Assessing Progress of Parkinson’s Disease Using Acoustic Analysis of Phonation

Jiri Mekyska*, Zoltan Galaz*, Zdenek Mzourek*, Zdenek Smekal*, Irena Rektorova†‡, Ilona Eliasova†‡, Milena Kostalova§, Martina Mrackova†‡, Dagmar Berankova‡, Marcos Faundez-Zanuy*, Karmele López-de-Ipiña∥ and Jesus B. Alonso-Hernandez∗∗

†Department of Telecommunications
Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic
Email: mekyska@feec.vutbr.cz
‡First Department of Neurology
St. Anne’s University Hospital, Pekarska 53, 65691 Brno, Czech Republic
§Applied Neuroscience Research Group
Central European Institute of Technology, Masaryk University, Komenskeho nam. 2, 60200 Brno, Czech Republic
∥Department of Systems Engineering and Automation
University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain
∗∗Institute for Technological Development and Innovation in Communications (IDeTIC)
University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain

Abstract—This paper deals with a complex acoustic analysis of phonation in patients with Parkinson’s disease (PD) with a special focus on estimation of disease progress that is described by 7 different clinical scales (e.g. Unified Parkinson’s disease rating scale or Beck depression inventory). The analysis is based on parametrization of 5 Czech vowels pronounced by 84 PD patients. Using classification and regression trees we estimated all clinical scores with maximal error lower or equal to 13 %. Best estimation was observed in the case of Mini-mental state examination (MAE = 0.77, estimation error 5.50 %). Finally, we proposed a binary classification based on random forests that is able to identify Parkinson’s disease with sensitivity SEN = 92.86 % (SPE = 85.71 %). The parametrization process was based on extraction of 107 speech features quantifying different clinical signs of hypokinetic dysarthria present in PD.

I. INTRODUCTION

Parkinson’s disease (PD) is the second most frequent neurodegenerative disease caused by a progressive loss of dopaminergic neurons (particularly in the substantia nigra pars compacta), however it can also result from anti-psychotic medications or frequent blows to head [1]. In 60–90 % of PD patients aged over 65 years it is possible to observe speech disorder called hypokinetic dysarthria (HD) [2]. HD manifests itself in areas of phonation, articulation, prosody, fluency and faciokinesis. Speech and voice disturbances are characterized by monotonous pitch and loudness, decreased stress and emphasis, breathy and harsh voice, reduced vocal intensity, variable rates including short rushes of speech or accelerated speech, consonant imprecision, impaired breath support for speech, reduction in phonation time, difficulty in the initiation of speech activities, and inappropriate pausing [3]–[6].

Clinical neurologists and psychologists use different scales to rate PD. Some of them were developed for complex assessment of PD (e.g. UPDRS – Unified Parkinson’s disease rating scale [7], NMSS – Non-motor symptoms scale [8]), and some of them are focused on specific clinical signs like depression (e.g. BDI – Beck depression inventory [9]), dyskinesia (e.g. AIMS – Abnormal involuntary movements scale [10]), gait (e.g. FOG – Freezing of gait questionnaire [11]), sleep disorders (e.g. RBDSQ – The REM sleep behavior disorder screening questionnaire [12]) or cognitive impairments (e.g. MMSE – Mini-mental state examination [13], ACE-R – Addenbrooke’s cognitive examination-revised [14]). Although significance of these scores have been robustly validated on large sets of patients, they have one disadvantage and that is the factor of subjectivity during examination. In all the cases we should count with possibility that two independent clinicians can rate the same patient with different scores.

Therefore scientists started to deal with new paraclinical methods of PD assessment that are objective, non-invasive, quick, low-cost and that significantly correlate with conventional rating scales. It is not considered that these methods will substitute clinicians during diagnosis and rating, they should rather bring new kind of biomarkers and generally parameters that will be used for objective and quick estimation of progress and that can be used for precise disease monitoring in time.

