Ann.Bull.RILP No.28(1994)45-50

Test of Vocal Controllability with Increasing Pitch and Intensity

H. Abdoerrachman*, S. Imaizumi** and S. Niimi**

*Center of Ear Care and Communicative Disorder, Dept. of Otolaryngology
Faculty of Medicine, University of Indonesia
Jl. Diponegeoro 71, Jakarta-Pusat, Indonesia
**Research Institute of Logopedics and Phoniatrics, Faculty of Medicine
University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo 113, Japan

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.

Furthermore, slow variations in vocal fundamental frequency and intensity 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). Slow variations in F_0 and vocal intensity should be analyzed in terms of the "controllability."

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 ability to maintain F_0 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.

In this study, controllability of pitch and intensity in patients with spastic dysphonia were investigated by increasing their vocal pitch and intensity relative

Method

The subjects for this study were 5 speakers, consisting of 2 patients with spastic dysphonia (SPD), and 3 healthy speakers (Hlth) without any pathology affecting phonation.

The speakers were requested to produce a sustained vowel /aa/ for 2 or 3 seconds at a comfortable pitch but low intensity (ComfLow) and then increase their voices to a higher pitch and higher intensity (LoudHigh) in fractional steps, following practice. Recordings were made using a DAT tape-recorder, with a constant distance of 15 cm from microphone to mouth, and performed in a sound treated booth.

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

Using the method developed by Imaizumi (1986, 1991, 1994), local maximum points which could correspond to vocal excitation epochs were detected successively, and then the fundamental frequency $F_0(i)$ and the maximum amplitude A(i) of i-th glottal period ware determined. Several acoustic properties were extracted: the mean F_0 (Hz), Pitch Perturbation Quotient (APQ), additive noise level (dB), F_0 overall variability, energy of slow and fast F_0 perturbations normalized by DC level (dB).

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 level of slow and fast fluctuations in $F_0(i)$, the power spectrum of the F0 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 $16<=f<$average $F_0/2$ were calculated. Finally, the logarithmic transformed values of them were normalized by DC level.

Analysis of variance (ANOVA) with two factors, voice condition and voice type, were performed to determine significance.

Results and Discussion

The present preliminary study yielded the following results.

The interaction bar plot for the Mean F_0 of all speakers shown in Fig. 1 indicated that from ComfLow to LoudHigh target, the subjects increased their pitch to almost the same height. The interaction line plot for log10(PPQ), shown in Fig. 2 revealed that PPQ for the SPD group decreases significantly ($p<0.05$), compared to normal speakers. The additive noise level (NLvl) in all groups (Fig. 3) showed a significant decrease in the NLvl for all speakers. ($p<0.0001$). The overall F_0 variability (Fig. 4) indicated a significant decrease for the SPD group ($p<0.001$), and about the same value as healthy group. The energy of the slow F_0 perturbations for all groups (Fig. 5) indicated a significance decrease in energy for the SPD group ($p<0.0001$), a slight decrease in the normals, and about the same level of strength for both. The energy of the fast F_0 perturbations (Fig. 6) showed a significant decrease in energy for the SPD group ($P<0.008$).

The results discussed above yielded the following:

(1) By increasing the pitch and intensity from comfortable-low to target loud-high voice, the SPD patients showed significant decreases in the scores of PPQ, APQ, F_0 variability, energy of slow and fast F_0 perturbations, and additive noise level. Thus, by increasing the frequency and amplitude of their voices when producing sustained vowels, the variability, stability and regularity of their voices were improved, indicating that their vocal controllability was better, and that their voices become more stable, regular and less breathy.

From the neuro-physiological aspect, the increases in the force of contraction of the muscles were brought about by three mechanisms: a) an increase in the number of motor units activated, b) an increase in the frequency of impulses in each motor unit, and c) a synchronization of different motor units (Hirono, 1981).

