STUDIA PHONOLOGICA IX (1975)

Pitch Perturbation in Normal and Pathologic Voice

Kazutomo KITAJIMA, Masahiro TANABE
and Nobuhiko ISSHIKI

The otolaryngologist has long been used subjective evaluation of voice quality in the assessment of various laryngeal disorders. At the same time many have attempted to quantify various aspects of both the normal and abnormal voice, to extract objective data from acoustic signal that might provide criteria usable in the detection and evaluation of laryngeal disorders.

Lieberman defined pitch perturbation as the difference between periods of adjacent glottal pulses, suggesting that pitch perturbation factor (the percentage of perturbation exceeding 0.5 msec.) might be useful index in detecting laryngeal diseases. Smith & Lieberman extended this work and described the significant relationship between selected parameters of pitch perturbation and a number of pathologic conditions.

Proposing a "directional perturbation factor" which takes into account the algebraic sign rather than the magnitude of differences between adjacent glottal pulse intervals, Hecker & Kreul indicate a significant meanings of that factor in laryngeal cancer. Crystal and Jackson measured both of pitch and amplitude variation of various pathologic voices, and indicated that extracted data could be considered as potential indices in detection of laryngeal dysfunction. Applying the newly developed inverse filtering technique to speech produced by patients with various laryngeal diseases and normal subjects, Koike & Markel suggested the feasibility of applying this technique as an aid to laryngeal diagnosis.

In this study, as a first step toward pratical application of pitch perturbation in clinical field, a digital computer was used to simplify and facilitate the procedure.

In order to approximate the magnitude of perturbation obtained to the perception of the ear, semitone scale was used.

Subjects

The subjects were 10 men and 10 women with no laryngeal or pulmonary disorders, who served as normal or controls, and 13 men with laryngeal cancer. The age distribution was 18 to 77 years for the normals.

Kazutomo KITAJIMA (北嶋和智), Masahiro TANABE (田辺正博) and Nobuhiko ISSHIKI (一色信彦): Department of Otolaryngology, Faculty of Medicine, Kyoto University. Director: Prof. M. Morimoto.

METHODS

The system used in collecting and processing the data is schematized in Fig. 1. The discrete steps involved are as follows.

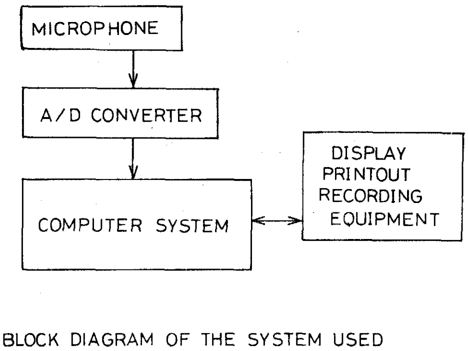

Fig. 1. Block diagram of the system used.

(1) Recording of voice

A voice recording for each subjects was made with a microphone placed on the anterior cervical skin in front of the second tracheal ring. The microphone (Sony ECM-51) has a adoptor tube connected at the head so that the air space is maintained between the cervical skin and the diaphragm of the microphone. The voice signal collected through this microphone is less affected by the resonance of vocal tract than that through a mouth microphone. This procedure of pretracheal voice recording facililated the detection of the fundamental frequency in voice signal.

(2) Type of phonation analyzed

Each subject was asked to phonate sustained vowel /a/. For some subjects the recordings of a phrase /Aou Umi/ or /blue sea/ were also made.

(3) Analog to digital conversion

About 500 msec duration of the voice signal was digitalized. A 500 msec was the maximum duration of the voice signal that could be processed by the computer. The signal was sampled at the rate of 12,315 per second and quantitized into nine bits. Measurements were made upon digitalized signal.

(4) Measurement

Each pitch period was measured automatically by the main peak detecting method as is described below. As an initial conditioning, the first pitch period was measured manually using a cursor which was projected on the display scope with voice signal. Measured period (suppose P_1) was memorized in the core for further processing. The computer then searches for the next main peak within the region of $P_1-P_1\cdot K\sim P_1+P_1\cdot K$. K is the constant which determines the range of search, and each time prior to execution of the program, K was defined. Except for very severe case of hoarseness, 0.1 was usually used as the value of K. The second pitch period, P_2, was thus determined automatically by detecting the following main peak. After this manner, 100 pitch periods were measured.

(5) Computation of pitch period

Prior to calculation each pitch period was transformed into frequency and then into semitones which was based upon 16.35 cps. The auditory sense is correlated with the magnitude of frequency in an approximately logarithmic relationship. Therefore semitone scale, which is the logarithmic difierence between two frequencies, is considered to be appropriate for this study.