One of the paraclinical methods is the acoustic analysis of dysarthric speech. Just a limited number of works focused on
TABLE I: DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PD PATIENTS

| Speakers | PD (females) | PD (males) |
|----------|--------------|------------|
| Number   | 36           | 48         |
| Age (years) | 68.47 ± 7.64 | 66.21 ± 8.78 |
| PD duration (years) | 7.61 ± 4.85 | 7.83 ± 4.39 |
| UPDRS III | 22.06 ± 13.73 | 26.85 ± 10.22 |
| UPDRS IV | 2.72 ± 3.01 | 3.15 ± 2.59 |
| RBDSQ | 3.42 ± 3.48 | 3.85 ± 2.99 |
| FOG | 6.94 ± 5.72 | 6.67 ± 5.57 |
| NMSS | 36.03 ± 26.72 | 38.19 ± 19.72 |
| BDI | 18.57 ± 23.94 | 9.69 ± 6.23 |
| MMSE | 27.38 ± 3.63 | 28.56 ± 1.05 |
| ACE-R | 78.00 ± 8.62 | 88.08 ± 7.16 |
| LED (mg) | 862.44 ± 508.3 | 1087.00 ± 557.47 |

1 UPDRS III – Unified Parkinson’s disease rating scale, part III: Motor Examination; UPDRS IV – Unified Parkinson’s disease rating scale, part IV: Complications of Therapy; RBDSQ – The REM sleep behavior disorder screening questionnaire; FOG – Freezing of gait questionnaire; NMSS – Non-motor symptoms scale; BDI – Beck depression inventory; MMSE – Mini-mental state examination; ACE-R – Addenbrooke’s cognitive examination-revised; LED – L-dopa equivalent daily dose.

PD assessment Parkinson’s disease exists in this field of science. Moreover scientists usually deal only with UPDRS (mainly part III: motor examination; part V: Modified Hoehn and Yahr Staging) [15]–[21]. Score estimation of the other, and still important, rating scales has not been investigated yet. Therefore, in this work, we aim to: 1) identify vowels whose analysis provides best estimation of particular clinical scores used for assessment of PD; 2) introduce a new concept of Parkinson’s disease progress quantification based on acoustic analysis of phonation; 3) propose a binary classifier of PD.

The rest of this paper is organized as follows. Sections II and III describe the dataset and methodology respectively. Section IV provides some preliminary results where we firstly performed binary classification (parkinsonic people vs. healthy controls) and Spearman’s correlation between clinical scores and particular speech features. Next, we employed classification and regression trees in order to achieve best estimation in terms of low mean absolute correlation and high Pearson’s correlation coefficient. The conclusion is given in sec. V.

II. DATA

We included in this study 84 PD patients (36 women, 48 men) and 49 (24 women, 25 men) age and gender matched healthy controls (HC) who were enrolled at the First Department of Neurology, St. Anne’s University Hospital in Brno, Czech Republic. The healthy participants had no history or presence of brain diseases (including neurological and psychiatric illnesses) or speech disorders. For more demographic characteristics of the PD group, see Table I. All participants signed an informed consent form that had been approved by the Ethics Committee of St. Anne’s University Hospital in Brno.

Each of the participants was firstly examined by clinical neurologist and psychologist who rated him according to 8 scales: Unified Parkinson’s disease rating scale (part III and IV), The REM sleep behavior disorder screening questionnaire, Freezing of gait questionnaire, Non-motor symptoms scale, Beck depression inventory, Mini-mental state examination and Addenbrooke’s cognitive examination-revised. In addition we processed duration of disease and LED – L-dopa equivalent daily dose.

After the clinical examination the participants uttered 4 sets of five Czech vowels ([a], [e], [i], [o], [u]): 1) s – short vowels pronounced with normal intensity; 2) l – sustained vowels pronounced with normal intensity; 3) ll – sustained vowels pronounced with maximum intensity; 4) ls – sustained vowels pronounced with minimum intensity, but not whispered.

