Journal of Speech and Hearing Research, Volume 30, 448–461, December 1987

A METHODOLOGICAL STUDY OF PERTURBATION AND ADDITIVE NOISE IN SYNTHETICALLY GENERATED VOICE SIGNALS

JAMES HILLENBRAND
RIT Research Corporation and Rochester Institute of Technology, NY

There is a relatively large body of research that is aimed at finding a set of acoustic measures of voice signals that can be used to: (a) aid in the detection, diagnosis, and evaluation of voice-quality disorders; (b) identify individual speakers by their voice characteristics; or (c) improve methods of voice synthesis. Three acoustic parameters that have received a relatively large share of attention, especially in the voice-disorders literature, are pitch perturbation, amplitude perturbation, and additive noise. The present study consisted of a series of simulations using a general-purpose formant synthesizer that were designed primarily to determine whether these three parameters could be measured independent of one another. Results suggested that changes in any single dimension can affect measured values of all three parameters. For example, adding noise to a voice signal resulted not only in a change in measured signal-to-noise ratio, but also in measured values of pitch and amplitude perturbation. These interactions were quite large in some cases, especially in view of the fact that the perturbation phenomena that are being measured are generally quite small. For the most part, the interactions appear to be readily explainable when the measurement techniques are viewed in relation to what is known about the acoustics of voice production.

Developments in signal-processing technology have made numerous techniques available to the voice scientist for analyzing the acoustic characteristics of human voices. Measurements of vocal characteristics have been applied to a wide range of practical problems. These include research aimed at both identifying individual speakers by their voice characteristics, and determining what acoustic characteristics of voices need to be transmitted in low bit-rate speech synthesis systems. There is also an extensive body of research directed toward the development of quantitative measures of the acoustic characteristics associated with laryngeal pathology.

The present paper focuses on measurement problems associated with three acoustic parameters that have received a relatively large share of attention in the voice-disorders literature: additive noise, pitch perturbation, and amplitude perturbation. The term additive noise is generally used to refer to the acoustic byproduct of turbulent air flow generated at the glottis during phonation. Additive noise is often represented on a decibel scale as a ratio of the amount of energy in the periodic or "harmonic" component to the amount of aperiodic energy. Pitch perturbation or "vocal jitter" refers to rapid, and generally relatively small, cycle-to-cycle variations in the fundamental period of the glottal source function. Amplitude perturbation, or "vocal shimmer" refers to analogous cycle-to-cycle variations in voice amplitude.

Although it is seldom stated explicitly, it is generally assumed that measured values of pitch perturbation, amplitude perturbation, and additive noise reflect fundamentally different sources of aperiodicity in the laryngeal vibratory pattern. It has also been assumed that these three parameters can be measured independent of one another in the voice waveform. The primary purpose of the present study was to use a series of computer simulations to determine whether pitch perturbation, amplitude perturbation, and additive noice can be measured independent of one another.

Measurement of Additive Noise

The presence of unusually large amounts of noise is thought to be associated with the perception of dysphonia. Poor signal-to-noise ratios in disordered voices are generally attributed to any one of a variety of physiological conditions (e.g., polyps, vocal nodules, tumors, vocal-cord paralysis) that result in air leakage during what ought to be the closed phase of the phonatory cycle (Haji, Horiguchi, Baer, & Gould, 1986; Kasuya, Ebihara, & Yoshida, 1984; Yanagihara, 1967; Zemlin, 1968). Detecting the presence of unusually large amounts of noise is a relatively simple matter because noise is easily visible on oscillograms, and on either wide- or narrow-band spectrograms. However, the precise quantification of noise levels has not proven to be a simple matter.

One of the first attempts to provide quantitative estimates of noise levels was a classification scheme devised by Yanagihara (1967) that was based on the visual rating of narrow-band spectrograms. Results from a 5-point rating scale correlated moderately ($r = 0.65$) with subjective hoarseness ratings derived from a panel of three trained listeners. Similar findings have been obtained using a variety of other noise measurement techniques (Deal & Emanuel, 1978; Emanuel & Sansone, 1969; Kojima, Gould, Lambaise, & Isshiki, 1980; Lively & Emanuel, 1970; Sansone & Emanuel, 1970; Yumoto, 1983; Yumoto, Gould, & Baer, 1982; Yumoto, Sasaki, & Okamura, 1984). For example, Kojima et al. developed a noise measurement technique based on Fourier analysis of sequences of pitch pulses extracted from sustained vowels. The authors reported signal-to-noise ratios rang-

0022-4685/87/3004-0448$01.00/0

ing from 15.4 to 23.4 dB (M = 19.5) for a group of 28 persons who spoke normally, and values ranging from −1.5 to 20.3 dB (M = 9.9 dB) for a group of speakers with a variety of laryngeal disorders. Kojima et al. also reported a strong correlation (r = 0.87) between signal-to-noise ratio measures and listener ratings of hoarseness severity for the disordered speakers.

A series of studies by Emanuel and his colleagues (Deal & Emanuel, 1978; Emanuel & Sansone, 1969; Lively & Emanuel, 1970; Sansone & Emanuel, 1970) used a noise measurement technique based on analysis of the output of a narrow-band wave analyzer. Estimates of noise levels correlated strongly with listener judgments of roughness severity for simulated rough voices produced by normal speakers and, in the case of the Deal and Emanuel study, for a group of speakers with laryngeal disorders. A recent series of studies by Yumoto and his colleagues (Yumoto, 1983; Yumoto et al., 1982, 1984) used a signal-averaging technique for measuring what the authors called "harmonics-to-noise ratio" (HNR), a measure of the relative amount of energy in the periodic and aperiodic components of a voice signal. As in previous studies, measurements with this technique showed large differences between normal and disordered subjects, as well as strong correlations with subjective ratings of hoarseness severity.

Several relevant pieces of information are unavailable for all of the noise-measurement procedures mentioned above: (a) to what extent do the techniques provide accurate estimates of additive noise levels; (b) to what extent are measures of noise level affected by variation in acoustic dimensions such as pitch perturbation, amplitude perturbation, and mean fundamental frequency (F_0); and (c) to what extent does the introduction of additive noise affect measured values of pitch perturbation and amplitude perturbation.

Measurement of Pitch and Amplitude Perturbation

A wide variety of methods have been used both to measure and calculate pitch and amplitude perturbation. The simplest calculation method that has been used to quantify pitch perturbation is mean jitter, which is the average absolute difference in fundamental period between adjacent pitch pulses. Because mean jitter tends to be proportional to the average fundamental period (Hollien, Michel, & Doherty, 1973; Horii, 1979; Lieberman, 1963), pitch perturbation is often represented as a percentage (mean jitter divided by the mean fundamental period, multiplied by 100). A wide variety of other methods have been used, however, including "jitter factor"—the percentage of cycle-to-cycle differences that exceed 0.5 ms (Hecker & Kruel, 1971; Lieberman, 1963); "directional jitter"—the percentage of cycle-to-cycle differences involving a change in algebraic sign (Davis, 1981; Hecker & Kruel, 1971; Murry & Doherty, 1980; Robbins, 1981); and several methods that calculate differences from a moving average (Kitajima, Tanabe, & Isshiki, 1975; Koike, 1973; Takahashi & Koike, 1975).

Mean shimmer is defined as the average absolute amplitude difference in decibels between adjacent pitch pulses. In most studies, measurements of either peak or peak-to-peak amplitude have been used, but in a few cases cycle-to-cycle differences in rms intensity have been calculated (Kempster, 1984; Kempster & Kistler, 1984; Robbins, 1981). As with pitch perturbation, a variety of other calculation methods have been used including directional shimmer (Robbins, 1981) and methods involving amplitude differences from a moving average (Davis, 1981; Kitajima & Gould, 1976; Takahashi & Koike, 1975).

Measurements of both pitch and amplitude perturbation require some method for locating the boundaries of individual pitch pulses. The signals that have been used for these measurements have been derived from standard air microphones, throat contact microphones, accelerometers, pneumotachometers, electroglottographs, and high-speed films. Measurement methods have included hand-marking of analog oscillograms, semi-automatic methods using interactive digital waveform editors, and both hardware and software automatic pitch trackers.

