SLOW AND FAST PERTURBATIONS IN VOICE
-A PRELIMINARY REPORT-

H. Abdoerrachman*, S. Imaizumi**, H. Hirose**, S. Niimi**
*Department of Otolaryngology, Fac. of Medicine, University of Indonesia
Jl. Diponegoro No.71 Jakarta-Pusat, Indonesia.
**Research Institute of Logopedics and Phoniatrics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113 Japan

Introduction

Although there are many acoustic studies of pathological fluctuations in voice focusing on rapid fluctuations in fundamental frequency (F0) or amplitude (Laver et al., 1992 and Imaizumi, 1986), slow variations have usually been disregarded as "trend" components, even though they could have given rich information about pathological conditions.

Some experts (Ludlow et al., 1988 and Aronson et al., 1992) have shown that respiratory and laryngeal muscle fiber contractions are implicated as causes of vocal tremor. For instance, Ludlow et al. (1988) conducted an investigation on the phonatory characteristics of vocal tremor in patients with benign essential tremor compared to normal controls and patients with phonatory tremor due to other neurological disorders. Ramig et al. (1987) performed comparative measures of vocal tremor on neurological conditions and vocal vibrato. Although there have been some investigations on perturbation and tremor in voice, attempts are still being made to enhance the voice analysis, for the sake of inventing a supportive diagnostic tool to characterize and estimate the sources of voice perturbation.

On the other hand, not only neurological patients but also patients with laryngeal disorders tend to complain that they can not control their vocal pitch and intensity flexibly enough for verbal communication. Thus, it is very important to assess the controllability of vocal pitch and amplitude. "Controllability" is the ability to produce desirable pitch, intensity and timber according to the speaker's intention. (Imaizumi et al., 1993).

In this paper, the "controllability" of pitch and intensity were investigated for patients with various voice problems. "Controllability" is defined here as the ability to keep the vocal fundamental frequency and amplitude as constant as possible when instructed to produce a sustained vowel. Possible sources of variations in voice which restrict "controllability" will be discussed as well.

Method

The subjects for this study were 246 patients with various diseases and controls; 51 modal voice samples produced by 51 normal/healthy speakers (MO), 46 tremorous voices (Tr), 17 cases of Reinke's edema (RE), 44 vocal cord polyps (VP), 37 cases of reccurent nerve paralysis (RP), 51 spastic dysphonia (SD). The normal/healthy voice samples (MO) were the modal voice samples produce at most comfortable level and pitch by healthy speakers without any pathology affecting phonation. Symbol 'F' or 'M' was attached to each goup name, for instace, TrF or TrM, to indicate the female or male subgroups of Tr.

The speakers were instructed to produce each of the Japanese five vowels for 2 or 3 seconds at their most comfortable pitch and intensity. Using a Sony DAT tape-recorder, the recording was made, with a constant distance from mouth to mic of 15 cm, following practice in a sound treated booth.

Voice samples of a sustained vowel /e/ were digitized through a 12-bit A/D converter at a sampling rate of 40 kHz and stored on a disk controlled by a computer. A one-second segment was extracted by excluding the initial and final portion from each sample.

Using the method developed by Imaizumi et al. (Imaizumi et al., 1991 and 1993) local maximum points which could correspond to vocal excitation epochs were detected successively, and then the fundamental frequency F0(i) and the maximum amplitude A(i) of i-th glottal period were determined. The following acoustic parameters were used in this paper: the mean F0 (Hz), the fractal dimension of F0(i) and A(i), the additive noise level (dB), overall variability, low and high frequency energy in F0(i) and A(i) normalized by DC level (dB).

Because the details of the method and the definitions of the parameters were discussed in another paper (Imaizumi et al., 1993), only the meanings of the parameters are explained briefly below.

The fractal dimension, FrctF0, is an index of irregularity calculated using Baken's method (Baken, 1990). FrctF0 has a value between 1 and 2; 1 means that F0(i) is regular or stable, and 2 irregular. Overall variability is the percentage of the standard deviation normalized by the average of F0(i). For the statistical analysis, a logarithmic transformation was used.

