# Research on Passive Assessment of Parkinson's Disease Utilising Speech Biomarkers

Daniel Kovac[1(✉)], Jiri Mekyska[1], Lubos Brabenec[2], Milena Kostalova[2,3], and Irena Rektorova[2,4]

[1] Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
xkovac41@vut.cz
[2] Applied Neuroscience Research Group, Central European Institute of Technology – CEITEC, Masaryk University, Brno, Czech Republic
[3] Department of Neurology, Faculty Hospital and Masaryk University, Brno, Czech Republic
[4] First Department of Neurology, Faculty of Medicine and St. Anne's University Hospital, Masaryk University, Brno, Czech Republic

**Abstract.** Speech disorders, collectively referred to as hypokinetic dysarthria (HD), are early biomarkers of Parkinson's disease (PD). To assess all dimensions of HD, patients could perform several speech tasks using a smartphone outside a clinic. This paper aims to adapt the parametrization process to running speech so that a patient is not required to interact actively with the device, and features can be extracted directly from phone calls. The method utilizes a voice activity detector followed by a voicing detection. The algorithm was tested on a database of 126 recordings (86 patients with PD and 40 healthy controls) of monologue mixed with noise with different signal-to-noise ratios (SNR) to simulate the real environment conditions. Pearson correlation coefficients show a strong linear relationship between speech features and patients' scores assessing HD and other motor/non-motor symptoms – p-value $< 0.01$ for the normalized amplitude quotient (NAQ) with Test 3F Dysarthric Profile (DX index) and Unified Parkinson's Disease Rating Scale (part III) in 20 dB SNR conditions, p-value $< 0.01$ for the jitter and shimmer with the Mini Mental State Exam (10 dB SNR). A model based on the Extreme Gradient Boosting algorithm predicts the DX index with a 10.83% estimated error rate (EER) and the Addenbrooke's Cognitive Examination-Revise (ACE-R) score with 13.38% EER. The introduced algorithm can potentially be used in mHealth applications for passive monitoring and assessment of PD patients.

This study was supported by the Quality Internal Grants of BUT (project KInG, reg. no. CZ.02.2.69/0.0/0.0/19_073/0016948; financed from OP VVV), by EU - Next Generation EU (project no. LX22NPO5107 (MEYS)), and by the Czech Ministry of Health (grant no. NU22J-04-00074).


A. Tsanas and A. Triantafyllidis (Eds.): PH 2022, LNICST 488, pp. 259–273, 2023.
https://doi.org/10.1007/978-3-031-34586-9_18

**Keywords:** Hypokinetic dysarthria · Parkinson's disease · Passive assessment · Running speech

## 1 Introduction

Even though Parkinson's disease (PD), the chronic neurodegenerative disorder [17], was described more than two hundred years ago [22], we still do not know its exact causes. Besides genetic predisposition and age, pesticide exposure, high caloric intake, or head injuries may increase the risk of its development [6]. In addition, we are still unable to cure it, but its alleviation is possible by using medication (e.g. levodopa) [7] or other, more invasive ways (such as deep brain stimulation [3]). Moreover, there are signs that the incidence is increasing over time [26]. Thus, early diagnosis has a crucial impact on the future course of the disease, and it is essential to have the means to diagnose and monitor it at its earliest possible stage, as it can improve the patient's life.

The first symptoms of PD include speech and voice disorders [23], referred to as hypokinetic dysarthria (HD) [11], manifested by deteriorated respiration, phonation, articulation, and prosody [24]. PD patients may have impaired speech in all or only some domains [25], but at least one speech domain is affected in up to 90% of cases [14]. Tremor [30], hoarseness [5], audible breath [28], hypernasality [15], speech disfluency [18], or inappropriate silences [13] are common symptoms of HD. The speech is also often quiet [1] and unintelligible [29] and can be followed by monopitch and monoloudness [8].

Ergo, acoustic speech and voice analysis is considered an objective, supportive but practical tool for PD detection and monitoring. The analysis is done by recording the patient's speech signal with its consequent digital processing. After obtaining acoustic features that quantify individual speech disorders, it is possible to compare them with those of healthy controls (HC) using statistics to assess the severity of HD. In order to examine all domains of HD, patients are asked to perform several speech tasks, which are most commonly a sustained phonation of the vowel [a], monologue, reading text, picture description and diadochokinetic task (DDK) consisting of repeating the syllables [pa]-[ta]-[ka].

Due to the clinical time pressure, number of patients, and the inability of some patients to travel for speech recording, telemedicine plays an important role here. Smartphone applications could allow neurologists to monitor speech impairment and PD progress remotely. Patients can then perform speech tasks from home, or it could even be possible to process the speech recorded during a phone call, which would have numerous advantages.

### 1.1 State of the Art

Zhan et al. (2018) sought to answer whether a smartphone can be used to quantify the severity of motor symptoms of PD. To do this, they used the HopkinsPD mobile app, on which patients performed a sustained phonation of the vowel [a], from which the signal amplitude was measured. In addition to this task, patients performed four others (finger tapping, walking, balance test and reaction time test). The results correlated with the current standard rating scales [32].

Horin et al. (2019) investigated the usability of smartphone apps to treat gait, speech, and dexterity in people with PD. Regarding the HD, they measured the maximum phonation time and the mean fundamental frequency from a sustained vowel [a] task and the maximum reading time and standard deviation of the number of semitones from the fundamental frequency from a text reading task. The tasks are part of the Beats Medical Parkinsons Treatment App mobile app. Statistical tests showed no significant correlation between these features and patients' UPDRS (Unified Parkinson's Disease Rating Scale) scores or the time over which they performed the tasks [16].

Orozco-Arroyave et al. (2020) presented the Apkinson mobile app that assesses and monitors the motor skills of PD patients in terms of speech articulation, gait regularity and rigidity, and finger tapping accuracy. Speech features were extracted from the following tasks: sustained phonation of the vowels [a], [i] and [u] (jitter), DDK (articulation rate as the number of voiced segments per time and the probability which phonemes belong to each phonemic group), text reading (error between the word read and the word recognized by the automatic speech recognition model) and picture description (standard deviation of the fundamental frequency). According to the Kruskal-Walis test, the jitter and articulation rate features showed significant differences between HC and PD subjects [21].

Arora et al. (2021) analyzed 4242 smartphone recordings of a sustained phonation of the vowel [a] collected in a clinic and at home from 92 HC, 112 patients with rapid eye movement sleep behavior disorder (iRBD), and 335 patients with PD. They used acoustic signal analysis (337 phonatory features) and machine learning. Using the leave-one-subject-out cross-validation method, they could distinguish PD patients from HC with sensitivity (SEN) of 59% and specificity (SPE) of 59% [2].

