Prosodic and Segmental Evaluation of Dysarthric Speech

Hiroki Mori†, Yasunori Kobayashi†, Hideki Kasuya†, Hajime Hirose‡ & Noriko Kobayashi‡

†Faculty of Engineering, Utsunomiya University
‡School of Allied Health Sciences, Kitasato University
hiroki@klab.jp

Abstract

We are investigating acoustical analysis for dysarthric speech, which appears as a symptom of neurologic disease, in order to elucidate its physiological and acoustical mechanism, and to develop aids for diagnosis and training, etc. In this report, acoustical characteristics of various kinds of dysarthrias are measured. As a result, shrinking of the F_0 range as well as vowel space are observed in dysarthric speech. We performed a perceptual experiment to clarify how such parameters affect so-called "monotonous" impression, and found that abnormality in the F_0 range affects the monotonous impression.

1. Introduction

Dysarthria is a symptom of neurologic diseases such as pseudo bulbar palsy (PBP), Parkinson's disease, spinocerebellar degeneration (SCD), amyotrophic lateral sclerosis (ALS), etc. The symptoms of dysarthrias often appear as prosodic disorders such as monopitch or monoloudness, as well as weak articulation or omission of segments.

There have been many reports on the acoustical characteristics of dysarthric speech[1, 2, 3, 4, 5, 6, 7, 8, 9]. Canter[1] reported a higher F_0 level and reduced F_0 range in speech of patients with Parkinson's disease. Although he did not find any significant difference in intensity measures between normal control and Parkinson patients, several later works indicated inconsistent results[2, 3]. Turner et al.[4] showed smaller vowel space areas in speech of ALS patients compared with neurologically normal subjects.

We have been developing several methods of acoustical analysis for dysarthric speech for the purpose of elucidating its physiological nature and developing the aid for the diagnosis and training of dysarthrias. Based on the above works, the present research focuses on the following topics: 1) To evaluate acoustical characteristics of dysarthrias by examining both prosodic and segmental features, and 2) To investigate which acoustical feature affects the monotonous impression of dysarthric speech by modifying prosodic and segmental parameters.

2. Acoustical evaluation

2.1. Method

2.1.1. Subjects

The speech samples subjected to the acoustical analysis were obtained from 16 adult male dysarthric patients consisting of 5 cases of pseudobulbar pulsy (PBP), 7 cases of Parkinson disease (PKN), and 4 cases of amyotrophic lateral sclerosis (ALS). As a control, speech samples were also obtained from 6 normal adult males (CNT).

2.1.2. Recording

The recordings were carried out in a soundproof room. Each subjects read an Aesop story "The North Wind and The Sun" (8 sentences) or "Sakura" passage (8 sentences), depending on recording date. A sound level meter (Ono Sokki LA-5111 with an electret condenser microphone MI-1233) was used for some subjects, to perform high-quality and level-calibrated recording. Table 1 shows the details. Each speech data was digitally recorded at the sampling frequency of 48 kHz using DAT, then applied a digital low-pass filter (cutoff 5500 Hz) and downsampled to 12 kHz.

2.1.3. Parameters

The F_0 range and F_0 minimum are used for prosodic evaluation, and vowel formant frequencies (F_1, F_2) are used for segmental evaluation.

F_0 range For each recorded sentence, its F_0 contour was extracted with the multiple window length method[10], then errors were corrected manually.

Each sentence was segmented into intonation phrases (IPs) according to the JToBI[11] framework. Some sentences were spoken disfluently (typically by PBP patients), and included self-corrections and repetitions. We discarded such disfluent portions from IPs.

For each IP, its F_0 range was calculated in the logarithmic domain.

F_0 minimum For each IP, its F_0 minimum was obtained as the lowest F_0 value in the IP.

Vowel formant frequencies Formant frequency contours were extracted automatically from whole utterances with the ARX speech analysis method, as described in later. Phoneme labeling was also performed manually. For each vowel, the first and second formant frequencies (F_1 and F_2) at the vowel center (50% point of the vowel duration) were extracted, which were then manually checked to avoid erroneous values.