III. METHODOLOGY

Speech samples were digitized by sampling frequency $f_s = 48$ kHz and consequently resampled to 16 kHz. The recordings were parametrized by NDAT (Neurological Disorder Analysis Tool) [6], [22], developed at the Brno University of Technology, and statistically processed in MATLAB.

A. Parametrization

The feature extraction process included calculation of wide range of speech features that we divided into several groups (description of all below mentioned features can be found in our recent article [23]):

1) Features describing phonation – $F_0$, 5 kinds of jitter and 6 kinds of shimmer [24], PPE (Pitch Period Entropy) [25], a measure of standard deviation (std) of the time that vocal folds are apart (GQ$_{open}$) and in collisions respectively (GQ$_{closed}$) [19], E (short-time energy), TKEO (Teager-Kaiser Energy Operator) [26], ME (4 Hz modulation energy) [27], MPSD (Median of Power Spectral Density) [28] and LSTER (Low Short-Time Energy Ratio) [29].

2) Features describing tongue movement – formants $F_1$–$F_3$ and their bandwidths $BW_1$–$BW_3$, VSA (vowel space area) [30] and InVSA (its logarithmic version) [30], FCR (formant centralization ratio) [30], VAI (vowel articulation index) [1] and $F_{2i}/F_{2u}$ (ratio of second forms of vowels [i] and [u]).

3) Features describing speech quality – ZCR (Zero-Crossing Rate), HZCRR (High Zero-Crossing Rate Ratio) [29], FLUF (Fraction of Locally Unvoiced Frames) [31], SF (Spectral Flux) [32], SDBM (Spectral Distance Based on Module), SDBP (Spectral Distance Based on Phase) [31], CPP (Cepstral Peak Prominence) [33], PECM (Pitch Energy Cepstral Measure) [31], VR (Variation in Ratio between the second/first harmonic within the derived cepstral domain) [31], HNR (Harmonic-to-Noise Ratio), NHR (Noise-to-Harmonic Ratio), NNE (Normalized Noise Energy) [34], GNE (Glottal-to-Noise Excitation ratio) [35], SPI (Soft Phonation Index) [36], VTI (Voice Turbulence Index) [36], SSD (Segmental Signal-to-Dysperiodicity ratio), MSER (Modulation Spectra Energy Ratio), MFP (Modulation Frequency
of Peak) [37], RPHM (Relative Peak Height of Modulation spectra) [23], ICER (Inferior Colliculus Energy Ratio) [23] and RPHIC (Relative Peak Height of Inferior Colliculus) [23].

4) Features based on bispectrum [31] and bicepstrum [23] – BII (Bicoherence Index Interference), HFEB (High Frequency Energy of one-dimensional Bicoherence), LFEB (low Frequency Energy of one-dimensional Bicoherence). BMII (Bisppectrum Module Interference Index), BPII (Bisppectrum Phase Interference Index), BCII (Bicoherence Index Interference), HFEBC (High Frequency Energy of one-dimensional Bicoepstral index), LFECB (Low Frequency Energy of one-dimensional Bicoepstral index), CMII (Bicoepstrum Module Interference Index), BCPII (Bicoepstrum Phase Interference Index), LCBCER (Low Cepstra/Bicoepstra Energy Ratio), HCB CER (High Cepstra/Bicoepstra Energy Ratio), LSBER (Low Spectra/Bispectra Energy Ratio), HSBER (High Spectra/Bispectra Energy Ratio), BCMD (Bicoepstral Module Distance) and BCPD (Bicoepstral Phase Distance).