Laganas et al. (2021) trained machine learning models on Mel Frequency Cepstral Coefficients and Bark-band Energies of HC and PD patients extracted from passive smartphone phone call recordings using the iPrognosis mobile application. Gender and age were added to the feature matrix, and groups of testing data were balanced in terms of these confounding factors. After leave-one-out cross-validation, the best-performing models provided an area under the curve (AUC) with the threshold operating characteristic (ROC) of 0.69/0.68/0.63/0.83 for English/Greek/German/Portuguese speaking subjects [20].

Simek and Rusz (2021) tested the effect of speech task and ambient noise (10 dB and 20 dB signal-to-noise ratio) on sensitivity to capture dysphonia of PD and iRBD patients using unsmoothed (CPP) and smoothed (CPPs) cepstral peak prominence features. There was a significant difference between PD patients and HC in sustained phonation of vowel [a] via the CPP ($p < 0.05$) and CPPS ($p < 0.01$) and the monologue via the CPP ($p < 0.01$) according to a one-way analysis of variance. The differences dropped with the addition of noise [27].

### 1.2 Objectives

The iPrognosis app is the only one known to have been used to detect PD from running speech recorded during phone calls. However, the employed speech

features do not quantify disorders in all domains of HD. It is clear that a new approach to parametrization is needed, as existing extraction is dependent on the speech task performed. The main aim of this paper is to explore a new approach to passive HD assessment based on acoustic analysis of running speech in a noisy environment. A new method of feature extraction will be proposed so that the patient is not required to interact actively with the device and perform speech tasks. Subsequently, the robustness of features to noise that may be present in recordings during a phone call will be determined.

## 2 Materials and Methods

### 2.1 Dataset

The test database (PARCZ [12]) contains a total of 126 recordings of Czech speech (40 HC and 86 patients with PD) recorded with a condenser microphone in a regular, non-soundproofed room. Table 1 describes the representation of males and females, and Fig. 1 shows the age distribution of the subjects. During the monologue recording the participants mainly talked about their hobbies, interests, family, or profession. The mean recording length of HC is 26.3 ± 13.5 s, and for people with PD, it is 28.0 ± 15.0 s. Recordings were downsampled from 44.1 kHz to 16 kHz sampling frequency.

**Table 1.** Demograpfhic data.

| | women | men | total |
|---|---|---|---|
| HC | 21 | 19 | 40 |
| PD | 37 | 49 | 86 |
| total | 58 | 68 | 126 |

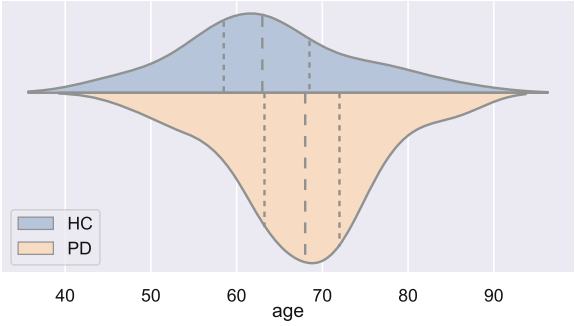


**Fig. 1.** Age distribution.

The participants also underwent clinical tests or questionnaires examining their motor and non-motor symptoms, sleep disorders, level of dysarthria, intelligence, depression and cognitive abilities. Results, along with the duration of the disease and medication, are shown in Table 2.

**Table 2.** Clinical data (mean ± std).

| | HC | PD |
|---|---|---|
| PD duration [year] | - | 14.1 ± 2.8 |
| LED [mg] | - | 987.1 ± 525.7 |
| UPDRS III | - | 24.6 ± 12.0 |
| UPDRS IV | - | 3.2 ± 2.7 |
| FOG | - | 7.1 ± 6.0 |
| NMSS | - | 39.3 ± 23.5 |
| RBDSQ | - | 3.8 ± 3.3 |
| faciokinesis | 27.9 ± 1.9 | 24.4 ± 3.5 |
| phonorespiration | 28.6 ± 1.4 | 23.8 ± 3.7 |
| phonetics | 29.5 ± 1.0 | 25.6 ± 3.7 |
| overall DX index | 86.0 ± 3.2 | 73.8 ± 9.3 |
| IQ | - | 106.7 ± 13.6 |
| BDI | - | 9.6 ± 5.9 |
| ACE-R | - | 86.4 ± 9.2 |
| ACE-R (attention and orientation) | - | 17.2 ± 1.3 |
| ACE-R (memory) | - | 19.5 ± 4.3 |
| ACE-R (fluency) | - | 10.4 ± 2.7 |
| ACE-R (language) | - | 24.9 ± 1.4 |
| ACE-R (visuospatial) | - | 14.9 ± 1.4 |
| MMSE | 28.3 ± 1.5 | 28.0 ± 2.5 |

[1] LED – Levodopa Equivalent Dose, UPDRS – Unified Parkinson's Disease Rating Scale (part III: Motor examination, part IV: Motor complications), FOG – Freezing of Gait, NMSS – Non-Motor Symptoms Scale, RBDSQ – REM Sleep Behavior Disorder Screening Questionnaire, DX index – Test 3F Dysarthric Profile (dysarthric index composed of faciokinesis, phonorespiration and phonetics), IQ – Intelligence Quotient, BDI – Beck Depression Inventory, ACE-R – Addenbrooke's Cognitive Examination-Revised, MMSE – Mini Mental State Exam

### 2.2 Simulation of Real Environment Conditions

The original recordings were mixed with three different types of ambient noise obtained from Freesound [10]:

- **Car**: the interior of a car during driving through a small city, the sound of an engine, rain and windscreen wipers.

- **Town square**: Union Square in San Francisco, a small art show, people chatting and moving with different objects, sometimes cars and birds.
- **TV**: ambient noise of an Indian television, channel changing, advertisements and music.

in 10 dB and 20 dB signal-to-noise ratio (SNR) in the following steps:

1. Resample the noise signal to the uniform 16 kHz sampling frequency.
2. Cut the noise signal to have an equal length as a speech recording.
3. Normalize both signals to have a maximum value equal to 1:

$$\mathbf{s} = \frac{\mathbf{s}}{\max(\mathbf{s})}, \tag{1}$$

where $\mathbf{s}$ stands for the noise or the speech signal.

4. Get the power $P$ of both signals:

$$P = \frac{\sum_{n=0}^{N-1} s[n]^2}{N}, \tag{2}$$

$s[n]$ stands for the $n$-th sample in a sampled audio signal of length $N$.

