Journal of Communication Disorders 76 (2018) 21–36

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis

Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease

J.C. Vásquez-Correa[a,b], J.R. Orozco-Arroyave[a,b,*], T. Bocklet[c], E. Nöth[b]

[a] Faculty of Engineering, University of Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia
[b] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[c] Intel Corporation, Germany

ARTICLE INFO

Keywords:
Parkinson's disease
Phonation
Articulation
Prosody
Frenchay Dysarthria Assessment
Longitudinal analysis

ABSTRACT

Background: Parkinson's disease (PD) is a neurological disorder that produces motor and non-motor impairments. The evaluation of motor symptoms is currently performed following the third section of the Movement Disorder Society – Unified Parkinson's Disease Rating Scale (MDS-UPDRS-III); however, only one item of that scale is related to speech impairments. It is necessary to develop a specific scale such that considers those aspects related to speech impairments of the patients.

Aims: (i) To introduce and evaluate the suitability of a modified version of the Frenchay Dysarthria Assessment (m-FDA) scale to quantify the dysarthria level of PD patients; (ii) to objectively model dysarthric speech signals considering four speech dimensions; (iii) to develop a methodology, based on speech processing and machine learning methods, to automatically quantify/predict the dysarthria level of patients with PD.

Methods: The speech recordings are modeled using features extracted from several dimensions of speech: phonation, articulation, prosody, and intelligibility. The dysarthria level is quantified using linear and non-linear regression models. Speaker models based on i-vectors are also explored.

Results and conclusions: The m-FDA scale was introduced to assess the dysarthria level of patients with PD. Articulation features extracted from continuous speech signals to create i-vectors were the most accurate to quantify the dysarthria level, with correlations of up to 0.69 between the predicted m-FDA scores and those assigned by the phoniatricians. When the dysarthria levels were estimated considering dedicated speech exercises such as rapid repetition of syllables (DDKs) and read texts, the correlations were 0.64 and 0.57, respectively. In addition, the combination of several feature sets and speech tasks improved the results, which validates the hypothesis about the contribution of information from different tasks and feature sets when assessing dysarthric speech signals. The speaker models seem to be promising to perform individual modeling for monitoring the dysarthria level of PD patients. The proposed approach may help clinicians to make more accurate and timely decisions about the evaluation and therapy associated to the dysarthria level of patients. The proposed approach is a great step towards unobtrusive/ecological evaluations of patients with dysarthric speech without the need of attending medical appointments.

* Corresponding author at: Faculty of Engineering, University of Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia.
E-mail addresses: jcamilo.vasquez@udea.edu.co (J.C. Vásquez-Correa), rafael.orozco@udea.edu.co (J.R. Orozco-Arroyave), tobias.bocklet@intel.com (T. Bocklet), elmar.noeth@fau.de (E. Nöth).

https://doi.org/10.1016/j.jcomdis.2018.08.002
Received 6 April 2017; Received in revised form 29 July 2018; Accepted 7 August 2018
Available online 20 August 2018

1. Introduction

Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the midbrain, producing several motor and non-motor impairments (Hornykiewicz, 1998). Motor symptoms include, bradykinesia, rigidity, resting tremor, micrographia, and different speech impairments. The disease progression in motor activities is currently evaluated with the third section of the Movement Disorder Society – Unified Parkinson's Disease Rating Scale (MDS-UPDRS-III) (Goetz, 2008). This section comprises 33 items and is administered by neurologist experts according to their own clinical criterion. Thus it could be highly subjective. The scale contains several items to evaluate different motor activities such as finger tapping, gait, speech, and facial expression. Although the majority of PD patients develop several speech disorders, only one item in the scale is related to speech (Logemann, Fisher, Boshes, & Blonsky, 1978). Those disorders are considered an early sign of further motor impairments (Hlavnicka et al., 2017; Rusz et al., 2013). The most common symptoms in the speech of PD patients include reduced loudness, monopitch, monoloudness, reduced stress, breathy, hoarse voice quality, and imprecise articulation. These impairments are grouped together and called *hypokinetic dysarthria* (Logemann et al., 1978).

Although the MDS-UPDRS-III evaluates motor skills including the movement of hands and arms, gait, and posture, among others, it is not suitable nor fair to assume that the scale can be accurately predicted only based on speech recordings. To evaluate the impact of PD on an aspect of communication, a scale for evaluating speech would be a valuable tool. Several studies have considered the application of scales to assess only the speech deficits of PD patients (Patel, Parveen, & Anand, 2016; Skodda, Visser, & Schlegel, 2011). For instance, the Frenchay Dysarthria Assessment (FDA) introduced by Enderby (1980) and revised by Enderby and Palmer (2008) was designed to assess dysarthria, which is also suffered by PD patients. The FDA scale includes several items to evaluate dysarthria such as reflexes, respiration, lips movement, palate movement, laryngeal capacity, tongue posture/movement, intelligibility, and others. This tool covers a wide range of aspects; however, it requires the patient to be with the examiner, which is not possible in many cases due to their reduced mobility.

The research community has addressed since several years the problem of reducing the subjectivity of clinical evaluations to guarantee their reproducibility. The main purpose is to provide additional information to the clinical expert to reduce subjectivity in the final diagnosis. For the specific case of pathological speech analysis, researchers work on two main aspects: the development of suitable and accurate acoustic measures to model the speech signals and the development of different signal processing and machine learning techniques to reduce the subjectivity in clinical evaluations, e.g., trying to predict the score of a scale. With the aim of contributing to these two challenges, this study introduces an objective and reproducible methodology to model speech signals and to quantify the dysarthria level of PD patients.

Several studies in the literature described the speech impairments of PD patients in terms of different dimensions such as phonation, articulation, prosody, and intelligibility (Orozco-Arroyave, 2016; Rusz, Cmejla, Ruzickova, & Ruzicka, 2011). Related studies describing the assessment of PD using each speech dimension are reviewed below and the features used to assess each dimension are described in Section 2.2.