2.2. Result

2.2.1. Prosodic characteristics

The F_0 range and the F_0 minimum of IPs obtained from each group of the subjects, CNT, PKN, ALS and PBP, were plotted in Figs. 1 to 4, respectively. The abscissa indicates F_0 minimum, while the ordinate indicates the F_0 range in semitones.

It was apparent that the F_0 range of dysarthric speech was generally narrower than that of the normal subjects, suggesting that their intonation pattern should be flat. This tendency

Table 1: *Speech materials.*

	CNT1	CNT3	CNT4	CNT5	CNT6	CNT7	PKN1	PKN2	PKN3	PKN4	PKN5
age	60s	50s	30s	60s	40s	60s	50s	40s	60s	50s	50s
passage	The North Wind and The Sun					Sakura	The North Wind and The Sun				
mic	dynamic					electret	dynamic				

	PKN6	PKN7	ALS1	ALS2	ALS3	ALS4	PBP1	PBP2	PBP3	PBP4	PBP5
age	60s	70s	50s	50s	50s	40s	60s	60s	50s	70s	50s
passage	The North Wind and The Sun					Sakura	The North Wind and The Sun				Sakura
mic	dynamic					electret	dynamic				electret

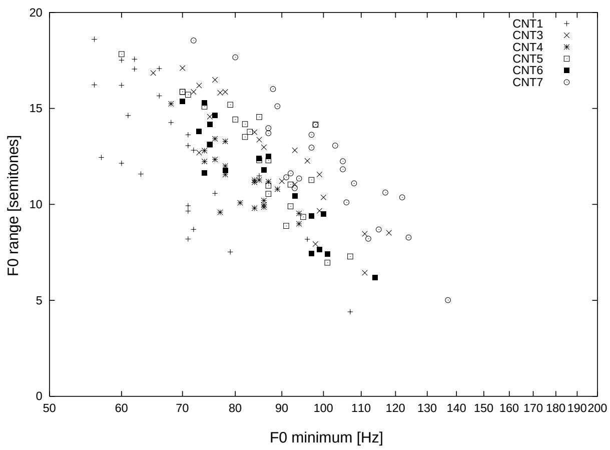

Figure 1: *Scatter plot of F_0 mininum vs. F_0 range (intact control).*

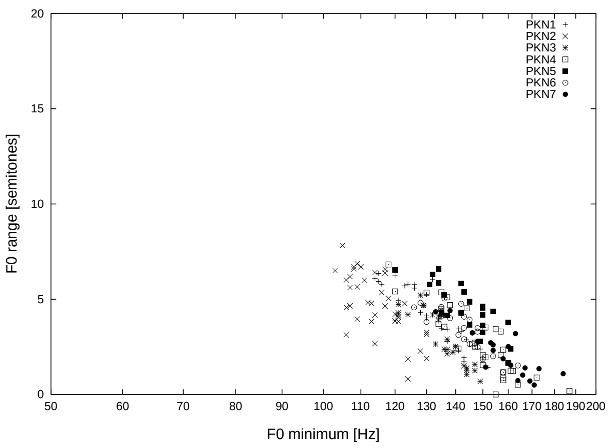

Figure 2: *Scatter plot of F_0 minimum vs. F_0 range (PKN).*

was most prominent in PKN. It was also apparent that the F_0 minimum in PKN was significantly higher than that of ALS or normal controls. A similar tendency was noted in some, but not all cases of PBP. From these results, it should be concluded that the flat intonation pattern was a common feature among dysarthrias, while the pattern of F_0 distribution reflected the difference in the type of dysarthric speech.

As for the physiological mechanisms underlying the above acoustical characteristics, it can be assumed that increased tension in the vocal folds due to rigidity resulted in higher F_0 level in PKN, while the lowering in vocal fold tension due to muscle weakness led to lower F_0 level in ALS. For PBP, the apparent bimodal distribution in F_0 range was most likely due to the different types of vocal manifestation in PBP, hypertensive and hypotensive, reported elsewhere[12].

2.2.2. Segmental characteristics

Vowel spaces in the F_1-F_2 plane are shown for CNT7 and PBP5, in Figs. 5 and 6, respectively. As shown in the figures, all vowels of PBP5 are overlapping the /i/ and /u/ region of CNT7. In the PBP case, especially low vowels, /a/ and /o/, occur at distant positions from those of the intact control. This fact suggests that movement to the low jaw position is incomplete.