5. Get the signal-to-noise ratio $SNR$ in dB:

$$SNR = 10 \cdot \log_{10} \frac{P_{\text{speech}}}{P_{\text{noise}}}, \tag{3}$$

where $P_{\text{speech}}$ and $P_{\text{noise}}$ is the power of the speech and noise signal, respectively.

6. Mix the normalized speech signal $\mathbf{s}_{\text{speech}}$ with the normalized noise signal $\mathbf{s}_{\text{noise}}$ attenuated by the coefficient $C$:

$$C = \sqrt{10^{\frac{(SNR - SNR_{\text{M}})}{10}}}, \tag{4}$$

$$\mathbf{s}_{\text{mix}} = \mathbf{s}_{\text{speech}} + C \cdot \mathbf{s}_{\text{noise}}, \tag{5}$$

where, $SNR_{\text{M}}$ is 10 or 20 (dB) and $\mathbf{s}_{\text{mix}}$ is the final mixed signal.

Figure 2 shows a clean (original) recording of the Czech word "cestování", meaning "travelling" in English, and the same signal mixed with noise using different SNRs.

Recordings were mixed with noise in a way so that a random third was mixed with car noise, a second third with square noise, and the last third with TV noise. That resulted in 3 datasets – a dataset of clean recordings, then noisy recordings with 20 dB SNR and finally with 10 dB SNR.

### 2.3 Feature Extraction

The parameterization algorithm is programmed in MATLAB environment. All features along with their description and the speech disorders they quantify are

