures for detecting laryngeal pathology. *Acta Otolaryngologica 84*, 105-117.

Lehiste, I. (1970) *Suprasegmentals*. Cambridge, Mass.: The M.I.T. Press.

Lieberman, P. (1961) Perturbations in vocal pitch. *Journal of the Acoustical Society of America 33*, 591-603.

Lieberman, P. (1963) Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America 35*, 344-353.

Lieberman, P. and Michaels, S.B. (1962) Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America 34*, 922-927.

Lorr, M. (1983) *Cluster Analysis for Social Scientists*. San Francisco: Josey-Bass.

Ludlow, C.L., Bassich, C.J., Connor, N.P., Coulter, D.P. and Lee, Y.J. (1987) The validity of using phonatory jitter and shimmer to detect laryngeal pathology. In *Laryngeal Function in Phonation and Respiration*. Eds. T. Baer, C. Sasaki, & K. Harris. Boston: Little, Brown & Company, 492-508.

Ludlow, C.L., Connor, N.P. and Coulter, D.P. (1984) A preliminary investigation into the validity of an optimum frequency for phonatory functioning in patients with laryngeal pathology. In *Transcripts of the Twelfth Symposium on the Care of the Professional Voice*. Ed. V.L. Lawrence. New York: The Voice Foundation, 159.

Ludlow, C.L., Coulter, D.C. and Gentges, F.H. (1983a) The effects of change in vocal fold morphology on phonation. In *Transcripts of the Eleventh Symposium on the Care of the Professional Voice*. Ed. V.L. Lawrence. New York: The Voice Foundation, Vol. 1, 77-89.

Ludlow, C.L., Coulter, D.C., and Gentges F.H. (1983b) The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies. In *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. Eds. D.M. Bless & J.H. Abbs. San Diego: College-Hill Press.

Metz, D.E., Whitehead, R.L. and Whitehead, B.H. (1984) Mechanics of vocal fold vibration and laryngeal articulatory gestures produced by hearing-impaired speakers. *Journal of Speech and Hearing Research 27*, 62-69.

Milenkovic, P. (1987) Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research 30*, 529-538.

Monsen, R.B., Engebretson, A.M. and Vemula, N.R. (1979) Some effects of deafness on the generation of voice. *Journal of the Acoustical Society of America 66*, 1680-1690.

Moore, G.P. (1976) Observations on laryngeal disease, laryngeal behavior and voice. *Annals of Otology, Rhinology, and Laryngology 85*, 553-564.

Moore, P. and Thompson, C.L. (1965) Comments on physiology of hoarseness. *Archives of Otolaryngology 81*, 97-102.

Moore, P. and von Leden, H. (1958) Dynamic variations in the vibratory pattern in the normal larynx. *Folia Phoniatrica 10*, 205-238.

Murry, T. (1980) Vocal jitter in sung and spoken vowels. In *Transcripts of the Eighth Symposium on the Care of the Professional Voice*. Eds. V. Lawrence & B. Weinberg. New York: The Voice Foundation, Vol. 1, 11-18.

Murry, T. and Large, J. (1979) Frequency perturbation in singers. In *Transcripts of the Seventh Symposium on the Care of the Professional Voice*. Ed. V. Lawrence. New York: The Voice Foundation, Vol. 1, 36-39.

Rabiner, L.R., Cheng, M.J., Rosenberg, A.E. and McGonegal, C.A. (1976) A comparative performance study of several pitch detection algorithms. *IEEE Trans. ASSP-24*, 399-418.

Ramig, L.A. and Ringel, R.L. (1983) Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech and Hearing Research 26*, 22-30.

Rozsypal, A.J. (1976) Digital gating of speech signals. *Language and Speech 19*, 57-74.

Rozsypal, A.J. and Millar, B.F. (1979) Perception of jitter and shimmer in synthetic vowels. *Journal of Phonetics 7*, 343-355.

Ryan, W.J. and Burke, K.W. (1974) Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders 7*, 181-192.

Scherer, K.R. (1986) Vocal affect expression: A review and a model for future research. *Psychological Bulletin 99*, 143-165.

Scripture, E.W. (1906) *Researches in Experimental Phonetics: The Study of Speech Curves*. Washington, D.C.: The Carnegie Foundation.

Shoup, J.E. and Pfeifer, L.L. (1976) Acoustic characteristics of speech sounds. In *Contemporary Issues in Experimental Phonetics*. Ed. N.J. Lass. New York: Academic Press, 171-224.

Simon, C. (1927) The variability of consecutive wavelengths in vocal and instrumental sounds. *Psychological Monographs 36*, 41-83.

Smith, W.R. and Lieberman, P. (1969) Computer diagnosis of laryngeal lesion. *Computers and Biomedical Research 2*, 291-303.

Sorensen, D. and Horii, Y. (1983) Frequency and amplitude perturbation in the voices of female speakers. *Journal of Communication Disorders 16*, 57-61.

Sorensen, D. & Horii, Y. (1984) Directional perturbation factors for jitter and shimmer. *Journal of Communication Disorders 17*, 143-151.

Sorensen, D., Horii, Y. and Leonard, R. (1980) Effects of laryngeal topical anaesthesia on voice fundamental frequency perturbation. *Journal of Speech and Hearing Research 23*, 274-283.

Stevens, K. and House, A. (1963) Perturbation of vowel articulations by consonant context: An acoustical study. *Journal of Speech and Hearing Research 6*, 111-128.

Takahashi, H. and Koike, Y. (1976) Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Oto-Laryngologica*, Supp. *338*, 3-24.

Titze, I.R., Horii, Y. and Scherer, R.C. (1987) Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research 30*, 252-260.

Von Leden, H. and Koike, Y. (1970) Detection of laryngeal disease by computer technique. *Archives of Otolaryngology 91*, 3-10.

Ward, P.H., Sanders, J.W., Goldman, R. and Moore, G.P. (1969) Diplophonia. *Annals of Otology, Rhinology, and Laryngology 78*, 771-777.

Wilcox, K.A. and Horii, Y. (1980) Age and changes in vocal jitter. *Journal of Gerontology 35*, 194-198.

Winer, B.J. (1971) *Statistical Principles in Experimental Design*. Second Edition. New York: McGraw Hill.

Wyke, B. (1969) Deus ex machine vocis--An analysis of the laryngeal reflex mechanisms of speech. *British Journal of Disorders of Communication 4*, 3-23.

Wyke, B. (1983) Neuromuscular control systems in voice production. In *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. Eds. D.M. Bless & J.H. Abbs. San Diego: College-Hill Press.

Zyski, B.J., Bull, G.L., McDonald, W.E. and Johns, M.E. (1984) Perturbation analysis of normal and pathologic larynges. *Folia Phoniatrica 36*, 190-198.