3. Parameter Conversion

Based on the above results, dysarthric speech can be changed as if its prosodic/segmental parameters were those of normal values, which may be a useful approach to distinguish those factors that might affect the monotonous impression.

In this paper, the ARX speech analysis-synthesis method[13] is used to change the speech parameters. It has been shown that speech resynthesized with the method is highly natural, and the method is robust to parameter modification.

For the PBP5 speech data, voice source parameters (F_0, source amplitude, open quotient, etc.) and formant parameters ($F_1, F_2, \ldots, F_6$, and their corresponding intensities) were extracted using the ARX analysis method. The ARX synthesis method reconstructs the original speech from the source and formant parameters.

3.1. F_0 range modification

As described above, dysarthric speech has a narrower F_0 range than normal speech. The F_0 contours of PBP5 were modified to have normal range by the following two methods.

F_0 range magnification (Method P1) The F_0 contours of PBP5 were linearly scaled according to the following formula:

$$\log F_0'(t) = \overline{\log F_0} + \alpha(\log F_0(t) - \overline{\log F_0}), \qquad (1)$$

where $\overline{\log F_0}$ denotes the mean $\log F_0$ value in IP.

The conversion was performed independently for each IP. The magnification coefficient α was determined as the ratio of the control group's mean F_0 range to PBP5's mean F_0 range, which were calculated from Figs. 1 and 4.