The additive noise level was defined as the difference in decibels between the energy of the non-harmonic and that of the harmonic components within 1-4 kHz frequency range. The harmonic and non-harmonic components were extracted from the original speech waveform using a comb filtering method (Lim et al., 1978).

To calculate low and high frequency energy in F0(i) and A(i), the power spectra of F0(i) and A(i) were calculated using the fast fourier transformation, and then energy in the frequency ranges between $0<f<16$, and $16<=f<$average F0/2 were calculated. f is frequency in Hz. Finally the logarithmic transformed values were normalized by DC level.

Analyses of variance with two factors, group and sex, were performed to determine significance.

Table I. The number of subjects.

Group|Sex	F	M	Totals
Hlth	22	29	51
Tr	33	13	46
RE	12	5	17
VCP	17	27	44
RNP	17	20	37
SPD	45	6	51
Totals	146	100	246

Results

The acoustic profile of a SD sample is shown in Figure 1. The profile consists of five panels, a) a control panel with display of analysis conditions and results, b) the voice waveform, c) the sound spectrogram, d) F0 and amplitude time series extracted by the acoustic analysis and e) the power spectrum of F0(i) and A(i). The SD sample in Fig. 1 reveals slow but large variations in F0 and amplitude. The noise level is not high in this case.

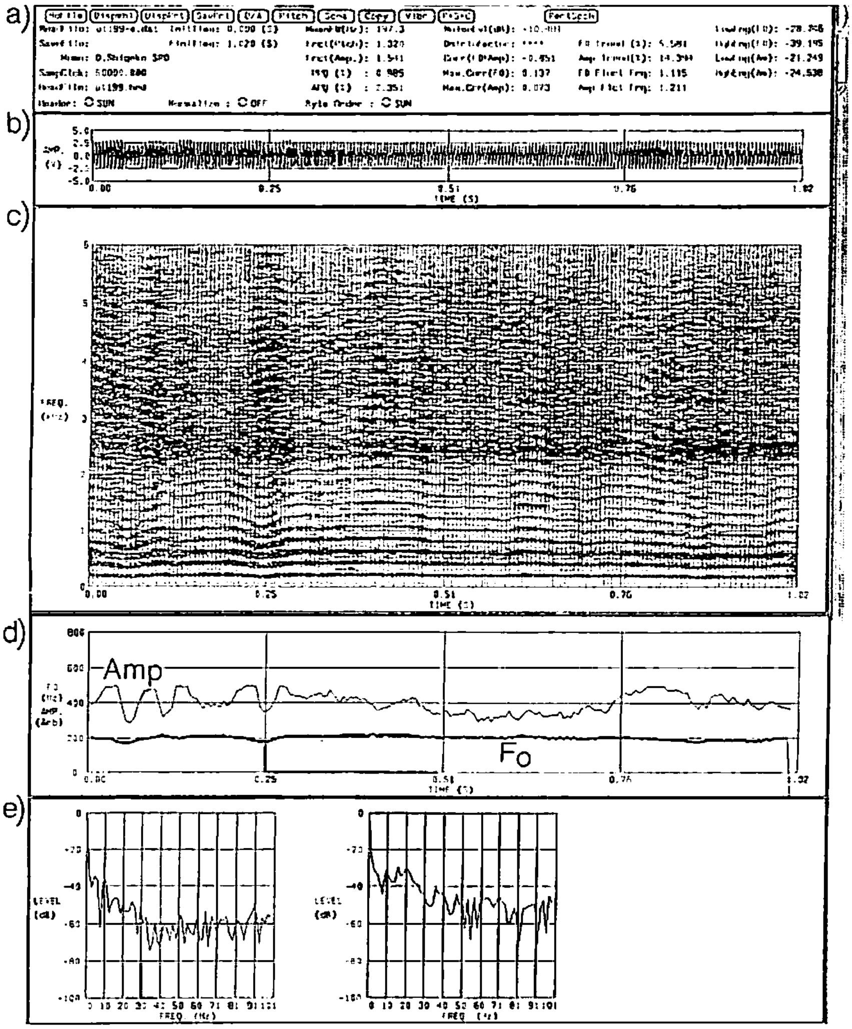

Fig. 1. The acoustic profile of a SPD voice analyzed by SONG. a) the control panel with display of analysis conditions and results, b) the voice waveform, c) the sound spectrogram, d) F0 and amplitude time series and e) the power spectrum of F0(i) and A(i).

