Evaluation of fundamental frequency ($F_0$) characteristics of speech in dysarthrias: A comparative study

Hiroki Mori$^{1,*}$, Yasunori Kobayashi$^1$, Hideki Kasuya$^1$, Noriko Kobayashi$^2$ and Hajime Hirose$^2$

$^1$Faculty of Engineering, Utsunomiya University, 7–1–2, Yoto, Utsunomiya, 321–8585 Japan
$^2$School of Allied Health Sciences, Kitasato University, 1–15–1, Kitasato, Sagamihara, 228–8555 Japan

(Received 11 July 2005, Accepted for publication 29 July 2005)

Keywords: Prosody, Dysarthria, Speech disorders, Speech therapy

PACS number: 43.70.Dn [DOI: 10.1250/ast.26.540]

1. Introduction

Dysarthria is regarded as a clinical symptom of neurological diseases including pseudobulbar palsy (PBP), Parkinson’s disease, spinocerebellar degeneration (SCD), amyotrophic lateral sclerosis (ALS), etc. Dysarthric speech is often characterized by an abnormality in its prosodic pattern such as monopitch or monoloudness as well as hoarse voice quality, associated with segmental abnormalities such as weak articulation or omission of segments.

There have been many reports on the acoustic characteristics of dysarthric speech [1–13]. However, their findings about the acoustic correlates to the speech disorders have been limited. Clinically, it is common belief that abnormalities in fundamental frequency ($F_0$) dynamics are associated with dysarthrias [14]. Despite of that, some studies analyzed sustained vowels of patients with dysarthrias and concluded that most voice source parameters including $F_0$ did not distinguish the patients from normal controls [7,8]. Canter [1] reported a higher $F_0$ level and reduced $F_0$ range in speech of patients with Parkinson’s disease, but there have been few comparative studies with regard to $F_0$ characteristics in running speech of dysarthric patients.

We have been developing several methods of acoustic analysis for dysarthric speech to elucidate its physiological nature as well as developing an aid for the diagnosis and training of dysarthrias. In the present research we focused on evaluating the acoustic characteristics of dysarthric speech by measuring the $F_0$ distribution. Based on prosodic labelling, our proposed method of acoustic analysis appeared to make good use of our limited number of speech samples. The method was successfully used to analyze a good number of recorded speech samples from different types of dysarthrias, and the results revealed that the prosodic characteristics corresponded well to the neurological nature of dysarthric patients.

2. Acoustic measurements

2.1. Method

2.1.1. Speech samples

The speech samples subjected to the acoustic analysis were obtained from 15 adult male dysarthric patients consisting of 6 cases of Parkinson’s disease (PKN), 4 cases of amyotrophic lateral sclerosis (ALS), and 5 cases of pseudobulbar pulsy (PBP). As a control, speech samples were also obtained from 6 normal male adults (CNT).

2.1.2. Recordings

The recordings were carried out in a soundproof room. Each subject read several paragraphs of the clinically used Japanese sentences “The north wind and the sun” or “Sakura (cherry blossom)” at his ordinary speed of speaking (Table 1). Each speech sample was digitally recorded at a sampling frequency of 48 kHz using a DAT recorder, and then low-pass filtered (cutoff 5,500 Hz) and downsampled to 12 kHz for further analysis.

2.1.3. Unit of analysis

The concept of the unit of analysis is an essential principle for speech analysis because the selection of the unit eventually affects the acoustic properties. Possible candidates for the unit of analysis include syntactic units such as sentence, and objective units delimited by, for example, pauses. In the present study, we partly adopted a prosodic labelling framework called JToBI (Japanese Tones and Break Indices) [16] to identify the unit of analysis. In the framework, utterances were segmented into phrases according to the prosodic hierarchy. The unit used in the analysis, i.e., Intonation Phrase (IP), is bounded by Break Index 3, which is the strongest intonational break and often corresponds to a syntactic boundary.

IPs, rather than sentences, were considered to be appropriate for acoustic analysis for the following reasons:

- Since IP is the prosodic domain within which the $F_0$ range is defined [16], it is a natural and minimum unit to define the $F_0$ range.
- When the amount of speech data of dysarthric patients is limited, IP is advantageous for statistical analyses because there are usually many more IPs than sentences.
- Applying JToBI labeling implies prior removal/correction of erroneous $F_0$ data, because the JToBI labels should not be marked for irrelevant acoustic events.