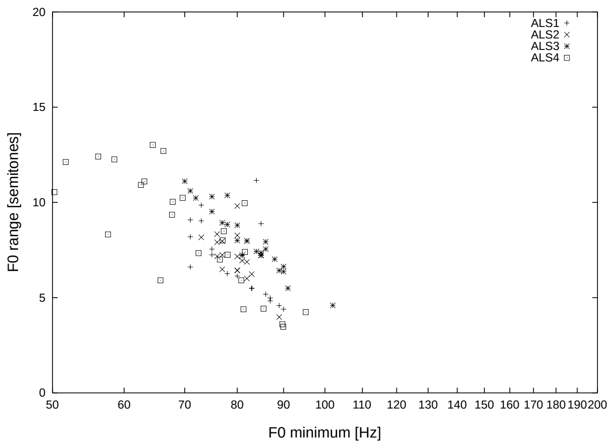

Figure 3: *Scatter plot of F_0 minimum vs. F_0 range (ALS).*

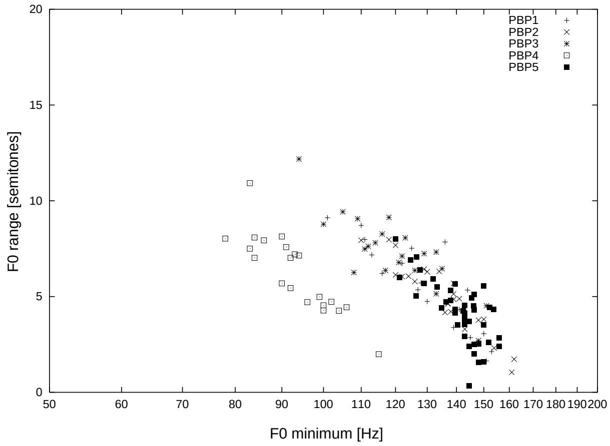

Figure 4: *Scatter plot of F_0 minimum vs. F_0 range (PBP).*

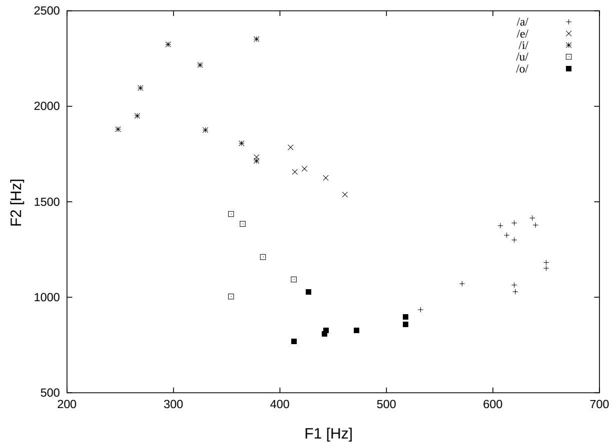

Figure 5: *Vowel space for CNT7.*

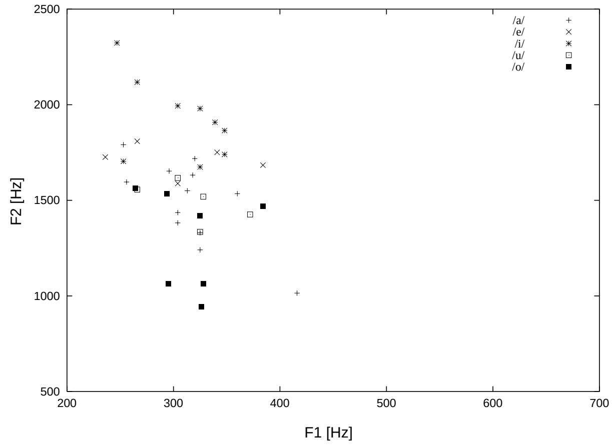

Figure 6: *Vowel space for PBP5.*

***F_0* replacement (Method P2)** One healthy male speaker (30s, speech researcher) read the "Sakura" passage, trying to speak with the same rhythm as PBP5, by hearing his utterances. F_0 contour extraction was performed after the recording, then the original F_0 contours of PBP5 were replaced by these F_0 contours.

3.2. Formant modification (Method S)

As shown in Fig. 6, the vowel space of PBP5 is narrowed. This segmental abnormality could cause a monotonous impression. To verify this hypothesis, formant frequencies of low vowels (/a/ and /o/) were shifted to their expected position.

The modification was done as follows: 1) Conversion factors (C_1, C_2) for (F_1, F_2) were calculated using the results shown in Figs. 5 and 6. We got (2.0, 0.9) for /a/ and (1.5, 0.75) for /o/. 2) Formant frequencies at the vowel center of /a/ and /o/ were converted according to the above factors. 3) Formant frequencies in the /a/ and /o/ region were interpolated so as to maintain continuity.

Formally, the converted i-th formant frequency contour is given by the following formula:

$$F_i'(t) = F_i(t) \cdot C_i^{\sin \frac{t-t_\mathrm{s}}{t_\mathrm{e}-t_\mathrm{s}}\pi} \qquad (t_\mathrm{s} \leq t \leq t_\mathrm{e}), \tag{2}$$

where t_s and t_e represent the start and end time of the vowel, respectively. A schematic picture of formant frequency conversion is shown in Fig. 7.

3.3. Auditory impression evaluation

Five speech therapists carried out an auditory impression evaluation for the resynthesized speech of PBP5 (4 sentences) using the Visual Analogue Scale method. The parameters were modified according to the combination of 3 prosodic (no conversion, Method P1 and Method P2) and 2 segmental (no conversion, Method S) conversion conditions. Thus six different versions of modified speech were presented for each sentence. Each version was presented four times, so the total number of sentences was $4 \times 6 \times 4 = 96$. The order of presentation of the stimuli was randomized. The listeners were asked to mark perceived monotonicity for each sentence on a printed scale.

The result is shown in Fig. 8. The vertical axis indicates the mean evaluated scores: 0 corresponds to no monotonous impression, whereas 200 corresponds to extreme monotone. This figure shows that concerning prosody, listeners A, C and E evaluated the monotonicity consistently: original speech is most monotonous, and F_0-replaced speech (Method P2) is least monotonous. This result suggests the importance of the F_0 range for the monotonous impression. However, the evalua-