Appendix A: RANDOM Program

```
{
  Name:       RANDOM
  Authors:    J.L. Adlington and A.J. Rozsypal
  Date:       September 1986
  Purpose:    This program randomizes a list of "NumbStim" stimuli
              into "NumbSeqc" sequences of "NumbStim" stimuli.
              Replication of the same stimulus in a single sequence
              is not permitted.
              The program is not sophisticated enough to randomize
              stimuli for "NumbSeqc" approaching "NumbStim".
              The random number seed is generated from Turbo Pascal.
  Input:      Random number seed
  Output:     Printed list of randomized stimuli
  Compiler:   Turbo Pascal 3.0
}

program RANDOM;

label    1;

const
   NumbStim                 = 9;
   NumbSeqc                 = 2;
   NumbRepl                 = 2;
   StimKwd                  : array[1..NumbStim] of String[5]
    = ( 'heed','hid','head','had','heard','Hud','who''d','hood','hod');
   StimCode                 : array[1..NumbStim] of String[2]
    = (  'IY',  'IH', 'EH',  'AE',  'ER',  'AH',   'UW',   'UH', 'AW');

var
  TempStim                  : integer;
  IndxStim,IndxSeqc         : integer;
  IndxRepl                  : integer;
  IndxCheck                 : integer;
  CheckRepet                : integer;
  KwdA,KwdB                 : integer;
  SentNumb                  : integer;

  SubjName                  : string[63];
  SubjNumb                  : string[63];
  SubjAge                   : string[63];
  SubjSex                   : string[63];
  SubjLang                  : string[63];
  SubjLing                  : string[63];
  SubjRes                   : string[63];
  SubjMusic                 : string[63];
  SubjSmok                  : string[63];
  SubjHear                  : string[63];
  OperName                  : string[63];
  RecorDate                 : string[63];
  RecorNote                 : string[63];
```

```
  StimList                  : array[1..NumbStim,1..NumbSeqc,
                                    1..NumbRepl] of integer;
  CheckStim                 : array[1..NumbStim] of integer;

  KybdResp                  : char;

{ Procedure to randomize one stimulus sequence                    }

Procedure RandSeqc;
  begin
    for IndxStim := 1 to NumbStim do CheckStim[IndxStim] := 1;
    if IndxSeqc=1
      then for IndxStim := 1 to NumbStim do
        begin
          repeat TempStim := Random(NumbStim)+1;
          until CheckStim[TempStim]=1;
          StimList[IndxStim,IndxSeqc,IndxRepl] := TempStim;
          CheckStim[TempStim] := 0
        end

        else for IndxStim := 1 to NumbStim do
          begin
            repeat TempStim := Random(NumbStim)+1;
            until CheckStim[TempStim]=1;
            for IndxCheck:=1 to IndxSeqc-1 do
              begin
                if TempStim=StimList[IndxStim,IndxCheck,IndxRepl]
                then
                  begin
                    CheckRepet:=1; exit end;
                    StimList[IndxStim,IndxSeqc,IndxRepl] := TempStim;
                    CheckStim[TempStim] := 0
              end;
          end;
   end;    { of procedure RandSeqc }

begin  { program Random }

  Randomize;              { set random number seed from Turbo Pascal }
  ClrScr;

  {  Enter Subject Information }

  WriteLn;
  WriteLn;
  WriteLn;
  WriteLn;
  WriteLn('Experiment:  Jitter and Shimmer in English Vowels');
  WriteLn;
  WriteLn;
  WriteLn('Enter the following information about the subject');
  WriteLn;
  Write('Name:              ');
```

```
  ReadLn(SubjName);
  Write('ID Number:        ');
  ReadLn(SubjNumb);
  Write('Age:              ');
  ReadLn(SubjAge);
  Write('Sex:              ');
  ReadLn(SubjSex);
  Write('Native Language:  ');
  ReadLn(SubjLang);
  Write('Other Languages:  ');
  ReadLn(SubjLing);
  Write('Places Lived:     ');
  ReadLn(SubjRes);
  Write('Music Training:   ');
  ReadLn(SubjMusic);
  Write('Smoking:          ');
  ReadLn(SubjSmok);
  Write('Hearing:          ');
  ReadLn(SubjHear);
  WriteLn;
  WriteLn;
  Write('Operator''s Name:  ');
  ReadLn(OperName);
  Write('Recording Date:  ');
  ReadLn(RecorDate);
  Write('Recording Note:  ');
  ReadLn(RecorNote);

  { Randomize all stimulus sequences }

1:                                            { label 1 }
  WriteLn('Randomization procedure begins');
  WriteLn;
  IndxRepl := 1;
  for IndxRepl := 1 to NumbRepl do
    begin
      IndxSeqc := 1;
      repeat begin
        (****)WriteLn('Sequence  ',IndxSeqc+1);
        CheckRepet:=0;
        RandSeqc;
      end;
          if CheckRepet=0 then IndxSeqc:=IndxSeqc+1;
      until IndxSeqc=NumbSeqc+1;

{ Display randomized stimulus list and the list of stimuli   }

        WriteLn;
        for IndxStim := 1 to NumbStim do begin
        for IndxSeqc := 1 to NumbSeqc do begin
          Write(StimList[IndxStim,IndxSeqc,IndxRepl]:5);
        end;                          { write stimulus lists }
        Write(IndxStim:10,StimKwd[IndxStim]:10,StimCode[IndxStim]:10);
        WriteLn;
```

```
        end;
     end;          { of replication loop }


{ Check last word of first replication not same as first
  word of second                                                }

  if (StimList[9,1,1] = StimList[1,1,2])
  or (StimList[9,2,1] = StimList[1,2,2])
     then goto 1;

{ Print list, Randomize again, or Abort                         }

  WriteLn;
  WriteLn;
  repeat
    Write('Select:  Print, Randomize, Abort:  ');
    Read(Kbd,KybdResp);
    until(KybdResp in ['P','p','R','r','A','a']);
    WriteLn;
  case KybdResp of
      'P','p'    :    ;
      'R','r'    :    goto 1;      { to randomization procedure }
      'A','a'    :    exit;
      end;
  WriteLn;
  WriteLn;

{ Print subject's list                                          }

        WriteLn;
        WriteLn;
        WriteLn(Lst, 'SUBJECT''S LIST');
        WriteLn(Lst);
        WriteLn(Lst);
        WriteLn(Lst,SubjName:10,SubjNumb:45);
        WriteLn(Lst);
        WriteLn(Lst);
        WriteLn(Lst,'1. Please say hoad not hide');
        WriteLn(Lst);
        WriteLn(Lst,'2. Please say hayed not hoad.');
        WriteLn(Lst);
        SentNumb := 2;
        for IndxRepl := 1 to NumbRepl do begin
        for IndxStim := 1 to NumbStim do begin
          KwdA:=StimList[IndxStim,1,IndxRepl];
          KwdB:=StimList[IndxStim,2,IndxRepl];
          SentNumb:=SentNumb+1;
          WriteLn(Lst,SentNumb,'. Please say ',StimKwd[KwdA],' not ',
                                              StimKwd[KwdB],'.');
          WriteLn(Lst);          { Stimulus items in sentence frame }
        end;
        end;
        WriteLn(Lst,'21. Please say hide not hoad.');
```

```
        WriteLn(Lst);

{ Print operator's list                                         }

        WriteLn(Lst,chr(12));
        WriteLn(Lst
        WriteLn(Lst,'OPERATOR''S LIST');
        WriteLn(Lst);
        WriteLn(Lst);
        WriteLn(Lst,'Subject''s Name:   ',SubjName:10);
        WriteLn(Lst);
        WriteLn(Lst,'Number:   ',SubjNumb:7);
        WriteLn(Lst,'Age:   ',SubjAge:10);
        WriteLn(Lst,'Sex:   ',SubjSex:10);
        WriteLn(Lst);
        WriteLn(Lst,'Native Language:   ',SubjLang:10);
        WriteLn(Lst,'Other Languages:   ',SubjLing:10);
        WriteLn(Lst);
        WriteLn(Lst,'Places Lived:   ',SubjRes:10);
        WriteLn(Lst,'Musical Training:   ',SubjMusic:10);
        WriteLn(Lst,'Smoking:   ',SubjSmok:10);
        WriteLn(Lst,'Hearing:   ',SubjHear:10);
        WriteLn(Lst);
        WriteLn(Lst,'Operator''s Name:   ',OperName:10);
        WriteLn(Lst,'Recording Date:   ',RecorDate:10);
        WriteLn(Lst,'Recording Notes:   ',RecorNote:10);
        WriteLn(Lst);
        WriteLn(Lst);

{ Print stimulus list as number, keywords, and code              }

        WriteLn(Lst);
        for IndxRepl := 1 to NumbRepl do begin
        for IndxStim := 1 to NumbStim do begin
        for IndxSeqc := 1 to NumbSeqc do begin
          KwdA:=StimList[IndxStim,1,IndxRepl];
          KwdB:=StimList[IndxStim,2,IndxRepl];
          Write(Lst,StimList[IndxStim,IndxSeqc,IndxRepl]:5);
        end;
          Write(Lst,StimKwd[KwdA]:15,StimKwd[KwdB]:10);
          Write(Lst,StimCode[KwdA]:12,StimCode[KwdB]:6);
        WriteLn(Lst);
        WriteLn(Lst);
        end;
        end;

end.        { of program Random }
```

Appendix B: Sentence List From RANDOM

SUBJECT'S LIST

Jane Doe

1. Please say hoad not hide.
2. Please say hayed not hoad.
3. Please say had not hood.
4. Please say Hud not had.
5. Please say hod not who'd.
6. Please say hood not Hud.
7. Please say hid not heed.
8. Please say who'd not hod.
9. Please say heed not head.
10. Please say heard not hid.
11. Please say head not heard.
12. Please say hood not hid.
13. Please say hod not heed.
14. Please say heed not head.
15. Please say who'd not Hud.
16. Please say hid not hood.
17. Please say Hud not hod.
18. Please say had not heard.
19. Please say head not who'd.
20. Please say heard not had.
21. Please say hide not hoad.