Figure 2 shows a box-whisker graph of the logarithmic transformed values of the mean F0 for the six groups of speakers. The ANOVA analysis showed that group (F=9.67, p<0.0001), sex(F=139.89, p<0.0001) and their interaction (F=4.86, p<0.0003) were significant. The RE group showed the lowest F0, and female RE speakers (REF) showed a lower mean F0 than the male speakers of other groups. The female speakers of SD tended to have a low F0 compared with the other pathological groups. The male SDs showed the largest variation in F0 among the male groups.

Figure 3 shows the additive noise level. The ANOVA analysis showed that group (F=25.5, p<0.0001), sex(F=10.81, p<0.002) and their interaction (F=4.22, p<0.001) were significant. Among the pathological groups, SDM and TrF showed a low noise level compared to REF, REM, RPF, and RPM. All the pathological groups revealed a higher noise level compared to the MO group.

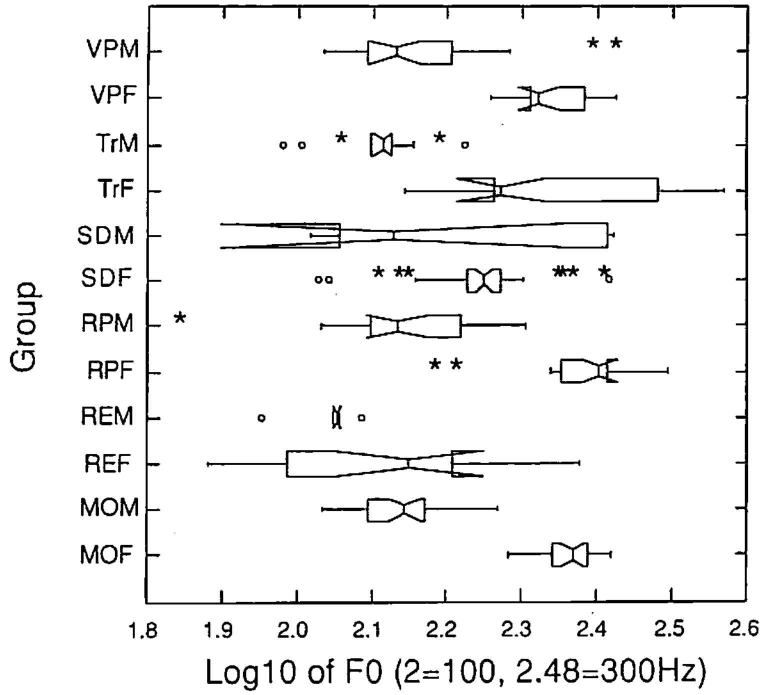

Fig. 2. Box-whisker graph of logarithmic transformation of mean F0 of six groups. (a) Female, (b) Male. The center vertical line indicates the median. The edges of the central box represents the lower and upper hinges. The whiskers represent most remote data points from the median which are inside of the lower and upper fences. Asterisks represent outside values and empty circles far outside values. The boxes are notched at the median and return to full width at the lower and upper confidence interval values of 95%.

