Evaluation of vocal controllability by an object oriented acoustic analysis system

Satoshi Imaizumi,* Hartono Abdoerrachman,** Seiji Niimi,* Hajime Hirose,* Haruhito Saida,*** and Yoko Shimura****

*RILP, Faculty of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113 Japan
**Department of Otolaryngology, Faculty of Medicine, University of Indonesia, J1, Diponegoro 71, Jakarta-Pusat, Indonesia
***Department of Otolaryngology, Asoka Hospital, 1-18-1 Shumiyoshi, Koto-ku, Tokyo, 135 Japan
****Department of Education, Saitama University, 225 Shimo-okubo, Urawa, Saitama, 338 Japan

(Received 15 June 1993)

Keywords: Vocal controllability, Pathological voice, Tremor, Vibrato, Acoustic analysis

PACS number: 43.70.Dn, 43.70.Gr

1. Introduction

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 pitch and amplitude. "Controllability" is the ability to produce desirable pitch, intensity, and timbre according to the speaker's intention.

Slow variations might be very important to assess the "controllability." Slow variations, however, have usually been disregarded as "trend" components, even though such variations might have given rich information about pathological conditions particularly for neurological patients as suggested by some experts.4,5)

This paper describes an object oriented acoustic analysis system to assess "controllability" of pitch and intensity. "Controllability" is defined here as the ability to keep the vocal fundamental frequency as constant as possible when instructed to produce a sustained vowel.

2. Method

The subjects were 246 patients with various diseases including 51 normal speakers: 51 normal/healthy voices (H1th), 46 tremorous voices (Tr), 51 spastic dysphonia (SPD), 17 cases of Reinke's edema (RE), 44 vocal cord polyps (VCP), 37 cases of recurrent 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.6) 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.,7,8) 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 $F_0(i)$ and $A(i)$, the fundamental frequency and the maximum amplitude of $i$-th glottal period, were determined. Although the system displays 16 acoustic parameters, only the following acoustic parameters were reported in this paper: the overall variability, the low and high frequency energy of time series $F_0(i)$ normalized by DC level (dB), the fractal dimension of $F_0(i)$.

The overall variability is the percentage of the standard deviation normalized by the average of $F_0(i)$. For the statistical analysis, the logarithmic transformed value of the percentage was used. To calculate the low and high frequency energy of time series $F_0(i)$, the power spectrum of the $F_0$ curve was approximated by the FFT power spectrum of time series $F_0(i)$, and then energy in the frequency ranges between $0 < f < 16$, $f = $ frequency, and $16 < f < \text{average } F_0/2$ were calculated. Finally, the logarithmic transformed values of them were normalized by DC level.

The fractal dimension Frct $F_0$ is an index of irregularity calculated using Baken's method. Frct $F_0$ has a value between 1 and 2; 1 means that $F_0(i)$ 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, SPD and...
vibrato voice samples. Analyses of variance (ANOVA) with two factors, group and sex, were performed to determine the usefulness of these parameters to assess the vocal controllability.

3. Results and discussion

The acoustic profiles of a SPD sample 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 \( F_0 \) and amplitude time series extracted by the acoustic analysis; and e) the power spectra of \( F_0(i) \) and \( A(i) \). The SPD sample in Fig. 1 reveals slow but large variations in \( F_0 \) and amplitude.

Figure 2 shows the box-whisker graph of the overall variability of \( F_0(i) \). For this parameter, group \( (p<0.0001) \) and the interaction between group and sex \( (p<0.0001) \) were significant. Compared with the H1th 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 H1th group. There was a significant difference between Vib and Str, which indicated the singers controlled the \( F_0 \) 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 \( F_0(i) \) in low \((DC<f<16\text{ Hz})\) and high frequency \( (16<f<F_0/2) \) ranges. For this parameter, 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 stronger energy both in the low and high frequency ranges than the H1th and Str groups. The Tr and SPD groups had the strongest energy in the low frequency range, while RNP and RE had the strongest energy in the high frequency range. The Vib had a stronger energy in the low frequency range than the H1th, Str and VCP groups, but a weaker energy than the Tr and SPD groups. The Vib samples had the weakest energy in the high frequency range.

Figure 4 shows the \( F_0 \) fractal dimension. Group \( (p<0.0001) \), six \( (p<0.0001) \) and the interaction \( (p<0.005) \) were significant. Compared with the H1th group, the RE, VCP and RNP groups showed a higher fractal dimension, while the Tr and SPD groups had lower values than the H1th. 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 Tr 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 stronger energy of rapid fluctuations than the H1th, Vib and Str samples. The Tr and SPD tend to have stronger energy of slow fluctuations than the other pathological groups. The Tr and SPD groups 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) Although some experts suggested that the voice tremor and the singer’s vibrato might be generated through a common mechanism, there were large differences in acoustic characteristics between the Vib and Tr groups as shown in Figures 2, 3 and 4.

4. Conclusion

An object oriented acoustic analysis system was developed to assess the controllability of vocal pitch and intensity. Using the system, slow and fast variations in the vocal fundamental frequency, F0, 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 singers controlled F0 variations in producing the vibrato and straight tones. There were significant differences in acoustic characteristics between them. These results indicate that the system is useful in analyzing vocal controllability for various voices covering pathological and artistic super-normal voices.

References

1) (For review) J. Laver, S. Hiller, and J. M. Beck, "Acoustic waveform perturbation and voice dis-