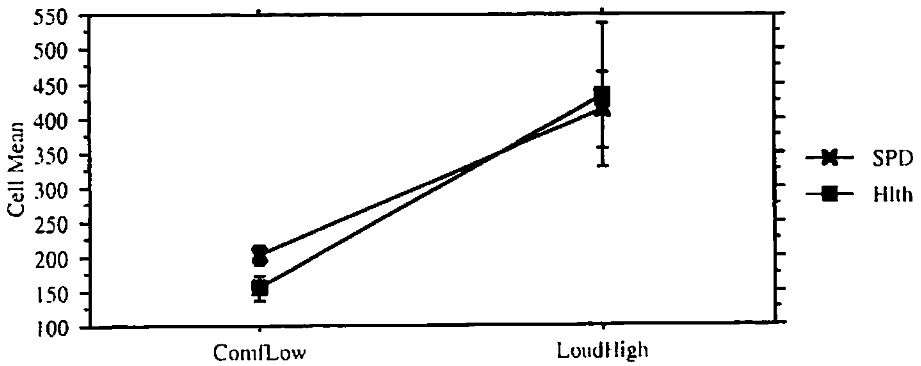

Fig. 1 shows the interaction bar plot for the Mean F_0 of all speakers, indicating that from ComfLow to LoudHigh target, the subjects increased their pitch to almost the same height.

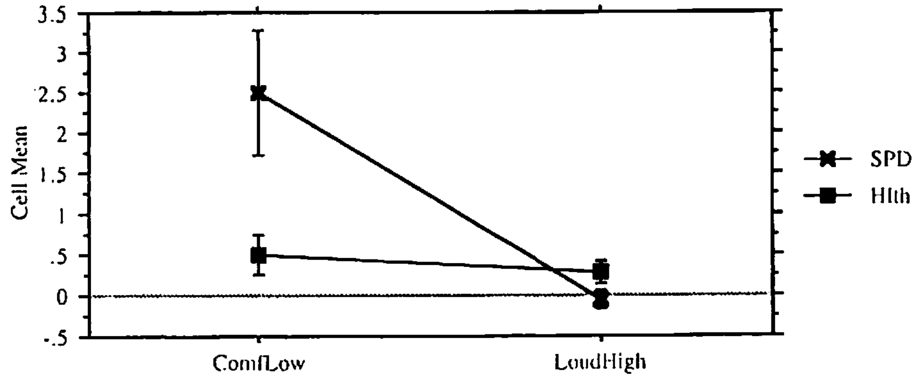

Fig. 2 shows the interaction line plot for log10(PPQ), revealing that PPQ for the SPD group decreases significantly ($p<0.05$), compared to normal speakers.

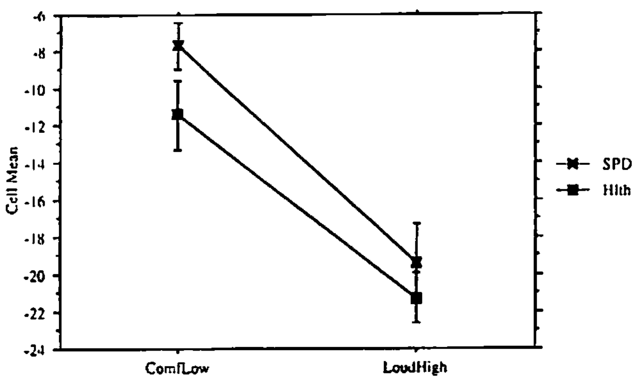

Fig. 3 presenting the interaction line plot of the additive noise level (NLvl) in all groups, indicating a significant decrease in the NLvl for all speakers. ($p<0.0001$)

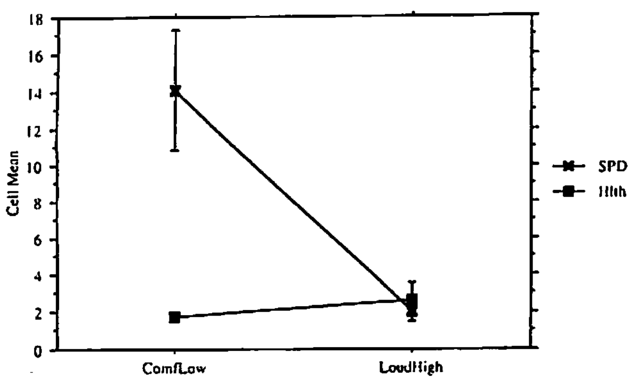

Fig. 4 showing the interaction line plot for overall F_0 variability, indicating a significant decrease for the SPD group ($p<0.001$), and about the same value as healthy group.