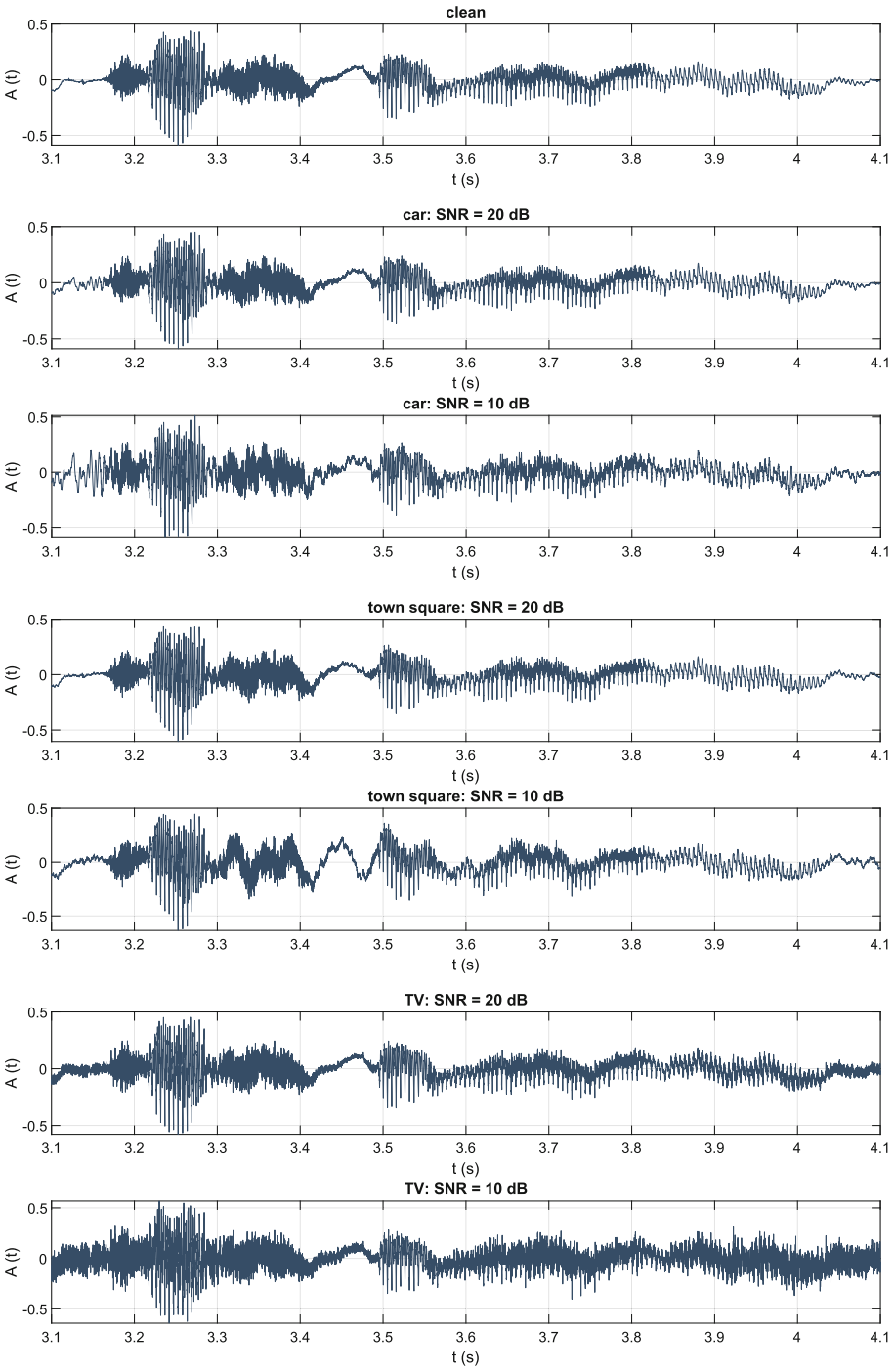


**Fig. 2.** A part of the clean and noisy monologue recording with different SNRs and types of noise.

**Table 3.** Speech features.

| Acoustic feature | Specific disorder | Expected change | Feature definition |
|---|---|---|---|
| PHONATION | | | |
| CPP | Increased hoarseness | ↓ | Cepstral peak prominence representing the dysphonia. CPP is defined as the difference between the cepstral peak representing the fundamental frequency and the linear regression line calculated from the magnitude-quefrency cepstra. |
| HRF | Increased breathness | ↓ | Harmonic richness factor, the amount of noise in the speech signal, mainly due to incomplete vocal fold closure. HRF is defined as the ratio between the sum of magnitudes of higher order harmonics and magnitude of the fundamental frequency. |
| NAQ | Increased voice harshness | ↑ | Normalised amplitude quotient, defined as A/(D*T0), where A is the amplitude of the glotal flow pulse, D is the negative peak amplitude of the glotal flow derivative and T0 is one period of glotal flow. |
| relNAQSD | Rigidity of vocal folds | ↑ | The standard deviation of normalised amplitude quotient relative to its mean. |
| QOQ | Increased voice harshness | ↑ | Mean quasi-open quotient, defined as the ratio between the time of opened phase and fundamental period (once cycle of the vocal fold). |
| relQOQSD | Rigidity of vocal folds | ↑ | The standard deviation of quasi-open quotient relative to its mean. |
| jitter | Microperturbations in frequency | ↑ | Frequency perturbation, extent of variation of the voice range. jitter is defined as the variability of F0 of speech from one cycle to the next. |
| shimmer | Microperturbations in amplitude | ↑ | Amplitude perturbation representing rough speech. shimmer is defined as the sequence of maximum extent of the signal amplitude within each vocal cycle. |
| ARTICTULATION | | | |
| RFA1 | Articulatory decay | ↓ | Resonant frequency attenuation defined as the distance in LPC spectrum between resonance of second formant and the local minima before this formant (in dB). |
| RFA2 | Articulatory decay | ↓ | Resonant frequency attenuation defined as the distance in LPC spectrum between resonance of second formant and the local minima after this formant (in dB). |
| #loc_max | Articulatory decay | ↓ | The number of local maxima in transfer function of the vocal tract representing the resonances. |
| relF1SD | Rigidity of tongue and jaw | ↓ | Standard deviation of first formant relative to its mean. |
| relF2SD | Rigidity of tongue and jaw | ↓ | Standard deviation of second formant relative to its mean. |
| PROSODY | | | |
| RSV | Reduced number of sustained vowels | ↓ | The ratio of vowels sustained for longer than 100 ms to the total number of vowels (grouped voiced segments). |
| relF0SD | Monopitch | ↓ | Pitch variation, defined as a standard deviation of F0 contour of voiced segments longer than 100 ms relative to its mean. |
| relSE0SD | Monoloudness | ↓ | Speech loudness variation, defined as a standard deviation of energy of voiced segments longer than 100 ms relative to its mean. |
| SPIR | Inappropriate silences | ↓ | Number of pauses (longer than 50 ms and shorter than 2 s) relative to total speech time. |
| DurMED | Longer duration of silences | ↑ | Median duration of silences longer than 50 ms and shorter than 2 s. |
| DurMAD | Higher variability of silence duration | ↑ | Median absolute deviation of silence duration (longer than 50 ms and shorter than 2s) |

listed in Table 3. It also describes the expected change in the feature for patients with increasing severity of HD. We used a Troparion [31] toolbox for extracting jitter and shimmer and Covarep [9] for the rest of the phonatory features.

The feature extraction is modified to extract the phonatory features directly from the monologue in order to examine all domains of HD together with articulatory and prosodic ones. First, a voice-activity-detector (VAD) is applied to the speech signal, followed by a voicing check. In both cases, the PRAAT [4] tool is used. If a voiced segment is longer than 100 ms, the phonatory features are extracted. Otherwise, the voiced signal is not long enough to obtain the features such as jitter or shimmer, as not enough vocal tract cycles are repeated. The fundamental frequency for someone may be less than 100 Hz, corresponding to 10 ms, and at least five cycles are needed to obtain these features. Fea-

tures measuring the pausing (SPIR, DurMED, DurMAD) neglect pauses shorter than 50 ms (articulatory pauses), but also neglect pauses longer than 2 s. This length is set based on the dissertation thesis written by Tyler S Kendall [19], which describes the variation in speech rate and silent pause duration by North American English speakers. Of the 22,000 measured pauses, only some outliers exceeded the pause length of 2 s, which probably include hesitation pauses and pauses that are purely for breathing. A speech signal fragment in Fig. 3 describes the particular parametrization process.

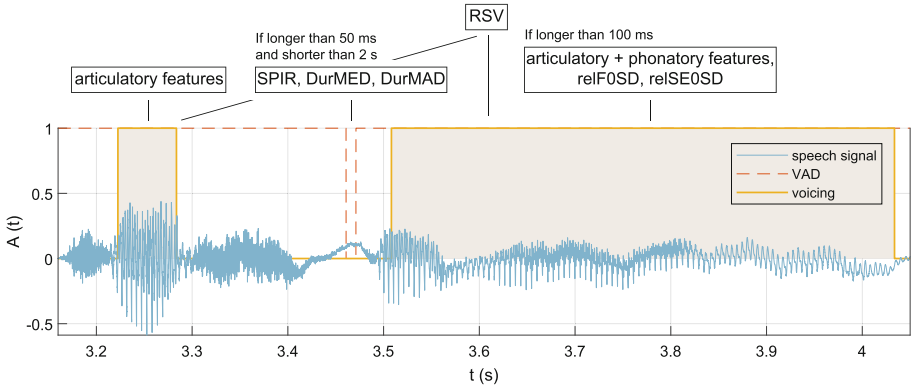


**Fig. 3.** Parametrization of the czech word "cestování". A zero value means that there is no speech according to the Voice Activity Detector (VAD) and/or no voiced segment.

Finally, to suppress the effect of confounding factors such as age and gender, we regressed them out. This was done using the Python programming language, which was also used for further statistical analysis and machine learning.

### 2.4 Statistical Analysis

To analyze the statistical relationship between speech features and clinical scores of PD patients, we calculated Pearson's correlation coefficients. The Benjamini/Hochberg test controlled the false discovery rate (FDR).

### 2.5 Machine Learning

By using the Extreme Gradient Boosting (XGBoost) algorithm, we mathematically modelled the extracted features during a prediction of clinical scores, or stratification the participants into the PD/HC groups. Hyperparameter tuning was included in the pipeline using a random search approach, and the models were validated using the stratified 10-fold cross-validation technique with 20 repetitions. The model's performance was evaluated by the area under the curve (AUC) of the receiver operating characteristics (ROC), sensitivity (SEN), and specificity (SPE) for the classification, and by the mean absolute error (MAE) and estimated error rate (EER) for the regression:

$$\text{EER} = \frac{\text{MAE}}{R}, \tag{6}$$

where $R$ represents the range of values (of a clinical scale) in the training set. Finally, each feature's importance in predicting each clinical score was obtained to measure how valuable the feature was in building the boosting decision tree. The importance coefficients of models trained on features of each dataset (clean, 20 dB, 10 dB) were multiplied to obtain global feature importances.

## 3 Results

Results of the correlation analysis can be found in Table 4. It focuses only on the clinical scores, which strongly correlate with at least one speech feature in any dataset (clean, 20 dB, 10 dB). Table 5 shows the regression model's performance in predicting the clinical scores. The three most important speech features in predicting each score are also listed. The classification results are illustrated in Fig. 4 (the ROC curve represents the model's performance trained on clean or noisy data with a signal-to-noise ratio of 20 dB and 10 dB; the curves and values shown are the averages of the results from the stratified cross-validation).

