Ann.Bull.RILP(1985)No.19, 179-190

ACOUSTIC MEASURES OF PATHOLOGICAL VOICE QUALITITES - ROUGHNESS -

Satoshi Imaizumi

INTRODUCTION

We have investigated acoustic measures of pathological voice qualities to develop an assessment system of vocal function in speech. Based on several acoustic analyses[1,2,3,4], we showed that the most important parameters for characterizing pathological voice samples are 1) extent of fundamental frequency fluctuation; 2) extent of amplitude fluctuation; 3) amount of the high frequency component; and 4) amount of noise. Via perceptual evaluation of pathological voice[5,6] using a "GRBAS" scale[7], we also found that these parameters correlate strongly with perceived voice qualities. Here, the "GRBAS" scale is a set of rating scales for evaluating hoarseness and consists of five characteristics: "grade (G)", "rough (R)", "breathy (B)", "asthenic (A)" and "strained (S)".

Through this research, we reached a hypothesis that "roughness" comes mainly from multiplicative variations or modulations in the pitch period, in the amplitude and/or in the waveform, whereas "breathiness" is mainly from additive noise components. The aim of this paper is to substantiate this hypothesis based on perceptual evaluation (Experiment I) and acoustic analysis (Experiment II). The special focus here is on the acoustic correlates of the "roughness".

EXPERIMENT I: PERCEPTUAL EVALUATION

Procedure of Perceptual Evaluation

Voice samples from 90 patients with various kinds of laryngeal pathology and 8 normal speakers were used. They were randomly selected from the voice samples of middle aged male speakers in our audiotape library.

Voice samples for /e/ were digitized through a 12-bit A/D converter at a sampling rate of 20 kHz and stored on a disk controlled by a computer. 0.5sec segments were extracted by excluding the initial and final portions from each sample. These segments were recorded on a listening tape in random order.

Six subjects with normal hearing served as the listeners. All of them were familiar with the "GRBAS" scale. The tape was presented through an audiometer at a comfortable level of about 50 dBSL.

To measure the degree of "roughness", for instance, each listener rated each voice sample on the scale shown in Fig. 1.

The listener marked a bar-type sign so that the position of the sign was in proportion to his perceptual degree of "roughness". The length from the left edge to the mark was measured and defined as the rating score for "roughness". We denote this score as R score in this paper. Similar rating scales were used for the other characteristics of the "GRBAS" scale.

Each subject performed the listening test five times at a rate of once per day. The relationships between the rating scores on the "GRBAS" scale were examined using a principal component analysis for each subject.

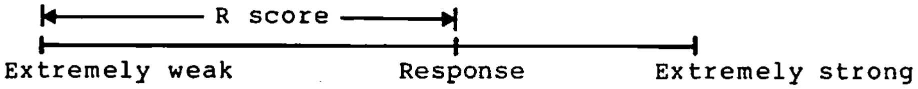


Fig. 1. The rating scale used to measure the degree of "roughness". The rating score was defined as the length of the bar drawn by the subject.

Results and Discussion of the Perceptual Evaluation

Two major components were extracted by the principal component analysis of the rating scores for each subject. The contribution of the first principal component was around 50%, and that of the second principal component was about 25%. Those of the third principal component or those above were each less than 10%. These rates were almost the same for all of the subjects.

The results of the pricipal component analysis are shown in Fig.2 for the three subjects 1, 2 and 3. The relationships among the five ratings by a subject on five characteristics (G, R, B, A, S) are represented by the vectors shown in these figures. Each one of the five vectors marked R, for instance, corresponds to one of the five ratings of R made by the subject. The horizontal (or vertical) coordinate of each vector indicates the standardized multiple regression coefficient of the corresponding ratings on the first (or second) principal component. A smaller angle between vectors indicates a closer correlation between the corresponding ratings.

