Harmonics-to-noise ratio as an index of the degree of hoarseness

Eiji Yumoto[a) and Wilbur J. Gould

Lenox Hill Hospital, Vocal Dynamics Laboratory, 100 East 77th Street, New York, New York 10021

Thomas Baer

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510

(Received 10 August 1981; accepted for publication 23 February 1982)

Degree of hoarseness can be evaluated by judging the extent to which noise replaces the harmonic structure in the spectrogram of a sustained vowel. However, this visual method is subjective. The present study was undertaken to develop the harmonics-to-noise (H/N) ratio as an objective and quantitative evaluation of the degree of hoarseness. The computation is conceptually straightforward; 50 consecutive pitch periods of a sustained vowel /ɑ/ are averaged; H is the energy of the averaged waveform, while N is the mean energy of the differences between the individual periods and the averaged waveform. Recordings of 42 normal voices and 41 samples with varying degrees of hoarseness were analyzed. Two experts rated the spectrogram of each voice sample, based on the amount of noise relative to that of the harmonic component. The results showed a highly significant agreement (the rank correlation coefficient = 0.849) between H/N calculations and the subjective evaluations of the spectrograms. The H/N ratio also proved useful in quantitatively assessing the results of treatment for hoarseness.

PACS numbers: 43.70.Bk, 43.70.Gr, 43.70.Ny

INTRODUCTION

Hoarseness is a general term used for describing a perceived abnormality of the voice. Most laryngologists rate the degree of hoarseness to assess the results of treatment for laryngeal disorders. However, these ratings are subjective, and there is no common scale among laryngologists. Judgments of degree of hoarseness clearly depend on properties of the acoustic signal. Therefore development of a common objective scale requires both understanding of the acoustic attributes of the hoarse voice and development of a procedure for quantifying these attributes.

Sound spectrographic analysis (Yanagihara, 1967) has revealed that sustained vowels perceived as hoarse have the following characteristics:

(1) Noise components in the main formants of various vowels.
(2) High-frequency noise components.
(3) Loss of the high-frequency harmonic components.

As the degree of judged hoarseness increases, more noise appears and replaces the harmonic structure. Yanagihara developed a technique for visually evaluating hoarseness based on the appearance of the spectrogram. In addition, Rontal *et al.* (1975) reported that spectrographic analysis is helpful in evaluating the pre- and post-treatment difference in the degree of hoarseness. From these studies, it can be inferred that the degree of perceived hoarseness can be evaluated by judging the extent to which noise replaces the harmonic structure in the spectrogram of a sustained vowel. However, this visual method is still subjective and therefore difficult to standardize.

Emanuel and his co-workers (Emanuel and Sansone, 1969; Sansone and Emanuel, 1970; Lively and Emanuel, 1970; Hanson and Emanuel, 1979) estimated noise levels in the spectra of sustained vowels, and found a relationship between the spectral noise level (SNL) and the perceived magnitude of the roughness of the voice (not the degree of hoarseness). However, their method is limited in application. The level of the harmonic components of the spectrum was not taken into account. To allow comparison of different measurements of SNL, subjects were required to phonate at an intensity of 75 dB SPL (±1 dB) for 7 s. This task is not feasible for most patients with laryngeal disorders.

Kojima *et al.* (1980) developed an objective measure of the degree of hoarseness which took the harmonic components of the spectrum into account. A series of discrete Fourier transforms, each based on three successive pitch periods, was used to estimate signal energy at the harmonic frequencies and noise energy between them. A ratio of the acoustic energy of the harmonic components to that of the noise (S/N ratio) was calculated. Their results showed a statistically significant correlation between this S/N ratio and psychophysical measurement of the degree of hoarseness. However, their method for quantifying the relationship between signal and noise components is complex and time consuming.

The purpose of this paper is to develop a similar measure, the H/N ratio, which is simpler to compute; and to examine whether the H/N ratio is useful in quantitatively assessing the results of treatment for hoarseness.