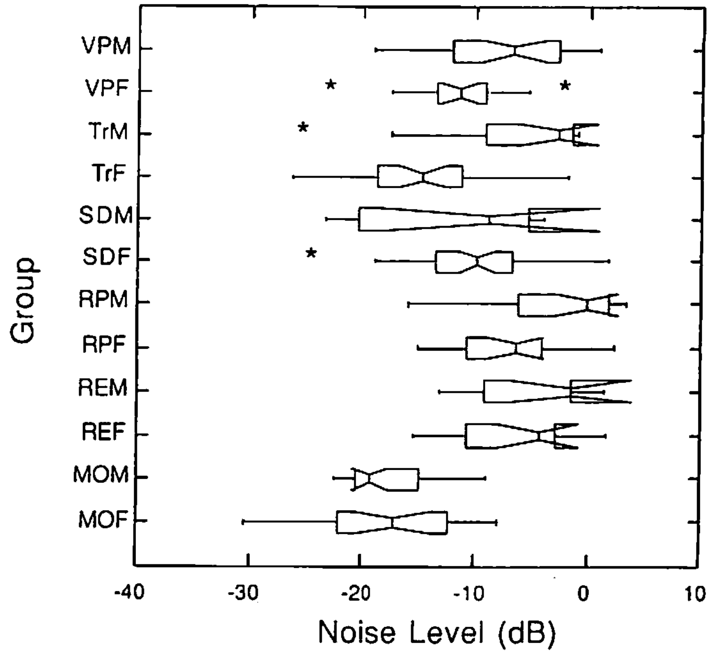

Fig. 3. Box-whisker graph of additive noise level.

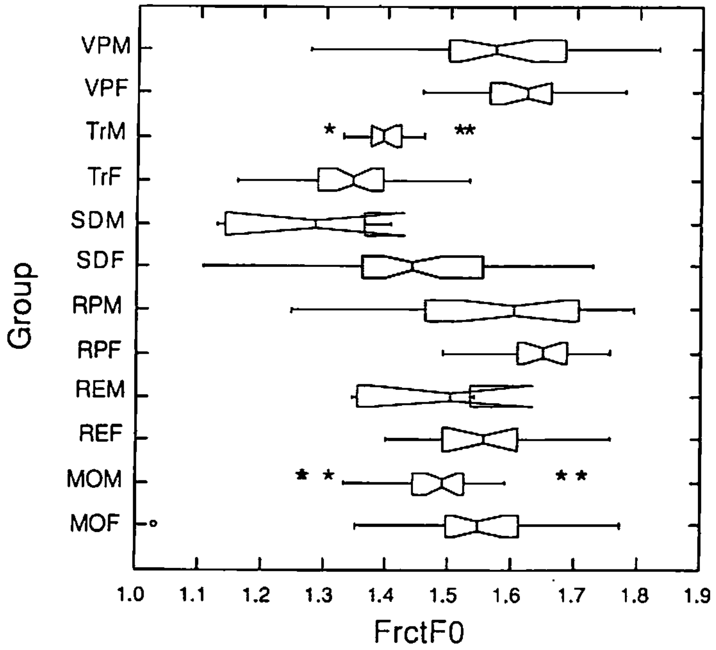

Fig. 4. Box-whisker graph of fractal dimensions of F0(i).

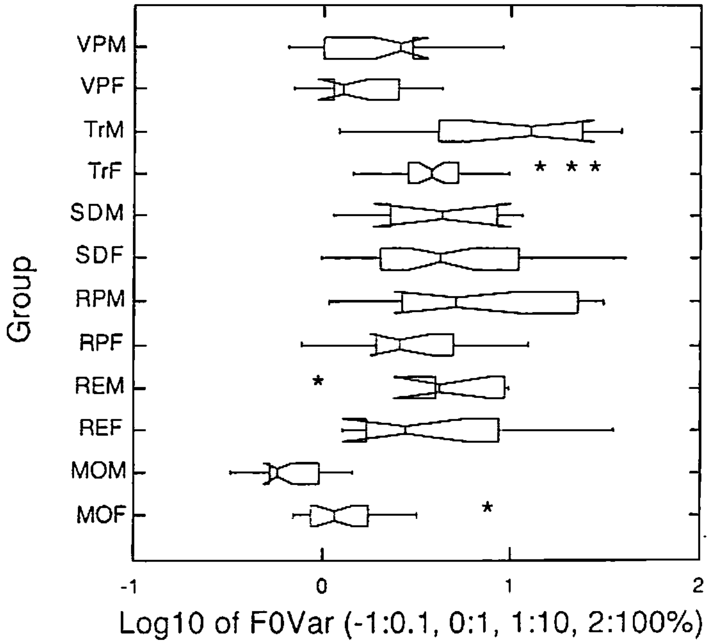

Fig. 5. Box-whisker graph of overall variability of F0(i).