Methodological Issues

Despite the great diversity in methods that have been used for data collection, data analysis, and calculation of perturbation, relatively little work has focused on methodological issues. One exception is the work of Heiberger and Horii (1982), who noted that the different analysis techniques used by various investigators have involved substantial variability in temporal resolution. The effects of this variability in temporal resolution were tested by digitizing voice signals at sample frequencies between 5 and 80 kHz. Results showed that jitter measurements were very strongly affected by time resolution, especially for sample frequencies below 20 kHz. A recent study by Titze, Scherer, and Horii (1987) addressed several methodological issues related to the measurement of pitch and and amplitude perturbation. Among their findings were: (a) the sensitivity of perturbation measurements to variations in time and amplitude resolution can be greatly reduced through the use of very simple interpolation techniques, (b) perturbation measurements based on zero crossings are not affected by low-pass filtering, (c) tape hiss and tape-speed variations introduced by analog tape recorders can inflate measures of pitch and amplitude perturbation, and (d) several tokens from a given speaker are needed to obtain stable perturbation measurements.

A wide variety of other methodological issues related to the measurement of pitch and amplitude perturbation have yet to be addressed. The primary purpose of the present study was to use measurements of synthetically generated voice signals to determine whether pitch perturbation, amplitude perturbation, and additive noise could be measured independently of one another.

SIMULATION 1: ACCURACY IN MEASURING HNR

The first simulation was designed to test the accuracy of the Yumoto et al. (1982, 1984) technique in measuring HNR for highly regular signals showing no pitch or amplitude perturbation. Yumoto et al. demonstrated that HNR measurements obtained with their procedure correlate with noise measurements estimated by the subjective rating of narrow-band spectrograms. The tests described below were designed to provide a more direct test of the precision of the Yumoto et al. technique.

METHODS

Analysis Software

A computer program called AVR (Hillenbrand, Biggam, & Wilde, 1984) was developed to measure HNR, pitch perturbation, amplitude perturbation, and mean fundamental frequency. The noise-measurement algorithm was an implementation of the signal-averaging technique described by Yumoto et al. (1982). The input to the program is a sustained vowel containing boundary markers to indicate the beginning of each pitch pulse. The pitch markers are entered using a semiautomatic method based on zero crossings. Successive 100 ms segments of the time-domain waveform are displayed on a high-resolution graphics terminal (Tektronix 4010) using a general-purpose waveform editor (Prall & Hillenbrand, 1981). For each pitch pulse, the user aligns a cursor until it is near the zero crossing preceding the first major peak of the pitch period. The program then makes a more precise determination of the zero crossing location by searching the waveform data points for a change in algebraic sign.[1]

The first step in the calculation of HNR is to average the individual pitch pulses. As in the Yumoto et al. technique, the size of the averaging window is determined by the longest pitch pulse in the signal. For periods that are shorter than this maximum, the interval between the end of the pitch pulse and the end of the averaging window is filled with zeros. If a sufficient number of periods are averaged, a large proportion of the noise is canceled. The rms energy of the average pitch pulse is used as the numerator in the HNR calculation. The amount of aperiodic energy is estimated by successive subtractions of the average pitch pulse from individual periods of the original vowel. The rms energy in the noise signal is used as the denominator in the HNR calculation. Represented on a decibel scale, HNR is defined as:

$$20 \log \frac{\text{rms (average)}}{\text{rms (noise)}}$$

AVR also uses the pitch boundary markers to calculate: (a) mean and standard deviation F_o, (b) mean and standard deviation pitch-pulse intensity, (c) mean jitter, (d) percent jitter, and (e) mean shimmer.

STIMULI

Test signals consisted of synthesized 5-formant vowels that were added point-for-point with synthesized formant-shaped noise. The periodic and aperiodic components were synthesized separately with an implementation of Klatt's (1980) formant synthesis program, with a 20 kHz sample frequency and 12 bits of amplitude resolution. Formant frequencies were set appropriate to [a]: F1 = 720, F2 = 1240, F3 = 2400, F4 = 3300, F5 = 3700 Hz. For the periodic signals, F_o was held constant at either 100, 130, 175, or 200 Hz. The noise signals were synthesized by passing the aspiration source through the same formant resonators that were used to generate the periodic signals, but with the amplitude of the voice source set to zero. The noise signals were then scaled appropriately and added point-for-point with the periodic signals to achieve HNRs varying in 3 dB steps from −22 dB to 32 dB. (For comparison, Yumoto et al., 1982, reported HNR values ranging from 7.0 to 17.0 dB for a group of normal talkers, and from −15.2 to 9.6 dB for a group of speakers with a variety of laryngeal disorders.) All test signals were 300 pitch periods in duration (1.5 to 3.0 s, depending on F_o). Shown in the Appendix are Fourier spectra of four representative stimuli from this continuum, as well as stimuli from the pitch- and amplitude-perturbation continua, described in later sections of the paper.

RESULTS AND DISCUSSION

Measurements with Automatic Pitch Marking

For the initial tests, pitch markers were entered automatically by the synthesis program, which was modified to enter a boundary marker each time a glottal impulse was generated. This method allowed a direct test of the precision of the algorithm itself, without the confounding influence of errors that might be introduced in locating the precise onsets of individual pitch pulses. The effects of pitch-marking errors will be considered in a separate set of tests.

The AVR program was used to measure HNR for the 76 test signals based on an averaging of all 300 periods. Table 1 compares known HNRs with those calculated by the AVR program. (The shimmer measurements will be

---

[1] It is often the case that voice signals do not actually cross the zero line immediately prior to the first major peak of the pitch pulse. A very common pattern is for the waveform to show very low-amplitude oscillations that do not quite cross the zero line, followed by the large pulse signaling the beginning of the pitch period. To address this problem we often find it necessary to introduce a relatively small DC offset (usually about 1–2%), shifting the waveform up or down so that a zero crossing always occurs prior to the beginning of the pitch pulse. The user selects the appropriate DC offset by aligning a cross-hair cursor after inspecting the waveform.

TABLE 1. Comparison of actual versus measured HNR for synthesized signals at several fundamental frequency levels. Pitch markers were entered automatically during synthesis. Values are also given for measured shimmer. All values are in decibels.

Actual HNR	F_o = *100 Hz* *Meas. HNR*	*Shim.*	F_o = *130 Hz* *Meas. HNR*	*Shim.*	F_o = *175 Hz* *Meas. HNR*	*Shim.*	F_o = *200 Hz* *Meas. HNR*	*Shim.*
32.0	32.1	0.07	32.1	0.08	31.8	0.07	31.6	0.10
29.0	29.1	0.09	29.1	0.11	29.0	0.09	28.9	0.14
26.0	26.1	0.13	26.1	0.17	26.1	0.13	26.1	0.19
23.0	23.1	0.19	23.1	0.24	23.1	0.19	23.1	0.27
20.0	20.1	0.26	20.1	0.33	20.1	0.26	20.1	0.38
17.0	17.1	0.37	17.1	0.46	17.1	0.37	17.2	0.54
14.0	14.1	0.51	14.1	0.64	14.1	0.51	14.2	0.75
11.0	11.2	0.70	11.1	0.88	11.1	0.70	11.2	1.03
8.0	8.1	0.94	8.1	1.17	8.2	0.94	8.2	1.38
5.0	5.0	0.94	5.2	1.49	5.1	0.94	5.2	1.75
2.0	2.1	1.20	2.2	1.78	2.1	1.20	2.2	2.03
−1.0	−0.9	1.46	−0.8	1.95	−0.8	1.46	−0.7	2.16
−4.0	−4.0	1.66	−3.8	2.02	−3.9	1.66	−3.7	2.16
−7.0	−7.0	1.78	−6.7	2.01	−6.8	1.78	−6.6	2.10
−10.0	−10.0	1.82	−9.7	1.98	−9.7	1.82	−9.5	2.04
−13.0	−13.1	1.82	−12.5	1.95	−12.6	1.82	−12.3	2.02
−16.0	−16.1	1.83	−15.4	1.93	−15.5	1.83	−15.0	2.00
−19.0	−18.9	1.80	−18.0	1.91	−17.9	1.80	−17.4	2.00
−22.0	−21.7	1.78	−20.8	1.91	−20.3	1.78	−19.4	1.99

explained below.) It can be seen that the program produced HNR measurements that were generally within a small fraction of a decibel of the actual HNR. The mean absolute measurement error, averaged across all HNRs and fundamental frequencies was 0.3 dB. The program tended to be somewhat less accurate at very poor HNRs, especially at high fundamental frequencies. For positive HNRs, the mean absolute measurement error was only 0.1 dB.

The shimmer measurements[2] in Table 1 were included to illustrate a very important point about the effect of additive noise on measured values of amplitude perturbation. These values indicate that measures of amplitude perturbation tend to increase systematically with increasing amounts of additive noise. The increase in measured shimmer values can be attributed to random fluctuations in the level of the noise component. As the amount of additive noise increases, these fluctuations make increasingly large contributions to pitch-pulse amplitude.