(i) Sustained phonation

The formula used for the calculation of pitch perturbation in sustained phonation was as follows.

$$\overline{\Delta F}=\frac{\sum_{i=1}^{N-i}|F_i-F_{i+1}|}{N-1} \qquad \cdots\cdots(1)$$

where: $\overline{\Delta F}$ is a magnitude of perturbation

F_i is semitones above 16.35 cps

N is a number of periods measured

In order to detect the whole system error which was resulted from the tape recorder and sampling rate etc,saw-tooth signal generated by oscillator was fed in the computer and the pitch perturbation was calculated. The signals of 100 cps and 200 cps were tried 10 times each. The result was 0.09 semitones (s.t.)∼0.13 s.t. for 100 cps, and 0.10 s.t.∼0.24 s.t. for 200 cps. Larger values in 200 cps could be resulted from the fixed sampling rate of A—D converter, because the number of the sampled data per one cycle is smaller in the signal of higher frequencies. It should be in mind that the measured value within these ranges might be a system error.

(ii) Phrase-reading

Formula (1) is not considered to be appropriate for the analysis of phrase-reading. The reason of this is as follows. During phrase-reading, slowly moving changes due to intonation and/or accent was commonly observed. If the formula (1) was used, the magnitude of abrupt frequency change would be masked or exaggerated by intonation and/or accent. Therefore the formula (1) was modified to eliminate these effects. At first the original series of frequencies were smoothed so as to find out a general trend which is considered to approximate the moving due to intonation and/or accent. Then the computation was performed by the formula described below.

$$\overline{\overline{\Delta F}}=\frac{\sum_{i=3}^{N-2}|F_i-\overline{F_i}|}{N-4} \qquad \cdots\cdots(2)$$

where: $\overline{\overline{\Delta F}}$ is a magnitude of perturbation

F is semitones above 16.35 cps

$$\overline{F_i}=C_{-2}\cdot F_{i-2}+C_{-1}\cdot F_{i-1}+C_0F_i+C_1\cdot F_{i-1}+C_2\cdot F_{i-2} \qquad \cdots\cdots(3)$$

C_{-2}, C_{-1}......C_2 are the constants which are calculated by least square fitting method.

N is a total number of periods measured. As can be noticed in formula (3), it is impossible to get the values of $\overline{F_1}$, $\overline{F_2}$, $\overline{F_{N-1}}$ and $\overline{F_N}$. Consequently i begins at 3 and ends at N-2.

RESULTS

(1) Sustained phonation

The $\overline{\overline{\Delta F}}$ values of the controls are listed in Table 1. Males have values between 0.11 s.t. and 0.21 s.t., whereas in female they distribute between 0.15 s.t.

Table 1. The values of $\overline{\overline{\Delta F}}$ in normal subject.

$\overline{\Delta F}$ in normal subjects

sustained phonation of vowel /a/

male	female
0.13 s.t.	0.22 s.t.
0.13	0.21
0.17	0.21
0.13	0.17
0.12	0.18
0.12	0.20
0.11	0.20
0.20	0.22
0.13	0.15
0.19	0.18
0.17	0.22
0.20	0.21
0.17	0.20
0.11	0.16
0.19	0.17
0.21	0.19
0.19	0.18
0.11	0.16
0.18	0.24
0.13	0.21

rejection interval (5%)

male	female
0.08 ~ 0.23	0.14 ~ 0.25

and 0.24 s.t..

The critical region calculated from these data was 0.08 s.t.∼0.23 s.t. for male, and 0.14 s.t.∼0.25 s.t. for female. The difference between male and female does not necessarily mean the larger perturbation in female, because as was pre-

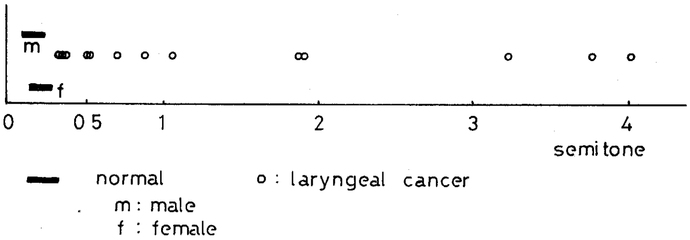

Fig. 2. $\overline{\Delta F}$ values of 13 cases of laryngeal cancer with critical region of normals. On the ordinate semitone is scaled.