Appendix C: JSPEAKS Program

```
{
  Name:       JSPEAKS
  Author:     J.L. Adlington
  Date:       May 1987
  Purpose:    This program takes as input integer arrays representing
              sampled waveforms.  For each peak marked by the user,
              it outputs the peak's sequential number, its estimated
              amplitude, and the estimated distance between the peak
              and the preceding peak in milliseconds.
              Signals are displayed in a 640 x 300 plane.
              This is the "maximum-peak" version of the program.
  Input:      Sampled waveform files
  Output:     PeakFiles, containing estimated period durations
              and peak amplitudes.
              One output file is produced for each speaker.
  Compiler:   Turbo Pascal 3.0
  Hardware[illegible]aphics Adaptor
}

program JS[illegible]
const
      ArraySize     = 6000;        { length of Signal Array }
      NumbFile      = 36;          { number of input files }
      DisMaxPts     = 639;         { screen display }
      BarLength     = 6;           { * 8 }
      SamplInt      = 1;
      SamplRate     = 20;          { sampling rate in KHz }
type
      ArrayType     = array[0..ArraySize] of integer;
      SegType       = array[1..NumbFile] of string[12];
      RegisterPack = record
            AX,BX,CX,DX,BP,SI,DI,DS,ES,Flags : integer
        end;
      KeyListType   = string[10];
      SubjIdType    = string[3];
var
      SigArray      : ArrayType;  { signal array }
      SigIndx       : Integer;    { index of SigArray }
      Count         : Integer;    { number of items in SigArray }
      ErrCode       : Integer;    { error codes }
      J             : Integer;
      IndxSeg       : Integer;    { index of current input file }
      SegName       : SegType;    { name of current input file }
      Rec           : RegisterPack;
      KeyList       : KeyListType;
      Subject       : SubjIdType;               { 3 characters }
      PeakFile      : Text;                      { output file }
      Xaddr         : array[0..639] of integer;
      Yaddr         : array[0..349] of integer;
      Point         : array[0..639] of integer;
      DisSig        : array[0..639] of integer;
```

```
   I            : integer;     { index counter }
   XBarBegin    : integer;     { position of bar on display }
   MemOffst     : integer;     { memory offset }
   Y1, Y2       : integer;     { Y position of bars }
   SigMax       : integer;     { max amplitude in signal }
   SigMin       : integer;     { min amplitude in signal }
   N            : integer;     { period number }
   IndxPeak1    : integer;     { index of first peak }
   IndxPeak2    : integer;     { index of second peak }
   PrecPeak     : real;        { index of peak before current peak }
   Which        : integer;     { index of current peak }
   Var80tol     : integer;     { decrements from 80 to 1 }
   ShiftBar     : integer;     { number of points to shift bar }
   ScaleBy      : integer;     { scale SigArray for display }
   Temp         : integer;     { temporary variable }
   Done         : boolean;     { finished one vowel }
   Confirm      : boolean;     { found peak }
   Invert       : char;        { inverts signal array }
   Inverted     : boolean;     { true if array inverted }
   Quit         : boolean;     { quit peak-picking }
   Reply        : char;        { read from keyboard }
   Reply2       : char;        { read from keyboard }

procedure GenSegName(var SegName : SegType;
                     var Subject : SubjIdType);
const
   NumbStim      = 9;          { number of stimuli types }
   NumbPos       = 2;          { number of positions in sentence }
   NumbRepl      = 2;          { number of replications }
   StimCode      : array[1..NumbStim] of String[2]
     = ('IY', 'IH', 'EH', 'AE', 'ER', 'AH', 'UW', 'UH', 'AW');
   PosCode       : array[1..NumbPos] of Char
                 = ('B', 'E');
   ReplCode      : array[1..NumbRepl] of Char
                 = ('1', '2');
var
   IndxStim     : Integer;     { index of stimulus loop }
   IndxPos      : Integer;     { index of position loop }
   IndxRepl     : Integer;     { index of replication loop }
   IndxSeg      : Integer;     { segment (vowel) index }
begin  { procedure GetSegName }
  ClrScr;
  WriteLn;
  WriteLn;
  Write('Enter Subject''s Sex and ID#:  ');
  ReadLn(Subject);
  WriteLn;
  IndxSeg := 1;
  for IndxStim := 1 to NumbStim do begin
     for IndxPos := 1 to NumbPos do begin
        for IndxRepl := 1 to NumbRepl do begin
           SegName[IndxSeg] := 'A:' + Subject + StimCode[IndxStim]
              + PosCode[IndxPos] + ReplCode[IndxRepl] + '.VN';
          WriteLn(SegName[IndxSeg]);
```

```
              IndxSeg := IndxSeg + 1;
           end;
        end;
     end;
end;                 {Procedure GenSegName}

procedure Loading(var SigArray : ArrayType;
   var Count, ErrCode, IndxSeg : integer;
                   var SegName : SegType);
var
      SegFile      : file of integer;         { input file }
      TooMany      : boolean;                 { arrays too small }
      Proceed      : char;                    { load segment }
begin  { procedure Loading }
     FillChar(SigArray, SizeOf(SigArray), 0);  { Initialization }
     WriteLn;
     Write('Type Y to load ' + SegName[IndxSeg] + '  ');
     Read(Kbd, Proceed);
     WriteLn;
     WriteLn;
     if (Proceed = 'Y') or (Proceed = 'y') then
       begin
         Write(SegName[IndxSeg]);
         Assign(SegFile,SegName[IndxSeg]);
{$I-}
         Reset(SegFile);
       end
     else
         begin
            Write('Enter New Segment Number; or 0 (zero) to Quit: ');
            ReadLn(IndxSeg);
            if IndxSeg = 0 then
              begin
                Close(PeakFile);
                exit;
            end;
            WriteLn;
            if (IndxSeg <= NumbFile) and (IndxSeg >= 1) then
              begin
               Write(SegName[IndxSeg]);
               Assign(SegFile,SegName[IndxSeg]);
               Reset(SegFile);
              end
            else
               begin
                 Write('Range is 1 to 36.  Please re-enter: ');
                 ReadLn(IndxSeg);
                 WriteLn;
                 Write(SegName[IndxSeg]);
                 Assign(SegFile,SegName[IndxSeg]);
                 Reset(SegFile);
               end;
         end;
     TooMany := False;
```

```
     while not eof(SegFile) do
        begin
           Count := Count + 1;
           if (Count > ArraySize) and (not TooMany) then
             begin
               TooMany := true;
               WriteLn('File', SegName[IndxSeg], ' too big.');
               WriteLn('Only 1st', ArraySize, ' loaded.');
               ErrCode := -2;
               Count := Count - 1
             end;
           if not TooMany then
             begin
               read(SegFile, SigArray[Count]);
               ErrCode := ioresult;
               if ErrCode <> 0 then
                 begin
                   WriteLn('Error during disk read.');
                   WriteLn('ErrCode = ', ErrCode, '(decimal)');
                   close(SegFile);
                   exit
                 end
           end
 {$I+}
           else
             begin
               close(SegFile);
               exit
             end
        end;
        close(SegFile);
        WriteLn('  ', Count, ' elements now in array.')
end;  { procedure Loading }

procedure FindSigMax(var SigArray: ArrayType; Count: integer;
                       var SigMax, SigMin: integer);
var
      Indx          :  integer;    { index counter }
      IndxMax       :  integer;    { index of max amp in signal }
      IndxMin       :  integer;    { index of min amp in signal }
begin  {procedure FindSigMax }
    IndxMax := 1;
    IndxMin := 1;
    SigMax := SigArray[1];
    SigMin := SigArray[1];
    for Indx := 2 to Count do
      begin
        if (SigArray[Indx] > SigMax) and (SigArray[Indx] < 3000) then
          begin                           { needed if signal inverted }
            SigMax := SigArray[Indx];
            IndxMax := Indx;
          end;
        if (SigArray[Indx] < SigMin) and (SigArray[Indx] > -3000) then
          begin
```

```
            SigMin := SigArray[Indx];
            IndxMin := Indx;
          end;
      end;
    WriteLn('Maximum point is:   ', SigMax);
    WriteLn('Index of Max point is:  ', IndxMax);
    WriteLn('Minimum point is:  ', SigMin);
    WriteLn('Index of Min point is:  ', IndxMin);
end;  { Procedure FindSigMax }

procedure FlatPeak(var Which, XBarPlace, PeakMax, SameY : integer;
                   var A0, A1, A2, Point1, Point2, Point3,
                   XPeakEst, YPeakEst : real);

    { This procedure is entered only if there are three or more
      points with the same amplitude at the peak.
      NOTE: all possible combinations of points are NOT handled,
            only those points which appeared in the data.      }
var
      MaxIndxArr    : array[1..10] of integer; { indices of max points }
      YPeakArr      : array[1..10] of real;    { y points of same amp  }
      XPosArr       : array[1..10] of real;    { x positions of points }
      Num           : integer;       { number of items in MaxType    }
      Consec        : integer;       { number of consecutive points }
      MaxConsec     : integer;       { max number of consec points   }
      Peak          : integer;       { gives choice of peak    }
      Indx          : integer;       { index counter               }
      Temp1         : integer;       { index of possible peak }
      Temp2         : integer;       { index of possible peak }
      MaxY          : real;          { highest interpolated amplitude }
 begin
   Num := 1;
   for Indx := XBarPlace+1 to XBarPlace+(BarLength*8) do
     begin
       if SigArray[Indx] = PeakMax then
         begin
           MaxIndxArr[Num] := Indx;
           WriteLn(MaxIndxArr[Num]);
           Num := Num + 1;
         end;
     end;
   MaxConsec := 1;
   Consec := 1;
   for Indx := 1 to Num-2 do
     begin
       if MaxIndxArr[Indx]+1 = MaxIndxArr[Indx+1] then
         begin
           Consec := Consec + 1;
           if MaxConsec <= Consec then MaxConsec := Consec;
          WriteLn('Consec points =  ', MaxConsec);
         end
       else
         begin
           Consec := 1;
```

```
      end;
  end;
if (MaxConsec = 1) then
  begin
    for Indx := 1 to Num - 1 do
      begin
        Which := MaxIndxArr[Indx];
        Point1 := SigArray[Which-1];
        Point2 := SigArray[Which];
        Point3 := SigArray[Which+1];
        WriteLn(SigArray[Which-1]:6, SigArray[Which]:6,
                SigArray[Which+1]:6);
        A0 := Point2;
        A1 := (Point3 - Point1)/(2. * SamplInt);
        A2 := (Point1 - (2. * Point2) + Point3)
              /(2. * (SamplInt*SamplInt));
        XPosArr[Indx] := -A1/(2. * A2);
        YPeakArr[Indx] := A0 + (A1 * XPosArr[Indx])
                          + (A2 * (XPosArr[Indx] * XPosArr[Indx]));
        WriteLn(YPeakArr[Indx]:20:12);
     end;
     MaxY := YPeakArr[1];
     XPeakEst := XPosArr[1];
     Which := MaxIndxArr[1];
     for Indx := 2 to Num - 1 do
       begin
         if YPeakArr[Indx] > MaxY then
           begin
             MaxY := YPeakArr[Indx];
             XPeakEst := XPosArr[Indx];
             Which := MaxIndxArr[Indx];
           end;
       end;
     YPeakEst := MaxY;
     WriteLn('Index of peak is  ', Which);
   end
else if (MaxConsec = 2) then
  begin
    if SameY > 3 then
      begin
        Temp1 := MaxIndxArr[1];
        WriteLn(SigArray[Temp1-1]:5, SigArray[Temp1]:5,
                SigArray[Temp1+1]:5);
        Temp2 := MaxIndxArr[3];
        WriteLn(SigArray[Temp2-1]:5, SigArray[Temp2]:5,
                SigArray[Temp2+1]:5);
      end;
    WriteLn('Enter POSITION of peak (1,2,...)');
    ReadLn(Peak);
    if SameY > 3 then
      begin
        if Peak = 3 then Peak := 5;
        if Peak = 2 then Peak := 3;
      end;
```