I. METHOD

The rationale for our method is based on the assumption that the acoustic wave of a sustained vowel consists of two components: a periodic component that is the same from cycle to cycle and an additive noise component that has a zero-mean amplitude distribution. The

[a)Send reprint requests to Eiji Yumoto, Department of Otolaryngology, School of Medicine Ehime University, Ehime, 791-02, Japan.

0001-4966/82/061544-07$00.80

original wave, $f(t)$, can then be considered as the concatenation of the waves, $f_i(\tau)$, from each pitch period, where τ ranges over the duration of a pitch period. Therefore, when one averages a sufficiently large number, n, of these $f_i(\tau)$, the noise component is canceled. The resulting average wave

$$f_A(\tau)=\sum_{i=1}^{n}\frac{f_i(\tau)}{n} \tag{1}$$

is a good estimate of the signal that gives rise to the harmonic component.

Evidently, our assumption of a periodic signal does not strictly hold. Even the phonation of a normal subject exhibits cycle-to-cycle pitch perturbations (jitter) (Lieberman, 1961; Hollien *et al.*, 1973). Pathological voices often exhibit increased jitter (for example, Lieberman, 1963; Iwata and von Leden, 1970; Hecker and Kreul, 1971; Davis, 1976). The major factor in certain types of severe hoarseness can be jitter rather than additive noise components. Therefore it should be mentioned that the present method may not be applicable to extremely severe hoarseness. For the purpose of calculating $f_A(\tau)$, we have accounted for this departure from the ideal conditions by assuming that $f_i(\tau)$ is equal to zero in the interval between T_i, the duration of the ith period, and $\hat{T}$, the maximum of the T_i. The function $f_A(\tau)$ given in Eq. (1) is valid over the interval from $\tau=0$ to T.

The measure of the acoustic energy of the harmonic component of $f(t)$ is defined as

$$H=n\int_0^T f_A^2(\tau)\,d\tau. \tag{2}$$

On the other hand, the noise wave in the ith pitch period is equal to $f_i(\tau)-f_A(\tau)$ where τ ranges between 0 and T_i. Then, the acoustic energy of the noise component of $f(t)$ is defined as

$$N=\sum_{i=1}^{n}\int_0^{T_i}[f_i(\tau)-f_A(\tau)]^2\,d\tau. \tag{3}$$

The ratio, H/N, is determined from Eqs. (2) and (3) as

$$H/N=n\int_0^T f_A^2(\tau)\,d\tau\Big/\sum_{i=1}^{n}\int_0^{T_i}[f_i(\tau)-f_A(\tau)]^2\,d\tau. \tag{4}$$

As noted by Emanuel and Sansone (1969), jitter affects the spectrum of a sustained vowel by reducing the amplitudes of the harmonics and adding noise between them. Because of our method for calculating $f_A(\tau)$ [i.e., assuming $f_i(\tau)=0$ for $T_i \leq \tau \leq T$], the presence of jitter contributes to the magnitude of the noise component, N, even in the absence of any additive noise.

Analysis was performed on phonation of the sustained vowel /ɑ/. One consideration in the choice of this vowel is that the acoustic effects of the vocal tract are likely to remain stable during a sufficiently long interval, since the transfer function for /ɑ/ is relatively insensitive to small articulatory movements (Stevens, 1972). This is important because changes in vowel quality during the analysis interval would otherwise contribute artificially to the noise component. In addition, the moment of greatest acoustic excitation within the glottal cycle is easily identified for this vowel, as shown in Fig. 1. This ensures that the interval over which the time integration is performed ends in a relatively low-amplitude portion of the glottal cycle, so that the contributory effects of jitter on the noise component are relatively small.

II. TEST PROCEDURES

A. Subjects

Twenty-two males and twenty females without laryngeal or pulmonary complaints served as normals. Their ages varied from 19 to 60 with a mean of 36. Analysis was also performed on 41 phonatory samples of 12 males and 8 females with various laryngeal disorders pre- and postoperatively. Their ages varied from 21 to 68 with a mean of 46. These samples were classified into the pre- and postoperative groups. Table I shows the sex and the diagnosis of each subject for these two groups.