The effect of additive noise on measures of amplitude perturbation is disturbing since it is generally assumed—implicitly if not explicitly—that shimmer measurements reflect variability in the amplitude of the glottal source. Because the test signals that were used did not contain any variability in source amplitude, the perturbation values shown in Table 1 suggest that shimmer measurements may, in fact, reflect additive noise rather than source-amplitude variability. In other words, current measurement techniques based on the analysis of output waveforms can not differentiate between additive noise and source-amplitude variability.

Measurements Without Automatic Pitch Marking

The HNR measurements reported in Table 1 represent the most favorable estimate of the precision of the noise-measurement algorithm, in part because the test signals (unlike naturally occurring voices) contained no source pitch or amplitude variability, but also in part because the pitch-boundary markers were entered automatically during synthesis. In measuring a naturally occurring voice, it is necessary to locate the onset of each pitch pulse using either a waveform editor or some type of automatic, pitch-synchronous fundamental frequency measurement technique. It can not be assumed that the precise onsets of individual pitch pulses can be located without error. Therefore, the tests described below were designed to determine what influence pitch-marking errors have on the accuracy of both HNR and perturbation measurements.

The stimuli for these tests were derived from a subset of the 130 Hz HNR series described above, with HNR varying in 3 dB steps from 32 dB to −4 dB. Tests were restricted to this range of HNRs because preliminary work suggested that pitch marking became very unreliable for signals with HNRs lower than about −4 dB. To facilitate pitch marking for noisy signals, all stimuli were digitally low-pass filtered at 500 Hz (see Figure 1). After low-pass filtering, pitch boundaries were located visually using the waveform-editor method described previously

---

[2] The shimmer values that are listed in Table 1, and throughout the paper, were calculated using cycle-to-cycle differences in rms rather than peak intensity. Both methods have been used in the literature, with the peak method being more common. The AVR program reports shimmer values using both the peak and rms methods. The choice of rms intensity was arbitrary; in all cases the pattern of results was virtually identical using both methods.

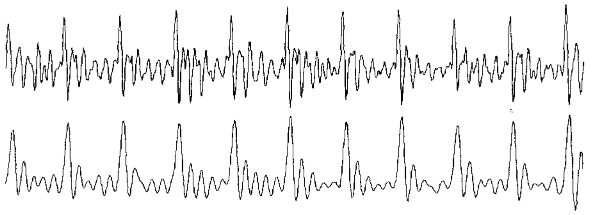

FIGURE 1. Unfiltered voice signal (top) and the same signal after digital low-pass filtering at 500 Hz (HNR = 8 dB).

under "Analysis Software." A separate program was then used to transfer the pitch-boundary markers from the low-pass filtered signals to the original waveforms. Perturbation and noise measurements were then made using the AVR program.

The measured HNR values shown in Table 2 indicate quite clearly that the technique loses accuracy when pitch-boundary markers are entered by hand. The mean HNR measurement error was 3.5 dB—more than ten times greater than the average error for the same set of signals in which pitch markers were entered automatically at the beginning of each pitch pulse. The jitter values in Table 2, which are given in both relative and absolute terms, can be taken as an indication of the degree of pitch-marking error. In terms of glottal source characteristics, none of the signals contained any variability in pitch. It can be seen that measures of pitch perturbation increase dramatically as a function of additive noise.

Setting aside for the moment the effect of pitch-marking errors on HNR measurements, these results provide yet another indication of the failure of this measurement approach to differentiate between perturbation and additive noise. The data reported in Table 1 suggested that a large shimmer value could indicate either a large degree of cycle-to-cycle variability in the amplitude of the glottal source, or a large amount of additive noise, or some combination of the two. The results shown in Table 2 suggest that a large jitter value could be due to pitch perturbation, additive noise, or some combination of the two.

TABLE 2. Comparison of actual versus measured HNR. Pitch markers were entered by hand using a semi-automatic waveform-editor method. Since the test stimuli contained no variability in the fundamental period of the glottal source function, the jitter measurements are an indication of the pitch-marking error that results when noise is added to periodic waveforms.

Actual HNR	*Measured HNR*	*HNR Meas. Error*	*Mean Jitter (μs)*	*Percent Jitter*
32.0	27.6	4.4	2.7	0.03
29.0	26.1	2.9	3.4	0.04
26.0	22.2	3.8	10.8	0.14
23.0	18.5	4.5	27.8	0.36
20.0	15.5	4.5	46.3	0.60
17.0	13.6	3.4	55.4	0.72
14.0	10.7	3.3	74.8	0.97
11.0	7.3	3.7	106.4	1.38
8.0	3.8	4.2	200.2	2.61
5.0	1.1	3.9	233.3	3.03
2.0	−1.9	3.9	266.2	3.46
−1.0	−3.5	2.5	275.3	3.57
−4.0	−5.0	1.0	355.7	4.62

It is very important to note that the source of the pitch-marking errors is primarily *the noise itself* and not the human operator. Recall that the operator of the waveform editor locates pitch boundaries by adjusting a cursor in the approximate location of the zero crossing that immediately precedes the first major peak of each pitch pulse. The program then locates the zero crossing by searching the waveform data points for a change in algebraic sign. This method produces pitch perturbation measurements that are highly repeatable (Kempster, 1984; Robbins, 1981). However, the data reported in Table 2 suggest that this method does not necessarily ensure measurement *validity*.

The effect of additive noise on pitch perturbation measures can be attributed to the random changes in zero-crossing locations that occur when noise is added to the periodic component. Assume that a signal is synthesized (or produced by a theoretically "perfect" voice) in which all glottal source pulses are exactly 10 ms in duration, identical in amplitude, and to which no noise has been added. Waveform editor measurements with automatic zero-crossing detection should show pitch boundary markers every 10 ms, within the resolution of the sample period. However, if noise is added to this signal, many of the data points which had been located at zero crossings will be shifted either upward or downward by the addition of the noise component. The larger the noise component, the larger will be the shift. Waveform editor measurements will accurately reflect these changes in zero-crossing locations, which will show up as random variability in pitch-pulse duration. It should also be noted that the same kinds of random changes will occur in the locations of other oscillographic landmarks such as voltage peaks.

Sensitivity to Pitch Measurement Error

From one point of view, the data in Table 2 are puzzling, at least at first inspection. While the degree of pitch-marking error clearly increases as HNR decreases (as reflected by increases in measured jitter values), the size of the HNR measurement error remains relatively constant at about 3.5 to 4.5 dB. For example, the number of pitch-marking errors for the 32 dB signal was very small—inspection of the signal showed that, of the 300 pitch markers in the waveform, only four were located incorrectly, and in each case the marker was off by a single sample period (50 μs). However, the size of the HNR measurement error was roughly equivalent to that of the 8 dB signal in which pitch-marking errors were both more numerous and larger in size.

One possible explanation for this pattern of results is that the signal-averaging algorithm may be more sensitive to pitch-marking errors at more favorable HNRs. To test this possibility, a program was written to introduce between 1 and 21 pitch-marking errors into the 13 test signals listed in Table 2. The pitch-marking errors were

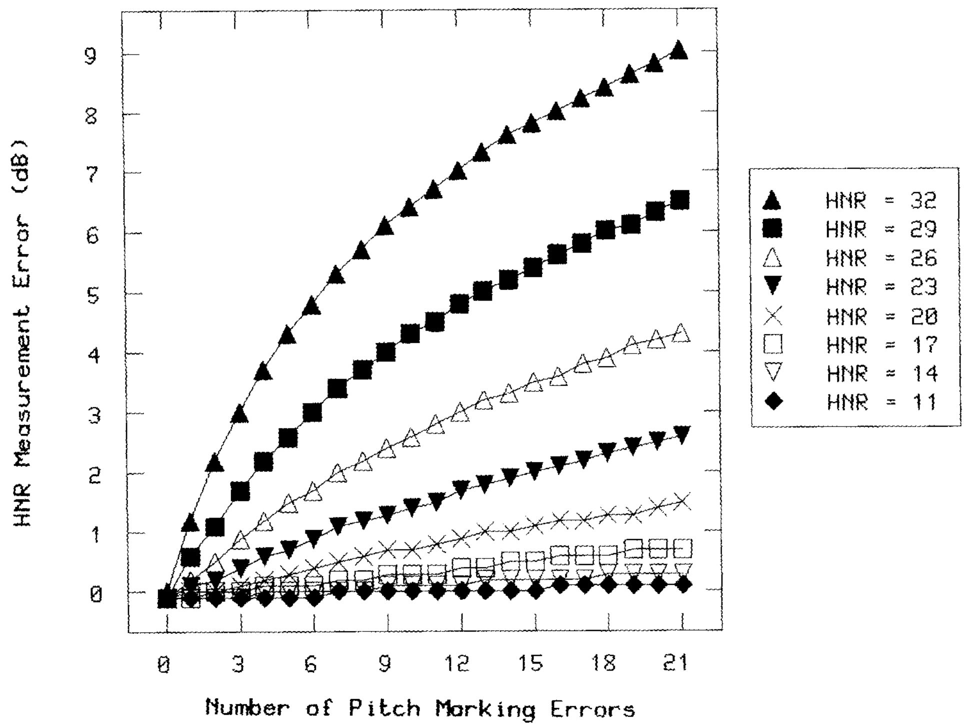

FIGURE 2. HNR measurement error as a function of the number of pitch-marking errors for signals with varying amounts of additive noise.