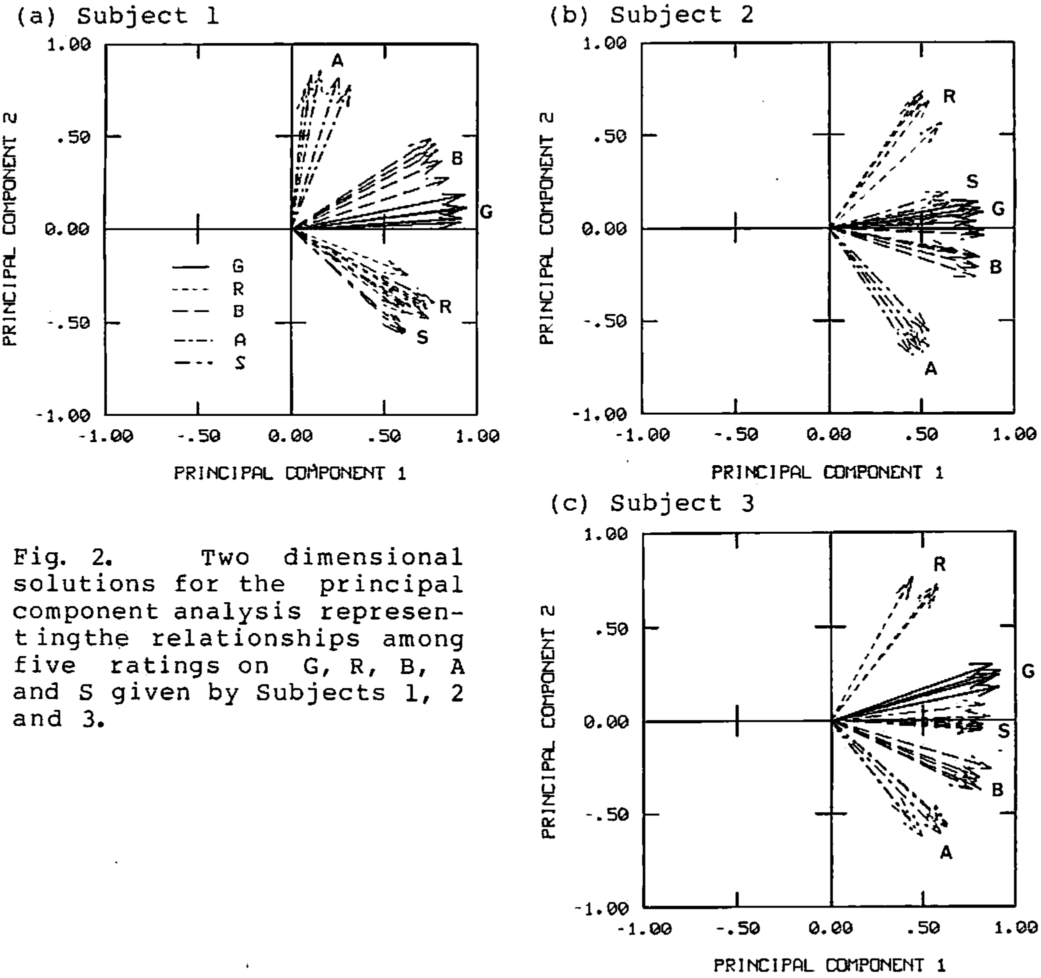
The polarity of the second principal component in Fig. 2 (a) is opposite that in Fig 2. (b) and (c). In each figure, however, the vectors with the same sign gather together to make a group. This indicates that the reproduciblity of the five ratings by each subject was rather high and that the perceptual evaluation was reliable.

The mutual relationships between the vectors R, G, B and A are almost the same for all subjects as shown in Fig. 2. The vectors S, however, show different relationships from the others among the subjects. This indicates that the subjects have a common interpretation for R, G, B and A, but not for S.

The vectors R and B each constitute a group. and they meet at almost right angles in Fig. 2. This means that the R scores and the B ones are almost independent of each other. In other words, the subjects clearly discriminated the "roughness" from the "breathiness". There were some voice samples which were consistently assigned large R scores and small B ones, while others which were rated conversely.

Conclusion of Perceptual Evaluation

From the results of the perceptual evaluation, it was found that the listening subjects clearly discriminated "roughness" from "breathiness". Among the voice samples used, some voice samples were assigned large R scores and small B ones, while others were rated conversely. Therefore, we can conclude that these voice samples compose a suitable set to investigate the acoustic correlates of "roughness" and "breathiness".



