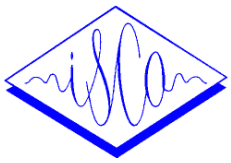


3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

CONTROLLABILITY OF VOICE QUALITY: EVIDENCE FROM PHYSIOLOGICAL AND ACOUSTIC OBSERVATIONS

Satoshi Imaizumi, Hartono Abdoerrachman**, Seiji. Niimi**

***Research Institute Logopedics Phoniatrics, Faculty of Medicine, University of Tokyo**
7-3-1 Hongo, Bunkyo-ku, 113 Tokyo
Email: imaizumi@tansei.cc.u-tokyo.ac.jp
****ENT Department, Faculty of Medicine, University of Indonesia**
J1. Diponegoro 71, Jakarta-Pusat, Indonesia

ABSTRACT

"Controllability" is defined as the ability to produce desirable pitch, intensity and timbre according to the speaker's intention. As one component of the "controllability", the ability to keep the vocal fundamental frequency, F0, and intensity as constant and as close to a target as possible when instructed to produce a sustained vowel was tested. Using an object oriented acoustic analysis system, magnitude of the slow and fast fluctuations in F0, fractal characteristics and some other voice-quality-related parameters were analyzed for singing, normal/modal and pathological voices. All the pathological groups showed larger variations in F0, or lower controllability, than the normal controls. The ability to keep vocal F0 and intensity as constant as possible was dependent on the target conditions which speakers intended to produce particularly for the neurological disorder patients. These results suggest that vocal controllability can be assessed quantitatively by the method proposed.

I. INTRODUCTION

"Controllability", we defined, is the ability to produce desirable pitch, intensity, and timbre according to the speaker's intention. Not only patients with neurological voice disorders 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 quality, pitch and amplitude.

In clinical examination of pathological voice, sustained vowel phonation have intensively been used to extract acoustic parameters such as jitter, shimmer and noise level which affect the voice quality. Sustained phonation, however, might be too simple to estimate the vocal "controllability" in daily conversation. For this purpose, conversational speech or even read tokens may be more useful. Precise acoustic analyses of such materials, however, are too complicated to be used in clinical examination.

Our strategy to solve this dilemma is to classify the "controllability" into several component abilities and to prepare proper measures to assess each component of the "controllability." This paper discusses the following two task-dependent abilities, C1 and C2.

C1: The ability to keep the vocal fundamental frequency, F0, and intensity as constant as possible when instructed to produce a sustained vowel.

C2: The ability to change F0 or intensity as precisely as possible when a target condition of phonation is specified.

Some other task-dependent abilities concerning the vocal "controllability" may be possible to define. For instance, the ability to control the timing of voicing gesture must be important for the production of distinctive voiced / unvoiced stop consonants. Those aspects are remained for the future works.

This paper describes an object oriented acoustic analysis system which was developed to assess the "controllability", C1 and C2. Slow variations in F0 and vocal intensity are particularly focused in terms of the "controllability."

Slow variations, have usually been disregarded as "trend" components (Laver et al., 1992; Kasuya et al.,1986; S. Imaizumi, 1986), even though such variations might have given rich information about pathological conditions particularly for neurological patients as suggested by some experts (Ludlow et al., 1988; Ramig & Shipp, 1987).

II. METHOD

2.1 Experiment I.

In Experiment I, an object-oriented acoustic analysis system, nicknamed SONG, was developed to assess the ability to keep F0 and intensity as constant as possible when instructed to produce a sustained vowel. Fig. 1 shows some of the panels which SONG produces for a voice sample produced by a patient suffered from spastic dysphonia (a neurological disorder).

Voice samples of a sustained vowel /e/ were digitized through a 12-bit A/D converter at a sampling rate of 50 kHz and stored on a disk controlled by a computer. A 1 s segment was extracted by excluding the initial and final portion from each sample. Using the method described by Imaizumi et al. (1986, 1991, 1994) local maximum points of the voice waveform which could correspond to vocal excitation epochs of each glottal cycles were detected successively, and then two time series F0(i) and A(i), the fundamental frequency and the maximum amplitude of i-th glottal period, were determined. Although the system displays 16 acoustic parameters which are useful to describe various aspects of voice quality, only the following acoustic parameters were reported in this paper: the overall variability, the level of slow and fast fluctuations in F0(i) normalized by DC level (dB), the fractal dimension of F0(i).

The overall variability is the percentage of the standard deviation normalized by the average of F0(i). For the statistical analysis, the logarithmic transformed value of the percentage was used.