B. Voice recording

The subjects were placed in a sound-treated booth and were asked to sustain the vowel /ɑ/ for a few seconds at a comfortable pitch and loudness level. Care was taken to keep pitch and loudness at as constant a level as possible. The mouth-to-microphone distance was approximately 20 cm. The voice recording was made through a high fidelity microphone/tape-recorder system.

C. Analog to digital conversion

High-frequency pre-emphasis was applied to the voice signal. This signal was low-pass filtered at 10 kHz and digitized with 12-bit precision at a sampling rate of 20 kHz for 3 s. Input amplitudes were adjusted to utilize nearly the full range of the A/D converter. The most stable region of about 600 ms duration was selected from the digitized waveform on the CRT screen and stored for further analysis. The initial and terminal portions of the phonation were excluded.

D. Pitch extraction

The sampled waveforms were subjected to a 500-Hz low-pass digital filter for pitch extraction. A CRT

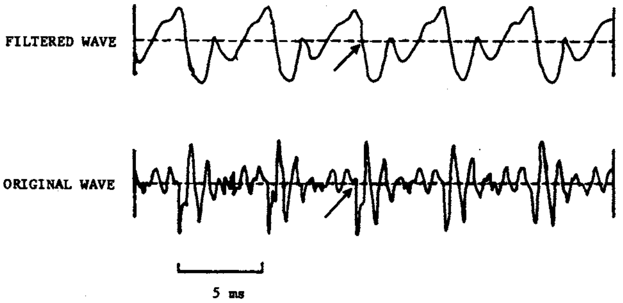

FIG. 1. Examples of the original acoustic wave (bottom) and its low-pass filtered wave (top). The arrow on the filtered wave shows the easily recognizable zero-crossing point. The arrow on the original wave shows the zero-crossing point before the prominent peak of the corresponding pitch period.

TABLE I. The H/N ratio and the spectrographic classification related to the sex and the diagnosis of each patient. Δ represents the difference in the H/N ratio or the spectrographic classification between the pre- and postoperative data.

Subject number	Sex	Diagnosis		H/N ratio (dB)			Spectrographic classification	
			Pre	Post	Δ	Pre	Post	−Δ
1	M	Keratosis	6.4	9.6	3.2	1	1	0
2	F	Nodules	9.5	14.5	5.0	1	0	1
3	M	Leukoplakia	2.5			3		
4	F	Polyp	−3.9	15.2	19.1	3.5	0	3.5
5	F	Polypoid degeneration	−8.0	7.9	15.9	3.5	0.5	3
6	F	Recurrent nerve paralysis	−15.2	10.9	26.1	4	0	4
				8.2	23.4		0.5	3.5
7	M	Polyp	6.5	13.5	7.0	2	0.5	1.5
8	M	Polyp		8.2			0	
				9.5			0	
9	F	Laryngeal web	7.4	11.9	4.5	1	1	0
				10.6	3.2		0.5	0.5
10	M	Atrophied vocal fold	0.0	9.1	9.1	3	1	2
11	M	Polyp		10.6			0	
12	M	Polyp	3.1			2		
13	M	Carcinoma	6.9	16.1	9.2	1	0	1
				14.0	7.1		0	1
				9.6	2.7		1	0
14	F	Hyperfunctional dysphonia	−7.2	5.9	13.1	3.5	2	1.5
				11.9	19.1		0	3.5
				9.6	16.8		0	3.5
				8.8	16.0		1	2.5
15	F	Polyp	0.7	16.6	15.9	2	0	2
				17.6	16.9		0	2
16	M	Keratosis	8.2	10.1	1.9	0	0	0
17	M	Polypoid degeneration	5.0			2		
18	F	Polyp	9.6			1		
19	M	Hemilaryn gectomized	−7.2			3.5		
20	M	Polyp	5.5			1		

screen displayed both the filtered wave and the original wave (Fig. 1). Pitch periods were marked on the filtered wave, using either a manual or semi-automated method based on zero crossings. The arrow on the filtered wave in Fig. 1 shows the easily recognizable zero-crossing point. The arrow on the original wave shows the zero-crossing point before the prominent peak of the corresponding pitch period. The phase difference between these two arrows was adjusted in such a way that the mark on the filtered wave coincided with the corresponding point on the original wave. Pitch periods were, then, extracted from the original wave and used for further processing. The filtered wave was used in order to facilitate the pitch extraction procedure for severely hoarse voices. As far as our samples are concerned, we could demarcate pitch periods even in the hoarse voices by this procedure. There may be severely hoarse voices beyond our consideration, in which no periodic components can be identified.

