UNOBTRUSIVE MONITORING OF SPEECH IMPAIRMENTS OF PARKINSON’S DISEASE PATIENTS THROUGH MOBILE DEVICES

T. Arias-Vergara1, J.C. Vásquez-Correa1,2, J.R. Orozco-Arroyave1,2, P. Klumpp2, and E. Nöth2
1 Faculty of Engineering, Universidad de Antioquia UdA, Calle 70 No. 52-21, Medellín, Colombia
2 Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

Introduction

- Automatic classification of Parkinson’s Disease (PD) patients is performed considering speech recordings collected in non-controlled acoustic conditions during normal phone calls in an unobtrusive way.
- A speech enhancement algorithm is applied to improve the quality of the signals.
- Two different classification approaches are considered: the classification of PD patients and healthy speakers (HC) and a multi-class experiment to classify patients in different stages of the disease.

Methods

- Feature extraction: Phonation analysis is performed by extracting voiced segments from the utterance. The articulatory capability of the patients is evaluated with information from the onset/offset transitions. The prosody features are based on duration, the F0 and the energy contour.
- Articulatory impairments are also modeled with a deep learning approach based on Convolutional Neural Networks (CNNs). Chunks of signal are transformed into a time-frequency representation using the short-time Fourier transform (STFT) and used as input to a CNN (Figure 1).
- The automatic classification of PD patients and HC subjects is performed with a Support Vector Machine (SVM) with margin parameter C and a Gaussian kernel with parameter γ.

Data

- Train data: Speech utterances of 68 PD (Figure 2) patients and 50 HC subjects balanced in age and gender are considered. The participants pronounce a monologue according to their daily activities, with an average duration of 79.1 ± 43.8 seconds.
- Test data: The speech of 17 PD patients was recorded using the Apkinson mobile application. The participants were asked to make a phone call and sustain an spontaneous conversation. The average duration of the recordings is 62.9 ± 49.9 seconds.

Experiments and results

- PD vs HC classification: The highest improvement is obtained for prosody features. The results improve in up to 21% (absolute) when the recordings are processed with the SE algorithm (Table 1).
- Multi-class experiment: For the test set, all of the speakers from Class 0 are classified correctly, which indicates that the system is capable of identifying healthy speech (Table 2).

Conclusions

- The variations of the speech during a free conversation, which are intended to be assessed with the phone calls are suitable to assess the speech deficits of PD patients.
- Data collection using Apkinson is still ongoing, thus in the near future we expect to perform more experiments for further development of the mobile application.

Table 1: Results for classification of PD patients and HC subjects.

<table>
<thead>
<tr>
<th>Features</th>
<th>ACC</th>
<th>SEN</th>
<th>SPE</th>
<th>AUC</th>
<th>Speech enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonation</td>
<td>66</td>
<td>76</td>
<td>42</td>
<td>0.59</td>
<td>71</td>
</tr>
<tr>
<td>Articulation</td>
<td>58</td>
<td>68</td>
<td>20</td>
<td>0.61</td>
<td>71</td>
</tr>
<tr>
<td>Prosody</td>
<td>58</td>
<td>70</td>
<td>28</td>
<td>0.59</td>
<td>79</td>
</tr>
<tr>
<td>Fusion</td>
<td>66</td>
<td>88</td>
<td>14</td>
<td>0.66</td>
<td>62</td>
</tr>
<tr>
<td>CNN</td>
<td>61</td>
<td>82</td>
<td>0</td>
<td>0.53</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix obtained for the classification with a multi-class SVM. Feature space was reduced to only prosody features computed after applying speech enhancement.

Acknowledgments

This work was financed by CODI from University of Antioquia by the grant Number PRV16-2-01 and 2015-7683. This work also contributes to the research project DysarTrain which aims to provide an automatic therapy tool for patients suffering from dysarthric speech. T. Arias-Vergara and J.C. Vásquez-Correa acknowledge the Training Network on Automatic Processing of Pathological Speech (TAPAS) funded by the Horizon 2020 programme of the European Commission.