Speaker Variability Caused by Emotional States

Slobodan T. Jovičić, Miodrag M. Đorđević, Zoran Šarić

Faculty of Electrical Engineering
University of Belgrade, Serbia and Montenegro
jovicic@etf.bg.ac.yu

Abstract

The pitch and formants were analyzed in neutral and emphatic speech on the texts spoken by a male speaker. The results showed high difference in all parameters of both pitch distributions except in statistical mode. The scatter distributions of first four formants are similar and show high stability of first and second formant, and tendency of movement of third and fourth formant towards higher frequencies for emphatic speech.

1. Introduction

Speech and speaker recognition are highly dependent on stability of acoustical features detected in speech signal. The stability of these features is frequently influenced by speaker’s emotional states [1]. As a consequence, each acoustical feature has its own “variation field” with same perceptual meaning in identification the linguistic information, but different emotional states [2].

The purpose of investigation presented in this paper was to examine variations in pitch and formants in two forensically most often cases: neutral and emphatic speech. The neutral speech was considered as speech without emotions (with elements of submissiveness and depression) and the emphatic speech as speech of dominance, threat, and command.

2. Experiment

Two types of text were prepared and one male speaker pronounced them with neutral and emphatic voice. Both texts were recorded in computer in .wav format with sampling frequency of 11.025 Hz. The table 1 shows main characteristics of texts.

Table 1: Description of two types of text.

<table>
<thead>
<tr>
<th>TEXTS</th>
<th>Neutral speech</th>
<th>Emphatic speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of text</td>
<td>157 words</td>
<td>189 words</td>
</tr>
<tr>
<td>Vocal number</td>
<td>335</td>
<td>334</td>
</tr>
<tr>
<td>/a/ Distribution</td>
<td>68</td>
<td>71</td>
</tr>
<tr>
<td>/e/</td>
<td>69</td>
<td>62</td>
</tr>
<tr>
<td>/i/</td>
<td>69</td>
<td>69</td>
</tr>
<tr>
<td>/o/</td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>/u/</td>
<td>64</td>
<td>61</td>
</tr>
<tr>
<td>Length of records</td>
<td>78.5 seconds</td>
<td>86 seconds</td>
</tr>
</tbody>
</table>

Figure 1 shows the waveforms of both texts. There are evident differences in amplitude and segmental distribution.

Figure 1: The waveforms of both texts: upper is neutral and lower is emphatic speech.

3. Experimental results

In experiment the following parameters were measured: mean, standard deviation and histogram of pitch, and mean, standard deviation and distribution of first four formants.

3.1. Pitch measurements

Figure 2 and 3 show the histograms of neutral and emphatic speech with numerical values of mean pitch and standard deviation.

Figure 2: Histogram of neutral speech.

The histogram of emphatic speech is broader in comparison with histogram of neutral speech, and mean pitch is considerably higher. It is interesting that statistical mode is same in both histograms. Does it is specific case for these two speech materials and used speaker, and does the statistical mode even on the local segment of histogram can indicate the “natural” (without emotions) vocal cord frequency, remains for further examination?

Th.P3.11  
IV - 3311
3.2. Formant measurements

The formants were measured on the middle point of all vowels using the software Praat. The scatter diagrams for F1 and F2 formants show very similar distributions for both types of speech, figure 4 and 5.

![Figure 4: F1 – F2 scatter for neutral speech.]

![Figure 5: F1 – F2 scatter for emphatic speech.]

Statistical tests for mean of F1 and F2 (t- and F-test) don’t show statistical differences. However, the second formants for vowels /e/ and /i/ show tendency toward the higher frequencies. The standard deviations are different as evident from figure 4 and 5.

F3 – F4 scatter diagram shows formant shifts toward higher frequencies, except for neutral vowel /a/. These shifts are in accordance with results for emotional [3] and Lombard [4] speech.

![Figure 6: F3 – F4 scatter for mean formants.]

4. Conclusions

The results in this investigation show that forensic examination of speaker’s voice in different emotional state must be carefully performed, regarding to the results in this experiment. The first formant is most stable feature for speaker identification and possible role of “local” statistical mode was demonstrated. Further examinations of emotional speech are now in progress [4].

5. Acknowledgements

This work was partially supported by Ministry of Science, Technology and Development of Republic Serbia under Grant number OI-1784.

6. References