(a) Subject 1

(b) Subject 2

(c) Subject 3

Fig. 2. Two dimensional solutions for the principal component analysis representingthe relationships among five ratings on G, R, B, A and S given by Subjects 1, 2 and 3.

EXPERIMENT II: ACOUSTIC ANALYSIS

Procedure for Acoustic Analysis

The voice samples used in Experiment I were analyzed. First, pitch periods were extracted from a low-pass filtered waveform[8] as shown in Fig. 3. Using a peak picking method[9], local maximum points which could correspond to vocal excitation epochs were detected successively. Here, we write L(i) for the i-th pitch location, A(i) for the amplitude at L(i), and X(n), L(i)-1<n<L(i+1) for the original voice waveform within the i-th pitch period.

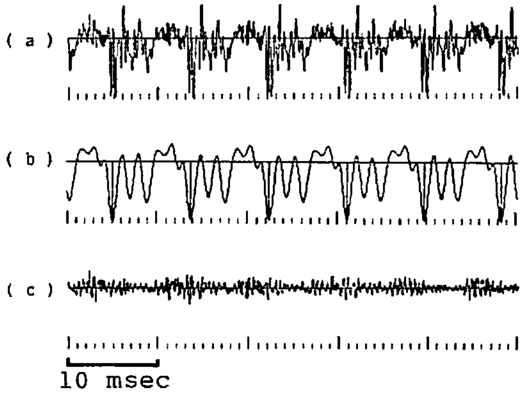


Fig. 3.

(a) A voice sample with a large R score.
(b) Lowpass filtered waveform and extracted pitch epochs (vertical lines).
(c) Additive noise component. See text.

Then, we calculated the correlation coefficient R(i,j) between the original voice waveform within the i-th pitch period and that within the j-th pitch period, that is,

$$
\begin{aligned}
&R(i,j)=\mathrm{CORR}(X(ni),X(nj),K)\\
&ni=L(i),L(i)+1,...,L(i)+K, \qquad (1)\\
&nj=L(j),L(j)+1,...,L(j)+K,\\
&K=\min(L(i+1)-L(i),L(j+1)-L(j))-1.
\end{aligned}
$$

Here, CORR(X(n1),Y(n2),K) indicates the correlation coefficient between the two variables of length K, X(n1) and Y(n2). R(i,j) was calculated for i=1,2,...,I-1, and j=i+1,i+2,...,min(i+10,I) where I was the total number of the pitch periods.

As shown in Fig.·4 (b) and Fig. 5 (b), R(i,j) showed more or less a periodical variation. We could detect the minimum Rmin(i,jmin) at j=jmin, and the maximum Rmax(i,jmax) at j=jmax where jmin<jmax. To measure the periodicity of the waveform variation in the range over several pitch periods, the waveform modulation index WMI was defined as

$$
\mathrm{WMI}=\frac{1}{\mathrm{NIp}}\sum_{i\in \mathrm{Ip}}(\mathrm{Rmax}(i,\mathrm{jmax})-\mathrm{Rmin}(i,\mathrm{jmin})) \qquad (2).
$$

Here, Ip was a set of the pitch periods for which the maximum and the minimum were detected, and NIp was the number of elements in the set Ip. Then, the waveform modulation frequency WMF was defined as

$$WMF = \frac{1}{NIp} \sum_{i \in Ip} SF/(L(jmax)-L(i)) \qquad (3),$$

where SF was the sampling frequency, that is, 20000.

To measure the periodicity of the pitch period perturbation and that of the amplitude perturbation, correlograms[10) Cp(m) and Ca(m) were calculated as follows. That is,

$$Cp(m)=CORR(P(i),P(i+m),I-m) \qquad (4),$$

and

$$Ca(m)=CORR(A(i),A(i+m),I-m) \qquad (5).$$

where $i=1,2,\ldots,I$, $P(i)=L(i+1)-L(i)$, and $m=1,2,\ldots,40$.