5) Features based on empirical mode decomposition [19]. [23] – IMF-SNR_{TKEO} (SNR based on Teager-Kaiser Energy Operator extracted from intrinsic mode functions), IMF-SNR_{SEO} (based on Squared Energy Operator), IMF-SNR_{SE} (based on Shannon Entropy), IMF-SNR_{RE} (based on second-order Rényi Entropy), IMF-SNR_{ZCR} (based on Zero-Crossing Rate), IMF-NSR_{TKEO}, IMF-NSR_{SEO}, IMF-NSR_{SE}, IMF-NSR_{RE}, IMF-FD (Fractal Dimension extracted from the 1st IMF), IMF-CPP (Cepstral Peak Prominence extracted from the 1st IMF) and IMF-GNE (Glottal-to-Noise Excitation ratio based on the 1st IMF).

6) Non-linear dynamic features – CD (Correlation Dimension) [38], FD (Fractal Dimension) [38], ZL (Ziv-Lempel complexity) [39], HE (Hurst Exponent) [40], SHE (Shannon Entropy), RE (second-order Rényi Entropy), CE (Correlation Entropy) [41], RBE1 (first-order Rényi Block Entropy) [42], RBE2 (second-order Rényi Block Entropy) [42], AE (Approximate Entropy) and SE (Sample Entropy) [43] with 8 different kernels [23], PE (Permutation Entropy), FMMI (First Minimum of Mutual Information function) [42] and LLE (Largest Lyapunov Exponent) [40].

If the feature has been represented by vector, we have employed transformation to scalar value based on median, standard deviation (std), 1st percentile (1p), 99th percentile (99p) and interpercentile range (ir) defined as 99p – 1p. Considering all the possible combinations, we have extracted approximately 350 features for each vowel.

B. Preliminary analysis

To accomplish the 3rd goal of this work (binary classification) we firstly employed RF (random forests) classifier along with sequential forward feature selection (SFFS). Precision of PD identification was tested in terms of classification accuracy (ACC), sensitivity (SEN), specificity (SPE) and trade-off between sensitivity and specificity (TSS) defined as:

\[ TSS = 2 \sin \left( \frac{\pi \cdot \text{SEN}}{2} \right) \sin \left( \frac{\pi \cdot \text{SPE}}{2} \right). \]

We considered two scenarios: 1) individual vowel analysis; 2) classification within each vowel set (see sec.II). In both cases we used leave-one-out validation.

Next, we performed Spearman’s rank correlation between particular feature vector and clinical information (PD duration, UPDRS III and IV, RBDSQ, FOG, NMSS, BDI, MMSE, ACE-R and LED) in order to find a possible candidate for preliminary PD assessment.

| Vowels | ACC [%] | SEN [%] | SPE [%] | TSS | No. |
|--------|---------|---------|---------|-----|-----|
| a (s)  | 75.19   | 79.76   | 86.63   | 1.7948 | 2   |
| e (s)  | 79.70   | 84.52   | 81.63   | 1.9059 | 9   |
| o (s)  | 87.22   | 89.29   | 83.67   | 1.9367 | 11  |
| u (s)  | 79.70   | 79.76   | 79.59   | 1.9680 | 5   |
| a (l)  | 75.68   | 71.43   | 77.55   | 1.7999 | 6   |
| e (l)  | 88.72   | 91.67   | 83.67   | 1.9440 | 8   |
| i (l)  | 81.95   | 83.33   | 79.59   | 1.8878 | 7   |
| o (l)  | 72.93   | 72.62   | 73.47   | 1.7791 | 2   |
| u (l)  | 76.69   | 78.57   | 73.47   | 1.8189 | 7   |
| a (I)  | 75.94   | 77.38   | 73.47   | 1.8116 | 7   |
| e (I)  | 83.46   | 86.90   | 77.55   | 1.8904 | 10  |
| i (I)  | 74.44   | 73.81   | 75.51   | 1.8020 | 3   |
| o (I)  | 81.95   | 83.33   | 79.59   | 1.8878 | 9   |
| u (I)  | 75.19   | 78.57   | 69.39   | 1.7861 | 3   |
| a (L)  | 73.68   | 77.38   | 67.35   | 1.7616 | 2   |
| e (L)  | 78.20   | 77.38   | 79.59   | 1.8529 | 5   |
| i (L)  | 81.20   | 83.33   | 77.55   | 1.8745 | 5   |
| o (L)  | 84.96   | 83.33   | 87.76   | 1.9293 | 3   |
| u (L)  | 81.95   | 86.90   | 73.47   | 1.8598 | 7   |
| a (s)  | 73.68   | 77.38   | 67.35   | 1.7616 | 2   |
| e (s)  | 78.20   | 77.38   | 79.59   | 1.8529 | 5   |
| i (s)  | 81.20   | 83.33   | 77.55   | 1.8745 | 5   |
| o (s)  | 84.96   | 83.33   | 87.76   | 1.9293 | 3   |
| u (s)  | 81.95   | 86.90   | 73.47   | 1.8598 | 7   |
| a (l)  | 84.21   | 82.14   | 87.76   | 1.9228 | 9   |
| e (l)  | 83.46   | 83.33   | 83.67   | 1.9110 | 3   |
| a (I)  | 90.23   | 92.86   | 85.71   | 1.9572 | 11  |

