Objective assessment of hoarseness: Psychophysical measurement and acoustic analysis

Eiji Yumoto and Hiroshi Okamura

Department of Otolaryngology, School of Medicine, Ehime University
454, Shizukawa, Shigenobu-cho, Onsen-gun, Ehime, 791-02 Japan

(Received 19 December 1983)

As the degree of judged hoarseness increases, the noise component of the spectrogram increases and replaces the harmonic structure. The relationship between these two components has been quantified as harmonics-to-noise (H/N) ratio. In this study, the contribution of cycle-to-cycle pitch perturbations (jitter) to the measured noise energy was analyzed. Moreover, the relationship between the acoustic parameters (H/N ratio and jitter) and the perceived abnormalities of hoarseness was studied. Eight laryngologists rated their auditory impressions for G, R and B factors of 87 pathologic voices. G (Grade) represents the degree of hoarseness; R (Rough) refers to rumbling and heaviness; B (Breathy) is related to air leakage during phonation. Correlations of R factor with the H/N ratio or jitter were not so high as those of G or B factor with the H/N ratio (p<0.05). These findings suggest that the H/N ratio can be a quantitative index of B factor as well as of G factor, and that neither of the acoustic parameters are good indicators of R factor.

PACS number: 43.66.Lj, 43.70.Gr, 43.70.Ny

1. INTRODUCTION

Hoarseness is a general term used for describing perceived abnormality of the voice. Sound spectrographic analysis has revealed that as the degree of hoarseness judged by listeners increases, the noise component appears to a greater degree and replaces the harmonic structure.

One of the authors has developed a new method of calculating a ratio of the acoustic energy of the stable harmonic components to that of the noise component (H/N ratio), as a practical index of the degree of hoarseness. This H/N ratio has been proved to be highly correlated with the subjective evaluation of the spectrogram. This H/N ratio method is computationally less complex and less time consuming than previous methods. Thus, the procedures for obtaining H/N measures can be performed in a clinic with the aid of a small, inexpensive computer. Moreover, the H/N ratio method is able to detect cycle-to-cycle pitch perturbations (jitter). One of the purposes of this paper is to examine the contribution of jitter to the obtained H/N ratios.

Most laryngologists rate the degree and quality of hoarseness to assess the results of treatment for laryngeal disorders. Such subjective judgments without psychophysical measurements have yielded many confusing terminologies about the qualities of hoarseness—hoarse, harsh, rough, creaking, rumbling, breathy, and so on. The Japan Society of Logopedics and Phoniatrics (JSLP) has advocated evaluating the following five factors: Grade (G), Rough (R), Breathy (B), Strained (S), and Asthenic (A). “Grade” represents the degree of hoarseness or how far the voice deviates from a normal voice, while the other four factors represent the different qualities of hoarseness. “Rough” voice is one perceived as rumbling, dull and heavy; “Breathy” voice is related to air leakage during phonation;
“Strained” voice is characterized by the adjectives “tight, tense and creaky”; “Asthenic” refers to a quality of voice described as “thin, poor and light.” A second purpose of this paper is to investigate the relationship between acoustic parameters and a psychophysical evaluation of hoarseness. An $H/N$ ratio and jitter were obtained as acoustic parameters, while G, R and B factors were rated as the psychophysical evaluation of hoarseness. We chose R and B factors from the four qualities of hoarseness, because they are used more often than others by laryngologists.

2. H/N RATIO METHOD

Since the procedure in detail for obtaining $H/N$ measures has been reported previously, it will be only briefly described in this paper.

It was assumed 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 original voice signal, $f(t)$, was separated into pitch periods, $f_i(t)$, as shown in Fig. 1 ($1 \leq i \leq n$). Each period is defined to begin at a zero-crossing point (shown by arrows in Fig. 1) before its prominent peak.

For the purpose of averaging consecutive periods, $f_i(t)$ was assumed to be equal to zero in the interval between the duration of $i$-th period, $T_i$, and the maximum of $T_i$, $T$ (shown by thicker lines in Fig. 1). The resulting average wave, $f_\text{A}(t)$, is a good estimate of the signal that gives rise to the harmonic component when $n$ is sufficiently large. Evidently, another significant acoustic factor of hoarseness is jitter. Our method for calculating $f_\text{A}(t)$ allows jitter to contribute to the obtained acoustic energy of the noise component.

The measure of the acoustic energy of the harmonic component, $H$, is defined as

$$H = n \sum_{0}^{T} f_\text{A}^2(t) \, dt \quad (0 \leq \tau \leq T) \tag{1}$$

where

$$f_\text{A}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i(t)$$ \tag{2}

The energy of the noise component, $N$, is equal to the mean energy of the differences between the individual periods and the average waveform, and is expressed as

$$N = \sum_{0}^{T} (f_i(t) - f_\text{A}(t))^2 \, dt \quad (0 \leq \tau \leq T) \tag{3}$$

The ratio of $H$ to $N$ is derived from Eqs. (1) and (3). As has been reported previously, the noise derived from the measurement system did not contribute significantly to the obtained $H/N$ ratio.