introduced at random locations throughout the signal and in each case consisted of moving a pitch marker one sample period to the left or right. The results of these tests are shown in Figure 2 for HNRs between 11 and 32 dB (for HNRs below 11 dB, the functions are essentially flat). The results indicate that the technique is, in fact, considerably more sensitive to pitch-marking errors at favorable HNRs. These findings suggest that the approximate constancy of HNR measurement error in Table 2 is the result of the offsetting influence of two factors. At favorable HNRs, the number and size of the pitch-marking errors are relatively small, but the algorithm is very sensitive to the errors. However, as HNR decreases, the algorithm becomes less sensitive to measurement error, but number and size of the errors tend to increase. These two influences apparently offset one another to produce an approximately constant HNR measurement error of 3.5 to 4.5 dB.

SIMULATION 2: EFFECTS OF PITCH PERTURBATION

In terms of glottal source characteristics, all of the signals used in the series of simulations reported above were perfectly periodic. The purpose of the next set of tests was to determine the effects of pitch perturbation on measures of harmonics-to-noise ratio and amplitude perturbation.

STIMULI

As in Experiment 1, the stimuli were generated using an implementation of the Klatt (1980) formant synthesis program. The basic approach was to introduce specific amounts of variability in the column of numbers that control fundamental frequency, which are updated every 5 ms in the Klatt program. An initial sequence of 200 fundamental frequency values was derived by measuring a sustained [a] vowel produced by a normal adult male using the waveform editor method described previously. Measurements revealed a mean fundamental frequency of 122 Hz, with a standard deviation of 0.89 Hz. A constant was then added to each fundamental frequency value to produce a distribution with a mean of 130 Hz. This sequence of numbers served as the input to another program which retained the 130-Hz mean but altered the standard deviation. (The change in standard deviation was accomplished simply by increasing or decreasing the deviation of each value from the mean by some constant proportion.) This program was used to generate 22 number sequences with means of 130 Hz and standard deviations ranging from 0 to 8 Hz. These number sequences were then used to control fundamental frequency in the Klatt synthesis program. The variations in the standard deviation of the fundamental frequency control column produced variations in jitter ranging from 0 to 500 μs in absolute terms, or from 0 to 6.4% in relative terms. (See Appendix for examples of Fourier spectra of stimuli from the pitch perturbation continuum.)

RESULTS AND DISCUSSION

After synthesis, all signals were scaled to maximum amplitude and analyzed by the AVR program. The results

TABLE 3. Effects of pitch perturbation on measures of HNR and shimmer. Jitter values are in μs, HNR and shimmer values are in dB. Values are given for calculations based on pitch markers entered automatically during synthesis, and for pitch markers entered manually using a waveform editor. The third column of HNR values is for manually measured signals that were corrected for the amplitude variability that is introduced as a side effect of pitch perturbation.

Jitter (auto)	*Jitter (manual)*	*HNR (auto)*	*HNR (manual)*	*HNR (corrected)*	*Shimmer (auto)*	*Shimmer (manual)*
0.0	0.0	55.3	54.7	55.2	0.00	0.00
2.3	2.3	40.1	27.3	45.1	0.02	0.02
19.2	20.0	35.7	24.8	37.8	0.08	0.10
39.2	41.5	31.7	21.3	32.7	0.13	0.14
46.9	49.2	32.8	21.4	33.0	0.12	0.12
90.0	134.6	26.4	13.6	27.3	0.28	0.30
92.3	146.2	26.0	13.4	27.0	0.29	0.37
125.4	146.1	24.4	13.2	26.0	0.39	0.39
141.5	163.0	23.5	12.8	25.1	0.42	0.42
170.8	215.4	25.0	12.5	26.1	0.33	0.48
185.4	234.6	22.1	12.3	24.1	0.55	0.56
217.7	274.6	21.4	12.3	23.9	0.64	0.60
251.5	282.3	20.9	13.5	22.8	0.62	0.63
281.5	316.2	19.9	12.1	21.2	0.72	0.70
294.6	370.8	20.0	10.2	22.2	0.71	0.74
335.4	393.9	19.3	10.5	22.2	0.82	0.80
360.8	418.5	19.0	11.7	22.1	0.81	0.83
403.1	446.2	18.8	10.9	21.6	0.81	0.84
420.0	457.7	18.4	10.7	21.1	0.83	0.84
455.5	473.1	18.1	09.7	20.8	0.89	0.82
495.4	484.6	18.2	09.3	20.2	0.80	0.79

of HNR, jitter and shimmer measurements of these stimuli are shown in Table 3. Values are given for measurements based on pitch markers that were entered automatically during synthesis, and for pitch markers that were entered using the waveform editor method. It can be seen that measures of HNR are strongly affected by variations in pitch perturbation. Data from the very low jitter values are probably misleading since naturally produced voices do not achieve these very high levels of periodicity. However, HNR values for the stimuli with jitter values above about 40 μs (values that are likely to be found in natural speech) vary over a range greater than 10 dB both for the stimuli with automatically located pitch boundaries, and for the stimuli that were marked manually.

The shimmer values reported in Table 3 are also of interest. It can be seen that shimmer measurements are also strongly affected by changes in pitch perturbation. From a measurement point of view the changes in measured shimmer values are troublesome since the test stimuli showed no variability in source amplitude. It is important to note that the increase in shimmer values can not be attributed to errors in locating the onsets of pitch pulses since the effect is seen on both the manually marked stimuli and stimuli for which pitch markers were entered automatically during synthesis.

The increase in measured shimmer values as a side effect of increasing pitch perturbation probably occurs for at least two reasons. First, the intensity of a given pitch pulse in the output waveform is determined not only by the intensity of the glottal source waveform, but also by the relationship between harmonics of the glottal source and the location of resonances in vocal-tract transfer function. These relationships will become more variable as increasing amounts of pitch perturbation are introduced, resulting in greater cycle-to-cycle variability in output intensity (see Fant, 1968, & House, 1960, for similar comments on related measurement problems.) The second, and probably more important reason, has to do with the overlap in energy that occurs between adjacent pitch pulses. When a train of glottal pulses is generated, it is generally the case that a given glottal pulse will be generated before the previous pulse is fully damped (Chandra & Lin, 1974; Fant, 1968; Makhoul & Wolf, 1972). For this reason, energy from the "tail" of a given pitch pulse will overlap with energy from the beginning of the next pitch pulse. The degree of overlap will be determined in part by the fundamental period, being greater for shorter fundamental periods. Therefore, the energy in a given pitch pulse in the output waveform will be determined in part by the intensity of the glottal source waveform, in part by the relationship between harmonics of the glottal source and the locations of vocal-tract resonances, and in part by the degree of overlap between that pitch pulse and the previous pitch pulse. The degree of overlap between adjacent pitch pulses will be more variable as pitch perturbation increases, resulting in greater cycle-to-cycle variability in the intensity of adjacent pitch pulses in the output waveform.