E. Extraction of the noise component

Figure 2 exemplifies several sample waveforms from one speaker. On the top is the average waveform and below it are the waveforms of three individual pitch periods. Subtraction of the average wave, $f_A(\tau)$, from the wave of the ith pitch period, $f_i(\tau)$, gives the noise component of the ith pitch period (where $0 \leq \tau \leq T_i$). These noise components, from the first pitch period to the 50th, were concatenated to form a noise sound comparable with 50 pitch periods of the original phonation.

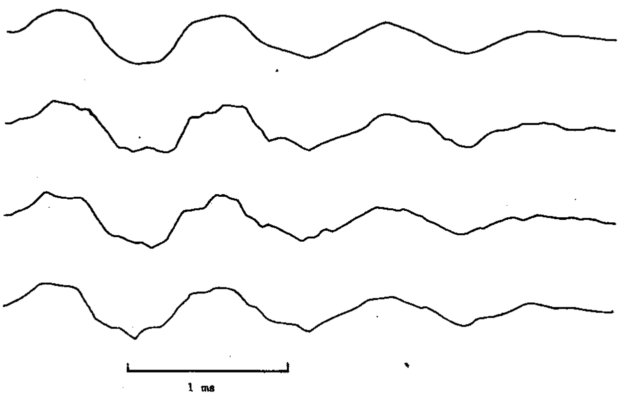

FIG. 2. Examples of averaged pitch-period waveform (top) and waveforms of individual pitch periods (others) from one speaker.

The original phonation and the extracted noise sound, which were stored in the computer, were repeated to obtain signals of 1.5-s duration. Spectrograms from the original phonation and the extracted noise sound were made using a narrow-band filter. We calculated the H/N ratio from 50 pitch periods using Eq. (4) and converted it to a decibel scale.

F. Rating of the spectrograms

Two experts rated the spectrograms of the original phonation of all cases in random order. This rating was based on the classification developed by Yanagihara (1967), in which spectrograms of hoarse voices were graded by the amount of noise relative to that of the harmonic components of vowels. The scale ranged from 1 for slight hoarseness to 4 for severe hoarseness. We added type 0 for clear phonation, free of the noise component. Before the rating, the judges discussed and unified the definition of the 5° of hoarseness in order to reduce effects of possible differences in interpretation. To examine the reliability of scores, 58 randomly selected spectrograms were rated again one month after the first rating. The overall inter- and intra-judge agreements were 85.0% and 85.3%, respectively. The variability of the rating expressed as the rms of differences between the ratings and the mean ratings of two occasions for the 58 re-rated samples was 0.263. These results justified an averaging of scores in order to obtain one rating for each spectrogram.

G. Accuracy of measurement

We recorded periodic sinusoidal waves of four different frequencies onto a magnetic tape. In order to evaluate the validity of the H/N ratio method, we processed these signals in the same fashion as the voice signal. The calculated H/N ratios were 37.4, 36.1, 32.1, and 30.7 dB for signals of 120, 160, 240, and 300 Hz, respectively. The greatest H/N ratio measured from the subjects was 17.6 dB. These results show that the error range of the measurement system did not contribute significantly to the calculated H/N ratios. Because of the systematic way these H/N values vary with signal frequency, it seems likely that the main source of measurement noise is due to the effects of discrete-time sampling. Other origins may be ascribed to the tape recorder or the signal generator.