3. PROCEDURES

3.1 Contribution of Jitter to the $H/N$ Ratio

The sinusoidal waves, pitch periods of which varied in random order within $\Delta t$ ms, were synthesized in a small computer (Hitachi, MB 6890). Average pitch periods of the sinusoidal waves were 10 ms, 6.67 ms, 5 ms, and 4 ms. $\Delta t$ varied every 0.05 ms from 0.05 ms to 0.75 ms. Amplitudes of pitch periods in a sinusoidal wave were kept con-
E. YUMOTO and H. OKAMURA: ASSESSMENT OF HOARSENESS

stant. Three different series of random numbers were used to arrange variation of pitch periods of the wave within $\Delta t$ ms. The mean of the differences between adjacent instantaneous frequencies was calculated from the synthesized wave and was divided by the mean frequency, $F_a$. The resulting value, $f$, was used as an index of jitter, where

$$f(\%) = 100 \times \frac{1}{n-1} \sum_{i=1}^{n-1} |F_i - F_{i+1}|/(n-1) \times F_a$$

(4)

The $H/N$ ratio of the synthesized wave was also ciphered out.

3.2 Phonatory Samples

Eighty-seven phonatory samples, which had been used by one of the authors in his earlier work, were analyzed. These samples consisted of pre- and post-operative voices and varied from nearly normal to severely hoarse.

3.3 Measurement of the Acoustic Parameters

High-frequency pre-emphasis was applied to the voice signal. This signal was low-pass filtered at 10 kHz (anti-aliasing filter) and digitized with 12-bit precision at a sampling rate of 20 kHz for 3 sec. 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 the terminal portions of the phonation were excluded. A laboratory computer (Digital Equipment Corp., PDP 11/45) was utilized for processing the voice signal. Either a semi-automated or manual method extracted consecutive pitch periods for averaging. We calculated the $H/N$ ratio from 50 pitch periods by using Eqs. (1) and (3), and converted it to a decibel scale. The index of jitter, $f$, was also calculated by using 50 pitch periods extracted for $H/N$ measures.

3.4 Psychophysical Evaluation of Hoarseness

The analyzed portion of each of the original phonatory samples, stored in the computer, was lengthened by repetition until the signal became 1.5 sec in duration, and was then recorded twice on to a magnetic tape for each of the 87 voices. The voices recorded on this tape were rerecorded to randomize their order. A blank between successive samples was about 5 sec, and the intensities of the different voices were adjusted to the same level.

The listening tape was reproduced through a cassette-recorder (SONY, TC 3000 SD), an amplifier (PIONEER, SA-8800 II), and a speaker (YAMAHA, NS-10 M) in a sound-treated booth. Eight trained laryngologists rated their auditory impressions for G, R and B factors, according to a 4 point equal-appearing-intervals scale (0: normal, 1: slight, 2: moderate, 3: severe). The judges listened to the standard tape available from JSLP before evaluating our tape.

4. RESULTS AND DISCUSSION

4.1 $H/N$ Ratio and Jitter

Figure 2 illustrates the contribution of jitter to an $H/N$ ratio of a synthesized sinusoidal wave. As jitter increased, the acoustic energy of the noise component grew more and, therefore, the obtained $H/N$ ratio became smaller. The shadowed portion in Fig. 2 represents a small variation of the result due to the different series of random numbers. Even if no additional noise component exists, the sinusoidal wave with jitter of 1.5%, for example, shows an $H/N$ ratio of about 10 dB.

Furthermore, we can obtain a ratio, $x$, of the energy of an additional noise component, $N_a$, to the noise energy derived from jitter, $N_j$, where

$$N = N_a + N_j$$

(5)

$$x = N_a/N_j$$

(6)

$N$: noise energy measured from an acoustic wave.

(under the assumption that $N_a$ and $N_j$ are not correlated)

From Fig. 2, for example, the phonatory sample with a jitter index of 0.71% (shown in Fig. 3, left) would have and $H/N$ ratio of 16 dB if an additional noise, $N_a$, were equal to zero.

$$H/N_j = 10^{0.8} = 39.8$$

(7)

However, the actual $H/N$ ratio is 0.24 dB. We have to take an additional noise, $N_a$, into account in order to explain the difference between the $H/N$ ratio predicted from Fig. 2 and the actual one.

$$H/(N_a + N_j) = 10^{0.08} = 1.06$$

(8)

Equations (6), (7) and (8) were arranged as

$$39.8/(1+x) = 1.06$$

(9)

Therefore, the ratio, $x$, of this phonatory sample is
Fig. 2 Contribution of jitter to the $H/N$ ratio.

Fig. 3 Spectrograms of two phonatory samples. The phonation shown in the left has the jitter index of 0.71% and the $H/N$ ratio of 0.24 dB, while the one in the right has the jitter index of 0.87% and the $H/N$ ratio of 5.63 dB. Ratios of an additional noise component to a noise derived from jitter were 36.5 and 7.7, respectively.