To calculate the level of slow and fast fluctuations in F0(i), the power spectrum of the F0 curve was

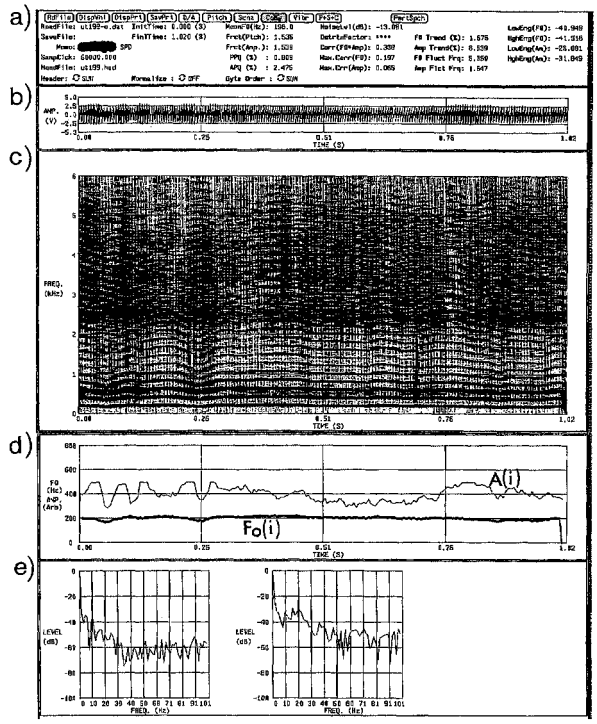


Fig. 1. The acoustic profile of a SPD voice analyzed by the system nicknamed as SONG. a)the control panel with display of analysis conditions and results, b) the voice waveform edit panel, c)the sound spectrogram panel, d)F0 and A(i) time series panel, and e) the power spectrum of F0(i) and A(i) panel.

approximated by the FFT power spectrum of time series F0(i), and then energy in the frequency ranges between 0<f<16, f=frequency, and 16<=f<average F0/2 were calculated. Finally, the logarithmic transformed values of them were normalized by DC level.

The fractal dimension FrctF0 is an index of irregularity calculated using Baken's method (1990). FrctF0 has a value between 1 and 2; 1 means that F0(i) is predictable, and 2 unpredictable or irregular.

These parameters were adopted to avoid defining "trend" components. It seems impossible to define what the "trend" is particularly for tremor, spastic dysphonia and vibrato voice samples.

The subjects were 246 patients with various diseases including 51 normal speakers: 51 normal/healthy voices (Hlth), 46 tremorous voices (Tr), 51 spastic dysphonia (SPD), 17 cases of Reinke's edema (RE), 44 vocal cord polyps (VCP), 37 cases of reccurent nerve paralysis (RNP). The Tr and SPD groups were the patients with neurological disorders. The RE, VCP and RNP were those with laryngeal disorders. Additionally, 57 samples of vibrato (Vib) and 29 samples of straight tones (Str) recorded from 4 singers (two soprano, 1 mezzo-soprano, and 1 baritone) were included to test if this system is useful to assess the vocal controllability for not only pathological but also artistic speakers.

The speakers were instructed to produce each of the Japanese five vowels for 2 or 3 s at their most comfortable pitch and intensity. Only voice samples of /e/ were analyzed according to the recommendation for clinical examination of voice by the Japan Society of Logopedics and Phoniatrics (1979).

Analyses of variance (ANOVA) with two factors, group and sex, were performed to determine the usefulness of these parameters to assess the vocal controllability.

2.2 Experiment II

The ability to change F0 or intensity as precisely as possible when instructed to produce a target value was tested for two patients of spastic dysphonia and three normal subjects. The target F0 and intensity values were set as the combinations between one intensity level (soft, comfortable and loud) and one F0 level (low, comfortable and high) for each subjects. The ability was assessed by measuring the changes in acoustic parameters observed when the subjects changed their voice from one target to another target.

III. RESULTS AND DISCUSSION

3.1 Experiment I

The acoustic profile of a SPD patient is shown in Fig. 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) the F0 and amplitude time series extracted by the acoustic analysis; and e) the power spectra of F0(i) and A(i). The SPD sample in Fig. 1(a) reveals slow but large variations in F0 and amplitude.