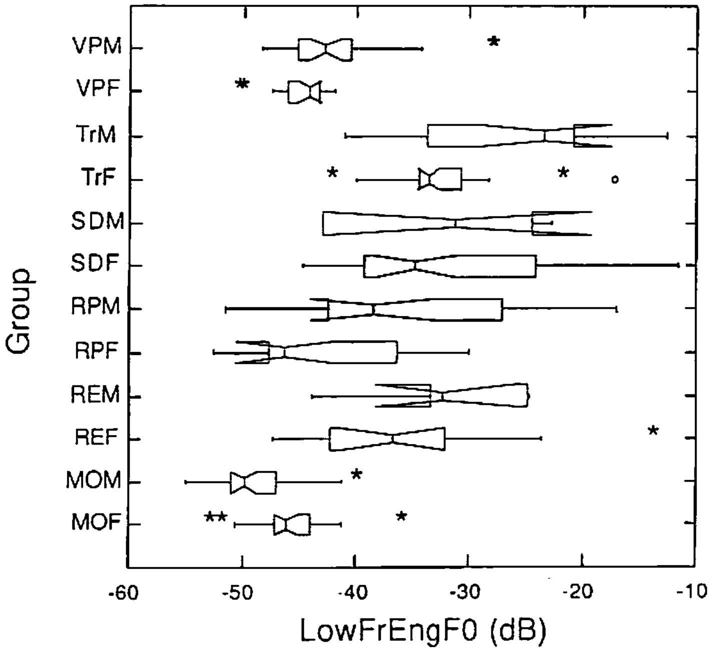

Fig. 6. Box-whisker graph of energy of the low frequency range of F0(i).

Figure 4 shows the F0 fractal dimension. The ANOVA analysis showed that group (F=24.84, p<0.0001), sex(F=12.85, p<0.004) and their interaction (F=3.04, p<0.01) were significant. Compared with the MO group, RP and VP showed a higher fractal dimension, while the Tr and SD groups had lower values. The SD revealed wider distribution compared to the Tr, which indicated that the SD had larger variations among tokens than the Tr group.

Figure 5 shows the overall variability of F0(i). For this parameter, group (F=29.91, p<0.0001) and the interaction between group and sex (F=4.83, p<0.0003) were significant. Compared with the MO group, the pathological groups showed a larger variability. Among these, although VP had the lowest variability (or highest stability) in producing sustained vowels with constant pitch, VPM had significantly larger variability compared with the MO group.

Figures 6 and 7 show the spectrum energy of F0(i) of low (DC<f<16Hz, f=frequency) and high frequency (16<f<=F0/2) ranges. As shown in these figures, all the pathological groups showed stronger energy both in the low and high frequency ranges than the MO group.

For the low frequency energy of F0 perturbation, group (F=36.88, p<0.0001), sex (F=6.65, p<0.01) and their interaction (F=3.21, p<0.008) were significant. SD and Tr showed a stronger energy than other pathological groups. The RE groups showed also relatively strong energy in the low freuency range.

For the high frequency energy of F0 perturbation, group (F=20.93, p<0.0001) and the interaction between group and sex (F=6.29, p<0.001) were significant. For this parameter, the female and male groups showed differnt tendencies. Amoung the male groups, all the pathological groups particulary TrM and RPM showed a stronger energy than others. Amoung the female groups, the differences between the MO group and the pathological groups were small compared to the male groups. The RPF and SDF showed a signicantly stronger energy than the MO group.