1.1. Phonation analysis

Phonation in PD patients is characterized by bowing and inadequate closure of vocal folds (Hanson, Gerratt, & Ward, 1984), which produce problems in stability and periodicity of the vibration. Phonation in PD was analyzed by Tsanas, Little, Fox, and Ramig (2014), who used features related to perturbation, noise content, and non-linear dynamics to evaluate the response of 14 PD patients to the Lee Silverman voice treatment as “acceptable” or “unacceptable”. The authors considered only information from the sustained vowel /a/, and reported accuracies close to 90% discriminating between “acceptable” vs. “unacceptable” utterances. Orozco-Arroyave et al. (2015) evaluated different characterization methods related to phonation analysis for the classification of PD patients and healthy control (HC) speakers. The authors used information of sustained vowels and evaluated four different characterization approaches: stability and periodicity, noise measures, spectral wealth, and non-linear dynamics. They reported accuracies of up to 84%, depending on the analyzed vowel and on the feature set. Recently, Hemmerling, Orozco-Arroyave, Skalski, Gajda, and Nöth (2016) proposed a novel phonatory analysis in PD patients based on the Hilbert-Huang transformation computed upon modulated (varying between low and high pitch) and sustained vowels. The authors analyzed the fundamental frequency and its range to assess monotonicity in PD speakers. The authors automatically discriminated PD and HC speakers and reported accuracies of up to 90%.

1.2. Articulation analysis

Articulation deficits in PD patients are mainly related to reduced amplitude and velocity of lip, tongue, and jaw movements (Ackermann & Ziegler, 1991). It has been studied in several works, for instance Skodda, Visser, et al. (2011) evaluated possible correlations between vowel articulation, global motor performance, and the stage of the disease. The data considered by the authors included speech recordings of 68 patients and 32 HC. The authors concluded that the vowel articulation index is significantly reduced in PD speakers. Novotný, Rusz, Čmejla, and Růžička (2014) modeled six different articulatory deficits in PD: vowel quality, coordination of laryngeal and supra-laryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors studied the rapid repetition of the syllables /pa-ta-ka/ pronounced by 24 Czech native speakers, and reported 88% accuracy discriminating between PD patients and HC speakers. Recently, Orozco-Arroyave (2016) proposed a method to model the difficulty of PD patients to start/stop the vocal fold vibration in continuous speech based on the energy content in the transitions from unvoiced to voiced and from voiced to unvoiced segments. The author addressed the automatic classification of PD

patients and HC speakers with speech recordings in three different languages (Spanish, German, and Czech), and reported accuracies between 80% and 94% depending on the language.

1.3. Prosody analysis

Prosody deficits in PD are manifested as monotonocity, monoloudness, and changes in speech rate and pauses (Rusz et al., 2011). Patel et al. (2016) calculated several measures to describe rate, phrasing, and stress in the speech of PD patients. Fundamental frequency and intensity were also estimated. The measures were compared to listener ratings performed by 12 naive listeners. The reported results using several statistical tests suggested that prosody is significantly impaired in PD. The authors also performed a linear regression to predict the listener ratings with the computed features. Skodda, Grönheit, and Schlegel (2011) studied different prosodic parameters in a large group of patients. A total of 169 PD and 64 HC speakers read the same set of sentences. The authors concluded that the variability of F_0 in male and female patients is reduced compared to HC. In addition, the authors observed a reduction in the percentage pause time within polysyllabic words in PD patients.

1.4. Intelligibility analysis

Clinicians have observed and reported reduced intelligibility in PD patients (Barnish, Whibley, Horton, Butterfint, & Deane, 2016; Logemann et al., 1978). De Letter, Santens, and Van Borsel (2005) analyzed possible correlations between the perceived intelligibility (in terms of correctly pronounced words) and the severity of the disease. The authors concluded that there is no correlation between the perceived intelligibility and the neurological state of the patients. A similar result was reported by Miller et al. (2007), where the intelligibility of 125 PD patients and the same number of age matched HC was evaluated. The authors investigated the correlation between the disease severity and the speech intelligibility (also in terms of correctly pronounced words) perceived by listeners unfamiliar with dysarthric speech. The authors concluded that although speech intelligibility is significantly reduced in people with PD, no correlations were found between perceived intelligibility and the disease severity. In recent years automatic intelligibility assessment considering speech of people with PD has captured the attention of the research community. For instance, Orozco-Arroyave, Vásquez-Correa, et al. (2016) and more recently Dimauro, Di-Nicola, Bevilacqua, Caivano, and Girardi (2017) used the API provided by Google cloud (2018) to transcribe speech recordings of people with PD. In both studies the authors consider the original texts with the speech of the patients and compared them with respect to the transcriptions automatically obtained from the speech recognizer. Based on those comparisons it is possible to measure several objective intelligibility measures like word error rate, word correctness, word accuracy, goodness of pronunciation, and others.

Most of the works reviewed in the literature have tried to predict/evaluate the disease severity according to the MDS-UPDRS scale; however, this is a general tool which was designed to evaluate general motor impairments in PD patients, but not to assess specific impairments of patients. This paper provides a methodology for the automatic assessment of speech impairments of PD patients according to a modified version of the FDA scale, namely m-FDA. This version can be administered based on speech recordings, hence the patient does not need to visit the examiner. The proposed approach covers four dimensions of speech: phonation, articulation, prosody, and intelligibility. The automatic evaluations to predict the m-FDA score are performed with features that model those speech dimensions. The results indicate that articulation features are the most accurate to estimate the dysarthria level of PD patients.

2. Methods

2.1. Subjects

An extended version of the PC-GITA database (Orozco-Arroyave, Arias-Londoño, Vargas-Bonilla, Gonzalez-Rátiva, & Nöth, 2014) is considered in this study. The data contain speech recordings of 68 PD patients (33 female) and 50 HC subjects (25 female). The mean age of the patients at the moment of the first recording session was 61.7 years (SD = 9.4) and the mean age of the control subjects was 61.0 years (SD = 9.1). Both groups are balanced in age [$t(99) = 0.39, p = 0.69$] and gender [$\chi^2(1) = 60.0, p = 0.99$]. At the time of the first recording session, the patients' mean years post-diagnosis was 10.3 (SD = 9.5). All of them are Colombian Spanish native speakers. The HC speakers were recorded once, while thirty-three of the patients were recorded in several sessions between 2012 and 2016. Most of the patients were recorded in two or three sessions and there was a set with six patients who were recorded in five sessions. All of the participants performed several speech tasks including sustained phonation of the five Spanish vowels, reading of ten isolated sentences, one read text, a monologue, and six diadochokinetic (DDK) exercises including the rapid repetition of the syllables /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/. Appendix A summarizes the speech exercises that the patients performed in this study. Further details of the recording protocol can be found in (Orozco-Arroyave et al., 2014) and details of the demographic and clinical information of the participants are provided in Table 1 . All participants gave written informed consent. The study was carried out in accordance with the Declaration of Helsinki and it was approved by the Ethical Research Committee of Antioquia University's Faculty of Medicine.