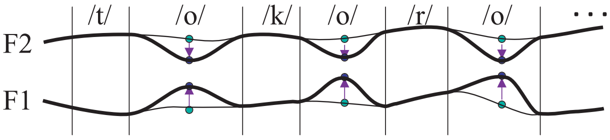

Figure 7: *Schema for formant frequency conversion.*

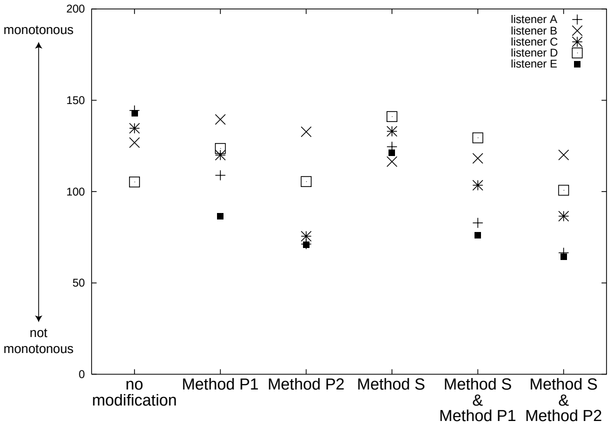

Figure 8: *Auditory impression evaluation of "monotonicity."*

tions by listeners B and D were not necessarily consistent. The effect of formant frequency conversion seems to be dependent on the listener, and the contribution of segmental features to the monotonous impression is not apparent from the result.

4. Conclusion

Acoustical analysis for dysarthric speech from prosodic and segmental aspects was discussed. It was revealed that the F_0 range and vowel space in F_1-F_2 are narrowed in dysarthric speech. We performed a perceptual experiment to clarify how such parameters affect so-called "monotonous" impression, and found that abnormality in the F_0 range affects the monotonous impression.

We think that intensity and rhythm are also important factors that affect the monotonous impression. In the future we will evaluate dysarthric speech subjectively using these measures, and assess their importance by the analysis-synthesis technique.

Acknowledgment

The authors would like to thank the speech therapists at Kitasato University for their cooperation.

5. References

[1] G. J. Canter, 1963. Speech characteristics of patients with Parkinson's Disease: Intensity, pitch and duration. *J. Speech Hear. Dis.*, 28(3), 221–229.

[2] C. L. Ludlow and C. J. Bassich, 1983. The results of acoustic and perceptual assessment of two types of dysarthria. In *Clinical Dysarthria*, W. R. Berry (ed.). San Diego: College-Hill Press, 121–153.

[3] L. O. Ramig, C. M. Bonitati, J. H. Lemke and Y. Horii, 1994. Voice therapy for patients with Parkinson disease: development of an approach and preliminary efficacy data. *J. Med. Speech-Language Path.*, 2, 191–210.

[4] G. S. Turner, K. Tjaden and G. Weismer, 1995. The influence of speaking rate on vowel space and speech intelligibility for individuals with Amyotrophic Lateral Sclerosis. *J. Speech Hear. Res.*, 38(5), 1001–1013.

[5] A. J. Caruso and E. K. Burton, 1987. Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. *J. Speech Hear. Res.*, 30(1), 80–87.

[6] I. Hertrich and H. Ackermann, 1999. Temporal and spectral aspects of coarticulation in ataxic dysarthria: an acoustic analysis. *J. Speech Lang. Hear. Res.*, 42(2), 367–381.

[7] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian and J. R. Duffy, 1999. Acoustic studies of dysarthric speech: methods, progress, and potential. *J. Commun. Dis.*, 32(3), 141–186.

[8] R. D. Kent, R. Netsell and J. Abbs, 1979. Acoustic characteristics of dysarthria associated with cerebellar disease. *J. Speech Hear. Res.*, 22, 627–648.

[9] W. Ziegler and D. von Cramon, 1986. Spastic dysarthria after acquired brain injury: an acoustic study. *British Journal of Disorders of Communication*, 21, 173–187.

[10] T. Takagi, N. Seiyama and E. Miyasaka, 1997. A method for pitch extraction of speech signals using autocorrelation functions through multiple window-lengths. *Trans. IEICE (A)*, J80-A(9), 1341–1350.

[11] J. J. Venditti, 1997. Japanese ToBI Labelling Guidelines. *Ohio State Univ. Work. Pap. Linguist.*, 50, 127–162.

[12] H. Hirose, S. Imaizumi and M. Yamori, 1995. Voice quality in patients with neurological disorders. In *Voice Quality Control*, O. Fujimura and M. Hirano (eds.). San Diego: Singular, 235–248.

[13] T. Ohtsuka and H. Kasuya, 2000. An improved speech analysis-synthesis algorithm based on the autoregressive with exogeneous input speech production model. *Proc. ICSLP 2000*, II, 787–790.