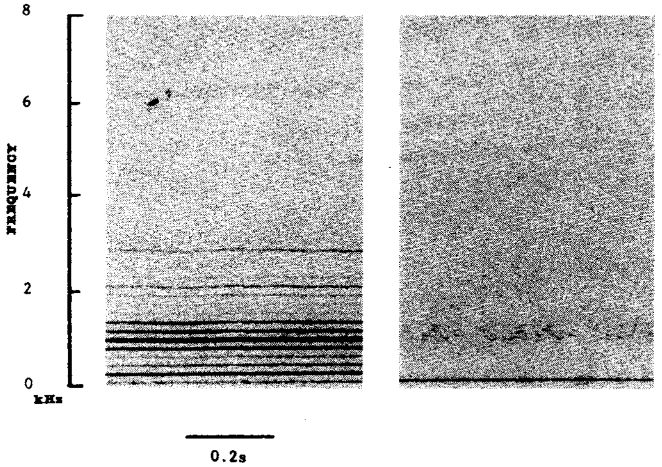

FIG. 3. Spectrograms made from the original phonation (left) and the extracted noise sound (right). The subject is a 45-year-old female free of hoarseness (type 0). The H/N ratio is 14.6 dB.

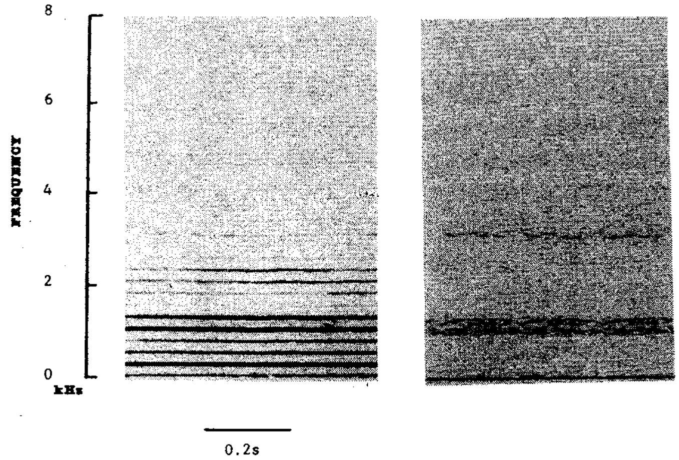

FIG. 4. Spectrograms arranged as in Fig. 3. The subject is a 38-year-old female with very slight hoarseness (type 1). The H/N ratio is 9.1 dB.

III. RESULTS

Figures 3–5 exemplify spectrograms made from the original phonation (left) and the extracted noise sound (right). Figure 3 is of a 45-year-old female with a normal voice quality free of hoarseness (type 0). The fundamental frequency is 190 Hz and the H/N ratio is 14.6 dB. There is a slight degree of noise in the main formants. Figure 4 is of a 38-year-old female with very slight hoarseness (type 1). The fundamental frequency is 260 Hz and the H/N ratio is 9.1 dB. More noise appears in the main formants than in the previous case. In addition, a faint shadow of noise is visible around 3 kHz. Figure 5 is of a 45-year-old female with severe hoarseness (type 3.5). The fundamental frequency is 250 Hz and the H/N ratio is − 7.2 dB. The noise component is dominant throughout the frequency range. Figures 3–5 show a highly satisfactory degree of extraction of the noise component from phonation even when the subject had severe hoarseness.

Figure 6 is a histogram of the H/N ratio for the normal, pre- and postoperative groups. The results of the H/N calculation and the spectrographic evaluation are shown in Table I for the pre- and postoperative groups.

The mean H/N ratio for the normal males was 12.2 dB and for the normal females 11.5 dB. However, this difference was not significant at the 5% level. Therefore we dealt with these values of normal subjects as samples drawn from a single population. The H/N ratio of the normal group ranges between 7.0 and 17.0 dB with a mean of 11.9 dB. We regard this as a normal distribution (S.D. = 2.32). We can expect 95% of normal subjects to have the H/N ratio greater than 7.4 dB (one-tailed test). The H/N ratio of the preoperative group ranges between − 15.2 and 9.6 dB with a mean of 1.6 dB. Three of 18 preoperative samples (16.7 %) had H/N ratios