Figure 2 shows the box-whisker graph of the overall variability of F0(i). For this parameter, group ($p<0.0001$) and the interaction between group and sex ($p<0.0001$) were significant. Compared with the Hlth group, the pathological groups showed a larger variability. Among these, although VCP had the lowest variability (or highest stability) in producing sustained vowels with constant pitch, VCP had significantly larger variability compared with the Hlth group. There was a significant

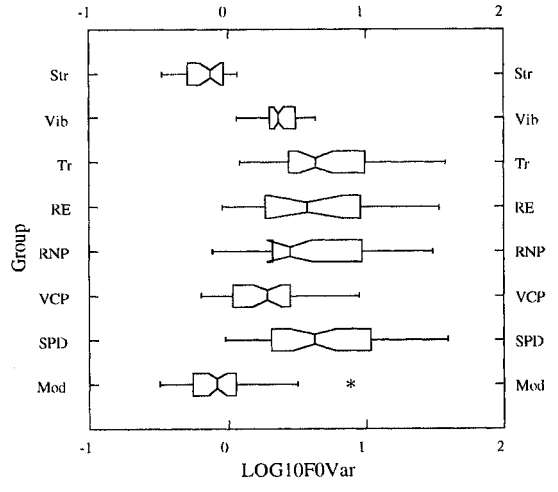
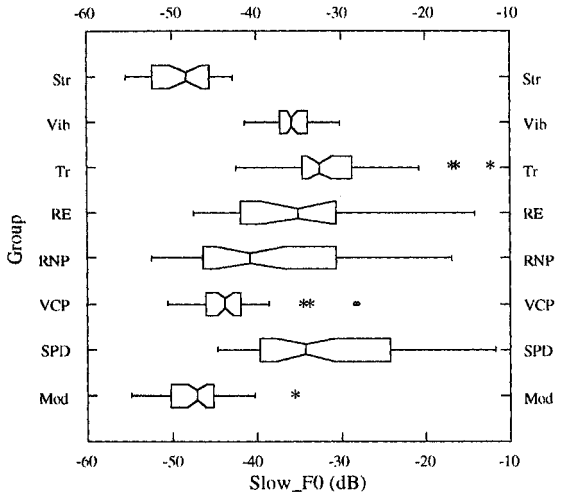
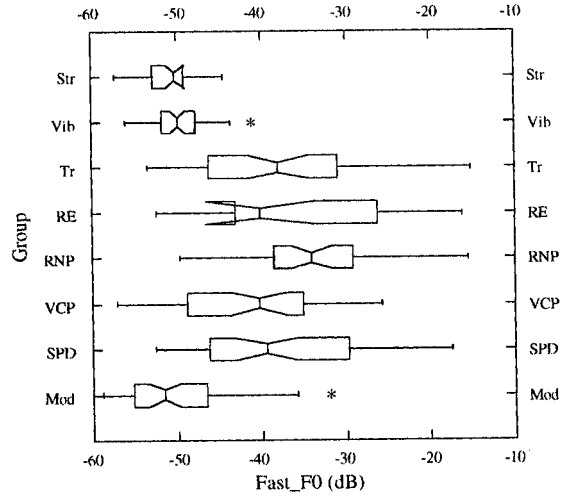


Fig. 2. Box-whisker graph of the overall variability of F0(i) in logarithmic scale. The center vertical line indicates the median. The edges of the central box represent the lower and upper hinges. The whiskers represent most remote data points from the median which are inside of the lower and upper 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%. If the confidence intervals around two medians do not overlap, the two population medians are significantly different at the 95% level.



(a)



(b)

Fig. 3. Box-whisker graph of the level of the slow (a) and the fast (b) F0 fluctuations.

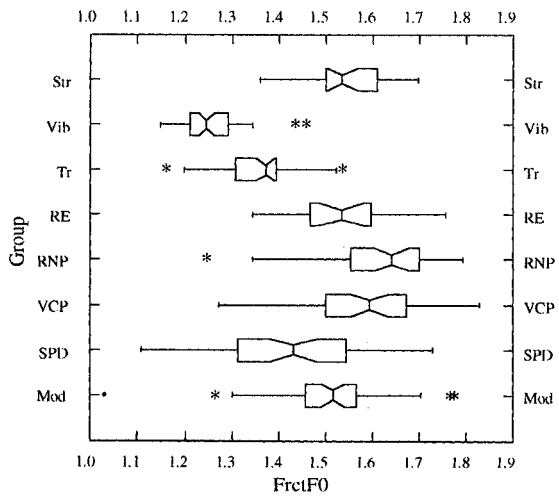


Fig. 4. Box-whisker graph of the fractal dimension of F0(i).

difference between Vib and Str, which indicated the singers controlled the F0 variability. The Vib group had a lower variability than the Tr, SPD, RE and RNP groups.