1 ACC – classification accuracy; SEN – sensitivity; SPE – specificity; TSS – trade-off between sensitivity and specificity; No. – number of selected features; s – short vowel pronounced with normal intensity; l – sustained vowel pronounced with minimum intensity (not whispering).

C. Classification and regression

After the preliminary analysis we moved further and trained classification and regression trees (CART) to estimate selected clinical information with higher accuracy measured by mean absolute error (MAE) and Pearson’s correlation coefficient \( \rho \). We used a two-step feature selection. Firstly we preselected 500 features using filtering method based on minimum redundancy and maximum relevance (mRMR). Then we preselected 500 features using filtering method based on minimum redundancy and maximum relevance (mRMR). Then we employed wrapping method based on SFFS to select the final feature subset. We considered the same two scenarios as in the previous analysis: 1) individual vowel analysis; 2) classification within each vowel set. We used the leave-one-out validation in both cases.
Finally we selected for each clinical information the best combination of vowels and speech features and calculated two kinds of estimation error to better describe ability of PD progress assessment. We measured these two estimation errors:

\[ \text{EE1} = \frac{\text{MAE}}{\text{range(CI)}}, \]  
\[ \text{EE2} = \frac{\text{MAE}}{\text{max(CI)}}, \]

where CI stands for particular clinical information (e.g. FOG). Function \( \text{range(CI)} \) calculates the range from clinical data available during the analysis, while function \( \text{max(CI)} \) returns the maximal score that can be theoretically reached in the specific scale.

### IV. EXPERIMENTAL RESULTS

Results from the binary classification can be found in Table III. In the first scenario (individual vowel analysis) the best results were observed in the case of classification based...
TABLE IV