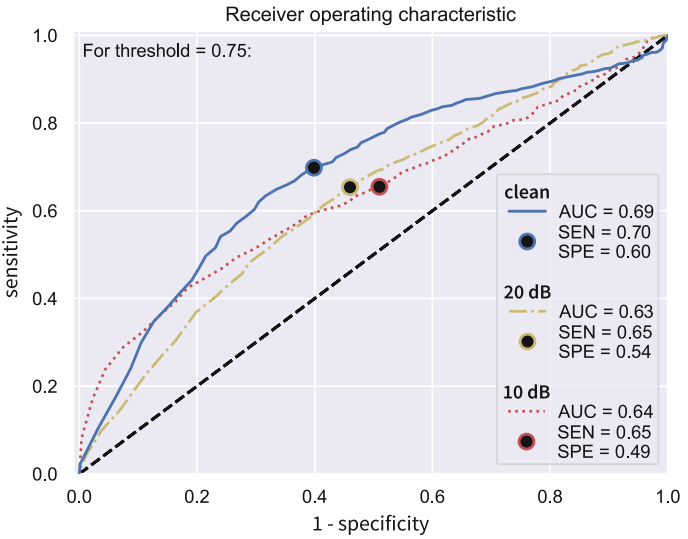


**Fig. 4.** Avarage ROC curve for the different signal-to-noise ratio scenarios.

**Table 4.** Coefficients and p-values after the FDR correction of Pearson's linear correlation between speech features and scores of clinical tests.