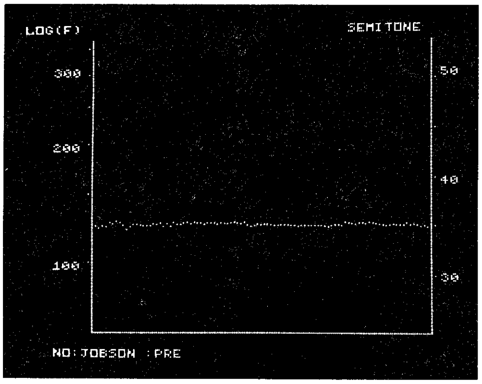

Fig. 3. Sustained phonation of normal subject. On the ordinates, both of frequency and semitone are scaled, left and right sides respectively. The number of sampled data is on the abscissa.

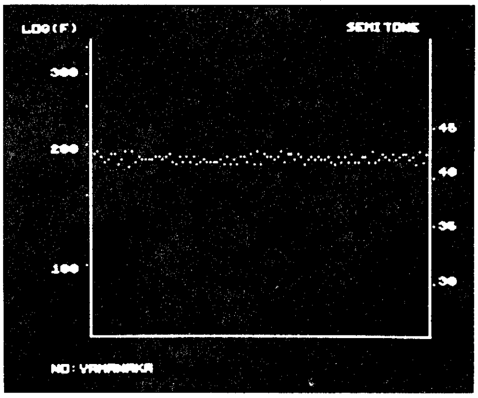

Fig. 4. Laryngeal cancer.

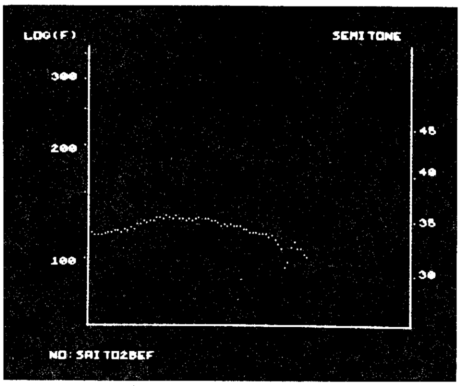

Fig. 5. / Aou Umi/ in normal subject pre-smoothing.

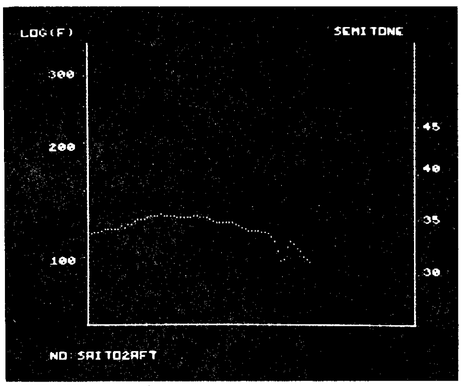

Fig. 6. After smoothing procedure of Fig. 5.

viously described the measurement error is depend upon the mean frequency of the voice to some extent.

The computed results of 13 cases of laryngeal cancers are shown in Fig. 2. All of them distribute outside the range of critical region of normals. Fig. 3 shows a series of fundamental frequencies which was projected on the display-scope. On the ordinate both of frequencies (left side) and semitones (right side) are scaled. Number of sample (i) is on the abscissa. This is a normal subjects. Notice that each fundamental frequency show almost steady value without remarkable variation. An example of laryngeal cancer was shown in Fig. 4. Of interest in this figure is that a series of frequencies is splitted into two levels especially in the latter section. The relatively high and low frequencies are observed alternately. Although this phenomenon was not observed during whole section of phonation, it was noticed that this "splitting" section well coincide with the perception of severe hoarseness.

(2) Phrase-reading

The display of a series of frequencies in phrase-reading is shown in Fig. 5. This is the normal subject. Smoothing procedure was performed on this series and the result is shown in Fig. 6. The abrupt variations in frequency, as can be noticed, were smoothed without affecting the geneal trend.

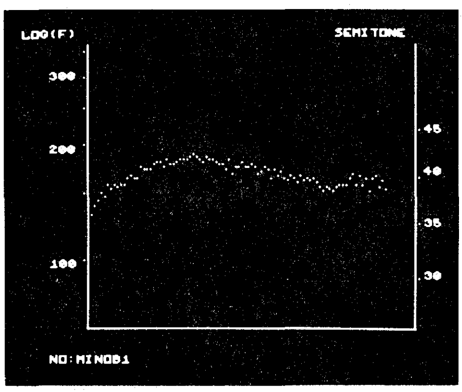

Fig. 7 /Aou Umi/ in laryngeal cancer

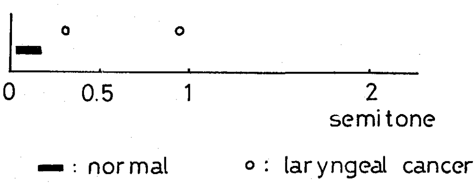

Fig. 8 $\overline{\overline{\Delta F}}$ values of 2 cases of laryngeal cancers with critical region of normals.