As shown in Fig. 4 (c) and Fig. 5 (d), Cp(m) and Ca(m) had the local minimums Cpmin and Camin respectively, and the local maximums Cpmax and Camax. Using these values and their locations, we defined the pitch modulation index PMI, the pitch modulation frequency PMF, the amplitude modulation index AMI and the amplitude modulation frequency AMF in a similar way as WMI and WMF were defined.

To measure the extent of the pitch period perturbation and that of the amplitude perturbation, the pitch perturbation quotient PPQ and the amplitude perturbation quotient APQ were calculated after Koike's formulation[11).

Results of the Acoustic Analysis

The results of the acoustic analysis for two voice samples rated as strongly "rough" are shown in Figs. 4 and 5. The waveforms of both examples vary periodically. The modulation cycle of the waveform in Fig. 4 (a) is around 3 or 4 pitch periods, and that of the waveform in Fig. 5 (a) is 2 pitch periods. These modulation cycles are clearly demonstrated by R(i,j) as shown in Figs. 4 (b) and 5 (b). As can be seen in Figs. 4 (c) and 5 (c), the correlograms Cp(m) and Ca(m) indicate the existence of these modulations in the pitch period and also in the amplitude. The correlogram Cp(m) in Fig. 4 (c) demonstrates the existence of another modulation which has a very long period.

The scatterplot for the 98 voice samples on the WMI (Waveform Modulation Index) versus the WMF (Waveform Modulation Frequency) plane is shown in Fig. 6. In this figure, the area within each circle is in proportion to the median of the five R

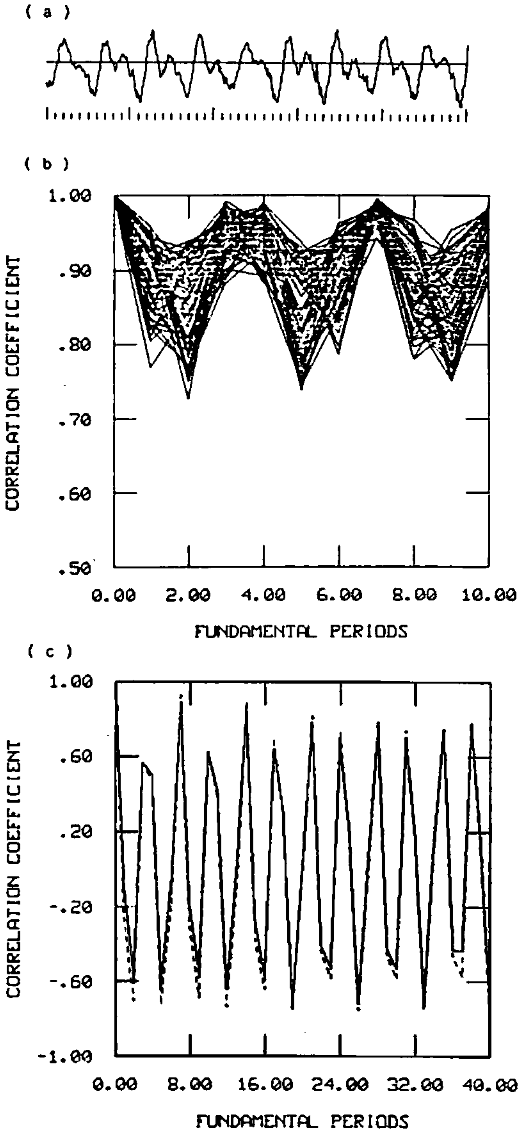


Fig. 4.
(a) A voice sample with a large R score.
(b) Correlation coefficient R(i,j) between the waveform within the i-th pitch period and that within the j-th pitch period.
(c) Correlogram of the pitch period Cp(m) (----) and that of the amplitude Ca(m) (——).