| | UPDRS III | | faciokinesis | | phonoresp. | | phonetics | | DX index | | MMSE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coeff | p-value | coeff | p-value | coeff | p-value | coeff | p-value | coeff | p-value | coeff | p-value |
| clean | | | | | | | | | | | | |
| RSV | −0.125 | 0.432 | 0.078 | 0.815 | 0.000 | 0.999 | 0.133 | 0.219 | 0.077 | 0.495 | −0.108 | 0.612 |
| CPP | 0.037 | 0.914 | 0.107 | 0.673 | 0.119 | 0.352 | 0.058 | 0.521 | 0.107 | 0.393 | 0.216 | 0.190 |
| HRF | −0.258 | 0.152 | 0.188 | 0.171 | 0.138 | 0.289 | 0.213 | 0.045* | 0.201 | 0.095 | −0.105 | 0.612 |
| NAQ | 0.387 | 0.001** | −0.309 | 0.001** | −0.262 | 0.028* | −0.353 | 0.001** | −0.345 | 0.001** | 0.171 | 0.336 |
| relNAQSD | −0.038 | 0.914 | 0.057 | 0.829 | 0.075 | 0.596 | 0.148 | 0.176 | 0.105 | 0.393 | −0.054 | 0.727 |
| QOQ | 0.197 | 0.219 | −0.217 | 0.095 | −0.192 | 0.122 | −0.269 | 0.019* | −0.254 | 0.025* | 0.020 | 0.883 |
| relQOQSD | 0.003 | 0.976 | 0.027 | 0.855 | 0.085 | 0.551 | 0.095 | 0.371 | 0.078 | 0.495 | −0.059 | 0.727 |
| jitter | −0.215 | 0.175 | −0.027 | 0.855 | −0.041 | 0.725 | 0.063 | 0.521 | −0.002 | 0.978 | −0.274 | 0.085 |
| shimmer | −0.217 | 0.175 | 0.038 | 0.855 | 0.029 | 0.791 | 0.151 | 0.175 | 0.081 | 0.495 | −0.275 | 0.085 |
| RFA1 | −0.008 | 0.976 | 0.104 | 0.673 | 0.087 | 0.551 | 0.169 | 0.125 | 0.134 | 0.255 | −0.015 | 0.883 |
| RFA2 | 0.023 | 0.933 | 0.231 | 0.085 | 0.283 | 0.019* | 0.252 | 0.019* | 0.288 | 0.010* | 0.135 | 0.475 |
| #loc_max | 0.129 | 0.432 | −0.037 | 0.855 | −0.042 | 0.725 | −0.105 | 0.326 | −0.069 | 0.528 | 0.216 | 0.190 |
| relF1SD | −0.114 | 0.469 | 0.003 | 0.969 | −0.053 | 0.707 | −0.087 | 0.393 | −0.052 | 0.629 | −0.145 | 0.456 |
| relF2SD | −0.156 | 0.413 | 0.022 | 0.855 | −0.058 | 0.707 | −0.060 | 0.521 | −0.037 | 0.717 | −0.200 | 0.213 |
| relF0SD | −0.050 | 0.914 | 0.085 | 0.815 | 0.151 | 0.269 | 0.255 | 0.019* | 0.184 | 0.130 | −0.054 | 0.727 |
| relSE0SD | −0.032 | 0.914 | 0.055 | 0.829 | 0.202 | 0.122 | 0.111 | 0.322 | 0.140 | 0.253 | 0.083 | 0.727 |
| SPIR | −0.217 | 0.175 | 0.112 | 0.673 | 0.197 | 0.122 | 0.244 | 0.023* | 0.209 | 0.090 | 0.068 | 0.727 |
| DurMED | 0.126 | 0.432 | −0.054 | 0.829 | −0.148 | 0.269 | −0.216 | 0.045* | −0.158 | 0.212 | 0.059 | 0.727 |
| DurMAD | 0.136 | 0.432 | −0.051 | 0.829 | −0.133 | 0.289 | −0.209 | 0.045* | −0.149 | 0.230 | 0.046 | 0.743 |
| SNR = 20 dB | | | | | | | | | | | | |
| RSV | 0.153 | 0.253 | −0.127 | 0.423 | −0.121 | 0.352 | −0.123 | 0.401 | −0.139 | 0.309 | 0.068 | 0.669 |
| CPP | −0.022 | 0.943 | 0.145 | 0.339 | 0.210 | 0.120 | 0.160 | 0.204 | 0.194 | 0.114 | 0.279 | 0.044* |
| HRF | −0.218 | 0.139 | 0.150 | 0.339 | 0.061 | 0.864 | 0.112 | 0.407 | 0.119 | 0.357 | −0.118 | 0.507 |
| NAQ | 0.400 | 0.001** | −0.276 | 0.038* | −0.242 | 0.066 | −0.376 | 0.001** | −0.334 | 0.001** | 0.204 | 0.177 |
| relNAQSD | 0.040 | 0.943 | 0.000 | 0.996 | 0.005 | 0.952 | 0.076 | 0.527 | 0.030 | 0.770 | −0.061 | 0.669 |
| QOQ | 0.234 | 0.139 | −0.234 | 0.057 | −0.192 | 0.122 | −0.302 | 0.010* | −0.271 | 0.013* | 0.054 | 0.669 |
| relQOQSD | 0.022 | 0.943 | 0.168 | 0.290 | 0.168 | 0.193 | 0.210 | 0.066 | 0.204 | 0.104 | −0.151 | 0.323 |
| jitter | −0.175 | 0.197 | −0.058 | 0.780 | −0.140 | 0.323 | −0.024 | 0.791 | −0.085 | 0.594 | −0.375 | 0.001** |
| shimmer | −0.236 | 0.139 | 0.084 | 0.669 | −0.010 | 0.952 | 0.102 | 0.415 | 0.064 | 0.651 | −0.394 | 0.001** |
| RFA1 | 0.194 | 0.192 | −0.110 | 0.467 | 0.017 | 0.952 | −0.073 | 0.527 | −0.060 | 0.651 | 0.050 | 0.669 |
| RFA2 | 0.031 | 0.943 | 0.245 | 0.057 | 0.294 | 0.019* | 0.239 | 0.044* | 0.292 | 0.010* | 0.152 | 0.323 |
| #loc_max | 0.121 | 0.393 | −0.035 | 0.780 | −0.048 | 0.908 | −0.113 | 0.407 | −0.073 | 0.651 | 0.217 | 0.177 |
| relF1SD | −0.223 | 0.139 | 0.075 | 0.698 | −0.032 | 0.916 | 0.033 | 0.754 | 0.026 | 0.770 | −0.153 | 0.323 |
| relF2SD | −0.274 | 0.104 | 0.111 | 0.467 | −0.033 | 0.916 | 0.063 | 0.542 | 0.050 | 0.684 | −0.200 | 0.177 |
| relF0SD | 0.007 | 0.955 | 0.046 | 0.780 | 0.014 | 0.952 | 0.064 | 0.542 | 0.046 | 0.684 | −0.002 | 0.987 |
| relSE0SD | −0.006 | 0.955 | 0.039 | 0.780 | 0.192 | 0.122 | 0.076 | 0.527 | 0.118 | 0.357 | 0.089 | 0.635 |
| SPIR | −0.189 | 0.192 | 0.011 | 0.957 | 0.045 | 0.908 | 0.101 | 0.415 | 0.059 | 0.651 | −0.106 | 0.544 |
| DurMED | 0.181 | 0.197 | −0.053 | 0.780 | −0.120 | 0.352 | −0.206 | 0.066 | −0.143 | 0.309 | 0.073 | 0.669 |
| DurMAD | 0.172 | 0.197 | −0.035 | 0.780 | −0.119 | 0.352 | −0.205 | 0.066 | −0.136 | 0.309 | 0.052 | 0.669 |
| SNR = 10 dB | | | | | | | | | | | | |
| RSV | 0.100 | 0.551 | −0.139 | 0.326 | −0.079 | 0.762 | −0.098 | 0.542 | −0.117 | 0.532 | 0.053 | 0.685 |
| CPP | 0.096 | 0.551 | 0.054 | 0.774 | 0.109 | 0.619 | 0.052 | 0.628 | 0.082 | 0.629 | 0.224 | 0.203 |
| HRF | −0.176 | 0.252 | 0.059 | 0.774 | 0.023 | 0.890 | 0.030 | 0.737 | 0.042 | 0.765 | −0.054 | 0.685 |
| NAQ | 0.302 | 0.076 | −0.166 | 0.300 | −0.174 | 0.269 | −0.248 | 0.057 | −0.220 | 0.076 | 0.149 | 0.380 |
| relNAQSD | 0.082 | 0.618 | −0.243 | 0.114 | −0.157 | 0.269 | −0.204 | 0.139 | −0.225 | 0.076 | −0.008 | 0.938 |
| QOQ | 0.227 | 0.152 | −0.176 | 0.300 | −0.155 | 0.269 | −0.244 | 0.057 | −0.215 | 0.076 | 0.069 | 0.657 |
| relQOQSD | 0.070 | 0.662 | −0.148 | 0.314 | −0.076 | 0.762 | −0.090 | 0.542 | −0.116 | 0.532 | −0.076 | 0.657 |
| jitter | −0.193 | 0.206 | −0.095 | 0.593 | −0.161 | 0.269 | −0.058 | 0.616 | −0.120 | 0.532 | −0.362 | 0.001** |
| shimmer | −0.271 | 0.076 | 0.028 | 0.874 | −0.028 | 0.890 | 0.072 | 0.542 | 0.025 | 0.778 | −0.351 | 0.001** |
| RFA1 | 0.214 | 0.152 | −0.157 | 0.300 | 0.006 | 0.949 | −0.102 | 0.542 | −0.092 | 0.608 | 0.086 | 0.657 |
| RFA2 | −0.018 | 0.872 | 0.200 | 0.228 | 0.260 | 0.057 | 0.183 | 0.190 | 0.243 | 0.076 | 0.140 | 0.384 |
| #loc_max | 0.133 | 0.385 | −0.048 | 0.774 | −0.056 | 0.890 | −0.135 | 0.498 | −0.089 | 0.608 | 0.190 | 0.266 |
| relF1SD | −0.218 | 0.152 | 0.091 | 0.593 | −0.046 | 0.890 | 0.045 | 0.649 | 0.031 | 0.778 | −0.169 | 0.337 |
| relF2SD | −0.271 | 0.076 | 0.132 | 0.335 | −0.032 | 0.890 | 0.078 | 0.542 | 0.063 | 0.653 | −0.210 | 0.214 |
| relF0SD | −0.030 | 0.872 | 0.046 | 0.774 | 0.077 | 0.762 | 0.072 | 0.542 | 0.074 | 0.653 | −0.009 | 0.938 |
| relSE0SD | 0.021 | 0.872 | 0.016 | 0.907 | 0.187 | 0.269 | 0.079 | 0.542 | 0.109 | 0.537 | 0.070 | 0.657 |
| SPIR | −0.046 | 0.798 | 0.063 | 0.774 | 0.038 | 0.890 | 0.082 | 0.542 | 0.068 | 0.653 | 0.074 | 0.657 |
| DurMED | 0.136 | 0.385 | −0.025 | 0.874 | 0.047 | 0.890 | −0.095 | 0.542 | −0.025 | 0.778 | 0.162 | 0.337 |
| DurMAD | 0.154 | 0.329 | −0.005 | 0.957 | 0.009 | 0.949 | −0.119 | 0.542 | −0.042 | 0.765 | 0.112 | 0.549 |

* – $p < 0.05$; ** – $p < 0.01$

**Table 5.** Results of PD duration and clinical scores prediction in different SNR scenarios (mean values from the stratified cross-validation).