It is important to note that the summing effect from previous pitch periods will be more significant at high fundamental frequencies—the shorter the fundamental period, the greater will be the amplitude at the end of the pitch pulse, and therefore more energy will be added to the next pitch pulse. One would therefore expect a stronger measurement interaction between pitch pertur-

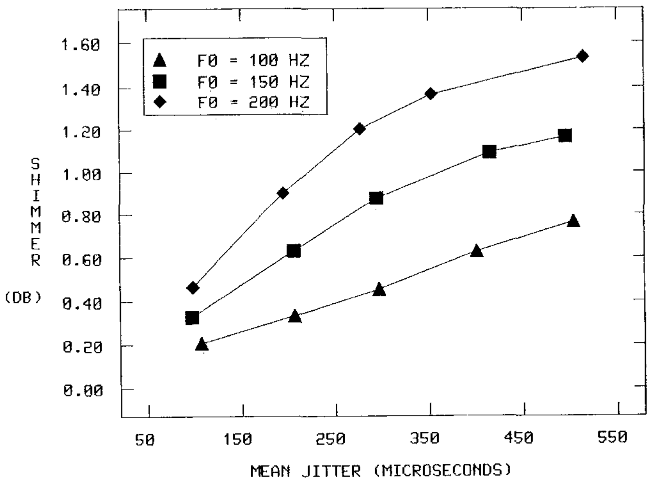

FIGURE 3. Measured shimmer values as a function of pitch perturbation for stimuli at three mean fundamental frequency levels.

bation and amplitude perturbation for voices with high fundamental frequencies. This effect is illustrated in Figure 3, which shows measured shimmer values for jitter stimuli produced with constant source amplitude at fundamental frequencies of 100, 150, and 200 Hz. As the figure shows, the "shimmer artifact" that occurs when pitch perturbation is introduced into a signal is larger at higher fundamental frequencies. Although not shown in Figure 3, it was also the case that HNR values for stimuli with a given jitter value were consistently lower for stimuli with higher fundamental frequencies.

Regardless of the precise reason for the increase in measured shimmer values as a function of pitch perturbation, this effect makes it more difficult to interpret the changes in HNR values shown in Table 3. It is possible that the decrease in HNR values is due to increases in pitch perturbation, or to increases in amplitude perturbation, or to some combination of the two. An attempt was made to separate these effects by removing amplitude variability from the 22 test stimuli. A program was written that measured the intensity of individual pitch pulses and scaled all pitch pulses to the same rms value. HNR measurements from these stimuli are shown in Table 3, and are compared with the uncorrected stimuli in Figure 4. It can be seen that in every case the corrected stimulus has a higher HNR value than the corresponding uncorrected stimulus. However, even after correction, stimuli with larger jitter values tend to have lower HNR values. These results suggest that jitter does, in fact, play an independent role in influencing HNR measurements, but that shimmer also plays a role in decreasing HNR values. The effects of amplitude perturbation will be studied in more detail in a separate set of simulations.

It can also be seen in Table 3 that there are very large differences in HNR, jitter, and shimmer measurements between stimuli for which pitch markers were entered automatically during synthesis versus those that were entered manually. For nearly all comparisons, measures of pitch and amplitude perturbation are higher, and measures of HNR are lower for the manually marked stimuli. These differences are particularly large for the HNR measurements, where the average difference between hand-measured and automatically-measured stimuli was 9.5 dB.

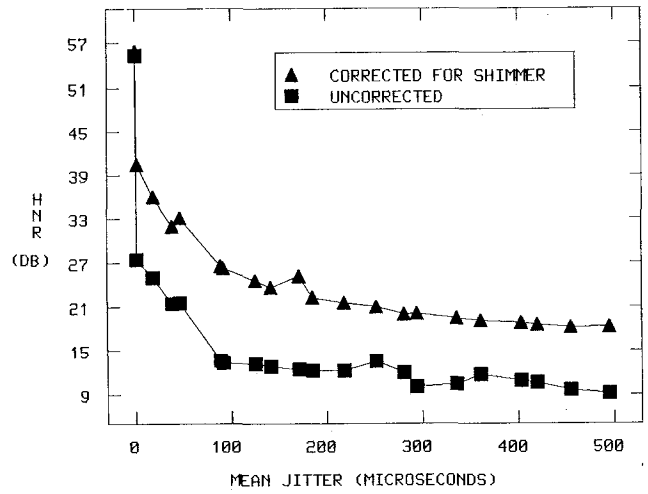

FIGURE 4. Measured HNR values for manually measured signals varying in pitch perturbation with and without correction for amplitude changes that are produced as a side effect of jitter.

SIMULATION 3: EFFECTS OF AMPLITUDE PERTURBATION

STIMULI

The effects of amplitude perturbation were studied by using the Klatt synthesis program to generate a set of stimuli differing in the amount of cycle-to-cycle variability in the amplitudes of the sequence of impulses that are used to generate glottal waveforms. A sequence of 200 pitch-pulse amplitude values was derived from the same male talker that was used to generate the fundamental-frequency control values for the pitch perturbation continuum. Using the same series of numerical manipulations that was used for the pitch perturbation continuum, this sequence of amplitude values was used to derive 21 sequences of numbers with mean shimmer values ranging from 0 to 2.6 dB. The maximum shimmer value of 2.6 dB is substantially larger than the figure of 0.17 dB reported by Horii (1982) for normal voices, but somewhat less than the 3.2 dB maximum shimmer value reported by Kitajima and Gould (1976) for a group of subjects with laryngeal polyps. The decision to restrict the shimmer continuum to the range below 2.6 dB was arbitrary to some extent and was based primarily on the increasingly bizarre perceptual quality of synthesized voice signals as shimmer values approach about 2 dB. (See Appendix for examples of Fourier spectra of stimuli taken from the amplitude perturbation continuum.)

After synthesis, all signals were scaled to maximum amplitude and analyzed by the AVR program. As in the previous tests, measurements were made using signals for which pitch markers were entered automatically during synthesis, and using signals for which pitch markers were entered after synthesis using a waveform editor.

RESULTS AND DISCUSSION

The results of these tests are shown in Table 4. It can be seen that the introduction of amplitude perturbation had a very large effect on HNR measurements. Excluding the perfectly periodic stimulus from consideration, stimuli varying in shimmer over a range of just less than 2.5 dB produced changes in HNR measurements totaling approximately 25 dB. It can also be seen that measured jitter values tended to increase as amplitude perturbation increased. This effect, however, is quite small for shimmer values below 0.5 dB and does not exceed 0.6% until shimmer values reach almost 2.0 dB.

Unlike results for the additive noise and jitter continua, measurements of the shimmer stimuli did not show large differences between manually and automatically marked signals. On the average, HNR values for the manually marked stimuli were only 0.7 dB lower than HNR values for the stimuli for which pitch markers were entered automatically during synthesis.

GENERAL DISCUSSION

Before discussing the results, two limitations of the present findings should be mentioned. An important aspect of the synthesis techniques that were used in the present study is that no attempt was made to model the momentary changes that occur in the waveshape of the glottal volume-velocity function. Although fundamental frequency and pitch-pulse amplitude were varied, the duty cycle and all other characteristics of the glottal waveform were identical for all pitch pulses.[3] The failure to model this source of variability is important because variations in glottal pulse shape would obviously affect measurements of all three parameters that were studied and would present an even more complex picture of measurement interaction. It is also significant that all of the simulations used the vowel [a]. It is quite possible that the magnitude of these measurement interactions would vary from one vowel to another.

TABLE 4. Effects of amplitude perturbation on measures of HNR and shimmer using both automatically and manually marked waveforms. Jitter values are given both in microseconds and as a percentage of the fundamental period. (Jitter values are given only for manually marked signals; for automatically marked signals, all jitter values were zero.)

Shimmer (auto)	*Shimmer (manual)*	*HNR (auto)*	*HNR (manual)*	*Mean Jitter (manual)*	*Percent Jitter (manual)*
0.00	0.01	55.3	53.4	0.1	0.00
0.18	0.19	24.7	24.2	0.4	0.01
0.22	0.23	21.9	21.6	0.4	0.01
0.37	0.38	18.4	18.2	0.4	0.01
0.50	0.50	16.2	13.9	36.4	0.47
0.67	0.69	14.1	12.6	35.4	0.46
0.73	0.72	12.2	11.2	36.4	0.47
0.90	0.92	10.9	9.9	39.1	0.51
1.01	1.03	9.6	8.8	42.2	0.55
1.28	1.27	8.5	7.9	43.7	0.57
1.32	1.31	7.7	7.2	42.2	0.55
1.46	1.50	6.5	6.1	43.7	0.57
1.62	1.62	5.3	5.0	42.2	0.55
1.69	1.68	4.6	4.3	37.1	0.48
1.94	1.93	3.7	3.5	47.3	0.61
2.09	2.06	2.9	2.6	49.9	0.65
2.12	2.11	2.4	2.0	56.6	0.73
2.26	2.25	1.8	1.5	60.9	0.79
2.47	2.47	1.1	0.7	76.2	0.99
2.59	2.54	0.3	0.0	69.4	0.90
2.65	2.60	−0.2	−0.6	66.1	0.86

Despite these limitations, the primary conclusion from the simulations reported above is that current time-domain approaches to the measurement of aperiodicities in voice signals are probably not capable of making a precise determination of the relative amounts of pitch perturbation, amplitude perturbation, and additive noise in a particular voice signal. For example, a relatively poor HNR value might be due to a large amount of additive noise, a large amount of pitch or amplitude perturbation, or some (unknown) combination of all three factors. Similarly, relatively large values of either pitch or amplitude perturbation could be attributed to a virtually unlimited number of combinations of glottal pitch and/or amplitude aperiodicities and additive noise.