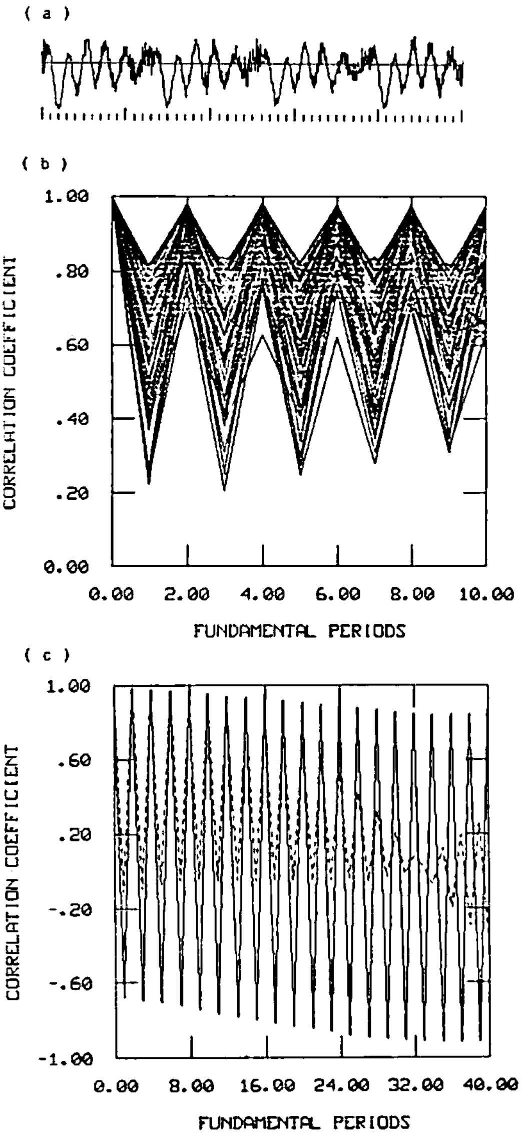


Fig. 5.
Another example similar to that in Fig. 4.

scores produced by subject 1. A larger circle represents a greater median for R scores. In Fig. 6, the large circles are in the region where WMI and WMF are large. This indicates that the voice samples with large values for WMI and WMF were strongly perceived as "rough". Some relatively large circles are found in the region where WMI is small. This means that some voice samples were perceived as "rough" to some extent even if their values for WMI were small.

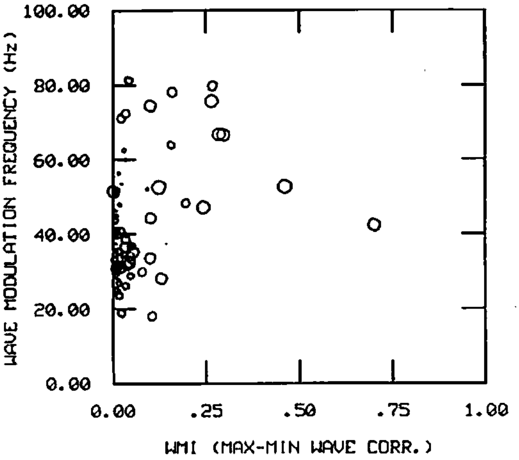


Fig. 6.
Scatterplot for the 98 voice samples on the WMI versus the WMF plane.
The median of the five R scores given by Subject 1 is represented by the size of the circle.

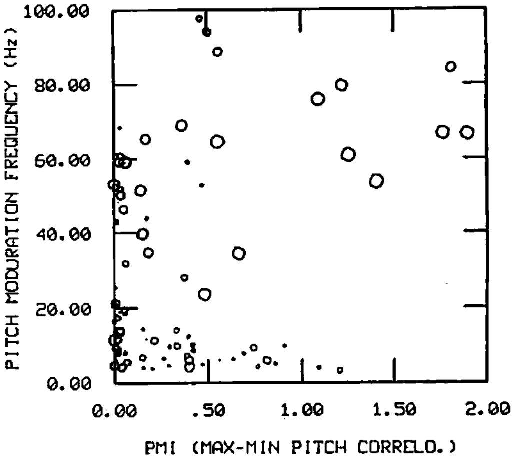


Fig. 7.
Scatterplot for the 98 voice samples on the PMI versus the PMF plane.
The median of the five R scores given by Subject 1 is represented by the size of the circle.