| | MAE | | | EER [%] | | | Important |
|---|---|---|---|---|---|---|---|
| | clean | 20 dB | 10 dB | clean | 20 dB | 10 dB | features |
| PD duration | 2.37 | 2.36 | 2.45 | 26.30 | 26.27 | 27.23 | shimmer, relF0SD, CPP |
| UPDRS III | 10.71 | 9.97 | 9.85 | 20.61 | 19.17 | 18.95 | RFA2, NAQ, relF2SD |
| UPDRS IV | 2.45 | 2.32 | 2.32 | 24.48 | 23.21 | 23.17 | #loc_max, NAQ, relSE0SD |
| FOG | 5.26 | 4.93 | 4.83 | 26.31 | 24.64 | 24.18 | RFA2, relSE0SD, shimmer |
| RBDSQ | 2.64 | 2.59 | 2.69 | 20.27 | 19.94 | 20.73 | NAQ, RSV, QOQ |
| faciokinesis | 3.01 | 2.82 | 2.79 | 14.32 | 13.41 | 13.29 | QOQ, HRF, SPIR |
| phonorespiration | 3.03 | 3.07 | 2.84 | 14.41 | 14.61 | 13.51 | jitter, relF2SD, RFA2 |
| phonetics | 2.56 | 2.76 | 2.69 | 14.21 | 15.32 | 14.97 | QOQ, shimmer, SPIR |
| overall DX index | 6.51 | 6.52 | 6.28 | 11.23 | 11.24 | 10.83 | QOQ, jitter, shimmer |
| BDI | 4.79 | 4.88 | 5.28 | 17.73 | 18.09 | 19.56 | relF0SD, QOQ, RFA2 |
| ACE-R | 8.68 | 7.25 | 7.09 | 16.38 | 13.67 | 13.38 | NAQ, RSV, relF0SD |
| ACE-R (attention and orientation) | 0.99 | 1.08 | 1.02 | 19.82 | 21.63 | 20.44 | relF1SD, relNAQSD, jitter |
| ACE-R (memory) | 4.01 | 3.47 | 3.62 | 17.45 | 15.09 | 15.76 | NAQ, relQOQSD, relF2SD |
| ACE-R (fluency) | 2.64 | 2.53 | 2.53 | 24.02 | 23.00 | 23.04 | NAQ, RSV, QOQ |
| ACE-R (language) | 1.15 | 1.04 | 1.04 | 19.10 | 17.33 | 17.34 | SPIR, relSE0SD, CPP |
| ACE-R (visuospatial) | 1.32 | 1.17 | 1.21 | 22.07 | 19.55 | 20.22 | relSE0SD, RFA2, relNAQSD |
| MMSE | 1.84 | 1.74 | 1.81 | 13.11 | 12.40 | 12.96 | CPP, NAQ, RFA2 |

## 4 Discussion

We tested a novel approach of acoustic speech feature extraction from a running speech on a database of 126 recordings (40 HC and 86 patients with PD). The algorithm was designed to be able to parametrize speech recordings obtained from phone calls, and, at the same time, to objectively assess the severity of HD in all speech domains using the obtained features. For testing purposes, the original recordings were mixed with a noise of a natural environment.

The results of the correlation analysis (Table 4) show a strong relationship between obtained features and scores of some clinical scales. These scores are of the motor skills test (UPDRS III), Test 3F Dysarthric Profile (DX index) and its parts, and cognitive function test (MMSE). The values of all significant correlation coefficients are consistent with the expected change in the feature at higher severity of HD described in Table 3. The analysis reveals that most features correlate with the results of the speakers' score of phonetics. Their increased breathiness due to incomplete vocal fold closure (HRF), increased voiced harshness (NAQ, QOQ), articulatory decay (RFA2), monopitch (relF0SD), inappropriate silences (SPIR), longer duration of silences (DurMED) and higher variability of silence duration (DurMAD) have a significant linear relationship with results of this test. The NAQ and RFA2 features strongly correlate with the overall dysarthric index and even with patients' clinically tested motor skills (UDPRS III). It is evident that with the increasing noise in the recordings, the correlation strength decreases. The phonatory feature NAQ, followed by RFA2 and QOQ, is the most robust in this sense. The Mini Mental State Exam (MMSE), which

mainly examines a patient's orientation in time and place, concentration, and short-term memory, is the only non-motor test that strongly correlates with some speech features. These features quantify increased voice hoarseness (CPP) and mainly perturbations in frequency (jitter) and amplitude (shimmer), strongly correlated even in 10 dB SNR conditions.

The ability of the regression model to predict scores of clinical scales is expressed in Table 5. The MAE metric gives us directly the average deviation of the prediction from the true value. In this respect, the model is able to predict the PD duration from the extracted features of the original recordings with an error of 2.37 years. However, we need to consider the range of values within which we operate. The EER metric that accounts for this range points out that this error is 26.30%, which climbed to 27.30% when predicted based on features extracted from noisy recordings (10 dB SNR). The best-performing prediction in this term is the overall DX index, where the EER reaches 10.83% in the scenario with the noisiest recordings. This model's most important speech features are QOQ, jitter and shimmer. This shows that the quantification of phonatory disorders plays an essential role in predicting the severity of HD, and it also implies that the adaptation of the parametrization to running speech works well. The prediction of MMSE is the second most successful, with an EER of 12.96%. The most important feature here quantifies increased voice hoarseness (CPP). The robustness to noise of this feature is consistent with the results of the study by Simek and Rusz [27]. The Addenbrooke's Cognitive Examination-Revise score can be predicted with an EER of 13.38%, and the Unified Parkinson's Disease Rating Scale (part II) score with an EER of 18.95%. The NAQ feature is important in both cases.

The classifier reaches AUC = 0.69 with SEN = 70% and SPE = 60%, and it is evident that noise affects the accuracy of stratifying speakers into HC and PD groups (Fig. 4). These results are similar to the ones reported by Arora et al. [2] and Laganas et al. [20]. However, comparisons are not very appropriate here because each study used a different database.

## 5 Conclusion

In this paper, we investigated the potential of passive speech analysis to predict clinical scores that quantify the severity of PD. We showed that asking patients to record the commonly used sustained vowel [a] is not necessary, because the phonatory features can be extracted directly from specific voiced segments of the running speech. In addition, our approach enables important quantification of HD in other dimensions, such as articulation and prosody.

This paper is the first that deals with an adaptation of the established HD parametrization process for passive monitoring purposes. The suggested algorithm, tested on noisy speech recordings, can be used in mHealth applications and facilitate passive monitoring and assessment of PD.

The work could be continued with a more in-depth statistical analysis, including statistical hypothesis tests. In addition, it would be interesting to observe

the number of patients deviating from the norms (given by healthy controls) in each speech feature. Nevertheless, the algorithm mainly needs to be validated in the wild.