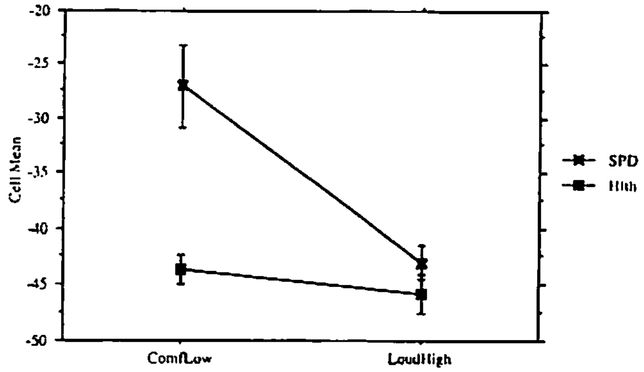

Fig. 5 shows the interaction line plot of the energy of the slow F_0 perturbations for all groups, indicating a significance decrease in energy for the SPD group ($p<0.0001$), a slight decrease in the normals, and about the same level of strength for both.

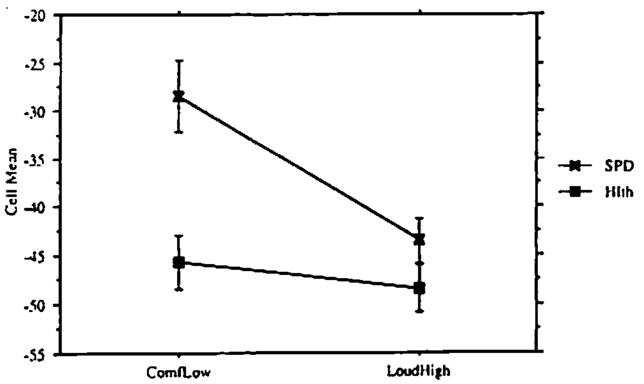

Fig. 6 shows the interaction line plot for the energy of the fast F_0 perturbations, with a significant decrease in energy for the SPD group ($P<0.008$).

These phenomena suggested that when producing target loud-high voices, the patients may have made an extra effort to reach the target pitch. From the mechanical point of view, the patients increased and maintained the subglottic pressure to raise loudness and also increased the laryngeal muscles contraction brought about by above mentioned mechanisms. These situations and conditions result in regulation and stabilization of the vocal fold's vibration.

In this study, the voices samples of comfortable-low and target low-high voices were recorded under the same situation and condition, so that effects of the emotional state of the patients could be minimized. The improvements in the voice conditions of the patients were mainly due to aero-dynamic and mechanical phenomena, although we believed these were not as simple as for speaking mode.

(2) At the given target loud-high phonation, the patients showed a decrease in some parameters, which were comparable with those of normal speakers.

Conclusion

Investigation on the controllability of vocal pitch and intensity was performed by analyzing the acoustic properties of 5 speakers' sustained vowel voice samples and the following results were obtained. (1) By increasing pitch and intensity, the SPD patients showed a higher ability to keep F_0 and amplitude stable. (2) The particular target-voice conditions enabled the patients to decrease several voice properties, which became comparable to those of normal speakers.

These results suggest that the analysis system used is useful in establishing the vocal controllability and speculating on its background mechanism. Future extended investigations involving more voice samples consisting of various types of speakers will be necessary to clarify the inquiry.

References

Bear, T. (1981). Investigation of phonatory mechanism. In Proceedings of the Conference on the Assessment of Vocal Patholigy: ASHA Report No. 11 (C. L. Luddlow & M. O'C. Hart, editors) pp. 38-47. Rockville, M. D. :American Speech-Language-Hearing Association. (as review).

Hachinsky, V. C. , Thomsen, I. V. & Buch, N. H. (1975). The nature of primary vocal tremor. Canadian Journal of Neurological Science, 2, 195-197. (as review).

Hirano, M. (1981). *Clinical Examination of Voice*. Disorder of Human Communication 5, pp. 11-24.

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., Hartono Abdoerrachman, Niimi, S., Hirose, H., Saida, H., and Shimura, Y. (1994) Evaluation of vocal controllability by an object oriented acoustic analysis system. J. Acoust. Soc. Jpn. (E) 15, 113-116.

Hartono Abdoerrachman, Imaizumi, S. , Hirose, H. , Niimi, S. (1993) Slow and Fast Perturbations InVoice-A Prelminary Report-Ann. Bull. RILP, 27, 125-134.

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.