Ann.Bull.RILP(1986)No.20, 211-216

CLINICAL APPLICATION OF THE ACOUSTIC MEASUREMENT OF PATHOLOGICAL VOICE QUALITIES

Satoshi Imaizumi

INTRODUCTION

In our previous papers[1-3], acoustic correlates of "roughness" and "breathiness"[4] in pathological voice were investigated based on acoustic analysis and perceptual evaluation. And the following results were obtained. The acoustic correlates of "rough" voice are not only the multiplicative variations which occur over several pitch periods, but also those which are synchronous with the vocal pitch period. As the acoustic correlates of "breathy" voice, the additive noise level within the frequency range above 1kHz, and also the weakness of the harmonics represented by the distortion factor, are very important. And also it was found[3] that the weakness of the harmonics is very important to describe the contrast between "asthenic" voice and "strained" voice.

In this paper, the clinical availability of acoustic correlates of pathological voice qualities was investigated. Extracted measures were 1) modulation indices representing periodical variations in the pitch period, in the amplitude and in the waveform; 2) a pitch perturbation quotient; 3) an amplitude perturbation quotient; 4) a distortion factor representing richness of high frequency component; and 5) the additive noise level. Using these measures, changes in voice quality resulting from clinical treatment were assessed for the patients suffered from Reinke's edema[5].

METHOD

Subjects

The voice samples analyzed were recorded from 38 patients, 22 males and 16 females, who were suffered from Reinke's edema. These patients were treated with the sucking technique[5] under the endolaryngeal microsurgery at the Department of Otolaryngolgy of Kinki University hospital. The details of these patients, clinical types, clinical features, surgical management and post-operative results were reported by H. Yonekawa, F. Ohta, S. Imaizumi and Y. Koike in another paper[5].

Recordings of voice samples

Voice samples were recorded from all of the patients at least twice; the first recordings were done at their first visit to the clinic and the second ones after surgery. The second recordings were carried out after the status of the patient's vocal folds became steady, and were done between 1 and 34 months with the mean of 6 months following surgical treatment. The recording was done in a sound proofed room. Each speaker produced a sustained Japanese vowel /e/ at his most comfortable manner for at least two seconds.

Voice samples for /e/ were digitized through a 12-bit A/D converter at a sampling rate of 20 kHz. Then 0.5sec segments were extracted by excluding the initial and final portions from each sample. These segments were analyzed.

Perceptual judgement

One well-trained listener judged the voice quality change from the aspects of "the grade of hoarseness," "breathiness," "roughness," "asthenicity" and "strained quality". Every pair of the voice samples recorded before and after the surgery was presented in a random order through an audiometer. The stimulus level was 60 dBSL. For each pair, the listener selected one category from those shown in Fig. 1 to judge the voice quality change of second voice sample comparing to first one within each pair. For instance, if the listener selected "extremely improved" for "roughness", it meant that the second voice sample was judged as extremely decreased in "rough" quality compared to the first voice sample within a pair.

```
   -3       -2       -1       0        1        2        3
    I--------I--------I--------I--------I--------I--------I
extremely         slightly same  .         slightly extremely
worsen    worsen  worsen            improved improved improved
```

Fig. 1. The categories and their numbers on the scale used in the paired comparison. The numbers were not displayed for the listener.

Acoustic analysis

First, pitch periods were extracted and the fundamental frequency, F_0, was calculated from a low-pass filtered waveform. Using a peak picking method, local maximum points which corresponded 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.

Then, to extract acoustic indices to represent the multiplicative variations observed in the waveform, 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,

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

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 pitch periods.

R(i,j) showed a more or less 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

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

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

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

where SF was the sampling frequency, that is, 20,000.

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

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

and

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

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

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 way similar to that in which 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[6].

To measure the richness of the high frequency component, RH was defined using the pitch synchronous spectrum as follows.

$$X_i(k)=\sum_{m=0}^{M_i-1} x(L(i)+m)\cdot\exp(-2\pi\cdot m\cdot k/M_i),$$

$$d_i = \sum_{k=2}^{M_i/2} |X_i(k)|^2 / |X_i(1)|^2,$$

$$RH = \frac{1}{I} \sum_{i=1}^{I} d_i \qquad (6).$$

Here, $M_i = L(i+1) - L(i)$ and $k=0,1,2..,M_i/2$. RH is the same index known as the distortion factor in the field of electronics.