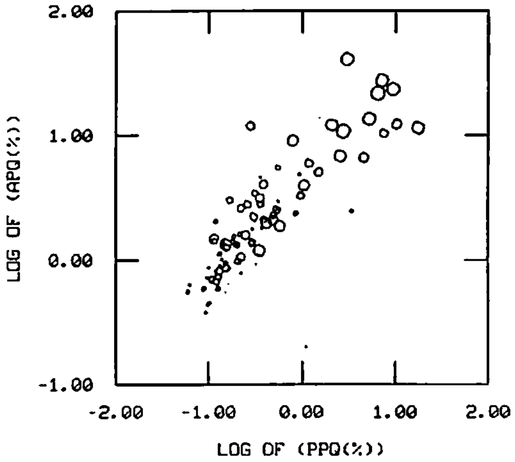


Fig. 8.
Scatterplot for the 98 voice samples on the PPQ versus the APQ plane.
The median of the five R scores givenby Subject 1 is represented by the sizeof the circle.

The scatterplot for the voice samples on the PMI (Pitch Modulation Index) versus the PMF (Pitch Modulation Frequency) plane is shown in Fig. 7. This figure shows a similar tendency to that in Fig. 6. That is, the voice samples with large values for PMI and PMF were perceived as strongly "rough", but others were perceived as "rough" even if PMI and PMF are small. In Fig. 7, some small circles are found in the region where PMF is less than 20 even if PMI is larger than, for instance, 0.5. This indicates that the voice samples which had a small PMF were not necessarily strongly perceived as "rough" even if they had a large PMI. A similar tendency was observed for AMI and AMF.

Fig. 8 shows the scatterplot for the voice samples on the PPQ (Pitch Perturbation Quotient) versus the APQ (Amplitude Perturbation Quotient) plane. Large circles are found in the region where PPQ and APQ are rather large. However, some relatively large circles can be found in the region where PPQ and APQ are small. This indicates that the voice samples with a large PPQ or APQ were strongly perceived as "rough". Some voice samples with a small PPQ and APQ, however, could be perceived as "rough" to some extent.

As in the results of the acoustic analysis described above, we could find some voice samples which were rated as "rough" to some extent even though they did not possess any significant periodical modulation in their pitch period, amplitude or waveform. Based on a careful inspection, the acoustic characteristics of these voice samples could be classified into three types: Type 1) voices with partial modulation; Type 2) voices with very low pitch; and Type 3) voices consisting of the alternative repetition of acoustically different segments.

The voice samples of Type 3 seemed to consist of the alternative repetition of two segments; one was a segment in which the formants' oscillations were dominant and the other was one in which the noise component was dominant. To confirm this conclusion, these voice samples were divided into their harmonic and additive noise components using a pitch synchronous averaging method[4,12]. An example is shown in Fig. 3. The additive noise component shown in Fig. 3(c) clearly shows an amplitude modulation which is synchronous with the pitch period of the voice. And this noise component is dominant in the segment in which the formants' oscillations are in decay.

Discussion of the Acoustic Analysis

The results of the acoustic analysis described above indicate that the voice samples with periodical variations in their pitch period, amplitude or waveform were actually strongly perceived as "rough", unless the frequencies of the variations were below about 20Hz.

Because the acoustic measures used here treat only the periodical variations observed over several pitch periods, variations which are synchronous with the vocal pitch period might be ignored. Actually, we found some voice samples which were rated as "rough" to some extent even though they did not possess any significant modulation observed over several pitch periods. Careful attention was paid to these samples, because their existence could disconfirm the hypothesis that "roughness" results mainly fróm multiplicative variations or modulations in the pitch period, amplitude and/or waveform. These voice samples could be classified into three types based on visual inspection.

The voice samples of Type 1 were partially modulated. The R scores for these voice samples can be determined mainly by such partial modulations. In other words, the modulated portion is more significant than others for perceptual evaluation. On the other hand, the acoustic parameters used here were the values averaged over the whole waveform. This could be the reason why the R scores for these voice samples are large although the modulation indices and the perturbation quotients are small.