2.1.1. The modified Frenchay Dysarthria Assessment (m-FDA)

Existing neurological scales aim to assess general motor impairments of PD patients; however, as it was shown by Nöth et al. (2016), the deterioration of the communication abilities suffered from PD patients is not properly evaluated in such scales. In order to

Table 1
Demographic and clinical information of the participants. The m-FDA scores per session (S1, S2, S3, S4, S5) are included. * Indicates patients recorded in the five sessions. ID: consecutive number assigned to each participant, G: gender, TD: time since diagnosis [years], calculated from the first recording session.

ID	G	Age	TD	m-FDA					ID	G	Age	TD	m-FDA					ID	G	Age	m-FDA	ID	G	Age	m-FDA
				S1	**S2**	**S3**	**S4**	**S5**					**S1**	**S2**	**S3**	**S4**	**S5**								
PD01	M	33	9	25	–	–	–	–	PD36	F	42	38	–	–	–	–	29	HC01	M	31	12	HC26	F	49	3
PD02	M	45	7	23	–	–	–	–	PD37	F	49	16	13	–	31	–	–	HC02	M	42	0	HC27	F	50	0
PD03	M	47	2	40	–	35	–	–	PD38*	F	51	41	15	21	10	14	20	HC03	M	42	0	HC28	F	50	6
PD04	M	48	12	35	–	–	–	–	PD39	F	51	10	45	–	–	–	–	HC04	M	50	0	HC29	F	55	3
PD05	M	48	15	–	24	–	–	16	PD40	F	53	1	–	28	–	–	–	HC05	M	51	4	HC30	F	55	15
PD06	M	50	7	24	–	–	–	–	PD41	F	54	7	23	22	14	–	–	HC06	M	54	17	HC31	F	57	2
PD07	M	50	17	39	–	–	–	–	PD42	F	55	12	29	24	–	–	–	HC07	M	55	5	HC32	F	57	2
PD08	M	54	4	35	–	–	–	–	PD43*	F	55	12	13	12	17	24	19	HC08	M	55	19	HC33	F	58	12
PD09	M	56	14	25	–	–	–	–	PD44	F	55	12	29	–	–	–	–	HC09	M	56	15	HC34	F	60	0
PD10	M	57	0.4	13	–	–	–	–	PD45*	F	55	43	23	28	21	23	21	HC10	M	61	23	HC35	F	61	13
PD11*	M	59	8	23	37	19	19	22	PD46	F	56	4	–	–	–	–	17	HC11	M	62	4	HC36	F	61	3
PD12	M	60	10	37	–	–	–	–	PD47*	F	57	37	29	35	27	26	34	HC12	M	62	0	HC37	F	61	3
PD13	M	60	8	–	–	24	24	22	PD48	F	57	17	39	26	–	–	–	HC13	M	63	23	HC38	F	62	5
PD14	M	61	10	–	29	–	–	–	PD49	F	58	1	29	38	–	–	–	HC14	M	63	13	HC39	F	62	3
PD15	M	64	3	31	15	21	19		PD50	F	59	14	25	27	24	–	–	HC15	M	64	6	HC40	F	63	2
PD16	M	64	3	25	–	–	–	–	PD51	F	59	17	29	–	–	–	–	HC16	M	65	5	HC41	F	63	3
PD17	M	64	0.6	–	–	19	37	25	PD52	F	60	7	33		–	–	–	HC17	M	67	2	HC42	F	63	7
PD18	M	65	12	38	32	38			PD53	F	61	4	31	22	–	–	–	HC18	M	67	10	HC43	F	63	13
PD19	M	65	19	33	–	–	–	–	PD54	F	61	1.5	–	22	–	23	–	HC19	M	67	9	HC44	F	64	6
PD20	M	65	10	–	–	20	–	24	PD55	F	62	12	29	24	–	–	–	HC20	M	67	17	HC45	F	65	13
PD21	M	66	8	–	20	14	–	–	PD56	F	64	3	21	–	–	25	–	HC21	M	68	15	HC46	F	65	2
PD22	M	67	4	21	–	–	–	–	PD57	F	65	8	27	–	–	–	–	HC22	M	68	7	HC47	F	68	15
PD23	M	67	5	–	33	31	38	31	PD58	F	65	5	–	36	–	–	–	HC23	M	71	0	HC48	F	73	24
PD24*	M	68	1	13	23	23	24	28	PD59	F	66	4	27	24	–	–	–	HC24	M	76	23	HC49	F	75	3
PD25	M	68	20	41	–	–	–	–	PD60	F	66	4	35	–	–	–	–	HC25	M	86	13	HC50	F	76	25
PD26	M	68	8	17	–	–	–	–	PD61	F	69	12	23	–	–	–	–								
PD27	M	69	5	32	–	–	–	–	PD62	F	70	12	37	40	–	–	–								
PD28	M	70	3	–	19	22	15	21	PD63	F	70	3	–	–	–	–	23								
PD29	M	70	7	–	42	–	–	–	PD64	F	71	0.5	–	37	32	32	–								
PD30	M	71	11	41	–	–	–	–	PD65	F	72	2.5	32	32	–	–	–								
PD31	M	74	12	27	–	–	–	–	PD66	F	73	4	25	17	–	–	–								
PD32	M	75	1	29	–	–	–	–	PD67	F	75	3	47	–	–	–	–								
PD33	M	75	16	38	–	–	–	–	PD68	F	75	14	–	–	–	37	37								
PD34	M	78	8	–	–	–	23	–																	
PD35	M	81	12	25	25	–	–	–																	

help clinicians, speech and language therapists, patients, and care givers to assess and monitor the communication abilities of PD patients, the first aim of this study is to introduce the m-FDA scale. The original version of the FDA considers several factors that are affected in people with dysarthria, such as reflexes, respiration, lips movement, palate movement, laryngeal capability, tongue posture/movement, intelligibility, and others. Although this tool covers a wide range of speech aspects, it requires the patient to visit the examiner. This may be difficult in many cases, because of PD patients' reduced mobility. In addition, most of the patients enrolled in our study lived in the country-side, which meant that travel to the clinic was more difficult. In order to contribute to the solution of this problem we developed the m-FDA scale, which can be administered based on speech recordings, therefore it is not necessary to make an appointment with the patient. This scale includes several aspects of speech including respiration, lips movement, palate/velum movement, larynx, tongue, monotonicity, and intelligibility.