```
       Which := MaxIndxArr[Peak];
       Point1 := SigArray[Which-1];
       Point2 := SigArray[Which];
       Point3 := SigArray[Which+1];
       WriteLn(SigArray[Which-1]:6, SigArray[Which]:6,
               SigArray[Which+1]:6);
       WriteLn('Index of peak is:  ', Which);
       A0 := Point2;
       A1 := (Point3 - Point1)/(2. * SamplInt);
       A2 := (Point1 - (2. * Point2) + Point3)
             /(2. * (SamplInt*SamplInt));
       XPeakEst := -A1/(2. * A2);
       YPeakEst := A0 + (A1 * XPeakEst) + (A2 * (XPeakEst*XPeakEst));
     end
   else
     begin
       if SameY <> MaxConsec then
         begin
           WriteLn('Enter NUMBER of peak point (1,2,3,...)');
           ReadLn(Peak);
           Peak := Peak - 1;
           Which := MaxIndxArr[Peak];
         end;
       Which := Which + trunc(MaxConsec / 2);
       XPeakEst := 0.0;
       YPeakEst := SigArray[Which];
       if (odd(MaxConsec) = false) then XPeakEst := XPeakEst - 0.5;
       WriteLn('Index of peak is:  ', Which);
     end;
end;  { procedure FlatPeak }

procedure FindPeakMax(var SigArray : ArrayType;
                      var XBarBegin, SigIndx, IndxPeak1, IndxPeak2,
                          ShiftBar, Which : integer;
                      var PrecPeak : real);
var
   PeakMax          : integer;      { peak point }
   Indx             : integer;      { index counter }
   XBarPlace        : integer;      { place of bar on signal }
   SameY            : integer;      { # of points with same amp }
   Which2           : integer;      { index of second max point }
   Position         : integer;      { allows selection of peak }
   Point1           : real;         { point before peak }
   Point2           : real;         { peak point }
   Point3           : real;         { point after peak }
   A0, A1, A2       : real;         { interpolation coefficients }
   XPeakEst         : real;         { -.5 to +.5 -- x position of peak }
   YPeakEst         : real;         { amplitude estimation }
   XPeakEst2        : real;         { x position of second max }
   YPeakEst2        : real;         { y estimation of second max }
   CurrentPeak      : real;         { index of current peak }
   PerLengthSam     : real;         { period length in samples }
   PerLengthMs      : real;         { period length in ms }
begin  { procedure FindPeakMax }
```

```
XBarPlace := XBarBegin + SigIndx - 640;
PeakMax := SigArray[XBarPlace];
Which := XBarPlace;
SameY := 1;
Which2 := 0;
for Indx := XBarPlace+1 to XBarPlace+(BarLength*8) do
  begin
    if SigArray[Indx] > PeakMax then
      begin
        PeakMax := SigArray[Indx];
        Which := Indx;
        SameY := 1;
      end
    else if SigArray[Indx] = PeakMax then
      begin
        SameY := SameY + 1;
        if SameY = 2 then Which2 := Indx;
      end;
end;
WriteLn('SameY is:  ', SameY);
WriteLn('Peak Index is:  ', Which);
WriteLn('Which2 is:  ', Which2);
if SameY < 3 then
  begin
    Point1 := SigArray[Which-1];
    Point2 := SigArray[Which];
    Point3 := SigArray[Which+1];
    WriteLn(SigArray[Which-1]:6, PeakMax:6, SigArray[Which+1]:6);
    A0 := Point2;
    A1 := (Point3 - Point1)/(2. * SamplInt);
    A2 := (Point1 - (2. * Point2) + Point3)
          /(2. * (SamplInt*SamplInt));
    XPeakEst := -A1/(2. * A2);
    YPeakEst := A0 + (A1 * XPeakEst) + (A2 * (XPeakEst*XPeakEst));
    if (SameY = 2) and (Which+1 <> Which2) then
      begin
        Position := 0;
        Point1 := SigArray[Which2-1];
        Point2 := SigArray[Which2];
        Point3 := SigArray[Which2+1];
        WriteLn(SigArray[Which2-1]:6, SigArray[Which2]:6,
                SigArray[Which2+1]:6);
        A0 := Point2;
        A1 := (Point3 - Point1)/(2. * SamplInt);
        A2 := (Point1 - (2. * Point2) + Point3)
              /(2. * (SamplInt*SamplInt));
        XPeakEst2 := -A1/(2. * A2);
        YPeakEst2 := A0 + (A1 * XPeakEst2)
                     + (A2 * (XPeakEst2*XPeakEst2));
        WriteLn('YPeakEst is: ', YPeakEst:16:12);
        WriteLn('YPeakEst2 is: ', YPeakEst2:16:12);
        if YPeakEst2 = YPeakEst then
          begin
            WriteLn('Enter POSITION of peak (1,2,...)');
```

```
                ReadLn(Position);
              end;
            if (YPeakEst2 > YPeakEst) or (Position = 2) then
              begin
                Which := Which2;
                YPeakEst := YPeakEst2;
                XPeakEst := XPeakEst2;
              end;
            WriteLn('Peak Index is: ', Which);
          end;
        end
    else FlatPeak(Which, XBarPlace, PeakMax, SameY, A0, A1, A2,
                  Point1, Point2, Point3, XPeakEst, YPeakEst);
    WriteLn(YPeakEst:16:12);
    WriteLn(XPeakEst:16:12);
    if N = 0 then
      begin
        PrecPeak := Which + XPeakEst;
        PerLengthMs := 0;
        PerLengthSam := 0;
      end
    else
      begin
        CurrentPeak := Which + XPeakEst;
        PerLengthSam := CurrentPeak - PrecPeak;
        PerLengthMs := PerLengthSam / SamplRate;
        PrecPeak := CurrentPeak;
      end;
    WriteLn;
    WriteLn(PerLengthSam:16:12);
    WriteLn(PerLengthMs:16:12);
    WriteLn(PeakFile, N:5, PerLengthMs:20:12, YPeakEst:20:12, '  ',
            SegName[IndxSeg]);
    WriteLn(Lst, N:4, SigArray[Which-1]:8, SigArray[Which]:5,
            SigArray[Which+1]:5, Which:8, PerLengthMs:16:8,
            YPeakEst:16:8, '   ', SegName[IndxSeg]);
    if N = 0 then IndxPeak1 := Which;
    if N = 1 then
      begin
         IndxPeak2 := Which;
         ShiftBar := IndxPeak2 - IndxPeak1 - round(BarLength*8 / 2);
      end;
    WriteLn;
    WriteLn('ShiftBar is: ', ShiftBar:3);
end;                  { Procedure FindPeakMax }

procedure InitGraphics;
var
  Indx                 : integer;
  CardSwitch           : byte;
  Info                 : byte;
  Mono                 : boolean;
begin  { procedure InitGraphics }
  CardSwitch:=MEM[$40:$88] AND $0F;
```

```
  Info:=MEM[$40:$87];
  Mono:=ODD((Info AND $02) SHR 1);
  if Mono then
    Rec.AX:=$000F
    else with Rec do
    case CardSwitch of
      6 : AX:=$0D;
      7 : AX:=$0E;
      8 : AX:=$10;
      9 : AX:=$10;
      else Rec.AX:=$0E;
    end; { case CardSwitch }
  INTR($10,Rec);
  for Indx:=0 to 349 do Yaddr[Indx]:=80*Indx;
  for Indx:=0 to 639 do Xaddr[Indx]:=Indx DIV 8;
  for Indx:=0 to 639 do Point[Indx]:=$80 SHR (Indx MOD 8);
end; { procedure InitGraphics }

procedure DrawHorAxis(Y1:integer);
var
  Xdrw0,MemOffst      : integer;
begin
  Xdrw0:=0;
  while Xdrw0<639 do
    begin
      MemOffst:=Xaddr[Xdrw0]+Yaddr[Y1];
      MEM[$A000:MemOffst]:=$FF OR MEM[$A000:MemOffst];
      Xdrw0:=Xdrw0+8;
    end;
end; { procedure DrawHorAxis }

procedure DrawVertLine(XPlace,YTop,YBottom:integer);
var
  Total               : integer;
begin
  while YTop<YBottom+1 do
    begin
      Total:=Xaddr[XPlace]+Yaddr[YTop];
      MEM[$A000:Total]:=Point[XPlace] OR MEM[$A000:Total];
      YTop:=YTop+4;
  end;
end; { procedure DrawVertLine }

procedure GetKey(KeyList:KeyListType; var Reply,Reply2:char);
begin
   KeyList := #27;
   Reply2 := chr(0);
   read(kbd,Reply);
   if (Reply = #27) and keypressed then
      read(kbd,Reply2);
end;  { procedure GetKey }

procedure DrawHorBar(Y1:integer);
var
```

```
  MemOffst                : integer;
  XBar                    : integer;
  Indx                    : integer;
begin
  XBar := XBarBegin;
  for Indx := 0 to BarLength do
    begin
      MemOffst:=Xaddr[XBar]+Yaddr[Y1];
      MEM[$A000:MemOffst]:=$FF;
      XBar:=XBar+8;
    end;
 end;  { procedure DrawHorBar}

procedure ClearHorBar(Y1:integer);
var
  Xdrw0,MemOffst          : integer;
  Indx                    : integer;
begin
  Xdrw0:=XBarBegin;
  for Indx := 0 to BarLength do
    begin
      MemOffst:=Xaddr[Xdrw0]+Yaddr[Y1];
      MEM[$A000:MemOffst]:=$00;
      Xdrw0:=Xdrw0+8;
    end;
end; { procedure ClearHorBar }

procedure PlotFrame(Xinit:integer);
var
  MemOffst             : integer;
begin
  for I:=0 to 639 do
  begin
    MemOffst:=Xaddr[I]+Yaddr[DisSig[I]];
    MEM[$A000:MemOffst]:=Point[I] OR MEM[$A000:MemOffst];
  end;
end; { procedure PlotFrame }

begin { program JSPeaks }
  WriteLn;
  GenSegName(SegName,Subject);
  Assign(PeakFile,Subject);
  Rewrite(PeakFile);
  IndxSeg := 1;
  while (IndxSeg>0) and (IndxSeg<NumbFile+1) do
    begin
      Count := 0;
      Loading(SigArray, Count, ErrCode, IndxSeg, SegName);
      if IndxSeg = 0 then exit;
      if ErrCode <> 0 then
        WriteLn('Unsuccessful file load. Code = ', ErrCode)
      else
        for J := 1 to 10 do WriteLn(J:4, SigArray[J]:14);
      FindSigMax(SigArray, Count, SigMax, SigMin);
```

```
Inverted := false;
WriteLn;
Write('Invert Signal (Y/N)?  ');
Read(Kbd, Invert);
WriteLn;
if (Invert = 'Y') or (Invert = 'y') then
  begin
    Inverted := true;
    for I := 0 to ArraySize do
      begin
        SigArray[I] := SigArray[I] * -1;
      end;
  end;
if Inverted = true then
  begin
    Temp := SigMax;
    SigMax := abs(SigMin);
    SigMin := Temp * -1;
  end;
WriteLn(Lst,'  Max is:', SigMax:5, '      Min is:', SigMin:5,
        '        Inverted is  ', Inverted);
ScaleBy := trunc((SigMax/150) * -1) - 1;
WriteLn;
WriteLn('ScaleBy is:  ', ScaleBy);
Delay(3000);
Y1 := round((SigMax/ScaleBy) + 160);
Y2 := round((140-Y1)/2 + Y1);
SigIndx := 0;
Which := 0;
N := -1;
IndxPeak1 := 0;
IndxPeak2 := 0;
XBarBegin := 80;
ShiftBar := 0;
Done := false;
while not Done do
  begin
    ClrScr;
    WriteLn('Signal Generation in Progress');
    if N = -1 then
      begin
        for I := 0 to 79 do
          begin
            DisSig[I] := 150;
          end;
      end
    else
      begin
        Var80tol := 80;
        for I := 0 to 79 do
          begin
            DisSig[I] := round((SigArray[Which-Var80tol]/ScaleBy)
                                +150);
            Var80tol := Var80tol-1;
```

```
      end;
    end;
  SigIndx := Which;
  for I:=80 to DisMaxPts do
    begin
      DisSig[I] := round((SigArray[SigIndx]/ScaleBy)+150);
      SigIndx := SigIndx + 1;
    end;
  InitGraphics;
  DrawHorAxis(0);
  DrawHorAxis(150);
  DrawHorAxis(300);
  DrawVertLine(79,0,300);
  DrawHorBar(Y2);
  PlotFrame(0);
  Confirm := false;
  Quit := false;
  while (not Confirm) and (not Quit) do
    begin
      DrawHorBar(Y1);
      GetKey(KeyList, Reply, Reply2);
      case Reply2 of
        #75      :  begin             { Left-Arrow }
                      ClearHorBar(Y1);
                      XBarBegin := XBarBegin - 8;
                      DrawHorBar(Y1);
                    end;
        #77      :  begin             { Right-Arrow }
                      ClearHorBar(Y1);
                      XBarBegin := XBarBegin + 8;
                      DrawHorBar(Y1);
                    end;
        #72      :  begin             { Up-Arrow }
                      N := N + 1;
                      FindPeakMax(SigArray,XBarBegin,SigIndx,
                      IndxPeak1,IndxPeak2,ShiftBar,Which,
                      PrecPeak);
                      XBarBegin := ShiftBar + 80;
                      Confirm := true;
                    end;
        #60      :  begin     { strike F2, End, any key }
                      Which := Which + 320;
                      WriteLn(XBarBegin:6,SigIndx-560:6,
                              Which-320:6);
                    end;
        #59      :  begin     { strike F1, End, any key }
                      Which := Which - 320;
                      WriteLn(XBarBegin:6,SigIndx-560:6,
                              Which+320:6);
                    end;
        #79      :  Quit := true;     { end key }
      end;  { case }
    end;
    GetKey(KeyList,Reply,Reply2);
```

```
            if Reply2 = #79 then
              begin
                WriteLn(PeakFile, -9999, 0:20, 0:20, ' ', 'End':12);
                WriteLn(Lst);
                WriteLn(Lst);
                Done := true;
              end;
           end;
        ClrScr;
        IndxSeg := IndxSeg + 1;
      end;
      Close(PeakFile);
end.  { program JSPEAKS }
```