Next, the additive noise level ANL(dB) was defined as follows. Using an adaptive comb filter with a window length of 7 [7,8,9], the original waveform x(n) was divided into its periodical component h(n) and non-periodical component n(n). Then their power spectra were calculated using FFT with a Hamming window of length 2,048 as $|H(k)|^2$ and $|N(k)|^2$. ANL_j for the frequency range from f_1 to f_2 was defined as

$$ANL_j = 10\log_{10}\{\sum_{k=k_1}^{k_2} |N(k)|^2 / \sum_{k=k_1}^{k_2} |H(k)|^2\} \qquad (7).$$

Here, f_1=1 kHz, f_2=5 kHz, $k_1 = (f_1 * 2048)/SF$, and $k_2 = (f_2 * 2048)/SF$. The sufix j indicates a window position wich was shifted every 1,024 sample points. The final ANL was acquired as an average of ANL_j with respect to j.

RESULTS AND DISCUSSION

The voice waveform of /e/ recorded from Case 1 before surgery is shown in Fig. 2(a), and that recorded after the surgery is in Fig. 2(b). Several acoustic indices extracted by the computer analysis system are indicated. The fundamental frequency, F_0, greatly ascended from 147 Hz to 185 Hz. The pitch period perturbation quotient, PPQ, decreased from 1.36 to 0.10, and the amplitude perturbation quotient, APQ, also decreased from 5.81 to 0.81. Furthermore, the additive noise level, ANL, also greatly decreased from 3.47 to -8.00. The richness of the high frequency component, RH, increased from 0.84 to 3.44. These results mean the following. The voice waveform after the surgery contained a smaller fluctuation in its amplitude and also in its pitch period as well as a smaller amount of additive noise and richer high frequency component compared to before the surgery.

According to our previous study[1-3], these results of acoustic analysis predict the voice before the surgery is very "breathy" and "asthenic", and that after the surgery must be rather improved in both qualities. The listening subject judged the voice change as "improved" in "breathiness", "asthenic" and "the grade of hoarseness".

The voice of Case 2 contains the periodical variations in the waveform before the surgery as shown in Fig. 2 (c), while after the surgery, the voice waveform was composed of a very stable repetition of fundamental periods as shown in Fig. 2 (d). The results of acoustic analysis predict the voice before the surgery is strongly "rough". And the voice after the surgery is almost normal. The listening subject judged "extremely improved" in "roughness".

Fig. 3 shows the changes in PPQ before and after the surgery. The values of PPQ for the voice samples below the solid line L decreased after the surgery. The number attached to each circle is the category number shown in Fig. 1 selected by the listening subject for "roughness". Large numbers can be observed for the voice samples whose values of PPQ decreased very much. This result indicates that PPQ is useful to describe the change in the "rough" voice quality to some extent. A similar tendency was observed between APQ and "roughness", ANL and "breathiness".

CONCLUSION

In this paper, the clinical availability of acoustic correlates of pathological voice qualities was investigated. Using acoustic measures representing the degree of modulation, richness of high frequency component and additive noise, changes in voice quality resulting from clinical treatment were assessed for the patients suffered from Reinke's edema. The results indicate that these acoustic measures are useful for the quantitative evaluation of the change in voice quality observed during clinical treatment or voice therapy.

Acknowledgement

We would like to express my gratitude to Prof. F. Ohta, Prof. Y. Koike, and Dr. H. Yonekawa for their help.

References

1. S. Imaizumi:"Acoustic Measures of Pathological Voice Qualities - Roughness -," Ann. Bull. RILP, Univ. of Tokyo, 19, 179-190 (1985).
2. S. Imaizumi:"Acoustic Measurement of Pathological Voice Qualities for Medical Purposes," IEEE, Proceedings of ICASSP 86, Tokyo, 677-680 (1986).
3. S. Imaizumi:"Acoustic Correlates of Pathological Voice Qualities," Folia Phoniatrica, Abstracts of XXth IALP,309 (1986).
4. M. Hirano, Clinical Examination of Voice, Wien, Springer-Verlag, Chap.6, 81-84, 1979.
5. H. Yonekawa, F. Ohta, S. Imaizumi and Y. Koike;"A Clinical Study of Reinke's Edema," Folia Phoniatrica, Abstracts of XXth IALP,375 (1986).
6. Y. Koike:"Application of Some Acoustic Measures for the Evaluation of Laryngeal Dysfunction," Studia Phonologica, VII, 17-23, (1973).