2.1.2. Speech tasks and aspects considered in the m-FDA scale

Different speech tasks are considered depending on the evaluated aspect. Respiratory capability (Aspect: Respiration) is evaluated with sustained phonations of vowel /a/ and DDK tasks. Strength and control of lips closing (Aspect: Lips) are evaluated with DDK tasks and the read text, respectively. Nasal escape and velar movement (Aspect: Palate/Velum) are evaluated with the read text and DDK tasks, respectively. Phonatory capability and effort to produce speech (Aspect: Laryngeal) are evaluated with the sustained vowel /a/ and the read text. Correctness and velocity in the tongue movement (Aspect: Tongue) are evaluated with DDK tasks (specially the repetition of the syllable /ta/). Finally, intelligibility and monotonicity are evaluated with the read text. The m-FDA scale has a total of 13 items, each of them ranges from 0 (normal or completely healthy) to 4 (very impaired). Thus the total score of the scale ranges from 0 to 52. Details of the speech exercises and items included in the evaluation performed to the participants are included in Appendix A.

2.1.3. Scoring process

The process of scoring the speech recordings was performed by three phoniatricians. At the beginning of the process they were

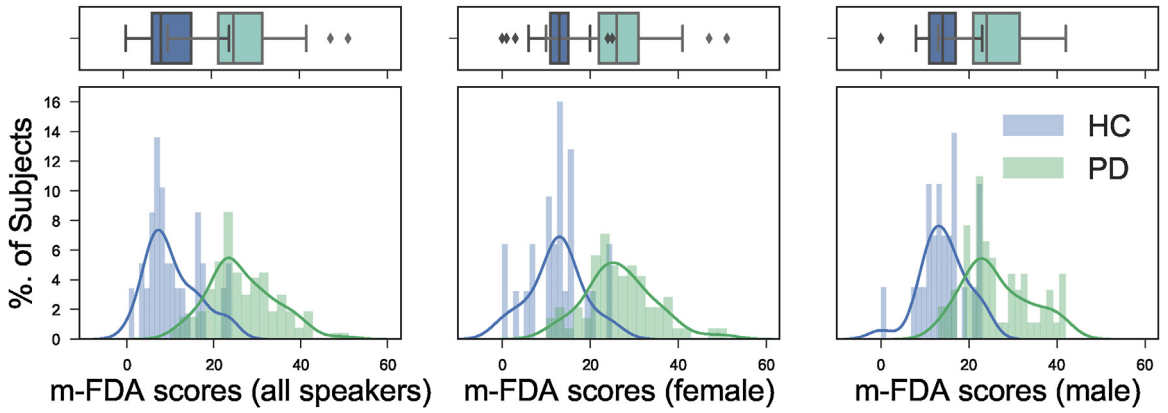

Fig. 1. Distribution of the m-FDA scores for all HC speakers and PD patients (left). m-FDA scores for female speakers (middle). m-FDA scores for male speakers (right). Significant differences between the scores of HC and PD speakers are found in all cases ($p < 0.001$).

asked to agree on the evaluations of the first ten speakers. Afterwards the experts evaluated the remaining recordings independently. The inter-rater reliability among the phoniatricians was 0.75, which was computed calculating the average Spearman's correlation between all possible pairs of raters. The median among the scores assigned by the phoniatricians was considered as the score per speaker. Fig. 1 displays the distribution of the scores assigned by the phoniatricians to all of the speakers (left), female speakers (middle), and male speakers (right). The associated box-plots are also included in the same figure. The statistical difference among scores per group (PD and HC) is evaluated by an analysis of variance (ANOVA), where the independent variables were the groups and the dependent variables were the m-FDA scores. The results indicate that there is statistical difference between the scores assigned by the phoniatricians to PD vs. those assigned to HC speakers, i.e., $F = 175.49$, $p < 0.001$ for all speakers, $F = 66.81$, $p < 0.001$ for female speakers, and $F = 52.13$, $p < 0.001$ for male speakers. Table 1 includes the median values of the m-FDA scores assigned by the three phoniatricians in the different recording sessions.

2.2. Acoustic modeling

The second aim of this work is to objectively model speech signals such that the dysarthria level of PD patients, in terms of the m-FDA, is robustly evaluated. Four speech dimensions are considered to achieve this purpose: phonation, articulation, prosody, and intelligibility. The feature sets extracted to model each dimension are introduced below.

2.2.1. Phonation

Phonation features model abnormal patterns in the vocal fold vibration. These features are computed upon voiced segments (where there is vibration of the vocal folds), hence they can be extracted from sustained vowels and continuous speech signals. Typically, phonation impairments have been analyzed in terms of features related to stability measures such as: *Jitter and shimmer*, which describe the variation of the fundamental period and amplitude of the voice signal, respectively; *pitch perturbation quotient (PPQ)*, which models long-term perturbations of the fundamental period; *amplitude perturbation quotient (APQ)*, which models long-term variations in the peak-to-peak amplitude of the signal. Additionally, the first and second derivatives of the fundamental frequency (F_0) were included along with the energy content of the signal. Further details of the methods can be found in (Orozco-Arroyave et al., 2015). APQ and PPQ were computed upon 100 ms frames with a time-shift of 80 ms. The rest of phonation features were computed upon frames of 20 ms with a time-shift of 15 ms. Four statistical functionals were calculated per feature (mean, standard deviation, skewness, and kurtosis), forming a 28-dimensional feature vector per utterance.