Discussion

Comparing the above parameters, the following tendencies were obtained.

1) Not only Tr and SD, but also the other pathological groups examined in this paper showed larger variations than the MO group in both F0 and amplitude as shown in Figure 5. This indicates that all the pathological groups have a lower ability to keep F0 and amplitude stable in producing sustained vowels. VP showed the greatest stability among the pathological groups.

2) The high and low frequency energy of F0 and the amplitude perturbations shown in Figures 6 and 7 indicate that, Tr and SD had stronger energy of slow fluctuations than others. Although there was not a simple tendency for the high frequency energy of F0 fluctuatios, both the female and male RP groups (RPF, RPM) showed significantly stronger energy of rapid fluctuations.

3) The Tr and SD groups showed lower F0 and amplitude fractal dimensions, although the SD group showed larger variations than Tr. This tendency and result 2 indicate that variations in the F0 and amplitude in the Tr and SD voice samples were relatively slow and large, but not necessarily irregular.

4) The additive noise level in the frequency range between 1 and 4 kHz was higher for all the pathological groups than the MO group. The TrF and SDM showed a overlapped distribution with the MO group. The RP and RE groups had a high noise level compared with VP. Clinically VP appears localized and smaller in size, while the RE is larger and diffuse with an aggravated phonatory mechanism. Among the pathological conditions, RPM had the largest score for noise

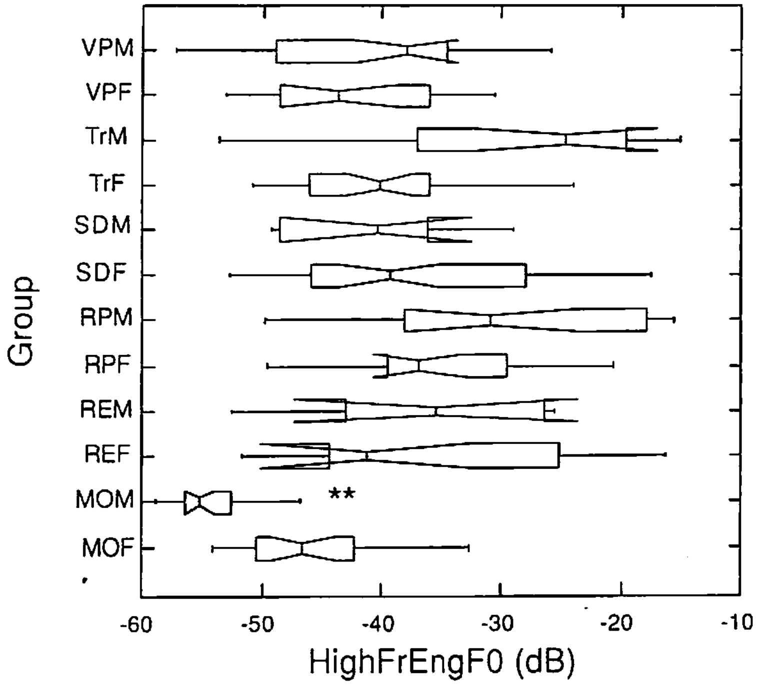

Fig. 7. Box-whisker graph of energy of the high frequency range of F0(i).

level. This could be due to an incomplete closure of the glottis, caused by the adductor paralysis of the cord.

These results indicate that a) Large but somewhat regular variations in F0 and amplitude with relatively strong fluctuation energy in the low frequency range were found mainly for SD and Tr voices, although SD seems changeable between tokens. b) Irregular fluctuations with a relatively high noise level and relatively strong fluctuation energy in the high frequency range were found in RP and RE voices.

By combining the clues from this evidence from a statistical point of view, we might speculate that characteristic a) may be mainly due to neurological sources, while b) have morphological sources, such as irregular vibrations of the vocal cords and irregular aerodynamic phenomena.