36.5. The spectrogram shown in Fig. 3, right, represents a phonation with a jitter index of 0.87%. The $H/N$ ratio of this phonation is 5.6 dB. Calculating in the same manner as the previous example, the ratio, $x$, of this phonation is 7.70. This method of obtaining a ratio, $x$, is based on Fig. 2 which illustrates the contribution of jitter to the $H/N$ ratio of a synthesized sinusoidal wave. But, the actual voice signal does not consist of a single sinusoidal wave. In this sense, the obtained ratio, $x$, may not be exactly equal to the actual value. However, the $H/N$ ratio method is the first one by which we can infer the relation of an additional noise component to a noise derived from jitter.
4.2 Reliabilities of the Psychophysical Measurement of Hoarseness

The inter-judge reliability of the evaluation was tested by calculating the Spearman's rank correlation coefficients between judges. The coefficients regarding G factor ranged between 0.511 and 0.785; the ones regarding R factor ranged between 0.364 and 0.731; the ones regarding B factor ranged between 0.416 and 0.672; all were statistically significant \((p < 0.01)\).

Two randomly selected judges were asked to rate G, R and B factors again one month after the first rating. Twenty-two voices, the rating of which were variable among the judges, were chosen for computing the Kendall's coefficients of concordance. The coefficients among eight judges for G, R and B factors were 0.252, 0.349 and 0.286, respectively, and all were statistically significant \((p < 0.001)\). The coefficients between two occasions for two judges were 0.832 and 0.893 for G factor (intra-judge reliability) and were statistically significant \((p < 0.05)\); the coefficients for R factor were 0.773 \((p < 0.05)\) and 0.909 \((p < 0.02)\); the coefficients for B factor were 0.698 \((p < 0.10)\) and 0.845 \((p < 0.05)\). It was suggested that the rating for B factor might be less reliable than the one for G and R factors. We added eight scores to get one rating for each factor of the phonatory sample. This means that 24 points represent the severest maximum of each factor while 0 represents the clearest one.

4.3 Relationship between the Psychophysical Evaluation of Hoarseness and the Acoustic Parameters

Figures 4-9 illustrate scatter diagrams of G, R or B factors versus the H/N ratio or jitter. A logarithmic transformation of the jitter index, \(f\), was used in these figures. In order to help to describe the relation between these measurements, correlation coefficients \((\text{Pearson } r)\) were calculated and were summarized in Table 1. All were statistically significant at the 0.001 level. When we compare these correlation coefficients with each other, we need to allow for the fact that the auditory evaluation of hoarseness (G, R and B factors), or the acoustic parameters (H/N ratio and jitter) are not independently distributed, but correlated. The analysis, taking the above fact into account, revealed the following: the correlations of G, R or B factors with the H/N ratio were significantly higher than those with jitter \((p < \ldots\).
Fig. 6 Scatter diagram of the $H/N$ ratio versus the rating of R factor.

Fig. 7 Scatter diagram of jitter versus the rating of R factor.

Fig. 8 Scatter diagram of the $H/N$ ratio versus the rating of B factor.

Fig. 9 Scatter diagram of jitter versus the rating of B factor.
0.05). The correlation of G factor with the H/N ratio was not significantly different from that of B factor with the H/N ratio (p = 0.05). Both of the correlations were significantly higher than that of R factor with the H/N ratio (p < 0.05). The correlation of G factor with jitter was not significantly different from that of B factor with jitter (p = 0.05). The latter was not significantly different from the correlation of R factor with jitter either (p = 0.05). But, the correlation of G factor with jitter was significantly higher than that of R factor with jitter (p < 0.05).

These results show that an H/N ratio can be a quantitative index of B factor as well as of G factor. This is not unexpected because of the following two reasons. First, an H/N ratio has been originally developed as a quantification of an additional noise component compared to a harmonic component. Secondly, G factor was reported to correlate with B factor much more than with other factors of hoarseness. Wendahl and Coleman reported that the perceived magnitude of roughness was strongly related to the magnitudes of stimulus jitter, that is, the correlation was near unity. The stimuli used by them were an electrical analog of the larynx, in which various physical properties of the stimuli were controlled. On the other hand, a number of physical properties of the sustained phonation in the present study were not under control. In this study, correlations of R factor with the H/N ratio or jitter were significant. But, these correlations were not so high as the correlation of G or B factors with the H/N ratio. Hence, we do not suggest that both acoustic parameters are good quantitative indicators of R factor. R factor may be related more with the acoustic energy in the lower frequency range of the noise selected from the original phonation. Thus, the usefulness of the H/N ratio method may be extended by further analysis of the relationship between the psychophysical evaluation of hoarseness and the spectral properties of the separated harmonic and noise components of the voice signal.

5. CONCLUSION

(1) The contribution of jitter to the measured noise energy was quantitatively analyzed in the synthesized sinusoidal wave. Based on this result, the relation of an additional noise to a noise derived from jitter was examined in the actual voice signal.

(2) Correlations of R factor with the H/N ratio or jitter were not so high as those of G or B factors with the H/N ratio (p < 0.05).

(3) The H/N ratio can be a quantitative index of B factor as well as of G factor, and that neither of the acoustic parameters (H/N ratio and jitter) are good indicators of R factor.

REFERENCES