F0 and Formant Frequency Distribution of Dysarthric Speech
— A Comparative Study —

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 F0 range as well as vowel space are observed in dysarthric speech. Also, from the comparison of F0 range and vowel formant frequencies it is suggested that speech effort to produce wider F0 range can influence vowel quality as well.

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 F0 level and reduced F0 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 evaluating acoustical characteristics of dysarthrias by examining both prosodic and segmental features.

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 palsy (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 F0 range and F0 minimum are used for prosodic evaluation, and vowel formant frequencies (F1, F2) are used for segmental evaluation.

F0 range For each recorded sentence, its F0 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 F0 range was calculated in the logarithmic domain.
Table 1: Speech materials.

<table>
<thead>
<tr>
<th></th>
<th>CNT1</th>
<th>CNT3</th>
<th>CNT4</th>
<th>CNT5</th>
<th>CNT6</th>
<th>CNT7</th>
<th>PKN1</th>
<th>PKN2</th>
<th>PKN3</th>
<th>PKN4</th>
<th>PKN5</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>60s</td>
<td>50s</td>
<td>30s</td>
<td>60s</td>
<td>40s</td>
<td>60s</td>
<td>50s</td>
<td>40s</td>
<td>60s</td>
<td>50s</td>
<td>50s</td>
</tr>
<tr>
<td>passage</td>
<td>The North Wind and The Sun</td>
<td>Sakura</td>
<td>The North Wind and The Sun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mic</td>
<td>dynamic</td>
<td>electret</td>
<td>dynamic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKN6</td>
<td>60s</td>
<td>70s</td>
<td>50s</td>
<td>50s</td>
<td>50s</td>
<td>50s</td>
<td>40s</td>
<td>60s</td>
<td>60s</td>
<td>50s</td>
<td>70s</td>
</tr>
<tr>
<td>passage</td>
<td>The North Wind and The Sun</td>
<td>Sakura</td>
<td>The North Wind and The Sun</td>
<td>Sakura</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mic</td>
<td>dynamic</td>
<td>electret</td>
<td>dynamic</td>
<td>electret</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Scatter plot of $F_0$ minimum vs. $F_0$ range (intact control).

$F_0$ 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[13]. The method requires high-quality recording, so it could be applied only for CNT7, ALS4 and PBP5. Phoneme labeling was 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. To avoid erroneous values, the formant frequencies were checked and corrected by visual inspection. The validity of the correction was also confirmed with resynthesized speech. Then, the formant frequency contours were smoothed with 9-point median filter. Finally, outlying formants were discarded with a priori knowledge.

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 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, ALS4 and PBP5, in Figs. 5 to 7, respectively. By comparing Fig. 5 and 6, it is understood that the vowel space of ALS4 are narrower than that of CNT7. Overlapping of different vowels are also observed in the distribution of ALS4. However, most vowels are located in their expected region, so overall formant frequency characteristics of ALS4 is relatively normal.

However, as shown in the Fig. 7, the vowel space is much narrower in the PBP5 case. 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.

It is worth noting that the $F_0$ range for PBP5 is also narrowed, as shown in Fig. 4. The $F_0$ range for ALS4 is narrower than that of healthy speakers, however, it is wider than other dysarthric cases. This implies a possi-
bility that there exists a correlation between $F_0$ range and formant frequency in the degree of narrowing. It is a common understanding among speech therapists that speech effort to produce wider $F_0$ range can influence vowel quality as well. This hypothesis should be investigated in future by means of the intonation emphasis training[14].

3. 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. Also, from the comparison of $F_0$ range and vowel formant frequencies it was suggested that speech effort to produce wider $F_0$ range can influence vowel quality.

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.

4. References