The sensitivity of HNR measurements to changes in pitch and amplitude perturbation will probably not come as a surprise to many investigators. In fact, Yumoto et al. (1982) suggested that their signal-averaging technique might be sensitive to pitch perturbation. The authors commented that, "The major factor in certain types of severe hoarseness can be jitter rather than additive noise. Therefore the present technique might not be applicable to extremely severe hoarseness" (p. 1545). However, Yumoto et al. claimed to have ". . . accounted for this departure from the ideal conditions [i.e., pitch perturbation] by assuming that [the signal] is equal to zero in the interval between T (i), the duration of the *i*th period, and T, the maximum of the T (i)" (p. 1545). The present results suggest that this procedure does not fully account for variations in jitter. This is due in part to the fact that cycle-to-cycle variations in output amplitude are produced as a side effect of jitter. The effect of pitch perturbation on HNR values was significantly reduced, although not entirely eliminated, when signals were corrected for these amplitude changes. As will be discussed

---

[3] It was not possible to introduce cycle-to-cycle variations in glottal waveshape without making a major modification to the Klatt (1980) synthesis program. While this synthesizer allows a good deal of control over the characteristics of the glottal signal, these parameters are "global" variables that are fixed for the duration of the signal. We have recently developed a pitch-synchronous vowel synthesizer that is based on addition of damped sinusoids which will allow glottal parameters to be changed from one pitch cycle to the next.

in more detail below, this finding suggests that a modification of the Yumoto et al. technique that normalizes for pitch-pulse amplitude differences might be better at separating the effects of perturbation and additive noise.

The difficulty in measuring jitter, shimmer, and additive noise independent of one another may partially explain the relatively strong inter-correlations that have been reported for these three variables (Davis, 1976; Deal & Emanuel, 1978; Horii, 1980; Kempster, 1984; Kempster & Kistler, 1984; Yumoto et al., 1984). For example, in a study of dysphonic speakers, Kempster (1984) reported correlations of 0.68 between jitter and shimmer, −0.71 between jitter and HNR and −0.73 between shimmer and HNR. Citing similar findings, Deal and Emanuel (1978) commented,

> "It is . . . germane that the PVI and AVI [nonsequential measures of pitch and amplitude perturbation] tended to be moderately and positively correlated; that is, they were apparently overlapping measures of wave variability. This may account for the observation that the multiple correlation of the obtained wave variability indices [pitch and amplitude variability] versus SNL [a frequency-domain measure of noise content] and versus roughness ratings did not exceed greatly the correlation of SNL or roughness ratings with the individual wave index manifesting the larger Pearson *r*." (p. 262)

While it would not be surprising to find that pitch and amplitude perturbation are actually correlated with one another, it also seems very probable that measurement interactions of the type described in the present study played some role in the inter-correlations reported by these and other investigators.

Are Measurement Interactions Important?

It is reasonable to ask whether it is important to be able to assign a given acoustic measurement to a specific type of aperiodicity at the glottal level. For some kinds of studies, these measurement problems are probably not important. For example, there is a large body of literature whose very practical goal is to find acoustic measures that can be used: (a) to detect laryngeal disease, (b) to sort voice patients into diagnostic categories, or (c) to evaluate a patient's progress throughout the course of treatment. For these practical purposes, the usefulness of a particular acoustic measure is determined by running the appropriate empirical tests, and not by running the kinds of validity tests described in the present study. In other words, for these purposes it is not important whether measured jitter values reflect pitch variability or additive noise, so long as the measurement makes clinically-useful discriminations among patients.

The present findings, however, impose certain limits on the ability to interpret measures such as jitter, shimmer, and harmonics-to-noise ratio in terms of underlying glottal events. This is particularly true for disordered voices that may contain large departures from perfect periodicity. To cite one example, recall that Kitajima and Gould (1976) reported shimmer values as high as 3.2 dB for a group of speakers with laryngeal polyps. While this kind of measurement might represent a valuable descriptive characteristic of the output waveform recorded from a particular subject, it should not necessarily be assumed that a consequence of laryngeal polyps is to produce a large amount of amplitude variability in the laryngeal vibratory pattern. In terms of underlying glottal events, it is quite possible that a large measured shimmer value resulted primarily from a combination of additive noise and pitch perturbation. In fact, our experience in synthesizing voice signals that vary in these three dimensions suggests that it is very unlikely that even a severely disordered voice would contain source amplitude variability approaching 3.2 dB. Informal listening tests suggest that stimuli which are synthesized with relatively high amplitude perturbation values (above about 1.7 to 2.0 dB) do not sound like convincing examples of naturally occurring voices. Listeners typically comment that the stimuli have a "popping" quality somewhat like static, or that the stimuli sound as though they were being played over a loudspeaker with a loose wire. By contrast, stimuli with large values of jitter (above about 5.0 to 6.0%) are perceived as very rough, but sound as though they could have been produced by a talker with a severely disordered voice. Similarly, stimuli with large amounts of additive noise sound very breathy but do not have the unnatural quality that is associated with large shimmer values.

Measurement interactions would also seem to be important for descriptive research that is aimed at establishing relationships between perceived vocal quality and specific acoustic attributes of voice signals. A number of studies have used correlational techniques in an effort to learn something about what acoustic variables or combinations of variables are associated with subjective aspects of vocal quality such as "roughness," "hoarseness," and "breathiness" (e.g., Deal & Emanuel, 1978; Kempster, 1984; Kojima et al., 1980; Murry, Singh, & Sargent, 1977; Prosek, Montgomery, Walden, & Hawkins, 1984; Smith, Weinberg, Feth, & Horii, 1978; Yanagihara, 1967; Yumoto et al., 1982). For these kinds of studies—especially those involving several acoustic dimensions—the possibility of measurement interactions among individual acoustic parameters would have important implications for interpreting the relationships between physical and subjective dimensions.

Can Steps be Taken to Reduce Measurement Interactions?

One question that might be posed about the present set of findings is whether the measurement techniques could be modified in ways that would reduce the degree of measurement interaction among the three variables. Although techniques probably do not exist that will entirely eliminate measurement interactions, there are a variety of approaches that might reduce the magnitude of these effects. For example, the "shimmer artifact" that results from the summing of energy from previous pitch pulses might be reduced by using a technique that was developed by Holden and Gulut (1976) in an effort to achieve more accurate measurements of formant frequencies, bandwidths, and amplitudes. Holden and Gulut's method attempted to "... extend the signal from the corrupting period into the next one using linear prediction techniques for forecasting [and then] subtract the extended signal from the signal in the corrupted period to obtain the actual impulse response" (p. 458). Although Holden and Gulut did not address the problem of perturbation measurements, it seems likely that their method would produce measurements of amplitude perturbation that are less affected by variations in pitch perturbation, especially at high fundamental frequencies.

The effects of both pitch and amplitude variability on HNR measurements could potentially be reduced by relatively simple modifications to the Yumoto et al. (1982) signal-averaging technique. As discussed previously, the effects of amplitude perturbation on HNR measurements can be significantly reduced, although not eliminated, simply by normalizing each pitch pulse for rms intensity. In results not reported in the present paper, we have also found that the effects of pitch perturbation on additive noise measurements can be significantly reduced by another minor modification to the Yumoto et al. technique. In the current technique, the size of the averaging window is determined by the longest period in the waveform. Preliminary tests with a version of the AVR program that uses the *smallest* period to set the size of the averaging window (ignoring the portion of each pitch pulse that exceeds this minimum value) suggest that this method is less sensitive to both pitch and amplitude perturbation.

Reducing the influence of additive noise on measured values of pitch and amplitude perturbation would seem to be a much more difficult problem. It is important to note that the jitter and shimmer artifacts resulting from additive noise were observed in spite of the fact that signals were low-pass filtered at 500 Hz. As was discussed previously, these kinds of artifacts should be expected in light of the changes that additive noise will produce in the locations of reference points such as zero crossings and peaks, and given the random amplitude fluctuations that occur in noise signals.

It is possible that the influence of additive noise on perturbation measures could be reduced by attempting to subtract the additive noise component from the voice signal. Computationally, this would be relatively simple since the Yumoto et al. technique attempts separate reconstructions of the periodic and aperiodic components. It remains to be determined, however, whether the reconstruction of the noise is sufficiently accurate to allow time-domain subtraction of the noise component.