STIMATION OF CLINICAL SCORES

| Vowels | MAE | No. | UPDRS III | ρ | No. | UPDRS IV | ρ | No. | RBDSQ | ρ | No. | FOG | ρ | No. |
|---------|-----|-----|-----------|---|-----|----------|---|-----|-------|---|-----|-----|---|-----|
| a (s)   | 3.21 | 0.4253 | 4 | 8.28 | 0.3956 | 7 | 1.42 | 0.6690 | 7 | 1.88 | 0.6002 | 7 | 3.32 | 0.5903 | 7 |
| e (s)   | 2.25 | 0.6891 | 19 | 6.61 | 0.7084 | 6 | 1.47 | 0.6600 | 15 | 2.09 | 0.6117 | 2 | 3.86 | 0.4692 | 7 |
| i (s)   | 2.89 | 0.5511 | 8 | 8.33 | 0.3934 | 14 | 1.59 | 0.6595 | 7 | 2.09 | 0.4493 | 3 | 3.99 | 0.4748 | 6 |
| o (s)   | 3.28 | 0.4860 | 2 | 7.59 | 0.5444 | 8 | 1.90 | 0.3706 | 3 | 1.83 | 0.6436 | 10 | 3.55 | 0.5479 | 8 |
| u (s)   | 3.13 | 0.4059 | 7 | 8.18 | 0.5121 | 4 | 1.69 | 0.5835 | 6 | 2.14 | 0.4644 | 6 | 4.42 | 0.3792 | 2 |
| a (l)   | 3.31 | 0.3952 | 1 | 8.02 | 0.5153 | 5 | 1.60 | 0.5926 | 7 | 2.23 | 0.4138 | 5 | 4.03 | 0.4104 | 5 |
| e (l)   | 2.40 | 0.6856 | 9 | 6.34 | 0.7210 | 11 | 1.87 | 0.5092 | 6 | 1.83 | 0.6034 | 8 | 3.40 | 0.6190 | 5 |
| i (l)   | 2.93 | 0.6077 | 3 | 8.82 | 0.4570 | 5 | 2.10 | 0.4181 | 4 | 1.83 | 0.5679 | 13 | 3.43 | 0.6131 | 3 |
| o (l)   | 2.38 | 0.6802 | 10 | 7.20 | 0.5874 | 5 | 2.01 | 0.3882 | 2 | 2.23 | 0.3710 | 2 | 3.64 | 0.4920 | 5 |
| u (l)   | 3.00 | 0.5580 | 3 | 8.44 | 0.5367 | 3 | 1.62 | 0.5958 | 6 | 1.54 | 0.7358 | 8 | 3.90 | 0.4861 | 5 |
| a (ll)  | 2.75 | 0.6185 | 7 | 6.59 | 0.6540 | 7 | 2.00 | 0.4339 | 5 | 2.47 | 0.2644 | 1 | 4.22 | 0.5371 | 3 |
| e (ll)  | 2.05 | 0.5214 | 7 | 7.32 | 0.6340 | 7 | 1.76 | 0.5336 | 5 | 1.79 | 0.5391 | 9 | 4.29 | 0.3551 | 1 |
| i (ll)  | 3.23 | 0.4638 | 5 | 5.70 | 0.7987 | 15 | 1.94 | 0.4579 | 4 | 1.86 | 0.6190 | 6 | 2.67 | 0.7409 | 19 |
| o (ll)  | 2.60 | 0.6527 | 9 | 6.72 | 0.6696 | 15 | 1.85 | 0.4738 | 4 | 1.92 | 0.5799 | 3 | 3.21 | 0.6097 | 11 |
| u (ll)  | 3.18 | 0.4933 | 7 | 8.03 | 0.5401 | 5 | 1.75 | 0.5824 | 7 | 1.85 | 0.6242 | 5 | 3.83 | 0.5074 | 2 |
| a (ls)  | 2.54 | 0.6806 | 11 | 7.52 | 0.6078 | 6 | 1.45 | 0.7059 | 10 | 2.35 | 0.3313 | 2 | 3.28 | 0.5803 | 4 |
| e (ls)  | 3.53 | 0.1794 | 9 | 9.41 | 0.3471 | 2 | 1.90 | 0.5051 | 4 | 2.10 | 0.5335 | 3 | 3.73 | 0.5206 | 7 |
| i (ls)  | 2.78 | 0.5168 | 9 | 8.84 | 0.4878 | 2 | 1.42 | 0.6689 | 12 | 2.19 | 0.4407 | 2 | 3.04 | 0.6527 | 7 |
| o (ls)  | 2.69 | 0.5940 | 11 | 7.50 | 0.6224 | 9 | 1.48 | 0.6105 | 6 | 1.70 | 0.5441 | 14 | 4.06 | 0.4493 | 1 |
| u (ls)  | 3.56 | 0.3257 | 2 | 7.27 | 0.6827 | 13 | 1.64 | 0.5695 | 7 | 1.70 | 0.6473 | 11 | 4.40 | 0.3483 | 3 |
| a (ls)  | 16.25 | 0.6472 | 9 | 4.58 | 0.9301 | 8 | 1.21 | 0.4788 | 1 | 4.35 | 0.6023 | 9 | 2.24 | 0.8232 | 11 |
| e (ls)  | 2.27 | 0.6850 | 12 | 8.07 | 0.5449 | 5 | 1.30 | 0.6768 | 12 | 1.72 | 0.6912 | 7 | 3.75 | 0.4187 | 7 |