## References


1. Adams, S.G., Dykstra, A., Jenkins, M., Jog, M.: Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. J. Med. Speech-Lang. Pathol. **16**(4), 165–173 (2008)
2. Arora, S., Lo, C., Hu, M., Tsanas, A.: Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and Parkinson's disease. IEEE Access **9**, 44813–44824 (2021)
3. Baudouin, R., Lechien, J.R., Carpentier, L., Gurruchaga, J.M., Lisan, Q., Hans, S.: Deep brain stimulation impact on voice and speech quality in Parkinson's disease: a systematic review. Otolaryngol. Head Neck Surg. 01945998221120189 (2022)
4. Boersma, P., Weenink, D.: PRAAT: doing phonetics by computer [computer program]. version 5.3. 51 (2013). http://www.praat.org/retrieved. Accessed 12 (2013)
5. Cernak, M., Orozco-Arroyave, J.R., Rudzicz, F., Christensen, H., Vásquez-Correa, J.C., Nöth, E.: Characterisation of voice quality of Parkinson's disease using differential phonological posterior features. Comput. Speech Lang. **46**, 196–208 (2017)
6. Chade, A., Kasten, M., Tanner, C.: Nongenetic causes of Parkinson's disease. Parkinson's Dis. Relat. Disord. **70**, 147–151 (2006)
7. Connolly, B.S., Lang, A.E.: Pharmacological treatment of Parkinson disease: a review. JAMA **311**(16), 1670–1683 (2014)
8. Darley, F.L., Aronson, A.E., Brown, J.R.: Differential diagnostic patterns of dysarthria. J. Speech Hear. Res. **12**(2), 246–269 (1969)
9. Degottex, G., Kane, J., Drugman, T., Raitio, T., Scherer, S.: COVAREP-A collaborative voice analysis repository for speech technologies. In: 2014 IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP), pp. 960–964. IEEE (2014)
10. Font Corbera, F., Roma Trepat, G., Serra, X.: Freesound technical demo. In: MM 2013. Proceedings of the 21st ACM International Conference on Multimedia; 21–25 Oct 2013 Barcelona, Spain. New York: ACM; 2013, p. 411–412. ACM Association for Computer Machinery (2013)
11. Freed, D.B.: Motor Speech Disorders: Diagnosis and Treatment. Plural Publishing, San Diego (2018)
12. Galaz, Z., et al.: Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. Comput. Methods Programs Biomed. **127**, 301–317 (2016)
13. Hammen, V.L., Yorkston, K.M.: Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria. J. Commun. Disord. **29**(6), 429–445 (1996)
14. Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L., Gates, S.: Speech impairment in a large sample of patients with Parkinson's disease. Behav. Neurol. **11**(3), 131–137 (1998)
15. Hoodin, R.B., Gilbert, H.R.: Nasal airflows in parkinsonian speakers. J. Commun. Disord. **22**(3), 169–180 (1989)

16. Horin, A.P., McNeely, M.E., Harrison, E.C., Myers, P.S., Sutter, E.N., Rawson, K.S., Earhart, G.M.: Usability of a daily mhealth application designed to address mobility, speech and dexterity in Parkinson's disease. Neurodegenerative Dis. Manage. **9**(2), 97–105 (2019)
17. Hornykiewicz, O.: Biochemical aspects of Parkinson's disease. Neurology **51**(2 Suppl 2), S2–S9 (1998)
18. Juste, F.S., Sassi, F.C., Costa, J.B., de Andrade, C.R.F.: Frequency of speech disruptions in Parkinson's disease and developmental stuttering: a comparison among speech tasks. PLoS ONE **13**(6), e0199054 (2018)
19. Kendall, T.S.: Speech Rate, Pause, and Linguistic Variation: An Examination Through the Sociolinguistic Archive and Analysis Project. Duke University, Durham (2009)
20. Laganas, C., et al.: Parkinson's disease detection based on running speech data from phone calls. IEEE Trans. Biomed. Eng. **69**(5), 1573–1584 (2021)
21. Orozco-Arroyave, J.R., et al.: Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement. Neurodegenerative Dis. Manage. **10**(3), 137–157 (2020)
22. Parkinson, J.: An essay on the shaky palsy. London: Sherwood, Neely and Jones, pp. 1–6 (1817)
23. Poewe, W.: Global Scales to Stage Disability in PD: the Hoehn and Yahr scale. Rating Scales Parkinsons Disease, pp. 115–122 (2012)
24. Rohl, A., Gutierrez, S., Johari, K., Greenlee, J., Tjaden, K., Roberts, A.: Chapter 7 - speech dysfunction, cognition, and Parkinson's disease. In: Narayanan, N.S., Albin, R.L. (eds.) Cognition in Parkinson's Disease, Progress in Brain Research, vol. 269, pp. 153–173. Elsevier (2022). https://doi.org/10.1016/bs.pbr.2022.01.017, https://www.sciencedirect.com/science/article/pii/S0079612322000176
25. Rusz, J., Tykalova, T., Novotny, M., Ruzicka, E., Dusek, P.: Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. npj Parkinson's Dis. **7**(1), 1–8 (2021)
26. Savica, R., Grossardt, B.R., Bower, J.H., Ahlskog, J.E., Rocca, W.A.: Time trends in the incidence of Parkinson disease. JAMA Neurol. **73**(8), 981–989 (2016)
27. Šimek, M., Rusz, J.: Validation of cepstral peak prominence in assessing early voice changes of Parkinson's disease: Effect of speaking task and ambient noise. J. Acoust. Soci. Am. **150**(6), 4522–4533 (2021)
28. Thijs, Z., Watts, C.R.: Perceptual characterization of voice quality in nonadvanced stages of Parkinson's disease. J. Voice **36**(2), 293.e11-293.e18 (2020)
29. Tjaden, K., Wilding, G.: Effects of speaking task on intelligibility in Parkinson's disease. Clin. Linguist. Phonetics **25**(2), 155–168 (2011)
30. Tykalová, T., Rusz, J., Švihlík, J., Bancone, S., Spezia, A., Pellecchia, M.T.: Speech disorder and vocal tremor in postural instability/gait difficulty and tremor dominant subtypes of Parkinson's disease. J. Neural Transm. **127**(9), 1295–1304 (2020)
31. Vashkevich, M., Petrovsky, A., Rushkevich, Y.: Bulbar ALS detection based on analysis of voice perturbation and vibrato. In: 2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pp. 267–272. IEEE (2019)
32. Zhan, A., et al.: Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol. **75**(7), 876–880 (2018)