Figures 3 (a) and (b) show the box-whisker graphs of the spectrum energy of F0(i) in low (DC<f<16Hz) and high frequency (16<=f<=F0/2) ranges. These parameters represent the level of slow and fast F0 fluctuations. For these parameters, group ($p<0.0001$) and the interaction between group and sex ($p<0.0002$) were significant.

As shown in these figures, all the pathological groups showed higher levels both of the slow and fast F0 fluctuations than the Hlth and Str groups. The Tr and SPD groups had the highest in the slow F0 fluctuation, while RNP and RE had the highest level of the fast F0 fluctuation. The Vib had a higher level of the slow F0 fluctuation than the Hlth, Str and VCP groups, but a lower level than the Tr and SPD groups. The Vib samples had the lowest level of the fast F0 fluctuation.

Figure 4 shows the F0 fractal dimension. Group ($p<0.0001$), sex ($p<0.0001$) and the interaction ($p<0.005$) were significant. Compared with the Hlth group, the RE, VCP and RNP groups showed a higher fractal dimension, while the Tr and SPD groups had lower values than the Hlth. The SPD group revealed wider distribution compared to the Tr, which indicated the SPD had larger variations among tokens than Tr group. The Str samples had larger values than the Vib samples which had the lowest values.

Comparing the above mentioned parameters, the following tendencies were suggested.

1) Not only the Tr and SPD groups, but also the laryngeal pathological groups examined in this paper showed larger variations than the Hlth in F0 as shown in Fig. 2. This indicates that all the pathological groups have a lower controllability to keep F0 and amplitude stable in sustained vowels. The VCP showed the greatest stability among the pathological groups.

2) All the pathological groups tended to have higher levels of the fast F0 fluctuations than the Hlth, Vib and Str samples. The Tr and SPD tend to have higher levels of the slow F0 fluctuations than the other pathological groups. The Tr and SPD groups showed lower F0 fractal dimension, although the SPD group showed larger variations than the Tr. The variations in F0 in the Tr and SPD voice samples were relatively slow and large, but not necessarily irregular.

3) There were large differences in acoustic characteristics between the Vib and Tr groups as shown in Figures 2, 3 and 4. This result suggests that the mechanism generating the slow F0 fluctuations might be different between the opera singers and the patients with tremor or spastic dysphonia.

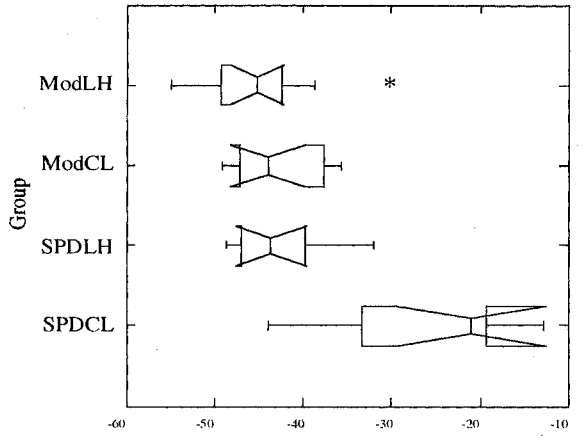
3.2 Experiment II

For the results of Experiment II, analyses of variance (ANOVA) with two factors, group and target, were performed.

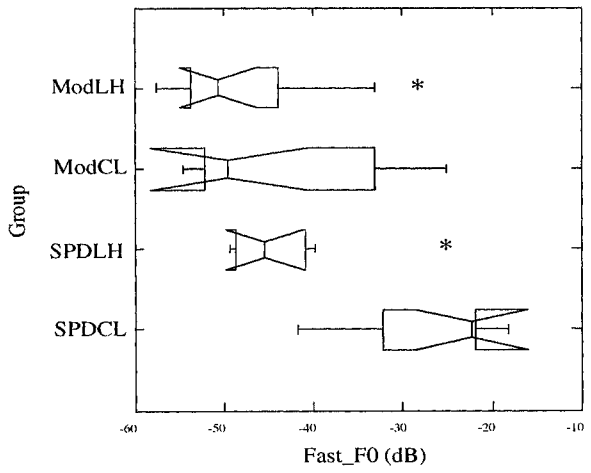
Figures 5 (a) and (b) show the box-whisker graph of the slow and fast F0 fluctuations measured at two target levels, a comfortable intensity with a low F0 versus a loud intensity with a high F0.