Output Sample: M01IYB1

```
 0   212  255  255    162    0.00000000   260.37500000
 1   230  282  274    308    7.29333333   286.03333333
 2   279  294  273    454    7.27750000   294.12500000
 3   280  299  260    601    7.34554597   299.86206896
 4   243  266  253    749    7.41556513   266.34722222
 5   240  252  231    897    7.38623737   252.30681818
 6   236  260  228   1046    7.45324675   260.14285714
 7   250  255  210   1195    7.43357142   259.00000000
 8   262  263  225   1344    7.44628205   267.38782051
 9   272  289  248   1493    7.46337312   290.24137931
10   221  286  285   1642    7.48458725   293.75757575
11   259  304  272   1793    7.52997835   304.27435064
12   237  296  274   1945    7.60719897   298.11265432
13   264  277  221   2100    7.72300054   280.34963768
14   246  247  205   2256    7.79174250   251.88662791
15   231  258  222   2413    7.87026578   258.16071429
16   173  235  234   2571    7.92777778   242.38293651
17   202  232  211   2746    8.73020542   232.19852941
18   220  223  199   2902    7.77614379   225.04166667
19   206  223  212   3066    8.22480159   223.16071429
20   200  222  215   3232    8.30757389   222.96982759
21   211  212  184   3401    8.41379310   215.14224138
22   206  218  197   3569    8.41645768   218.30681818
23   220  231  204   3739    8.49629187   231.84210526
24   221  237  213   3911    8.60552632   237.20000000
25   221  237  230   4084    8.66478261   237.44021739
26   226  242  230   4256    8.59378882   242.07142857
-9999                                 0              0
```