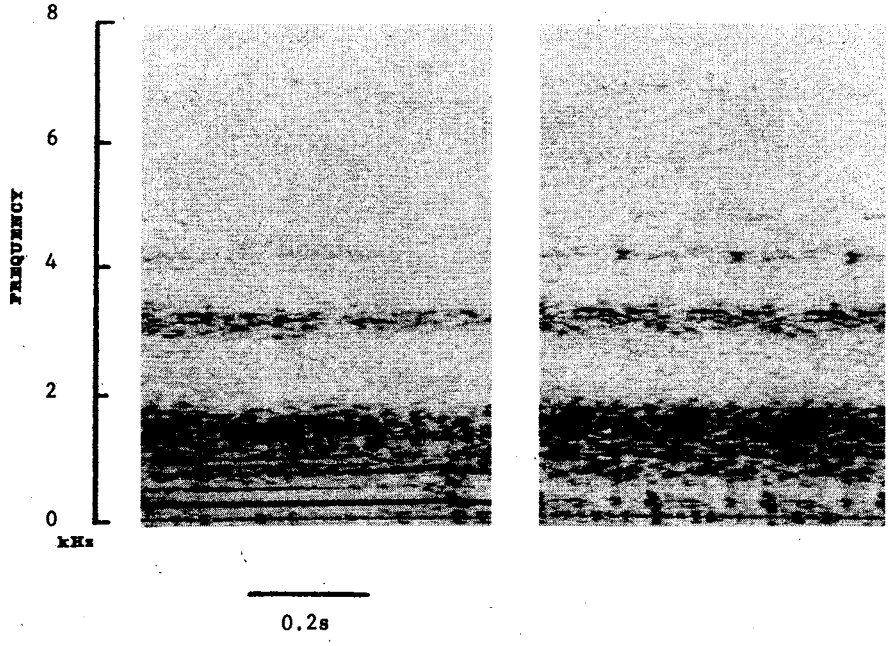

FIG. 5. Spectrograms arranged as in Fig. 3. The subject is a 45-year-old female with severe hoarseness (type 3.5). The H/N ratio is –7.2 dB.

greater than 7.4 dB (subjects 2, 16, and 18). This overlap is not unexpected. The preoperative phonatory samples included varying degrees of hoarseness, and some of the normal subjects showed slight hoarseness. The H/N ratio of the postoperative group ranges between 5.9 and 17.6 dB. Twenty-two of 23 postoperative samples (95.7%) had H/N ratios greater than 7.4 dB (the exception was subject 14). Indeed, their mean H/N ratio of 11.3 dB is nearly the same as the normals' mean of 11.9 dB. Figure 7 shows, however, that their distributions have greater spread (S.D.=3.13) than normals' (S.D.=2.32).

We compared the difference in the H/N ratio with the difference in the spectrographic classification of each patient pre- and postoperatively. The data are shown at the third column of the H/N ratio (labeled as Δ) or the spectrographic classification (labeled as – Δ) on Table I. Figure 7 is a scatter diagram relating these two variables. The abscissa represents the difference in the spectrographic classification while the ordinate represents the difference in the H/N ratio. These two variables appear highly correlated. Spearman's rank correlation coefficient, corrected for tied ratings, is 0.944 and is significant at the p =0.001 level. The line in the figure is the least-squares regression line fit of all the data points with the regression coefficient of – 5.1. These findings reveal that the H/N ratio is a useful tool for quantitative comparison of the degree of hoarseness of a post-treatment voice with that of a pre-treatment voice.

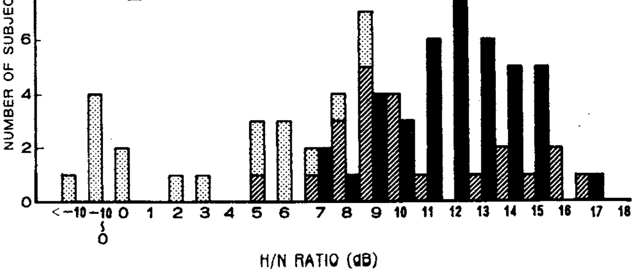

FIG. 6. Histogram of the H/N ratio for the normal, pre- and postoperative groups.