2.2.2. Articulation

Articulatory features are designed to model changes in the position of the tongue, lips, velum, and other articulators involved in the speech production process. Different articulation phenomena can be analyzed in sustained vowels and in continuous speech signals. In this study sustained vowels were modeled with 55 descriptors such as: *Formant frequencies* to represent resonances in the vocal tract and the capability of the speaker to hold the tongue in a certain position while producing a sustained phonation. The first two formants (F_1 and F_2) and their first two derivatives were also considered; the *Teager energy operator* was calculated to model non-linear effects that appear during the speech production. Besides, twelve Mel frequency cepstral coefficients (*MFCCs*) and their corresponding first and second derivatives were computed as a smooth representation of the voice spectrum that considers the human auditory perception (Godino-Llorente, Gómez-Vilda, & Blanco-Velasco, 2006). We are aware of the fact that MFCCs convey perceptual information of the speech signals; however, the computation of these coefficients is quantitative and completely reproducible, which clearly contributes to the aims of this study. Articulation in sustained vowels is modeled by 220-dimensional feature vectors which are formed with four statistical functionals calculated upon the introduced articulation measures.

The articulation model in continuous speech is performed considering features that represent the difficulties exhibited by PD

patients to start/stop the movement of the vocal folds (Orozco-Arroyave, 2016). The model was based on computing the energy content in the transition from unvoiced to voiced segments (onset), and from voiced to unvoiced (offset). The detection of voiced/unvoiced segments was based on the computation of F_0 which was performed using the software Praat (Boersma, 2002). Those segments where F_0 was found are considered as voiced, conversely segments where no F_0 was found were labeled as unvoiced (Orozco-Arroyave, Hönig, et al., 2016). Once the borders are detected, 40 ms of the signal were taken to the left and to the right of such borders. The spectrum of the transitions was distributed into 22 critical bands according to the Bark scale, and the Bark-band energies were calculated upon the onsets and offsets. A 384-dimensional feature vector per utterance was formed to represent articulation in continuous speech signals.

2.2.3. Prosody

This speech dimension is included to model timing, intonation, and loudness during the production of natural speech. Prosody has been typically evaluated with measures derived from the F_0 and energy contours. The prosody-based features considered here were computed with the Erlangen prosody module (Zeißler et al., 2006). In this case voiced segments were used as the speech unit. The set of features comprises a total of 95 measures, 19 of them were based on duration and include the number and length of voiced frames, duration of pauses, and others. Another 36 features were based on the F_0 contour and included mean values, standard deviation, and jitter. The energy-based features included measures of the energy within voiced frames, shimmer, and the position of the maximum energy. The features were grouped into one feature vector and the four functionals were also computed to form 380-dimensional feature vectors per utterance. Further details of the extracted features to model prosody can be found in Haderlein (2007).

2.2.4. Intelligibility

This dimension is related to the capability of a person to be understood by another person or by a computer. In general, one patient with PD is less intelligible than a healthy subject (De Letter et al., 2005; Miller et al., 2007). Although extensively reported, impairments in speech intelligibility of PD patients have been analyzed through perceived intelligibility. This approach is expensive, time consuming and highly subjective, making it difficult to be reproduced. In this paper we performed the intelligibility analysis based on the performance of an external automatic speech recognizer (ASR), i.e., we used the commercial ASR system provided in Google cloud (2018).

Two features were calculated: word accuracy (WA) and a similarity measure based on the dynamic time warping (DTW) distance computed between the recognized and the original utterances. These measures were introduced by Orozco-Arroyave, Vásquez-Correa, et al. (2016) and Vásquez-Correa, Orozco-Arroyave, and Nöth (2016) to model intelligibility deficits of PD patients.

The WA was defined as the number of words correctly recognized by the system relative to the total of words in the original string. It is computed using Eq. (1).

$$\mathrm{WA} = \frac{\#\,\text{words correctly recognized}}{\#\,\text{of total words}} \tag{1}$$

The similarity measure is based on the DTW distance which was originally introduced to analyze differences between two time-series that differ in time and number of samples. The time-warping procedure performs a time-alignment between the two sequences and the distance is computed between the predicted string and the original sentence at a grapheme level. In this case the distance is transformed into a similarity score by using Eq. (2). If original and recognized sequences are the same, the DTW_distance will be zero, and the similarity will be 1. Conversely, when the strings are very different, the distance will be high, and the similarity will be close to zero.

$$\mathrm{similarity} = \frac{1}{1 + \mathrm{DTW\_distance}} \tag{2}$$

2.2.5. Speaker model based on i-vectors

An additional approach was considered to model speaker traits related with the dysarthria level of the patients. The i-vector (Dehak, Kenny, Dehak, Dumouchel, & Ouellet, 2011) model was used to extract utterance-dependent fixed-length vectors. i-vectors are low-dimensional representations that contain information related to different traits of a speaker. An i-vector contains information such as the identity of the subject (Dehak, Kenny, et al., 2011), the spoken language (Dehak, Torres-Carrasquillo, Reynolds, & Dehak, 2011), gender, age, and others. We hypothesize that i-vectors may contain also information related to speech impairments like those developed by PD patients (Garcia, Orozco-Arroyave, D'Haro, Dehak, & Nöth, 2017).

A total of 512-dimensional Gaussian Mixture Models were employed to extract sufficient statistics per utterance. We used these statistics to model 200-dimensional i-vectors (Dehak, Kenny, et al., 2011). This can be regarded as an extension to our previous work, where we based the analysis on GMM-super vectors in combination with support vector machines for classification and regression (Bocklet, Nöth, Stemmer, Ruzickova, & Rusz, 2011; Bocklet, Steidl, Nöth, & Skodda, 2013). First, an utterance specific i-vector was created. Due to the fact that there are multiple speaking/reading tasks per patient, we also extracted multiple i-vectors per speaker. We then either used utterance specific i-vectors per speaker or we averaged the i-vectors per speaker in order to create speaker-specific i-vectors. The combination on i-vector level achieved significantly better results than extracting one i-vector per speaker by combining features on MFCC-level. We were evaluating different kinds of inter-session variability; our intention was to find a speaker modeling approach that purely focuses on preserving information about the dysarthria level of the patient.