Sample output from JSPEAKS. For each period, the program lists the period number, the three peak sample points, the array index of the maximum point, the interpolated duration of the period, and the interpolated peak amplitude.

Appendix D: JSEXTR Program

```
{
 Name:       JSEXTR
 Author:     J.L. Adlington
 Date:       May 1987
 Purpose:    To compute mean jitter and shimmer in a vowel segment.
             To plot the signed jitter and shimmer values for each
             period in the vowel.
             To cross-correlate signed jitter with signed shimmer
             for each period, and to autocorrelate signed jitter
             and signed shimmer.
             Include Files:  JSPLOT.PAS (graphics)
                             JSTRBGFX.PAS (graphics)
 Input:      Shimmer and Maximum-Peak Jitter File
             Characteristic-Peak Jitter File
 Output:     Mean Jitter and Shimmer File
             Cross-Correlations File
             Autocorrelations File
 Compiler:   Turbo Pascal 3.0
 Hardware:   EGA Graphics Adapter
}

program JSEXTR;
const
      ArraySize     = 80;          { length of input arrays + 12 }
      NumbSeg       = 36;          { number of vowel segments    }
      StringSizeGlb = 80;
      CharFile      : string[StringSizeGlb] = '4x6.fon';
      MaxProcsGlb   = 27;
      MaxErrsGlb    = 7;
type
      RegisterPack  = record
                        AX,BX,CX,DX,BP,SI,DI,DS,ES,Flags : integer
                      end;
      IntArrType    = array[0..ArraySize] of integer;
      RealArrType   = array[0..ArraySize] of real;
      AutoArrType   = array[-ArraySize..ArraySize] of real;
      SegType       = array[0..ArraySize] of string[9];
      WrkString     = string[StringSizeGlb];
      CharArray     = array[32..126] of char;
var
      Rec           : RegisterPack;
      Xaddr,Point   : array[0..639] of integer;
      Yaddr         : array[0..349] of integer;
      ErrorProc     : array[0..MaxProcsGlb] of ↑WrkString;
      ErrorCode     : array[0..MaxErrsGlb] of ↑WrkString;
      XTextGlb,
        YTextGlb    : integer;
      MessageGlb,
        BrkGlb      : boolean;
      DrawFlag,
        FillFlag    : boolean;
```

```
      GrafModeGlb    : boolean;
      ColorGlb       : byte;
      ErrCodeGlb     : byte;
      PcGlb          : string[40];
      GrafBase       : integer;
      CharSet        : CharArray;
      PerNumb        : IntArrType;      { sequential period number }
      XLength        : RealArrType;     { length of period in ms }
      YPeak          : RealArrType;     { amplitude of period peak }
      DeltaJitter    : AutoArrType;     { jitter value for period }
      DeltaShimmer   : AutoArrType;     { shimmer value for period }
      SegName1       : SegType;         { segment name }
      SegName2       : SegType;         { segment name }
      PeakFile       : Text;            { input file }
      JSInFile       : String[6];       { name of input file }
      FPFile         : Text;            { one peak input file }
      FPInFile       : String[6];       { name of peak input file }
      JSFile         : Text;            { jitter/shimmer output file }
      JSOutFile      : String[6];       { name of j/s output file }
      ACorrFile      : Text;            { autocorrelations output file }
      ACorrOutFile   : String[6];       { name of autocorr. output file }
      CCorrFile      : Text;            { crosscorrelation output file }
      CCorrOutFile   : String[6];       { name of cr-corr output file }
      Blanks         : String[2];
      OnePeak        : char;            { Y if segment in FPFile }
      Indx           : integer;         { index counter }
      Count          : integer;         { number of periods in arrays }
      ErrCode        : integer;         { error code number }
      J              : integer;         { index counter }
      JSMax          : real;            { max of jitter or shimmer }

procedure Load(var XLength, YPeak : RealArrType;
                     var PerNumb : IntArrType;
           var SegName1, SegName2 : SegType;
               var Count, ErrCode : integer; var OnePeak : char);
var
      TooMany        : Boolean;
      PeriodNumb     : integer;
      PerLength      : real;
begin  { procedure Load }
    FillChar(PerNumb, SizeOf(PerNumb), 0);
    FillChar(XLength, SizeOf(XLength), 0);
    FillChar(YPeak, SizeOf(YPeak), 0);
    WriteLn('Segment is in FPFile? (Y/N)');
    Read(OnePeak);
    TooMany := false;
    while not eof(PeakFile) or not eof(FPFile) do
      begin
        Count := Count + 1;
        if (Count > ArraySize) and (not TooMany) then
          begin
            TooMany := true;
            WriteLn('File', JSInFile, ' too big.');
            WriteLn('Only first', ArraySize, ' records loaded.');
```

```
              ErrCode := -2;
              Count := Count - 1;
            end;
          if not TooMany then
            begin
              if (OnePeak = 'Y') or (OnePeak = 'y') then
                begin
                  ReadLn(FPFile, PeriodNumb, XLength[Count],
                         Blanks, SegName2[Count]);
                  ReadLn(PeakFile, PerNumb[Count], PerLength,
                         YPeak[Count], Blanks, SegName1[Count]);
                end
              else
                ReadLn(PeakFile, PerNumb[Count], XLength[Count],
                       YPeak[Count], Blanks, SegName1[Count]);
              ErrCode := ioresult;
              if ErrCode <> 0 then
                begin
                  WriteLn('Error during disk read.');
                  WriteLn('ErrCode = ', ErrCode, '(decimal)');
                  Close(PeakFile);
                  Close(FPFile);
                  exit;
                end
              else if PerNumb[Count] = -9999 then exit
            end
    {$I+}
          else
            begin
              Close(PeakFile);
              Close(FPFile);
              exit;
            end;
          end;
end;                    { procedure Load }

{$I JSPLOT.PAS   }     { include files with graphics procedures }
{$I JSTRBGFX.PAS}

procedure JSPlots(var DeltaJitter, DeltaShimmer : AutoArrType;
                  var JSMax : real);
const
      MaxNumbPeriods      = 100;              { for display }
      YofXAxis            = 191;
      XofYAxis            =  54;
      ScaleStep           =  96;
var
      I,Xpos              : integer;
      YposJitter,
        YposShimmer       : integer;
      DisplStepX          : integer;
      Response            : char;
      Temp                : integer;
      ScaleBy             : integer;
```

```
      YLabel             : real;
begin {procedure JSPlots}
    DrawFlag:=true;
    FillFlag:=true;
    ScaleBy:=1;
    if ((ScaleBy+1) * ScaleStep * JSMax < 100) then
      begin
        repeat
          ScaleBy := ScaleBy + 1;
        until ScaleBy*ScaleStep*JSMax > 100;
      end;
    YLabel := 1 / ScaleBy;
    ScaleBy := ScaleBy * ScaleStep;
    WriteLn('ScaleBy is:  ', ScaleBy:6);
    Delay(1000);
    InitGraphic;
    GraphicsMode;
    if Count>MaxNumbPeriods then
      begin
        Count:=MaxNumbPeriods;
        GotoXY(2,43); Write('ONLY FIRST',MaxNumbPeriods:4,
                            ' PERIODS DISPLAYED');
      end;
    GotoXY(2,1);  Write('SIGNAL:');
    GotoXY(10,1); Write(SegName1[1]);
    GotoXY(32,1); Write('JITTER & SHIMMER EXTRACTION');
    GotoXY(62,43);  Write('JITTER');
    GotoXY(72,43);  Write('SHIMMER');
    PlotSqreMark(474,342);
    PlotCrclMark(554,342);
    GotoXY(68,1); Write('DELTA VALUES');
    DrawFrame;
    DrawHorizDotLine(8,631,YofXAxis);
    DisplStepX:=585 div Count;
    if DisplStepX>28 then DisplStepX:=28;
    for I:=1 to 1 do
      begin
        Xpos:=XofYAxis+10+((I-2)*DisplStepX);
        GotoXY(2,24); Write('  0.0-');
        GotoXY(2,12); Write(YLabel:5:3,'-');
        DrawVertDotLine(XofYAxis,50,300);
      end;
    for I:=3 to Count-1 do
      begin
        Xpos:=XofYAxis+10+((I-2)*DisplStepX);
        YposJitter:=round(YofXAxis+ScaleBy*(-1)*DeltaJitter[I]);
        PlotSqreMark(Xpos,YposJitter);
        YposShimmer:=round(YofXAxis+ScaleBy*(-1)*DeltaShimmer[I]);
        PlotCrclMark(Xpos,YposShimmer);
        if YposShimmer>YposJitter then
          begin
            Temp:=YposJitter;
            YposJitter:=YposShimmer;
            YposShimmer:=Temp;
```

```
          end;
        DrawVertDotLine(Xpos+2,YposShimmer,YposJitter);
      end;
    GotoXY(2,43);
    Write('SIGNAL CONTAINS',Count:4,' PERIODS           ');
    Delay(2000);
    GotoXY(2,43); Write('DO YOU REQUEST A PRINTED COPY?   [Y/N]');
    Read(Kbd,Response);
    if ((Response='Y') or (Response='y')) then
      begin
        GotoXY(2,43); Write('SIGNAL CONTAINS',Count:4,
                            ' PERIODS           ');
        HardCopy(1);
      end;
    LeaveGraphic;
end;  { procedure JSPlots }

procedure AutoCalc(var DeltaJitter, DeltaShimmer : AutoArrType);
   { This procedure autocorrelates the DeltaJitter and DeltaShimmer
     arrays, and cross-correlates DeltaJitter with DeltaShimmer for
     a specified number of lags.                                 }
var
      Lag           : integer;     { lag time for correlation }
      MaxLag        : integer;     { greatest lag value }
      I             : integer;     { index counter }
      SumJJLag      : real;        { sum of jitter * lagged jitter }
      SumSSLag      : real;        { sum of shimmer * lagged shimmer }
      SumJS         : real;        { sum of jitter * lagged shimmer }
      SumJ          : real;        { sum of DeltaJitter values }
      SumS          : real;        { sum of DeltaShimmer values }