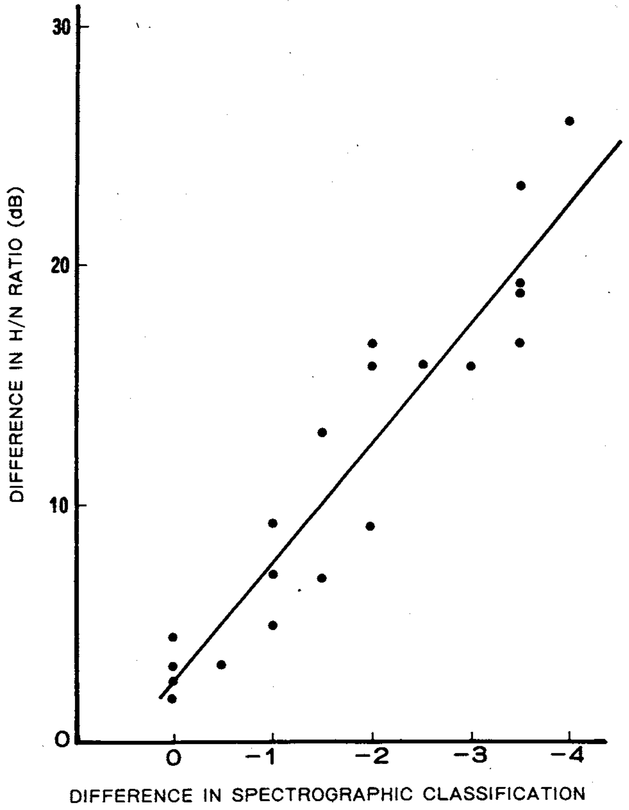

FIG. 7. Scatter diagram of the difference in the H/N ratio versus the difference in the spectrographic classification of each patient pre- and postoperatively. Spearman's rank correlation coefficient is 0.944. The line is the least-squares regression line fit of all the data points. The regression coefficient is –5.1.

IV. DISCUSSION

As illustrated on Figs. 3–5 this simple method of extraction successfully assessed the noise component of phonation. We can regard an H/N ratio smaller than 7.4 dB as pathological, based on the distribution of normal subjects (5% false-alarm rate). Moreover, 15 (83.3%) of our 18 preoperative samples had H/N ratios below 7.4 dB.

The S/N ratio reported by Kojima *et al.* (1980) ranged between 15.0 and 23.5 dB with a mean of 19.5 dB for 28 normal subjects. The S/N ratio was greater than the H/N ratio obtained in the present study. The major reason for this discrepancy seems to be a methodological difference between the two studies. Their results, based on discrete Fourier transforms of the signals during three pitch periods, might have changed if a different number of pitch periods had been employed. Their method measures as noise only those components between the harmonic components of the spectrum, while the present method additionally differentiates noise components which overlap the harmonic components.

Figure 8 is a scatter diagram of 18 preoperative and 23 postoperative samples. The abscissa represents the spectrographic classification while the ordinate represents the H/N ratio. These two variables appear to be highly correlated. Spearman's rank correlation coefficient, corrected for tied ratings, is 0.849 and is significant at $p = 0.001$. This confirms that the H/N ratio is a useful index to quantitatively assess the noise component relative to the harmonic component of the vowel. Further study is necessary to scrutinize the relationship between the H/N ratio and the psychophysical measurement of the degree of hoarseness.

In summary, the H/N ratio, developed in this study, provides an objective method for detecting vocal pathology by evaluating hoarseness, and for assessing the results of treatment for hoarseness. The H/N ratio is computationally less complex and less time consuming than previous methods. The procedures for obtaining H/N measures could be performed in the clinic or laboratory with the aid of a small, inexpensive computer. Moreover, the usefulness of the technique may be extended by further analyzing the spectral properties of the separated harmonic and noise components of the voice signal. It should be noted, however, that this H/N ratio method may not be applicable to extremely severe hoarseness.

ACKNOWLEDGMENTS

This work was supported in part by NIH grant NS-13870 to Haskins Laboratories. Eiji Yumoto carried out this research during a leave of absence from Ehime University.