Figure 1 shows an example utterance of a healthy speaker used in the analysis. The utterance is composed of six IPs. The figure shows that BI 3 (IP boundary) ‘resets’ the $F_0$ range for the next phrase.

For extremely monotonous speech, it is hard to perceive...
Table 1 Speech materials. CNT, PKN, ALS and PBP stand for control group, Parkinson’s disease, amyotrophic lateral sclerosis and pseudobulbar pulsy, respectively. CNT2 and PKN5 are deleted IDs.

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>CNT1</th>
<th>CNT3*</th>
<th>CNT4*</th>
<th>CNT5*</th>
<th>CNT6*</th>
<th>CNT7*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>60</td>
<td>50s</td>
<td>30s</td>
<td>60s</td>
<td>40s</td>
<td>30s</td>
</tr>
<tr>
<td>Passage</td>
<td>The North Wind and The Sun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>PKN1</th>
<th>PKN2</th>
<th>PKN3</th>
<th>PKN4</th>
<th>PKN6</th>
<th>PKN7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>58</td>
<td>49</td>
<td>68</td>
<td>59</td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td>Passage</td>
<td>The North Wind and The Sun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>ALS1</th>
<th>ALS2</th>
<th>ALS3</th>
<th>ALS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>54</td>
<td>71</td>
<td>61</td>
<td>47</td>
</tr>
<tr>
<td>Passage</td>
<td>The North Wind and The Sun</td>
<td>Sakura</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker ID</th>
<th>PBP1</th>
<th>PBP2</th>
<th>PBP3</th>
<th>PBP4</th>
<th>PBP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>66</td>
<td>67</td>
<td>58</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>Passage</td>
<td>The North Wind and The Sun</td>
<td>Sakura</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(“from "Speech databases by Grant-in-Aid for Scientific Research" [15])

Fig. 1 An example utterance of a healthy speaker (“tabibitoga” was a mispronunciation of “tabibito-no.”).

pitch accents (e.g. Fig. 2). In such cases, therefore, the BI 3 was given where a strong temporal disjuncture such as a long pause was perceived.

Some utterances include disfluent portions, such as hesitation, repetition and self-correction. Figure 3 shows a disfluent portion, which was uttered by a PBP patient. The speaker hesitated when speaking “himarayanimo.” To clarify such disfluent portions, corrected segments (such as “hima” or “ni”) were marked with parentheses. In the succeeding analysis, parenthesized segments (and following pauses, if any) were ignored: the whole utterance was treated as if uttered fluently. Although this treatment of disfluency does affect temporal properties in the acoustic analysis such as speech rate, it was considered to be less sensitive to F0 statistics.

2.1.4. Parameters

F0 range For each recorded sentence, its F0 contour was extracted with the multiple window length method [17], then

errors were corrected manually.

For each IP, its F0 range was defined as the difference between the highest and lowest F0 values in the IP, both of which were calculated in the logarithmic domain.

F0 minimum For each IP, its F0 minimum was obtained as the lowest F0 value in the IP.

2.2. Results and discussion

The F0 range and the F0 minimum of IPs obtained from each group of subjects, CNT, PKN, ALS and PBP, were plotted in the panels in Fig. 4.

It is apparent that the F0 range of dysarthric speech is generally narrower than that of the normal subjects, suggesting that their intonation pattern should be flat. This tendency was most prominent in PKN. It is also apparent that the F0 minimum in PKN is higher than that of ALS or normal controls. A similar tendency is noted in some, but not all, cases of PBP. From these results, it can be concluded that the flat intonation pattern is a common feature among dysarthrias, while the pattern of F0 distribution reflects the difference in the type of dysarthric speech.

As for the physiological mechanisms underlying the above acoustic characteristics, it is considered that increased tension in the vocal folds due to rigidity resulted in higher F0 level in PKN, while the lowering in vocal fold tension due to muscle weakness led to lower F0 level in ALS. For PBP, the apparent bimodal distribution in the F0 level is most likely due to the different types of vocal manifestation in PBP, hypertensive and hypotensive, reported elsewhere [18]. However, this hypothesis should be verified by further
investigation with additional examples of dysarthric speech, which might include consecutive observations of individuals.

3. Conclusion

The result of acoustic analysis of dysarthric speech indicated that the $F_0$ range became narrower in different types of dysarthrias, while the pattern of $F_0$ distribution appeared to differ depending on the difference in the nature of underlying neurological disorders.

Acknowledgements

A part of this research was supported by Grants-in-Aid for Scientific Research of Priority Areas (12132201, 16091104).

References