Table 2 The values of $\overline{\overline{\Delta F}}$ in normal subject.

$\overline{\overline{\Delta F}}$ in normal subjects

/aoi umi/

male	female
0.11 s.t.	0.14 s.t.
0.11	0.15
0.16	0.17
0.10	0.15
0.15	0.19
0.17	0.18
0.08	0.17
0.16	0.16
0.11	0.14
0.16	0.12
0.10	0.12
0.11	0.15
0.12	0.11
0.12	0.15
0.14	0.12
0.15	0.13
0.12	0.12
0.10	0.13
0.11	0.12
0.16	0.13

rejection interval (5%)

male	female
0.07 ~ 0.18	0.09 ~ 0.19

0.08 ~ 0.19

An example of laryngeal cancer (pre-smoothing) is shown in Fig. 7. Remarkable irregularities is observed as compared with Fig. 5.

The values of 20 normal subjects are shown in Table 2. Critical region of normals based upon these data was 0.08 s.t.~0.19 s.t. (No statistical difference was found between male & female). In other words, the value over 0.19 s.t. can be considered as pathologic. In Fig. 8 two cases of laryngeal cancers are shown with the normal region.

Discussion

(1) Technical difficulty in extraction of pitch

It is commonly observed that the voice of patient shows more roughness during speech than in a sustained phonation. In case of the minor lesion of the vocal cord, roughness is only noticed during speech. Therefore the analysis of pitch perturbation during speech is believed to give a useful information for early detection of laryngeal lesion. Extraction of fundamental frequency during speech, however, is rather difficult and troublesome work. Previous investigators have measured each frequency by hand under visual inspection of the voice signal.

From clinical point of view, this method is far from practical use. It needs much time and energy for the analysis of one patient.

In this study, as a first step toward more practical use, small laboratory computer was used in the analysis of pitch perturbation. The measurement was done automatically after feeding the computer with some information of the initial condition. Although the phrase analyzed has only vowels and many problems to be solved are lying before the analysis of ordinary centences, it was confirmed that the approach in this study could be extended toward early detection of laryngeal lesion.

(2) Quantitization of pitch perturbation

Several attempts have been made to find out a useful index for detection of laryngeal lesion. Lieberman's perturbation Factor is based upon the absolute difference between adjacent pitch periods. Wendahl, on the other hand, has described that the perception of hoarseness is more closely related to the value $\Delta F/F$ rather than ΔF or ΔT (absolute difference between consequtive frequency or pitch). In this study, semitone scale was used based upon the reasons previously described. It was believed that the quantitization of magnitude in semitone could well correspond to auditory perception of the ear.

Summary

(1) Pitch perturbation was analyzed both in sustained phonation and in phrase-reading

(2) In sustained phonation, the values over 0.23 s.t. for male and 0.25 s.t. for female was defined as pathologic.

(3) In the phrase-reading over 0.19 s.t. was abnormal

(4) To fascililate and simplify the procedure, laboratory computer was used.

REFERENCES

1. Lieberman P: Perturbation in vocal pitch. J Acoust Soc Am 33: 597–603, 1961
2. Lieberman P: Some acoustic measures of the fundamental periodicity of normal and pathologic larynges.. J Acoust Soc Am 35: 344–353, 1963
3. Smith WR, Lieberman P: Computer diagnosis of laryngeal lesion. Comput Biomed Res 2: 291–303, 1969
4. Hecker MHL, Kreul EJ: Description of the speech of patient with cancer of the vocal folds. Part 1: Measures of fundamental frequency. J Acoust Soc Am 49: 1275–1282, 1971
5. Crystal TH, Jackson CL: Extracting and processing vocal pitch for laryngeal disorder detection. J Acoust Soc Am 48: 118
6. Wendahl RW: Laryngeal analog synthesis of jitter and shimmer Auditory parameters of harshness. Folia phoniat 18: 98–108, 1966
7. Koike Y, Markel J: Application of inverse filtering for detecting laryngeal pathology. Ann Otol 84: 117–124, 1975

(Aug. 31, 1975, received)