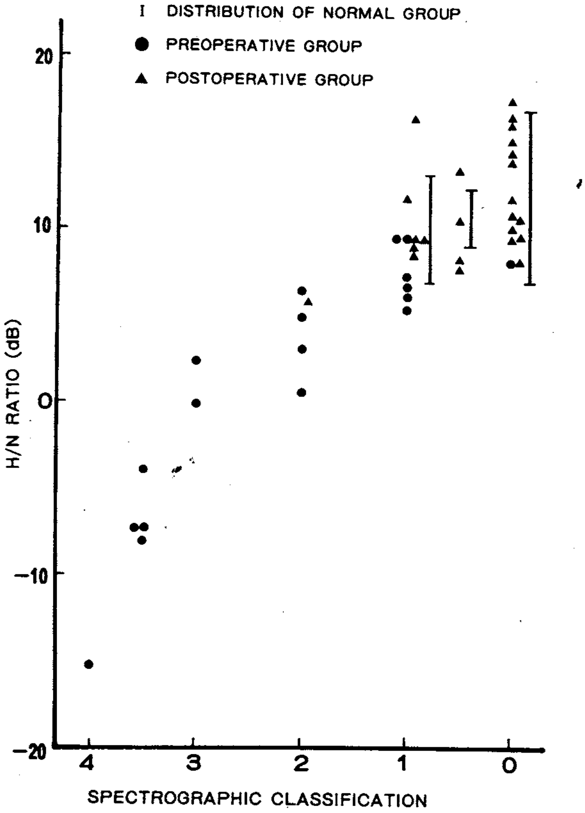

FIG. 8. Scatter diagram of the H/N ratio versus the spectrographic classification for pre- and postoperative subjects with varying degrees of hoarseness. Spearman's rank correlation coefficient is 0.849. Three vertical lines indicate the ranges of the measurements for normal subjects.

Davis, S. B. (1976). "Computer evaluation of laryngeal pathology based on inverse filtering of speech," Monograph 13, 1–247, Speech Communications Research Laboratory, Santa Barbara, CA.

Emanuel, F. W., and Sansone, F. E., Jr. (1969). "Some spectral features of 'Normal' and simulated 'Rough' vowels," Folia Phoniat. 21, 401–415.

Hanson, W., and Emanuel, F. W. (1979). "Spectral noise and vocal roughness relationships in adults with laryngeal pathology," J. Commun. Disord. 12, 113–124.

Hecker, M. H. L., and Kreul, E. J. (1971). "Descriptions of the speech of patients with cancer of the vocal folds. Part 1: Measures of fundamental frequency," J. Acoust. Soc. Am. 49, 1275–1282.

Hollien, H., Michel, J., and Doherty, E. T. (1973). "A method for analyzing vocal jitter in sustained phonation," J. Phon. 1, 85–91.

Iwata, S., and von Leden, H. (1970). "Pitch perturbations in normal and pathologic voices," Folia Phoniat. 22, 413–424.

Kojima, H., Gould, W. J., Lambiase, A., and Isshiki, N. (1980). "Computer analysis of hoarseness," Acta Otolaryngol. 89, 547–554.

Lieberman, P. (1961). "Perturbations in vocal pitch," J. Acoust. Soc. Am. 33, 597–603.

Lieberman, P. (1963). "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," J. Acoust. Soc. Am. 35, 344–353.

Lively, M. A., and Emanuel, F. W. (1970). "Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females," J. Speech Hear. Res. 13, 503–517.

Rontal, E., Rontal, M., and Rolnick, M. (1975). "Objective evaluation of vocal pathology using voice spectrography," Ann. Oto-Rhino-Laryngol. 84, 662–671.

Sansone, F. E., Jr., and Emanuel, F. W. (1970). "Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males," J. Speech Hear. Res. 13, 489–502.

Stevens, K. N. (1972). "The quantal nature of speech: Evidence from articulatory–acoustic data," in *Human Communication, A Unified View*, edited by E. E. David, Jr. and P. B. Denes (McGraw–Hill, New York), Chap. 3, pp. 51–66.

Yanagihara, N. (1967). "Significance of harmonic changes and noise components in hoarseness," J. Speech Hear. Res. 10, 531–541.