      SumJLagged    : real;        { sum of lagged jitter }
      SumSLagged    : real;        { sum of lagged shimmer }
      SumJSqr       : real;        { sum of squared jitter }
      SumSSqr       : real;        { sum of squared shimmer }
      SumJLagSqr    : real;        { sum of squared lagged jitter }
      SumSLagSqr    : real;        { sum of squared lagged shimmer }
      Numerator     : real;
      Denominator   : real;
      R             : real;        { correlation value }
begin  {procedure AutoCalc }
             { Find number of Lags:  MaxLag will be the number of
               DeltaJitter values divided by 3, to a maximum of 12 }
    MaxLag := round((Count-3)/3);
    if MaxLag > 12 then MaxLag := 12;
    if MaxLag < 1 then MaxLag := 1;
    WriteLn('MaxLag is:  ', MaxLag);
    for Lag := 1 to MaxLag do   { Autocorrelate the DeltaJitter  }
      begin                     { values, for lags up to MaxLag  }
        SumJJLag := 0;
        SumJ := 0;
        SumJLagged := 0;
        SumJSqr := 0;
        SumJLagSqr := 0;
        for I := 1 to Count - 3 do
```

```
        begin
          SumJJLag := SumJJLag + DeltaJitter[I] * DeltaJitter[I-Lag];
          SumJ := SumJ + DeltaJitter[I+Lag];
          SumJLagged := SumJLagged + DeltaJitter[I-Lag];
          SumJSqr := SumJSqr + sqr(DeltaJitter[I+Lag]);
          SumJLagSqr := SumJLagSqr + sqr(DeltaJitter[I-Lag]);
        end;
        Numerator := SumJJLag - (SumJ * SumJLagged
                    / (Count - 3 - Lag));
        Denominator := sqrt((SumJSqr - (sqr(SumJ)
                         / (Count - 3 - Lag)))
                         * (SumJLagSqr - (sqr(SumJLagged)
                         / (Count - 3 - Lag))));
        R := Numerator / Denominator;
        WriteLn(ACorrFile, Lag:8, R:20:8, SegName1[1]:20);
    end;
  WriteLn(ACorrFile);
  for Lag := 1 to MaxLag do    { Autocorrelate the DeltaShimmer }
    begin                      { values, for lags up to MaxLag  }
      SumSSLag := 0;
      SumS := 0;
      SumSLagged := 0;
      SumSSqr := 0;
      SumSLagSqr := 0;
      for I := 1 to Count - 3 do
        begin
          SumSSLag := SumSSLag + DeltaShimmer[I]
                      * DeltaShimmer[I-Lag];
          SumS := SumS + DeltaShimmer[I+Lag];
          SumSLagged := SumSLagged + DeltaShimmer[I-Lag];
          SumSSqr := SumSSqr + sqr(DeltaShimmer[I+Lag]);
          SumSLagSqr := SumSLagSqr + sqr(DeltaShimmer[I-Lag]);
        end;
      Numerator := SumSSLag - (SumS * SumSLagged
                  / (Count - 3 - Lag));
      Denominator := sqrt((SumSSqr - (sqr(SumS)
                       / (Count - 3 - Lag)))
                       * (SumSLagSqr - (sqr(SumSLagged)
                       / (Count - 3 - Lag))));
      R := Numerator / Denominator;
      WriteLn(ACorrFile, Lag:8, R:20:8);
  end;
  WriteLn(ACorrFile);
  for Lag := -1 to 1 do        { Cross-correlate the DeltaJitter   }
    begin                      { and DeltaShimmer values, for Lags }
      SumJS := 0;              { of -1, 0, and 1                   }
      SumJ := 0;
      SumSLagged := 0;
      SumJSqr := 0;
      SumSSqr := 0;
      for I := 1 to Count - 3 do
        begin
          SumJS := SumJS + DeltaJitter[I] * DeltaShimmer[I-Lag];
          SumJ := SumJ + DeltaJitter[I+Lag];
```

```
            SumSLagged := SumSLagged + DeltaShimmer[I-Lag];
            SumJSqr := SumJSqr + sqr(DeltaJitter[I+Lag]);
            SumSSqr := SumSSqr + sqr(DeltaShimmer[I-Lag]);
          end;
        if Lag = 0 then
          begin
            Numerator := SumJS - (SumJ * SumSLagged / (Count - 3));
            Denominator := sqrt((SumJSqr - (sqr(SumJ) / (Count - 3)))
                               * (SumSSqr - (sqr(SumSLagged)
                               / (Count - 3))));
          end
        else
          begin
            Numerator := SumJS - (SumJ * SumSLagged / (Count - 4));
            Denominator := sqrt((SumJSqr - (sqr(SumJ)
                                  / (Count - 4)))
                                  * (SumSSqr - (sqr(SumSLagged)
                                  / (Count - 4))));
        end;
        R := Numerator / Denominator;
        WriteLn(CCorrFile, Lag:8, R:20:8, SegName1[1]:20);
      end;
      WriteLn(CCorrFile);
end;  { procedure AutoCalc }

procedure JSCalculation(var DeltaJitter, DeltaShimmer : AutoArrType;
                        var JSMax : real);
       { This procedure calculates average jitter and shimmer, and
         average fundamental frequency; it also finds the DeltaJitter
         and DeltaShimmer values (positive and negative deviations
         from a three-point linear trend line) which are correlated
         in the AutoCalc procedure.                                   }
var
   Period              : integer;      { index counter for periods }
   I                   : integer;      { index counter }
   SumDeltaJitter      : real;         { total of absolute DeltaJitter }
   SumDeltaShimmer     : real;         { total of absolute DeltaShimmer }
   SumPeriod           : real;         { total of period lengths }
   AvgJitter           : real;         { average jitter for segment }
   AvgShimmer          : real;         { average shimmer for segment }
   AvgFo               : real;         { average fundamental f }
   JitterMax           : real;         { maximum jitter value }
   ShimmerMax          : real;         { maximum shimmer value }
begin  { procedure JSCalculation }
    SumDeltaJitter := 0;
    SumPeriod := 0;
    JitterMax := 0;
    for Period := 3 to Count - 1 do
      begin
        DeltaJitter[Period] := (XLength[Period] - (XLength[Period-1]
                               + XLength[Period+1]) / 2 )
                               / ((XLength[Period-1] + XLength[Period]
                               + XLength[Period+1]) / 3 );
        if abs(DeltaJitter[Period]) > JitterMax then
```

```
           JitterMax := abs(DeltaJitter[Period]);
        SumPeriod := SumPeriod + XLength[Period];
        SumDeltaJitter := SumDeltaJitter + abs(DeltaJitter[Period]);
      end;
    SumPeriod := SumPeriod + XLength[2] + XLength[Count];
    AvgJitter := SumDeltaJitter / (Count - 3) * 100;
    AvgFo := 1 / (SumPeriod / (Count-1)) * 1000;
    WriteLn;
    WriteLn('Jitter is:  ', AvgJitter:16:12);
    Write(JSFile, AvgJitter:20:8);
    WriteLn('Average Fo is:  ', AvgFo:16:12);
    WriteLn('JitterMax is:  ', JitterMax:16:12);
    WriteLn;
    SumDeltaShimmer := 0;
    ShimmerMax := 0;
    for Period := 3 to Count - 1 do
      begin
        DeltaShimmer[Period] := (YPeak[Period] - (YPeak[Period-1]
                                 + YPeak[Period+1]) / 2 )
                                 / ((YPeak[Period-1] + YPeak[Period]
                                 + YPeak[Period+1]) / 3 );
        if abs(DeltaShimmer[Period]) > ShimmerMax then
           ShimmerMax := abs(DeltaShimmer[Period]);
        SumDeltaShimmer := SumDeltaShimmer + abs(DeltaShimmer[Period]);
      end;
    AvgShimmer := SumDeltaShimmer / (Count - 3) * 100;
    WriteLn;
    WriteLn('Shimmer is:  ', AvgShimmer:16:12);
    WriteLn('ShimmerMax is:  ', ShimmerMax:16:12);
    WriteLn(JSFile, AvgShimmer:16:8, AvgFo:17:8, SegName1[1]:14);
    WriteLn(Lst, AvgJitter:20:8, JitterMax:16:8, AvgFo:20:8, Count:4);
    WriteLn(Lst, AvgShimmer:20:8, ShimmerMax:16:8, SegName1[1]:20);
    WriteLn(Lst);
    if ShimmerMax > JitterMax then JSMax := ShimmerMax
      else JSMax := JitterMax;
    JSPlots(DeltaJitter, DeltaShimmer, JSMax);
    for I := 1 to Count - 3 do
      begin
        DeltaJitter[I] := DeltaJitter[I+2];
        DeltaShimmer[I] := DeltaShimmer[I+2];
      end;
    DeltaJitter[Count-1] := 0;
    DeltaJitter[Count-2] := 0;
    DeltaShimmer[Count-1] := 0;
    DeltaShimmer[Count-2] := 0;
    AutoCalc(DeltaJitter, DeltaShimmer);
end;  { procedure JSCalculation }

begin  { program JSExtr }
     WriteLn;
     WriteLn('Enter name of shimmer input file:  ');
     ReadLn(JSInFile);
     WriteLn;
     Assign(PeakFile, JSInFile);
```

```
{$I-}
    Reset(PeakFile);
    WriteLn('Enter name of jitter input file:  ');
    ReadLn(FPInFile);
    WriteLn;
    Assign(FPFile, FPInFile);
    Reset(FPFile);
    WriteLn('Enter name of jitter output file:  ');
    ReadLn(JSOutFile);
    WriteLn;
    Assign(JSFile, JSOutFile);
    Rewrite(JSFile);
    WriteLn('Enter name of autocorrelation output file:  ');
    ReadLn(ACorrOutFile);
    WriteLn;
    Assign(ACorrFile, ACorrOutFile);
    Rewrite(ACorrFile);
    WriteLn('Enter name of cross-correlation output file:  ');
    ReadLn(CCorrOutFile);
    WriteLn;
    Assign(CCorrFile, CCorrOutFile);
    Rewrite(CCorrFile);
    for Indx := 1 to NumbSeg do
      begin
        Count := 0;
        Load(XLength, YPeak, PerNumb, SegName1, SegName2,
             Count, ErrCode, OnePeak);
 {$I+}
        Count := Count - 1;
        if ErrCode <> 0 then
          WriteLn('Unsuccessful file load.  Code = ', ErrCode);
        if (SegName1[1] <> SegName2[1]) and ((OnePeak = 'y')
            or (OnePeak = 'Y')) then
          begin
            WriteLn;
            WriteLn('STOP!!!   ALARM!!!   ERROR!!!');
          end;
        WriteLn;
        WriteLn(SegName1[1]);
        WriteLn(SegName2[1]);
        WriteLn;
        WriteLn(Count, ' elements now in arrays.');
        WriteLn;
        YPeak[1] := 0;
        FillChar(DeltaJitter, SizeOf(DeltaJitter), 0);
        FillChar(DeltaShimmer, SizeOf(DeltaShimmer), 0);
        JSCalculation(DeltaJitter, DeltaShimmer, JSMax);
      end;
    Close(PeakFile);
    Close(FPFile);
    Close(JSFile);
    Close(ACorrFile);
    Close(CCorrFile);
end.  { program JSEXTR }
```