2.3. Prediction of the dysarthria level

The third aim of this paper is to introduce a methodology to objectively quantify and predict the dysarthria level of patients with PD. The regression approach consists of predicting the m-FDA score of a given speech recording based on information/features extracted from recordings previously evaluated by phoniatry experts according to the m-FDA scale. The extracted features are used to form a matrix $\mathbf{X}$ where the number of rows corresponds to the number of speakers and the number of columns corresponds to the number of features, i.e., $\mathbf{X} \in \mathbb{R}^{N \times d}$. The simplest linear model to predict the dysarthria level (let's name it $\hat{\mathbf{y}}$) can be expressed according to Eq. (3), where $\boldsymbol{\omega}$ is a weight vector that represents the slope of the linear predictions, and b is the bias term.

$$\hat{\mathbf{y}} = \boldsymbol{\omega} \cdot \mathbf{X}^T + b \tag{3}$$

It is not always possible to find accurate linear predictions. Thus a nonlinear function $\Phi(\mathbf{X})$ of the features is considered with the aim of mapping the feature matrix into a higher dimensional space, where a more accurate prediction can be performed. The function $\Phi(\mathbf{X})$ is called kernel. The predicted value of the score considering kernel functions is expressed by Eq. (4).

$$\hat{\mathbf{y}} = \boldsymbol{\omega} \cdot \Phi(\mathbf{X}^T) + b \tag{4}$$

In this paper we considered and compared four different regression methods to predict the dysarthria level of speakers with PD. These methods are briefly described below, if the reader wants to look into details of the techniques, we suggest to review the references.

2.3.1. Linear Regression with Regularization (LASSO)

LASSO is a linear model that estimates sparse coefficients that represent the weight vector $\boldsymbol{\omega}$. This method requires less hyper-parameters to be optimized than other methods, which reduces the complexity of the solution. The objective function to be minimized is expressed by Eq. (5), where α is the regularization hyper-parameter that penalizes the L_1-norm of the weight coefficients $||\boldsymbol{\omega}||_1$. The term $||\boldsymbol{\omega} \cdot \mathbf{X}^T - \mathbf{y}||_2^2$ represents the prediction error between real and predicted dysarthria levels.

$$\arg\min_{\omega} \frac{1}{2N} ||\boldsymbol{\omega} \cdot \mathbf{X}^T - \mathbf{y}||_2^2 + \alpha ||\boldsymbol{\omega}||_1 \tag{5}$$

2.3.2. Support vector regression (SVR)

Support vector regression allows to predict a real variable, like the m-FDA score, using a loss function $L(\mathbf{y}, \hat{\mathbf{y}})$ which has an insensitive parameter ε related to the maximum error allowed in the prediction. The loss function is expressed by Eq. (6).

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \begin{cases} 0 & \text{if } |\mathbf{y} - \hat{\mathbf{y}}| \leq \varepsilon \\ |\mathbf{y} - \hat{\mathbf{y}}| - \varepsilon & \text{in other cases.} \end{cases} \tag{6}$$

The effect of the loss function is observed in Fig. 2, where errors smaller than ε are ignored. The error is described only in terms of the variable ξ to measure the deviation of training samples outside the ε-insensitive zone (the zone between the dotted lines). In this study both linear and non-linear SVRs were considered. A Gaussian kernel (rbf-SVR) was considered for the non-linear regression (Schölkopf & Smola, 2001).

2.3.3. Linear ridge regression (LRR) and kernel ridge regression (KRR)

LRR was used to solve the optimization problem expressed in Eq. (7), where α is the hyper-parameter that controls the shrinkage of the weights.

$$\arg\min_{\omega} ||\boldsymbol{\omega} \cdot \mathbf{X}^T - \mathbf{y}||_2^2 + \alpha ||\boldsymbol{\omega}||_2 \tag{7}$$

For the kernel ridge regression method the weights $\boldsymbol{\omega}$ were also optimized with an L_2-norm criterion. A Gaussian kernel with bandwidth γ was used.

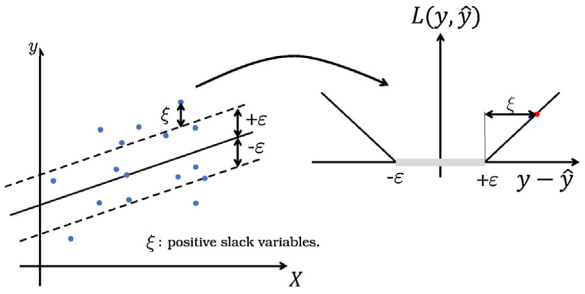

Fig. 2. Effect of the loss function $L(y, \hat{y})$ in the SVR.

2.3.4. Bayesian ridge regression (BRR)

This is a modified version of the linear regression, where additional information related to the prior probability distribution of the weights $\boldsymbol{\omega}$ is considered. Such a distribution is assumed to be a zero-mean Gaussian with covariance matrix $\boldsymbol{\lambda}$. When optimizing the weights $\boldsymbol{\omega}$, the optimal regularization hyper-parameters were also found. The predicted scores $\hat{\mathbf{y}}$ were assumed to follow a Gaussian distribution with mean $\boldsymbol{\omega} \cdot \mathbf{X}^T$.

$$p(\hat{\mathbf{y}}|\mathbf{X}, \omega, \alpha) = \mathcal{N}(\hat{\mathbf{y}}|\omega \cdot \mathbf{X}^T, \alpha) \tag{8}$$

2.3.5. Optimization of the hyper-parameters

As it was explained above, the aforementioned regressors have hyper-parameters that need to be optimized in order to train the system to predict the scores that better represent the dysarthria level of the speakers, i.e., the m-FDA scores. In this paper those hyper-parameters are optimized following a 10-fold cross-validation strategy (speaker independent), i.e., 90% of the speakers are considered to find the optimal parameters of the regressors, and the remaining 10% are used to test the accuracy of the model. This procedure is repeated ten times to assure that all of the speakers are tested. The speaker independence guarantees that utterances of one speaker are not included in the train and test sets simultaneously. Thus the regression models are not biased by speaker-specific characteristics. The performance of each regressor is evaluated in terms of the Spearman's correlation coefficient (ρ).

