Towards an Automatic Monitoring of the Neurological State of Parkinson’s Patients from Speech


Shanghai, 22.03.2016
Outline

● Introduction
● Aims of the work
● Methodology
● The data
● Experiments and results
● Discussion and future work
Introduction
Parkinson’s Disease: Motivation and Prevalence

• The second most prevalent neurological disease worldwide

• Four million patients worldwide

• Speech impairments are one of the earliest manifestations

• If the impact of the disease is delayed in 20%, the total cost during the life of the patient could decrease in up to $USD 60K
MDS-UPDRS: Standard Neurological Scale for PD

• MDS-UPDRS: Movement Disorders Society – Unified Parkinson’s Disease Rating Scale

• Section III (MDS-UPDRS III) is only about motor activities

• Ranges from 0 to 132 (33 different items)

• Speech is evaluated only in one of those items!
Characteristics of Parkinson’s Speech

- Reduced loudness
- Monotonic speech
- Breathy voice

Dysarthric Speech

- Imprecise articulation
- Accelerated or slowed
- Stutter-like

Czech

German

Colombian
Aim of this Study

To assess the neurological state of Parkinson’s patients (according to the MDS-UPDRS-III score) from speech in three different languages:

- Spanish
- German
- Czech
Patients and speech tasks
Three databases with recordings in three languages: Spanish, German, and Czech

Isolated words, sentences, read texts, and monologues
SPANISH (PC-GITA)

- 50 patients and 50 healthy controls
- Age around 60 years
- Sound-proof booth
- Speech tasks: 21 words, 6 sentences, read text, and monologue
• 85 patients and 85 healthy controls
• Age around 64 years
• No sound-proof booth
• Speech tasks: 6 words, 5 sentences, read text, and monologue
CZECH

• 20 patients (diagnosed at the recording session) and 15 healthy controls
• Age around 60 years
• No sound-proof booth
• Speech tasks: 7 words, 3 sentences, read text, and monologue
Methodology
Voice recordings

Spanish
German
Czech
Voice recordings

Preprocessing

- Amplitude normalization
- Re-sampling down to 16kHz.
- Mean Cepstral Subtraction
Characterization: Onsets & Offsets

- 12 MFCC: Mel Frequency Cepstral Coefficients
- 25 BBE: Bark Band Energies

Transitions

... uv v v uv ...

onset offset

... 80 ms 80 ms ...

80 ms
Characterization: Speech Intelligibility
(Monologues are not Considered)

• Off-the-shelf speech recognizer: Google Inc.® (API v2)

• es-CO: Colombian; de-DE: German; cs: Czech

• Correctly pronounced words are counted and Word Accuracy (WA) is computed
Voice recordings

Preprocessing

Characterization

Regression

$\varepsilon$-SVR with linear kernel
Experiments and Results
Experiment 1: Cross Validations and LOSO

Validation strategies

- Spanish & German: 10-fold cross validation
- Czech: LOSO (leave-one-speaker-out)

Optimization of SVR meta-parameters ($C$ and $\varepsilon$)

Grid search: $C \in \{10^{-4}, 10^{-3}, ..., 10\}$ and $\varepsilon \in \{1, 10, 20, 30\}$

Evaluation of the system

Spearman’s ($\rho$) correlation between the predicted values and the MDS-UPDRS-III labels
### Spearman's Correlations:

**Energy of onset transitions**

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
<th>Sentences</th>
<th>Read text</th>
<th>Monologue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>0.44</td>
<td>0.49</td>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>German</td>
<td>0.36</td>
<td>0.28</td>
<td>0.41</td>
<td>0.55</td>
</tr>
<tr>
<td>Czech</td>
<td>0.29</td>
<td>0.25</td>
<td>0.21</td>
<td>0.23</td>
</tr>
</tbody>
</table>