Appendix E: Outlier Token Values

		Jitter [%]	Shimmer [%]	Alternate Period Jitter [%]	Alternate Period Shimmer [%]
M02:	IYE2	16.68	16.09	1.34	4.06
	IHE2	8.27	(9.22)	.69	7.38
	EHE2	9.77	(12.24)	1.27	8.92
M03:	IHB1	5.76	16.89	1.83	11.93
	IHB2	10.86	35.69	2.42	9.03
	EHB2	5.12	(12.36)	2.59	8.95
	EHE1	12.11	34.77	2.01	8.66
	EHE2	14.61	53.45	1.49	6.19
	AEB2	5.75	(11.59)	3.35	13.36
	AEE1	17.25	52.87	2.67	16.50
	AEE2	4.00	18.68	1.87	6.11
	ERB2	5.51	(4.02)	.91	5.97
	ERE1	4.59	(9.54)	.90	6.40
	ERE2	(3.75)	15.11	1.64	7.92
	AHB2	13.14	29.42	4.90	3.81
	AHE2	11.81	36.42	4.67	13.51
	UWB2	4.45	(14.27)	1.42	5.02
	UHE2	15.16	39.47	8.30	20.75
	AWE1	8.40	22.99	5.17	16.75
	AWE2	4.84	(10.67)	1.25	11.68
M06:	ERE2	16.49	(13.33)	4.96	9.34
F08:	AEE1	17.20	84.43	1.20	9.74
F09:	IHB2	4.97	(6.43)	2.42	4.76
	EHB1	4.99	15.44	2.25	7.50
	AEB1	1[illegible]	23.32	3.72	8.49
	AEE1	5.76	(10.82)	3.00	7.67
	ERB1	4.32	(13.70)	1.92	6.21
	AHB1	4.36	(10.99)	4.41	11.11
	UWE1	5.31	(2.89)	2.21	2.48
	AWE2	10.35	15.78	7.64	9.60

Jitter, shimmer, and the corresponding measures taken from alternate periods, for tokens which are outliers for jitter or shimmer. Bracketed values are not considered outliers. The underlined values remain above the "normal" range when measured from alternate periods.