2.4. Analysis

This subsection briefly outlines how the methods considered in this study contribute to achieve its aims. The first aim is to introduce the m-FDA scale to assess the dysarthria level of PD patients. The m-FDA scale is administered considering only speech recordings which avoids the necessity of the patient to visit the clinician and attend a medical appointment. The reliability of the scale to differentiate PD patients and HC subjects is evaluated with an ANOVA test, which guarantees that the scores assigned to PD patients are significantly different than those assigned to the HC subjects. The second aim is to perform the automatic and objective modeling of dysarthric speech signals considering several speech dimensions including phonation, articulation, prosody, and intelligibility. The third aim of this study considers the automatic estimation of the dysarthria level of PD patients, according to the m-FDA scale, using speech processing and machine learning methods. This estimation also includes several analyses to know which speech dimensions are more affected due to the disease and which correlate the most with the proposed scale. The extracted features and several state-of-art regression algorithms are used to automatically predict the scores of the m-FDA scale. Each regressor offers specific advantages that may help improving results of the automatic prediction of the m-FDA scores. For instance, LASSO and LRR regressors provide linear predictions without any further transformation of the original feature space, which may help in the interpretation of the predictions. In addition, they provide less expensive computationally solutions. On the other hand, the kernel-based methods such as KRR and SVR may provide more accurate results, even though they are more complex because the feature space is transformed into a higher dimensional space where linear solutions can be obtained. Besides the feature extraction and regression approaches, we introduced the use of speaker models based on i-vectors, which describe specific traits of patients' speech in a low-dimensional representation. We validated the fact that i-vectors extracted from the speech of PD patients also convey information related to the dysarthria level of the speakers.

3. Experiments and results

3.1. Prediction of the total score of the m-FDA scale

The total score of the m-FDA scale was predicted considering features extracted per each dimension and also with their combination. All of the regression techniques introduced in Section 2.3 were tested.

3.1.1. Prediction using phonation features

The results of predicting the m-FDA scores with the phonation features are shown in Table 2. The Spearman's correlation coefficient (ρ) was computed between the real m-FDA scores and the predicted values. The rows in Table 2 indicate that there were no great differences in the correlation coefficient of different regressors per task; however, when comparing different speech tasks, large differences are observed, for instance the results with sentences are below those obtained with the DDK tasks where correlations of up to 0.56 are observed.

3.1.2. Prediction using articulation features

The results obtained using the articulation features are shown in Table 3. A Spearman's correlation coefficient of 0.63 was obtained with the rbf-SVR regressor. The evaluation of DDK tasks exhibited correlations between 0.43 and 0.57 and the combination of the seven DDK tasks improved the results up to 0.63. The highest correlation is obtained with the rbf-SVR regressor (0.63) which indicates that this method is the most appropriate and accurate to predict the dysarthria level based on the articulation features. Since the combination of all of the speech tasks also exhibited a correlation coefficient of 0.63, it indicates that maybe in clinical applications it is enough to include only DDK tasks, a monologue and a read text.

Table 2
Correlations between original and predicted m-FDA scores using phonation features. Sentences: indicates the combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: indicates the combination of the six DDK exercises. Average*: indicates the average computed over all of the speech tasks per regression technique.

Task	LASSO	Linear-SVR	rbf-SVR	LRR	KRR	BRR
Vowel /a/	0.31	0.37	0.33	0.35	0.20	0.32
Sentence 1	0.01	0.26	0.11	0.24	0.15	0.12
Sentence 2	0.25	0.19	0.23	0.31	0.05	0.30
Sentence 3	0.12	0.46	0.24	0.16	0.13	0.08
Sentence 4	0.25	0.11	0.16	0.28	0.14	0.26
Sentence 5	0.30	0.34	0.28	0.31	0.03	0.27
Sentence 6	0.33	0.14	0.20	0.30	0.09	0.22
Sentence 7	0.34	0.24	0.34	0.40	0.35	0.35
Sentence 8	0.22	0.57	0.23	0.33	0.24	0.31
Sentence 9	0.32	0.22	0.31	0.34	0.03	0.31
Sentence 10	0.38	0.43	0.38	0.35	0.18	0.30
DDK1	0.51	0.53	0.49	0.52	0.39	0.50
DDK2	0.56	0.55	0.48	0.55	0.45	0.50
DDK3	0.40	0.40	0.38	0.46	0.50	0.41
DDK4	0.39	0.51	0.32	0.36	0.40	0.32
DDK5	0.34	0.40	0.32	0.40	0.25	0.36
DDK6	0.41	0.38	0.21	0.28	0.32	0.22
Read text	0.36	0.11	0.14	0.34	0.30	0.30
Monologue	0.29	0.10	0.34	0.44	0.28	0.45
Average*	0.32	0.33	0.29	0.35	0.24	0.31
Sentences	0.29	0.31	0.22	0.23	0.23	0.32
DDKs	0.50	0.45	0.46	0.47	0.43	0.50
Sentences + DDKs + Read text + Monologue + Vowel /a/	0.48	0.40	0.44	0.32	0.41	0.44

Table 3
Correlations between original and predicted m-FDA scores using articulation features. Sentences: indicates the combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: indicates the combination of the six DDK exercises. Average*: indicates the average computed over all of the speech tasks per regression technique.

Task	LASSO	Linear-SVR	rbf-SVR	LRR	KRR	BRR
Vowel /a/	0.37	0.40	0.44	0.46	0.50	0.40
Sentence 1	0.43	0.44	0.41	0.38	0.50	0.46
Sentence 2	0.47	0.54	0.53	0.46	0.58	0.56
Sentence 3	0.47	0.49	0.47	0.37	0.49	0.47
Sentence 4	0.44	0.44	0.56	0.54	0.58	0.54
Sentence 5	0.44	0.44	0.48	0.45	0.44	0.46
Sentence 6	0.49	0.52	0.54	0.50	0.52	0.56
Sentence 7	0.48	0.50	0.49	0.47	0.55	0.53
Sentence 8	0.42	0.46	0.45	0.44	0.46	0.46
Sentence 9	0.42	0.42	0.39	0.28	0.40	0.37
Sentence 10	0.45	0.45	0.45	0.30	0.46	0.47
DDK1	0.53	0.54	0.51	0.42	0.55	0.49
DDK2	0.57	0.50	0.54	0.47	0.57	0.57
DDK3	0.52	0.55	0.55	0.43	0.55	0.52
DDK4	0.50	0.50	0.49	0.40	0.52	0.55
DDK5	0.52	0.47	0.45	0.43	0.52	0.49
DDK6	0.54	0.49	0.53	0.44	0.53	0.50
Read text	0.50	0.51	0.55	0.43	0.50	0.52
Monologue	0.52	0.50	0.50	0.49	0.49	0.50
Average*	0.48	0.48	0.49	0.43	0.51	0.50
Sentences	0.43	0.58	0.61	0.57	0.56	0.57
DDKs	0.60	0.59	0.63	0.60	0.59	0.57
Sentences + DDKs + Read text + Monologue + Vowel /a/	0.46	0.55	0.63	0.61	0.59	0.61

