Quantification of Speech Impairment in Parkinson’s disease

Taha Khan1, Jerker Westin1, Peter Funk2, Mark Dougherty2
1 Academy of Industry and Society, Computer Science, Dalarna University, Borlänge, Sweden, SE-78188;
2 School of Innovation, Design and Engineering, Mälardalen University, Vasteras, Sweden, SE-72123

Background
Vocal Impairment is an early indicator of Parkinson’s disease (PD) and 90% of patients suffer from speech and larynx anomalies. The impairment gets worse with disease progression. Muscular constrictions in articulatory and respiratory systems (figure 1a) cause ‘whispering voice’ and ‘short rushes’ in speech [1]. These symptoms are caused by effortful glottal closures at Trachea Bronchi which block the smooth flow of air pressure $P_a$ through the vocal tract. According to the source-filter model of speech production [2], a speech signal $s[n]$ is a convolution between an excitation signal $e[n]$ (originated by the air pressure $P_a$ expelled from lungs) and a filter signal $h[n]$ representing the vibration of vocal-folds to modulate $P_a$ (figure 1b and 1c). We hypothesize that the blockage of $P_a$ induces air turbulence at Trachea Bronchi which may result in increased energy in the excitation signal $e[n]$. The hindrances to air expulsions may result in reduced energy in the vocal-tract (filter) signal $s[n]$. 

Methods
A dataset of 855 speech recordings (consisting of 3 running speech tests, 4 sustained vowel phonation tests and 2 Laryngeal-diadochokinetik tests) having speech symptom ratings of 0, 1, 2 or 3 in UPDRS from 75 PD subjects and 20 healthy controls was processed. A novel dysphonia measure, Cepstral Separation Difference (CSD) between the speech source and filter signals is introduced. The speech signal is deconvoluted into low-time (vocal tract parameters) and high-time (excitation parameters) log-spectrums using speech cepstrum. Air-turbulence at trachea bronchi in impaired speech has been estimated using the residual log-spectrum which is computed by subtracting the magnitudes of high-time coefficients from the magnitudes of low-time coefficients. The mean absolute deviation (represented as $\delta_{\text{CSD}}$) among the magnitudes of residual coefficients showed increasing trend with the increase in symptom ratings (figure 2).

Figure 1. The Vocal System

![The Vocal System](image1)

(a) Anatomy of Speech [2]  
(b) Speech Production System [2]  
(c) Source-Filter Model of Speech

Objective
To score the degree of speech impairment in PD using signal processing for extracting established and novel dysphonic features and compare the obtained scores to Unified Parkinson’s Disease Rating Scale (UPDRS) speech ratings (in UPDRS part-III: Motor Examination of Speech).

Results
An overall accuracy of 82% is yielded for the ordered classification of dysphonic features between the 4 UPDRS speech categories. CSD in running speech tests was found to be the most correlated feature ($\rho = 0.76$) with the UPDRS target ratings. An ROC area of 88% supports the feasibility of obtained features modeled using KNN for tracking speech symptoms. The classification results obtained from the ordered KNN algorithm is shown in Table 1. Each matrix row represents the actual class instances while each matrix column represents the instances in a predicted class. The ordered classification resulted in high true positive rates for symptom classes 0 (75%), 1 (76%), 2 (82%) and 3 (100%).

Further spectral analysis of speech signals was applied to quantify increased number of pauses and reduced energy intensity in patients’ voice. A total of 35 vocal fundamental frequency features (e.g. pitch period estimations, Jitter, Shimmer and formants) were used in addition to CSD for speech quantification using the speech recordings. Spearman’s correlation coefficients between the features and the UPDRS speech ratings was calculated. Since the PD speech symptoms exhibit a monotonic order, an appropriate statistical approach is to classify the symptoms in an ordered fashion. The advantage of using an ordered classification approach is that, it transforms the non-monotonic properties of a classifier into monotonic coefficients between the features and the UPDRS speech ratings. The ordered classification method in conjunction with the K-Nearest Neighbor (KNN) classifier and 18-fold cross-validation was used to categorize each speech sample. Accuracy and Receiver Operating Characteristic (ROC) area were used for evaluating performance to estimate UPDRS speech ratings.

Table 1. Classification Matrix. The matrix presents system performance based on a comparison between the actual UPDRS-ratings of speech samples (y-axis) and the predicted UPDRS-ratings by the system (x-axis). The numbers in bold-italic depict a remarkable match between the system ratings and clinician ratings.

<table>
<thead>
<tr>
<th>Predicted Ratings by System</th>
<th>Normal ‘0’</th>
<th>Mild ‘1’</th>
<th>Moderate ‘2’</th>
<th>Severe ‘3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal ‘0’</td>
<td>18</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mild ‘1’</td>
<td>5</td>
<td>19</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Moderate ‘2’</td>
<td>2</td>
<td>3</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Severe ‘3’</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
</tbody>
</table>

References