The voice samples of Type 2 had very low fundamental frequencies. The large R scores for these samples can be interpreted as follows. The voice waveform of /e/ is the repetition of the formants' oscillations which decay at the rates determined by the formants' bandwidths. If a voice sample has a lower fundamental frequency or a longer pitch period, then the decay in one pitch period becomes larger. And the decay becomes more rapid when the glottis opens. Therefore, if the opened phase of the glottis is long enough, the formants' oscillations can even decay completely. In other words, the voice samples with a lower fundamental frequency have a deeper modulation in the amplitude envelope. This phenomenon can be interpreted as a periodical variation, which is synchronous with the vocal pitch period.

The voice samples of Type 3 consisted of the alternative repetition of two segments; one was a segment in which the formants' oscillations were dominat and the other was one in which the noise component was dominant. This phenomenon can also be considered a periodical variation, which is synchronous with the vocal pitch period.

As discussed above, these voice samples of Type 2 and 3 are not independent of periodical variations. These variations are synchronous with the vocal pitch period. Therefore, the existence of these voice samples does not disconfirm the hypothesis that "roughness" in voice results mainly from multiplicative variations or modulations in the pitch period, amplitude and/or waveform.

The existence of Type 3 voice samples is very interesting in terms of a consideration of the method of noise generation at the glottis. For the example shown in Fig. 3, the noise component clearly demonstrates an amplitude modulation which is synchronous with the pitch period of the voice. And this noise component is dominant in the segments in which the formants' oscillations have almost completely decayed. Since it is theoretically reasonable to consider these segments as corresponding to the opened or relatively opened phase of the glottis, the noise component must be amplified when the glottis is in its opened phase. Such a noise component might be generated by an insufficient glottic "opening" instead of by an insufficient glottic "closing". It should be noted that the additive noise component with an amplitude modulation can cause the "rough" sensation instead of the "breathy" sensation.

An interesting question for us is how and why periodical variations within a certain frequency range can cause the "rough" sensation. Many hearing researchers have investigated "roughness". According to the review given by Plomp and Levelt[13], Helmholtz investigated the degree of dissonance for simultaneous tones and concluded that it is determined by the "roughness" of rapid beats between the tones. He ascertained that this "roughness" has a maximum for a frequency difference of 30-40Hz. Plomp and Levelt partially agreed with Helmholtz's finding, although, they found that the "roughness" has a maximum for a frequency difference of about a quarter of the critical bandwidth. They also showed that the "roughness" appears only for tones at a frequency distance not exceeding the critical bandwidth and concluded that the total dissonance (or "roughness") of a complex tone is the sum of the dissonances of each pair of adjacent partials. Ohgushi[14] has recently shown that the degree of "roughness" increases if the number of the harmonics increases. Their point of view can be a good explanation for our results. Modulations in the pitch period or in the amplitude generate subharmonics, and then the voices with such modulations have similar spectra to those consisting of a lot of partials which are closer than the critical bandwidth to each other. Voice samples with very low fundamental frequencies are also in the same situation. For these voice samples, many partial tones are close enough to be at a frequency distance not exceeding the critical bandwidth, especially within the high frequency range.

Fastl[15] showed that the degree of "roughness" can be predicted from a model based on temporal masking patterns. He found that the degree of "roughness" for sinusoidally amplitude modulated broadband noise is in proportion to the value of the depth of modulation in the temporal masking pattern multiplied by its modulation frequency. Similar relationships were found for amplitude modulated tones[16] and also for frequency modulated ones[17]. Fastl and others have found that modulated tone or noise causes a "fluctuation" sensation instead of a "rough" sensation when the modulation frequency is below around 30Hz. Although the perceptual effects of the voice waveform might be much more complicated than those of simple tones or broadband noise, their results seem to support our interpretation of "roughness", especially for voice samples of Type 2 and 3. Voice samples of Type 2 contain modulations in the amplitude envelope. And those of Type 3 possess amplitude modulated noise.