**Energy of offset transitions**

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
<th>Sentences</th>
<th>Read text</th>
<th>Monologue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>0.46</td>
<td>0.46</td>
<td>0.53</td>
<td>0.74</td>
</tr>
<tr>
<td>German</td>
<td>0.37</td>
<td>0.24</td>
<td>0.36</td>
<td>0.31</td>
</tr>
<tr>
<td>Czech</td>
<td>0.24</td>
<td>0.18</td>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**Intelligibility: WA**

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
<th>Sentences</th>
<th>Read text</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>0.39</td>
<td>0.20</td>
<td>0.07</td>
<td>0.49</td>
</tr>
<tr>
<td>German</td>
<td>0.12</td>
<td>0.18</td>
<td>0.19</td>
<td>0.31</td>
</tr>
<tr>
<td>Czech</td>
<td>0.22</td>
<td>0.16</td>
<td>0.15</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Experiment 2 (only Spanish): Train with PC-GITA and Test with the INTERSPEECH Challenge 2015

<table>
<thead>
<tr>
<th>Train and Development sets</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC-GITA</td>
<td>INTERSPEECH Challenge 2015</td>
</tr>
<tr>
<td>Recorded in a sound proof booth</td>
<td>Eleven new patients</td>
</tr>
<tr>
<td>Mean MDS-UPDRS-III: 37</td>
<td>Mean MDS-UPDRS-III: 39</td>
</tr>
<tr>
<td>$\varepsilon$ and C are optimized on development</td>
<td>Different acoustic conditions</td>
</tr>
<tr>
<td>Only words, sentences, read texts, and monologues are considered</td>
<td>Only words, sentences, read texts, and monologues are considered</td>
</tr>
</tbody>
</table>

Evaluation of the systems: Spearman’s correlation.
Baseline: & Winners:
### Spearman’s Correlations: $\rho$

<table>
<thead>
<tr>
<th></th>
<th>Energy of onset transitions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Sentences</td>
<td>Read text</td>
<td>Monologue</td>
</tr>
<tr>
<td><strong>C; $\varepsilon$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train/Dev</td>
<td>0.28</td>
<td>0.44</td>
<td>0.48</td>
<td>0.42</td>
</tr>
<tr>
<td>Train/Test</td>
<td>0.33</td>
<td><strong>0.63</strong></td>
<td>0.23</td>
<td>0.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Energy of offset transitions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C; $\varepsilon$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train/Dev</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.62</td>
</tr>
<tr>
<td>Train/Test</td>
<td>0.22</td>
<td>0.52</td>
<td>-0.24</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Intelligibility: WA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C; $\varepsilon$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train/Dev</td>
<td>0.40</td>
<td>0.22</td>
<td>0.28</td>
<td>0.44</td>
</tr>
<tr>
<td>Train/Test</td>
<td>0.56</td>
<td><strong>0.31</strong></td>
<td>0.51</td>
<td><strong>0.69</strong></td>
</tr>
</tbody>
</table>
Experiment 3: Considering the Ranks of the Predictions

• The aim of the community is to develop technology to monitor the neurological state of the patients

• It is not imperative to predict the real value of the UPDRS score, but to state whether the patient's state is improved or not

• The same SVR with no further optimization is used

• Measurements: WA and Energy of onsets
Experiment 3 (cont.)

- The obtained results are ranked and correlated with the original UPDRS scores

\[ \rho_{WA} = 0.59 \quad \text{and} \quad \rho_{Onsets} = 0.69 \]

- When WA and Onsets are combined: \( \rho_{all} = 0.72 \)
Discussion and Future Work
• The energy content of the onset transitions is suitable to predict the neurological state of PD patients

• The results on Spanish recordings are better because those patients are in a more advanced stage of the disease
• An off-the-shelf Speech Recognizer can be used to develop an intelligibility-based test for telemonitoring PD patients

• The proposed approach is slightly better than the method proposed by the winners of the ComParE 2015
Future work

• Multi-modal analyses: speech, handwriting, and gait
• Longitudinal analyses
• Recordings before, during, and after surgery
Thank you for your attention!