1 UPDRS III – Unified Parkinson’s disease rating scale, part III: Motor Examination; UPDRS IV – Unified Parkinson’s disease rating scale, part IV: Complications of Therapy; RBDSQ – The REM sleep behavior disorder screening questionnaire; FOG – Freezing of gait questionnaire; NMSS – Non-motor symptoms scale; BDI – Beck depression inventory; MMSE – Mini-mental state examination; ACE-R – Addenbrooke’s cognitive examination-revised; LED – L-dopa equivalent daily dose; MAE – mean absolute error; ρ – Pearson’s correlation coefficient; No. – number of selected features; s–short vowel pronounced with normal intensity; ls – shortened vowel pronounced with minimum intensity (not whispering)
on parametrization of sustained vowel [e] (ACC = 88.72 %, SEN = 91.67 %, SPE = 83.67 %, TSS = 1.9440 %). 8 features were selected by SFFS. Although most of the researchers use sustained vowel [a] in order to diagnose PD from speech, no clear explanation for this selection has been published yet. One possible reason is that tongue goes to maximal vertical position during pronunciation of this vowel (see Hellwag triangle [44]), but on the other hand there is nearly no movement in horizontal one. Other publications also show that significance of vowel [a] in PD patients should not be dogma [15]–[17]. Therefore it is still necessary to come up with a robust and complex work that will clarify the significance of particular vowels based on testing on multilingual PD databases (theoretically the effect of culture and language on phonation is minimal, but this must be proved in the case of dysarthric speech as well).

In the second scenario we considered classification within each vowel set. The results show that the best discrimination power provide features extracted from sustained vowels pronounced with minimum intensity (ACC = 90.23 %, SEN = 92.86 %, SPE = 85.71 %, TSS = 1.9572 %, number of selected features: 11). This result hits another dogma and that is the analysis of sustained vowels pronounced with normal intensity. In our recent paper we have proved that sustained vowels pronounced with minimum intensity (not whispering) accent vocal tremor and they are more complicated for precise vocal fold vibration (more than in the case of normal sustained vowels where a speaker does not have to concentrate too much on precise voicing – he couldn’t whisper) [48].

Next, we tried to find a possible candidate for preliminary PD assessment using Spearman’s rank correlation between particular feature vector and selected clinical information. We have visualized the results using correlation graphs that can be seen on Fig. 1. Each graph contains a non-linear regression line defined by second-order polynomial. Although all correlation are significant (p < 0.01), the correlation coefficients in absolute values are not so high and it is clear that it is necessary to include more features in order to reach better clinical score estimation.

Therefore, in the next step, we employed CART along with SFFS. The obtained results are given in Table IV. Finally, to better evaluate accuracy of estimation we selected for each clinical information the best MAE and additionally computed EE1 and EE2 (see Table III). Regarding EE1 the lowest estimation error was observed in the case of Mini-mental state examination (possible range: 0 – 30, MAE = 0.77 and resulting EE1 = 5.50 %) and the highest one in the case of UPDRS IV (possible range: 0 – 23, MAE = 1.30, EE1 = 13.00 %). This type of error is probably more relevant, because it is quantified using the range of data that were really processed. However, we also provide estimation accuracy in terms of EE2 that is related to the maximal value that can be in specific scale reached. From this point of view the lowest estimation error was observed in the case of MMSE too (EE2 = 2.57 %), but the highest one was measured in the case of RBDSQ (possible range: 0 – 13, MAE = 1.54, EE1 = 11.85 %).