CONCLUSION

Acoustic correlates of "roughness" in pathological voice were investigated using acoustic analysis and perceptual evaluation. The following results were obtained. Voice samples which possessed multiplicative variations or modulations over several pitch periods were strongly perceived as "rough", if the modulation frequencies were higher than about 20 Hz. Voice samples which contained acoustically different segments in each pitch period and those with very low fundamental frequencies were also perceived as "rough" to some extent. This means that "roughness" is connected not only with the multiplicative variations which occur over several pitch periods but also with those which are synchronous with the vocal pitch period.

Acknowledgements

The experiment I described in this paper was started with the co-operation of Prof. F. Ohta, Dr. H. Yonekawa and Ms. H. Abe, at the Department of Otolaryngology, School of Medicine, Kinki University. I would like to express my gratitude to them for their help.

References

1. S. Hiki, S. Imaizumi, M. Hirano, H. Matsushita and Y. Kakita: "Acoustic Analysis for Voice Disorders," Conference Record of the 1976 ICASSP, (Canterbury Press Rome, NY, 1976), 613-616.
2. Y. Kakita, M. Hirano, H. Matsushita, S. Hiki and S. Imaizumi: "Acoustic Parameters Relevent to Diagnosis in Voice Disorders," Pract. Otol. (Kyoto), 70:4, 269-276 (1977).
3. S. Imaizumi, S. Hiki, M. Hirano and H. Matsushita: "Analysis of Pathological Voices with a Sound Spectrograph," J. Acoust. Soc. Jpn., 36:1, 9-16 (1980).
4. Y. Kakita, M. Hirano, S. Imaizumi and S. Hiki: "Discrimination of Pathological Voice Based on Acoustic Analysis," Medical Electronics and Biological Engineering, 18:6, 405-412 (1980).
5. S. Hiki and S. Imaizumi: "Perceptual Evaluation of Pathological Voices Using Sustained Vowel Phonation," in Clinical Examination of Voice, Jpn. Soc. Logopedics Phoniatrics, Eds. (Ishiyaku Shuppan, Tokyo, 1979) , Chap. 7, 181-209 (in Japanese).
6. S. Imaizumi, S. Boku, Y. Koike and F. Ohta: "Multidimensional Analysis of Alaryngeal Voice Quality," J. Acoust. Soc. Jpn(E)., 4:3, 139-148 (1983).
7. M. Hirano: "Clinical Examination of Voice", (Springer-Verlag, Wien, 1981), Chap. 6, 81-84.
8. J. Dobnowski, R. Schafer and R. Rabinar: "Real-time Digital Hardware Pitch Detector," IEEE, ASSP-24, 2-8 (1976).
9. H. Kasuya, S. Ebihara, T. Chiba and T. Konno: "Characteris-

tics of Pitch Period and Amplitude Perturbations in the Speech of Patients with Laryngeal Cancer," Trans. IECE of Jpn, J65-A:5, 423-430(1982).

10. Y. Koike: "Vowel Amplitude Modulations in Patients with Laryngeal Diseases," J. Acoust. Soc. Amer., 45, 839-844 (1969).
11. Y. Koike: "Application of Some Acoustic Measures for the Evaluation of Laryngeal Dysfunction," Studia Phonologica, VII 17-23 (1973).
12. E. Yumoto, W. Gould and T. Bear: "Harmonics-to-noise Ratio as an Index of the Degree of Hoarseness," J. Acoust. Soc. Amer., 71, 1544-1651 (1982).
13. R. Plomp and W.J.M. Levelt: "Tonal Consonance and Critical Bandwidth," J. Acoust. Soc. Amer., 38, 548-560 (1965).
14. K. Ohgushi: "Relationship between Attributes of Timbre and Unpleasantness," Proceedings of Acoust. Soc. Jpn. (1982.10), 213-214 (1982).
15. H. Fastl: "Roughness and Temporal Masking Patterns of Sinusoidally Amplitude Modulated Broadband Noise," in Psychophysics and Psychology of Hearing, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London, 1977), 403-414.
16. E. Terhardt: "On the Perception of Periodic Sound Fluctuations (Roughness)," Acoustica, 30, 201-213 (1974).
17. S. Kemp: "Roughness of Frequency-Modulated Tones," Acoustica, 50, 126-133 (1982).