It is clear from the discussion above that there are several modifications to existing analysis techniques that might reduce the magnitude of measurement interactions. However, one very significant problem that must be kept in mind is that the perturbation phenomena being measured are quite small in most cases. Even in dysphonic voices, shimmer values for sustained vowels are generally a small fraction of a decibel; jitter values are usually less than 1.0%, often considerably less (Kempster, 1984). This means that even small measurement artifacts can be quite significant in relation to the perturbation phenomena that are being measured. It is quite possible, then, that acoustic methods may simply be incapable of determining the precise sources of glottal aperiodicity that are associated with a particular voice signal. Until the issue of measurement interaction is adequately addressed, it might be more appropriate to view measures such as jitter, shimmer and additive noise as more-or-less generic measures that reflect the degree of aperiodicity in a voice, without attempting to interpret these measures in terms of specific glottal events.

SUMMARY AND CONCLUSIONS

The present study used a series of computer simulations that were designed primarily to determine whether pitch perturbation, amplitude perturbation, and additive noise could be measured independent of one another. The results suggested that there are strong measurement interactions among the three variables. For example, adding increasing amounts of noise to an otherwise perfectly periodic voice signal resulted not only in decreases in measured HNR values, but also substantial increases in measured values of pitch and amplitude perturbation. For these reasons, it may be very difficult to make a precise determination of the source of aperiodicity in voice waveforms. This would be especially true in the case of disordered voices, which may show large departures from perfect periodicity. There may be a number of relatively simple ways to modify existing measurement techniques to reduce the degree of measurement interaction among these three variables. However, until the validity of these techniques is established, caution should be exercised in interpreting measures of perturbation and noise in terms of specific aspects of the laryngeal vibratory cycle.

ACKNOWLEDGMENTS

A portion of this research was carried out while the author was on the faculty in Communication Sciences and Disorders at Northwestern University. A number of colleagues contributed advice and ideas to the present work, including Bill Martens, Marty Wilde, Ray Colton, Dale Metz, and Tom Edwards. This research was supported by NIH grant 1-R01-NS-22234-01 to Northwestern University and NIH grant 7-R01-NS-23703-01 to RIT Research Corporation.

REFERENCES

CHANDRA, S., & LIN, W. C. (1974). Experimental comparison between stationary and non-stationary formulations of linear prediction applied to voiced speech analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-22,* 405–415.

DAVIS, S. B. (1976). Computer evaluation of laryngeal pathology based on inverse filtering of speech. *SCRL Monograph, 13,* Speech Communication Research Library, Santa Barbara.

DAVIS, S. B. (1981). Acoustical characteristics of normal and pathological voices *ASHA Reports, 11,* 97–115.

DEAL, R. E., & EMANUEL, F. W. (1978). Some waveform and spectral features of vowel roughness. *Journal of Speech and Hearing Research, 21,* 250–264.

EMANUEL, F. W., & SANSONE, F. (1969). Some spectral features of 'normal' and 'simulated rough' vowels. *Folia Phoniatrica, 21,* 410–415.

Fant (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 173–277). Amsterdam: North Holland.

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

Hecker, M., & Kruel, 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., & Horii, Y. (1982). Jitter and shimmer in sustained phonation. In N.J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice, Vol. 7* (pp. 299–332). New York: Academic Press.

Hillenbrand, J., Biggam, D. F., & Wilde, M. D. (1984). *AVR: A computer program for the measurement of perturbation and signal-to-noise ratio in sustained vowels* [Computer Program]. Evanston, Illinois: Northwestern University.

Holden, A. D. C., & Gulut, Y. K. (1976). A new method for accurate analysis of voiced speech. *Proceedings of the 1976 IEEE International Conference on Acoustics, Speech and Signal Processing,* 458–461.

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

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.

House, A. S. (1960). A note on optimal vocal frequency. *Journal of Speech and Hearing Research, 2,* 55–60.

Kasuya, H., Ebihara, S., & Yoshida, H. (1984). Clinical screening of laryngeal pathology by voice. *Journal of the Acoustical Society of America, 76* (Suppl. 1), S60 (A).

Kempster, G. B. (1984). *A multidimensional analysis of dysphonia in two dysphonic groups.* Unpublished doctoral dissertation, Northwestern University.

Kempster, G. B., & Kistler, D. J. (1984). Perceptual dimensions of dysphonic voices. *Journal of the Acoustical Society of America, 75* (Suppl. 1), S8 (A).

Kitajima, K., & 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., & Isshiki, N. (1975). Pitch perturbation in normal and pathologic voice. *Studia Phonologica, 9,* 25–32.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67,* 971–995.

Koike, Y. (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica, 7,* 17–23.

Kojima, H., Gould, W. J., Lambaise, A., & Isshiki, N. (1980). Computer analysis of hoarseness. *Acta Oto-Laryngologica, 89,* 547–554.

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.

Lively, M. A., & Emanuel, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. *Journal of Speech and Hearing Research, 13,* 503–517.

Makhoul, J. E., & Wolf, J. J. (1972). *Linear prediction and the spectral analysis of speech.* Bolt, Beranek & Newman Report No. 230 (NTIS AD749066). Cambridge, MA.

Murry, T., & Doherty, E. T. (1980). Selected acoustic characteristics of pathologic and normal speakers. *Journal of Speech and Hearing Research, 23,* 361–369.

Murry, T., Singh, S., & Sargent, M. (1977). Multidimensional classification of abnormal voice qualities. *Journal of the Acoustical Society of America, 61,* 1630–1635.

Prall, C. W., & Hillenbrand, J. (1981). *AUDED: A general-purpose waveform editor* [Computer Program]. Evanston, IL: Northwestern University.

Prosek, R. A., Montgomery, A. A., Walden, B. E., & Hawkins, D. B. (1984). Some relations between voice-quality judgments and derived acoustic measurements. *Journal of the Acoustical Society of America, 75* (Suppl. 1), S8 (A).

Robbins, J. (1981). *A comparative acoustic study of laryngeal speech, esophageal speech, and speech production after tracheo-esophageal puncture.* Unpublished doctoral dissertation, Northwestern University.

Sansone, F., & Emanuel, F. W. (1970). Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. *Journal of Speech and Hearing Research, 13,* 489–502.

Smith, B., Weinberg, B., Feth, L., & Horii, Y. (1978). Vocal jitter and roughness characteristics of esophageal speech. *Journal of Speech and Hearing Research, 21,* 240–249.

Takahashi, H., & Koike, Y. (1975). Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Oto-Laryngologica* (Suppl. 228), 1–24.

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

Yanagihara, N. (1967). Significance of harmonic change and noise components in hoarseness. *Journal of Speech and Hearing Research, 10,* 531–541.

Yumoto, E. (1983). The quantitative evaluation of hoarseness: A new harmonics to noise ratio method. *Archives of Otolaryngology, 109,* 48–52.

Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America, 71,* 1544–1550.

Yumoto, E., Sasaki, Y., & Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *Journal of Speech and Hearing Research, 27,* 2–6.

Zemlin, W. R. (1968). *Speech and Hearing Science: Anatomy and Physiology.* Englewood Cliffs, NJ: Prentice-Hall.

Received August 27, 1986
Accepted March 3, 1987

Requests for reprints should be sent to James Hillenbrand, RIT Research Corporation, Rochester Institute of Technology, 75 Highpower Rd., Rochester, NY 14623-3435.

APPENDIX:

SPECTRAL EFFECTS OF PERTURBATION AND ADDITIVE NOISE

The signal-analysis procedures that were discussed in the main body of the paper were based on time-domain measurements of voice signals. The analyses that are described in this Appendix were designed to compare the effects of pitch perturbation, amplitude perturbation, and additive noise on frequency-domain representations of voice signals. Signals from the additive noise (F_0 = 130 Hz), pitch perturbation and amplitude perturbation continua were multiplied by a Hamming window and analyzed using a 1024-point (51.2 ms) FFT. In all cases, the left edge of the Hamming window was aligned with the beginning of the first pitch pulse in the signal.

Four representative spectra from the additive noise series are shown in Figure A-1. To provide better visual resolution, results are displayed only for spectral components below 5 kHz. Not surprisingly, the primary effect of additive noise is to reduce harmonic organization. In general, the smearing of harmonic structure seems to be due primarily, although not exclusively, to increases in energy levels between harmonic frequencies rather than decreases in harmonic levels. This effect seems to be particularly evident in the regions around vowel formants. It is not clear how much emphasis should be placed on this particular finding, however, because this result was basically assured by the method that was used to synthesize the aperiodic components for these signals; that is, by passing white noise through formant resonators. However, the synthesis method is well motivated on theoretical grounds, and the basic finding agrees well with Yanagihara's (1967) measurements of vowels produced by disordered speakers.