Since patients, or even healthy speakers, can produce voices with large F0 or amplitude fluctuations and a high noise level by changing laryngeal setting or physical conditions which determine vibratory characteristics of the vocal folds with or without intention, sources of perturbations may be complexly interdependent. The SD patients showed sudden and unpredictable changes in F0 or amplitude during phonation which might resulted in large variance for most of the acoustic parameters.

As shown in the results of the two factor ANOVA analysis, for most of the parameters, the interaction effect between group and sex was significant. There were significant pathology-dependent differences between the male and female speakers. This should be discussed in the future research.

It must be a very important issue to investigate which type of neural phathology causes what types of vocal problems and why. The results described here indicate that the acoustic analysis system described here can contribute to assess the vocal controllability and to search sources of various voice problems.

Conclusion

The controllability of vocal pitch and intensity was investigated analyzing the acoustic characteristics of 246 speakers' sustained vowel samples. The following results were obtained. 1) All the pathological groups examined here had a lower ability to keep F0 and amplitude stable in producing sustained vowels. 2) Voices with tremor showed relatively stable variations in F0 and amplitude with a strong magnitude of low frequency fluctuations, while voices produced by patients with morphological laryngeal pathology, particularly reccurent nerve paralysis, showed irregular fluctuations with a strong magnitude of high frequency fluctuations. 3) There were significant differences among the speaker groups, and significant pathology-dependent differences between the male and female speakers for all the parameters reported. These results suggest that the acoustic analysis system described here is beneficial in evaluating vocal controllability and in searching sources of various voice problems.

References

Aronson, A. et al.(1992). Rapid voice tremor, or flutter in Amyotrophic Lateral Sclerosis. *NCVS Status and Progress Report*, 2, 107-118.

Baken, R.J. (1990). Irregularity of vocal periode and amplitude: A first approach to the fractal analysis of voice. *J. of Voice*, 4, 185-197.

Hammarbeg, B., Fritzell, G., Gauffin, J., and Wedin, L.(1980). Perceptual and acoustic correlates of abnormal voice quali ties. *Acta Otolaryngol.*, 90, 441-451.

Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., Kikuchi, Y.(1988). Acoustic analysis of pathological voice. *Acta Otolaryngol.*, 105, 432-438.

Horii, Y.(1982). Jitter and shimmer differences among sustained vowel phonation. *J. Speech and Hearing Research.* 12-16.

Imaizumi, S.(1985). Acoustic measures of pathological voice quality. *J. Phonetics.* 457-462.

Imaizumi, S.(1986). Acoustic measurement of pathological voice qualities for medical purposes. *Proc. ICASSP*, 1, 677-680.

Imaizumi, S., Gauffin, J.(1991). Acoustical perceptual character istics of Pathological Voices: rough, creak, fry, and diplo phonia. *Ann. Bull. RILP*, 25, 109-119.

Imaizumi, S., Abdoerrachman, H., Niimi, Shimura, Y.,and Saida, H.(1993). Evaluation of vocal controllability by an object-oriented acoustic analysis system. *J. Acoust. Soc. Jpn (E).* (to be appeare).

Kasuya, H., Ogawa, S., Kikuchi, Y. and Ebihara, S.(1986). An acoustical analysis of pathological voice and its applica tion to the evaluation of laryngeal function. *Speech Communication*, 5, 171-181.

Koike, Y.(1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonetica*, 7, 17-23.

(For review) Laver, J., Hiller, S., Beck, J. M.(1992). Acoustic waveform perturbation and voice disorder. *J. of Voice*, 6, 115-126.

Lim, J., Oppenheim A.V. and Braida, L.D.(1978). Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition. *IEEE Trans. ASSP*, 26, 354-358.

Ludlow, C.L., Bassich, C.J., Connor, N.P. and Coulter.D.C. (1988). Phonatory characteristics of vocal fold tremor. *J.of Phonetics*, 14, 509-515.

Ramig, L.A., Shipp, T.(1987). Comparative measures of vocal tremor and vocal vibrato. *J. of voice*, 1, 162-167.