V. CONCLUSION

In this paper we performed a complex acoustic analysis of phonation in patients with Parkinson’s disease in order to estimate a degree of this disease measured by 7 different clinical scales. The analysis is base on parametrization of 5 Czech vowels pronounced by 84 PD patients. We achieved all goals of this work: 1) We identified vowels that are suitable for estimation of particular clinical scores (see Table III). 2) We proposed a concept of PD progress assessment that is based on CART and features coming from different domains (description of phonation, tongue movement, speech quality, bispectrum, bicepstrum, empirical mode decomposition, etc.). We evaluated the proposed concept using estimation of 10 different clinical scores (PD duration, UPDRS III and IV, RBDSQ, FOG, NMSS, BDI, MMSE ACE-R and LED). All scores were estimated with maximal estimation error lower or equal to 13 %. Best estimation was observed in the case of Mini-mental state examination (EE1 = 5.50 %). 3) We proposed a binary classification based on random forests that is able to identify Parkinson’s disease with sensitivity SEN = 92.86 % (SPE = 85.71 %). The classifier is fed by features extracted from sustained vowels pronounced with minimum intensity.

All the considered scores are given by clinicians who examine the patients subjectively. Therefore it would be interesting to calculate difference of estimation among approximately 10 clinical neurologists/psychologists and compare it to estimation error that was measured in this work. Other interesting idea for future work would be to deeper investigate relations between speech and depression or sleep disorders. It is well documented that speech disorders in PD patients significantly correlate with changes in gait [49]. However, according to estimation errors of BDI or RBDSQ (which are lower than in the case of FOG), we can suppose that some hidden relations between depression, sleep disorders and speech are far more stronger.

ACKNOWLEDGMENT

Research described in this paper was financed by the National Sustainability Program under grant LO1401 and by projects NT13499 (Speech, its impairment and cognitive performance in Parkinson’s disease), COST IC1206, project “CEITEC, Central European Institute of Technology”: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund, FEDER and Ministerio de Economía y Competitividad TEC2012-38630-C04-03 (Kingdom of Spain). For the research, infrastructure of the SIX Center was used.

REFERENCES

[1] S. Skodda, W. Visser, and U. Schlegel, “Short- and long-term dopaminergic effects on dysarthria in early Parkinson’s disease,” J Neural Transm, vol. 117, pp. 197–205, 2010.
[2] S. Sapir, L. Ramig, and C. Fox, “Speech and swallowing disorders in Parkinson disease,” Curr Opin Otolaryngol Head Neck Surg, vol. 16, no. 3, pp. 205–210, 2008.
[3] A. M. Goberman, “Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease,” Med Sci Monit, vol. 11, no. 3, pp. CR109–116, 2005.
[46] J. R. O. Arroyave, J. D. Arias-Londono, J. F. V. Bonilla, and E. Noth, “Analysis of speech from people with Parkinson’s disease through nonlinear dynamics,” in Advances in Nonlinear Speech Processing, ser. Lecture Notes in Computer Science, T. Drugman and T. Dutoit, Eds. Springer Berlin Heidelberg, 2013, vol. 7911, pp. 112–119.

[47] J. Orozco-Arroyave, J. Arias-Londono, J. Vargas-Bonilla, and E. Noth, “Perceptual analysis of speech signals from people with Parkinson’s disease,” in Natural and Artificial Models in Computation and Biology, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7930, pp. 201–211.

[48] Z. Smekal, J. Mekyska, Z. Galaz, Z. Mzourek, I. Rektorova, and M. Faundez-Zanuy, “Analysis of phonation in patients with Parkinson’s disease using empirical mode decomposition,” in 2015 International Symposium on Signals, Circuits and Systems (ISSCS), 2015, pp. 1–4, (in print).

[49] C. Moreau, C. Ozsancak, J.-L. Blatt, P. Derambure, A. Destee, and L. Defebvre, “Oral festination in Parkinson’s disease: Biomechanical analysis and correlation with festination and freezing of gait,” Mov Disord, vol. 22, no. 10, pp. 1503–1506, 2007.