The effects of pitch perturbation are shown for four representative signals in Figure A-2. Again, the primary effect of introducing departures from perfect periodicity is to decrease harmonic organization. However, the smearing of harmonic structure is pronounced primarily for signals with jitter values above about 2.5 to 3.0%. The spectral change seems to be most prominent for signal components above about 1.0 to 2.0 kHz. As in the additive noise continuum, the spectral change seems to be attributable primarily to increases in energy levels between harmonic frequencies, with harmonic levels changing by relatively small amounts.

Spectra for four signals that vary in amplitude perturbation are shown in Figure A-3. The primary spectral effects of amplitude perturbation seem to be both systematic decreases in harmonic levels and systematic increases in energy levels between harmonic frequencies. However, harmonic organization is quite evident even in the signal with the highest level of amplitude perturbation. This finding suggests that amplitude perturbation is not likely to be reflected in hoarseness-severity measures such as the one proposed by Yanagihara (1967), which is based on changes in harmonic structure in narrow-band spectrograms.

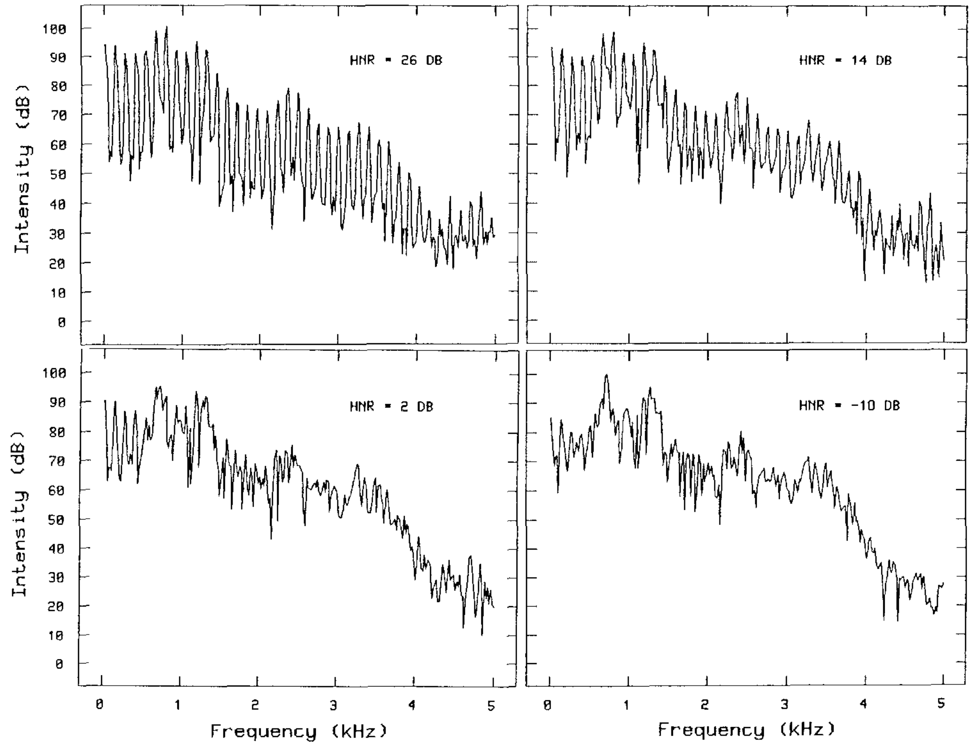

FIGURE A-1. Fourier analysis of four signals from the additive noise continuum.

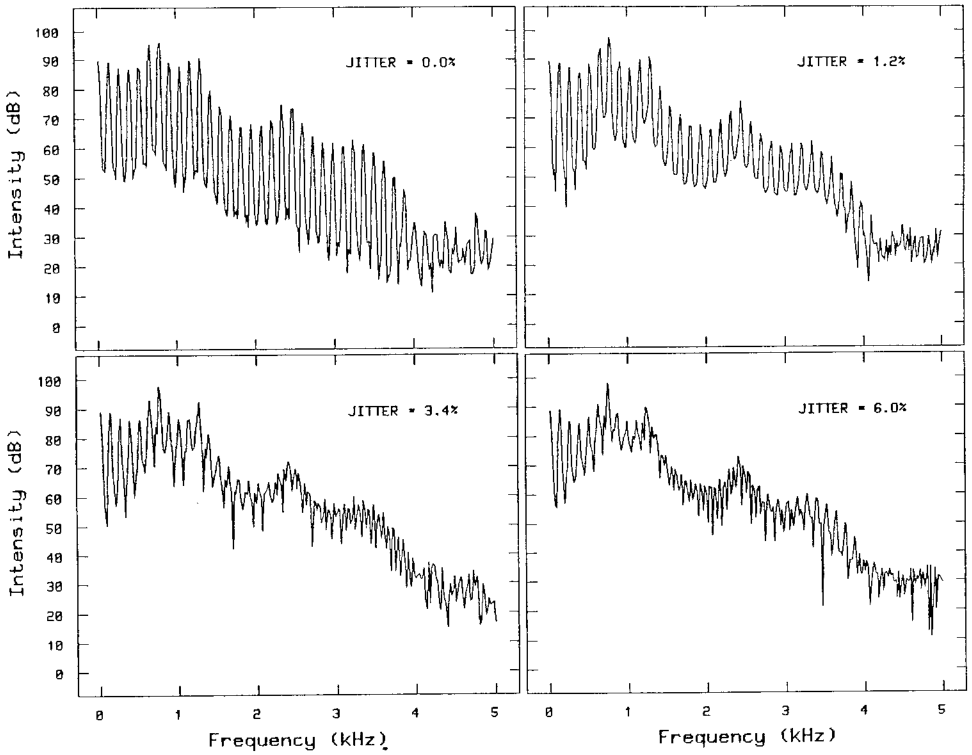

FIGURE A-2. Fourier analysis of four signals from the pitch perturbation continuum.

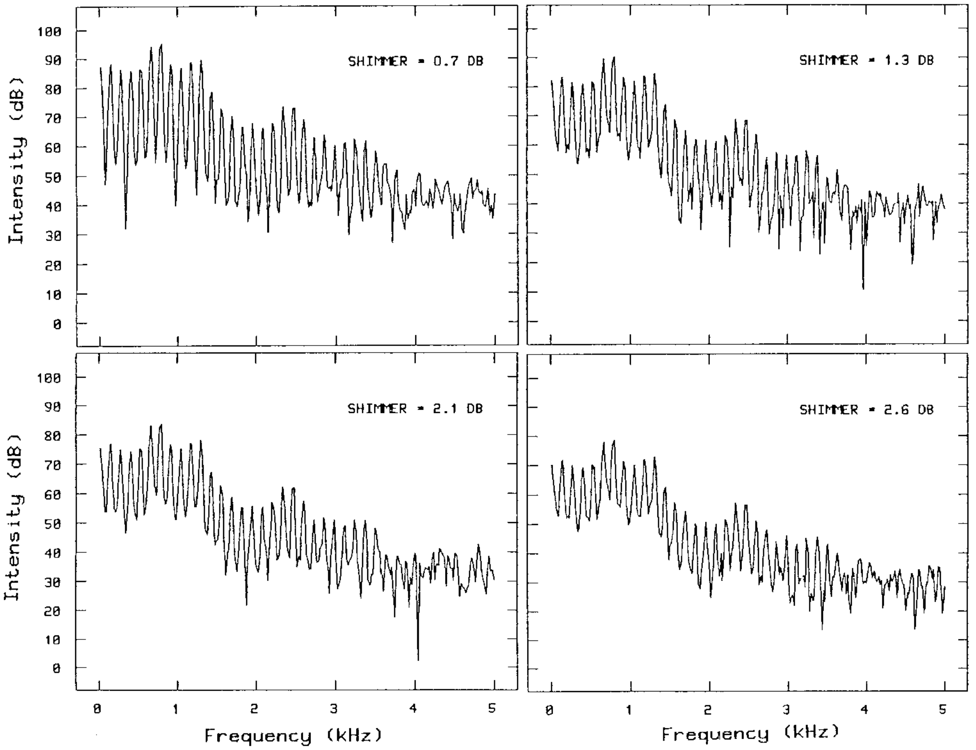

FIGURE A-3. Fourier analysis of four signals from the amplitude perturbation continuum.