7. J. S. Lim, A. V. Oppenheim and L.D. Braida:"Evaluation of an Adaptive Comb Filtering Method for Enhancing Speech Degraded by White Noise Addition," IEEE ASSP-26, 4, 354-358 (1978).
8. H. Kasuya, S. Ogawa, Y. Kikuchi:"An Adaptive Comb Filtering Method as Applied to Acoustic Analyses of Pathological Voice," IEEE, Proceedings of ICASSP 86, Tokyo, 669-672 (1986).
9. E. Yumoto, W. J. Gould and T. Baer:"Harmonics-to-noise Ratio as an Index of the Degree of Hoarseness," J. Acoust. Soc. Am., 71, 1544-1550 (1982).

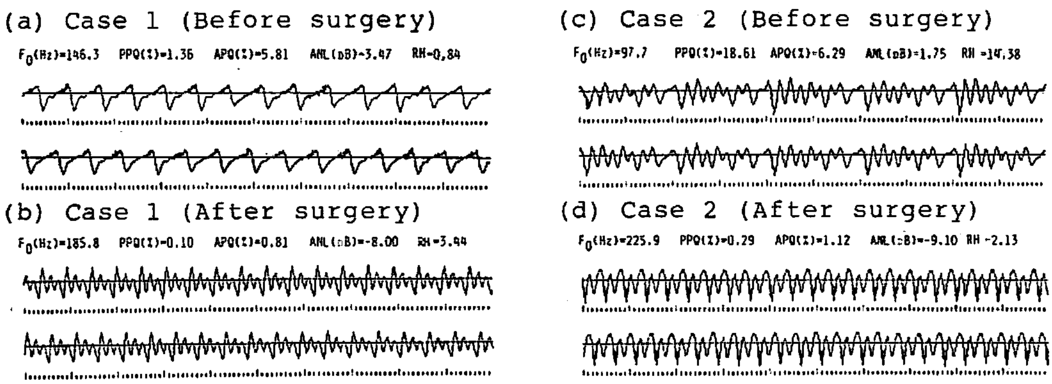

Fig. 1. The voice samples of Case 1 and Case 2 together with some acoustic measures.

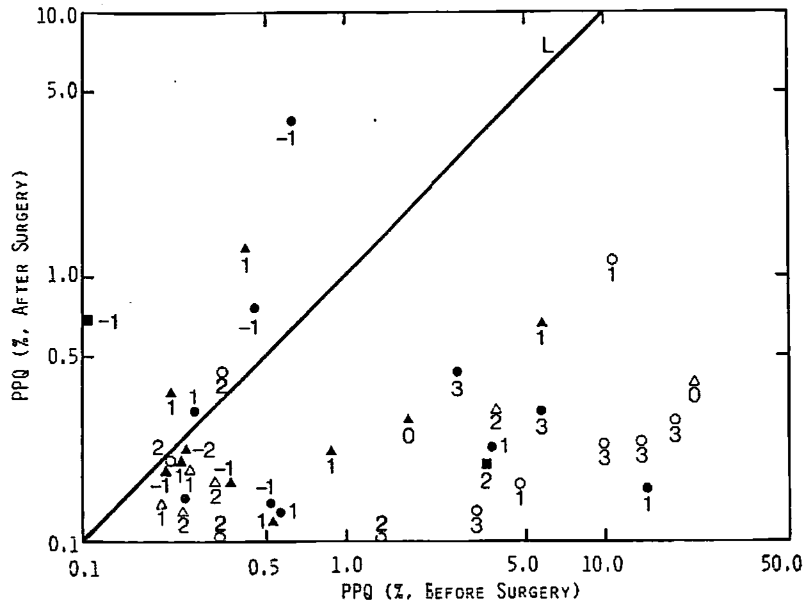

Fig. 2. Relationship between the changes in PPQ due to the surgery and the results of the perceptual judgements. The numbers represent the category numbers selected in the perceptual judgement for "roughness". The symbols (○,△,□) indicate female voice samples and (●,▲,■) male ones. For "the grade of hoarseness", the selected category number was indicated as positive (○,●), zero (△,▲), and negative (□,■).