For the slow F0 fluctuation, as shown in Fig. 5 (a), target ($p=0.0003$) and group ($p=0.0002$) and their interaction ($p=0.0047$) were significant. Fisher's PLSD test also revealed a significant difference between the target levels ($p=0.0021$) and between the groups ($p=0.0004$).

The SPD patients showed significant change in this parameter. When they uttered at a loud intensity with a high F0, they had low levels of the slow F0 fluctuation which were comparable with those of the normal subjects.



(a)



(b)

Fig. 5. Box-whisker graph of the level of the slow (a) and the fast (b) F0 fluctuations. CL: Comfortable intensity and low F0 phonation; LH:Loud and high phonation.

They had very high level of the slow F0 fluctuations at comfortable intensity with a low F0.

For the fast F0 fluctuation, as shown in Fig. 5 (b), target ($p=0.0036$) and group ($p=0.0011$) and their interaction ($p=0.0443$) were significant. Fisher's PLSD test also revealed a significant difference between the target levels ($p=0.0124$) and between the groups ($p=0.0016$). The fast F0 fluctuations also varied depending on the target particularly for the SPD patients.

These results suggest that the ability to keep vocal F0 as close to a target value as possible is dependent on the target condition particularly for the SPD patients. At a loud and high phonation, they showed very low level of fast and slow F0 perturbations which were comparable with those of the normal speakers.

Comparing some dysarthria groups such as Parkinson and pseudobulbar palsy patients with peripheral disease (RNP and VCP), Hirose et al (1994) suggested that the vocal controllability is affected not only by the neural processes but also the physiological conditions of peripheral vocal organ. Acoustic characteristics reported here should be and are analyzed by taking account of physiological constraint of the vocal organs of the patients.

IV. CONCLUSION

"Controllability" is defined as the ability to produce desirable pitch, intensity and timbre. The "controllability" is classified into several task-dependent abilities and the following two, C1 and C2, are discussed; C1: the ability to keep the vocal fundamental frequency, F0, and intensity as constant as possible when instructed to produce a sustained vowel; C2: the ability to change F0 as precisely as possible when instructed to follow a target value.

In Experiment I, an object oriented acoustic analysis system was proposed to assess C1. Using the system, the slow and fast F0 fluctuations were analyzed for singing, normal/modal and pathological voices. All the pathological groups showed larger F0 fluctuations, or lower controllability, than the normal controls. The singers could control the magnitude of the slow F0 fluctuations in producing the vibrato and straight tones.

In Experiment II, the ability to change F0 as precisely as possible when instructed to follow a target value was tested for two patients of spastic dysphonia and three normal subjects. At a loud and high phonation, the SPD patients showed very low level of the slow and fast F0 fluctuations which were comparable with those of the normal speakers, although they showed very high level of fluctuations at a comfortable loudness and low pitch phonation. These results suggest that the ability to keep vocal F0 and intensity as close to a target as possible is dependent on the target condition for the SPD patients.

These results suggest that vocal controllability can be assess quantitatively by classifying it into several task-dependent abilities. Slow fluctuations are very important to assess the "controllability," although those fluctuations have been disregarded as "trend" components.

References

Baken, R. J. (1990). "Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice," J.voice. 4 (3), 185-197.

Hirose, H., Imaizumi, S. and Yamori, M. (1994). "Voice Quality in Patients with Neurological Disorders," Proc. of Vocal Fold Physiol. Conf., Kurume, April, 1994.

Imaizumi, S. (1985). "Acoustic measures of pathological voice quality," J. Phonetics. 14, 457-462.

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

Imaizumi, S., and Gaufin, J. (1991). "Acoustical perceptual characteristics of Pathological Voices: rough, creak, fry, and diplophonia," Ann. Bull. RILP. 25, 109-119.

Imaizumi, S. et al. (1993). "Acoustic evaluation of vocal controllability-Characteristics of vocal registers and vibrato-" Technical Report of IEICE, SP93-67, 25-29.

Imaizumi, S., Abdoerrachman, H., et al. (1994). "Evaluation of vocal controllability by an object oriented acoustic analysis system," J. Acoust. Soc. Jpn(E), 15(2), 113-116.

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

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

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

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

The Japan Society of Logopedics and Phoniatrics, (1979). "Clinical Examination of Voice, (IshiYaku-Shuppan, Tokyo, 1979), p150 (in Japanese).