3.1.3. Prediction using prosody features

The results considering the prosody features are displayed in Table 4 . The highest correlation was obtained using LASSO with the combination of the DDK exercises ($\rho = 0.51$). Note that similar correlations (0.48 and 0.49) were obtained with the rbf-SVR in the read text and in the monologue, respectively. This result indicates that prosody features are, to some extent, accurate and stable to predict the dysarthria level of PD patients. We consider that continuous speech signals like the monologue and read text are more appropriate to evaluate prosody. These features extracted from continuous speech signals allow to find interpretable results about characteristics like speech rate, timing, tempo, and pausing.

Table 4
Correlations between original and predicted m-FDA scores using prosody features. Sentences: indicates the combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: indicates the combination of the six DDK exercises. Average*: indicates the average computed over all of the speech tasks per regression technique.

Task	LASSO	Linear-SVR	rbf-SVR	LRR	KRR	BRR
Sentence 1	0.28	0.30	0.28	0.32	0.03	0.38
Sentence 2	0.31	0.28	0.33	0.28	0.23	0.37
Sentence 3	0.38	0.15	0.28	0.27	0.11	0.34
Sentence 4	0.35	0.46	0.40	0.29	0.36	0.39
Sentence 5	0.34	0.32	0.32	0.31	0.33	0.37
Sentence 6	0.37	0.31	0.36	0.24	0.25	0.37
Sentence 7	0.16	0.20	0.33	0.27	0.23	0.26
Sentence 8	0.33	0.22	0.39	0.41	0.34	0.41
Sentence 9	0.21	0.22	0.29	0.32	0.31	0.35
Sentence 10	0.23	0.08	0.16	0.18	0.16	0.21
DDK1	0.41	0.52	0.37	0.42	0.18	0.40
DDK2	0.33	0.48	0.33	0.43	0.32	0.39
DDK3	0.27	0.31	0.07	0.29	0.13	0.23
DDK4	0.11	0.27	0.38	0.37	0.13	0.29
DDK5	0.20	0.14	0.06	0.26	0.07	0.26
DDK6	0.44	0.04	0.44	0.47	0.40	0.51
Read text	0.33	0.08	0.49	0.26	0.39	0.38
Monologue	0.43	0.38	0.48	0.34	0.32	0.45
Average*	0.30	0.26	0.32	0.32	0.24	0.35
Sentences	0.13	0.16	0.46	0.34	0.48	0.35
DDKs	0.51	0.49	0.37	0.33	0.41	0.40
Sentences + DDKs + Read text + Monologue	0.24	0.13	0.49	0.35	0.51	0.36

3.1.4. Prediction using intelligibility features

The results obtained with the intelligibility features are shown in Table 5 . As it was expected, in this case the highest correlations (between 0.32 and 0.44) were obtained with speech tasks that contain utterances more complex to be pronounced like read text and sentence 10. Note also that the combination of different speech tasks improved the results. Monologues were not included in these tests because we did not have the transcriptions of the conversations, hence there was not a reference to compute the DTW distances. Although the results here were below those obtained with other feature sets, we think that it is promising and it is worth to continue working on the development of dysarthria assessment methods based on ASR techniques.

3.1.5. Prediction using the combination of feature sets

Table 6 shows the correlations obtained when phonation, articulation, prosody, and intelligibility features are combined to predict the total score of the m-FDA scale. Correlations of up to 0.64 are obtained with the rbf-SVR in the DDK exercises. Note that these results are, in most of the cases, similar to those obtained with articulation features. This result suggests that the articulation measures are the most accurate to model continuous speech signals and DDK tasks. This result also indicates that the proposed approach is suitable to support and/or validate subjective observations of the clinicians without the need to dedicate long time or expensive equipment.

The consistency and reproducibility of the methods across different phoniatricians were evaluated considering the scores of the

Table 5
Correlations between original and predicted m-FDA scores using intelligibility features. Sentences: indicates the combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: indicates the combination of the six DDK exercises. Average*: indicates the average computed over all of the speech tasks per regression technique.

Task	LASSO	Linear-SVR	rbf-SVR	LRR	KRR	BRR
Sentence 1	0.00	0.03	0.09	0.10	0.00	0.00
Sentence 2	0.25	0.24	0.26	0.28	0.16	0.26
Sentence 3	0.22	0.30	0.30	0.27	0.22	0.25
Sentence 4	0.22	0.28	0.22	0.31	0.20	0.21
Sentence 5	0.28	0.20	0.27	0.30	0.33	0.28
Sentence 6	0.31	0.31	0.32	0.31	0.21	0.31
Sentence 7	0.36	0.37	0.37	0.39	0.33	0.38
Sentence 8	0.18	0.16	0.16	0.22	0.20	0.17
Sentence 9	0.06	0.05	0.00	0.12	0.00	0.04
Sentence 10	0.36	0.38	0.37	0.39	0.32	0.37
Read text	0.41	0.37	0.36	0.44	0.40	0.40
Average*	0.24	0.24	0.25	0.28	0.22	0.24
Sentences	0.38	0.42	0.41	0.43	0.44	0.44
Sentences + Read text	0.39	0.42	0.41	0.44	0.44	0.45

Table 6
Correlations between original and predicted m-FDA scores using the combination of phonation, articulation, and prosody features. Sentences: indicates the combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: indicates the combination of the six DDK exercises. Average*: indicates the average computed over all of the speech tasks per regression technique.

Task	LASSO	Linear-SVR	rbf-SVR	LRR	KRR	BRR
Sentence 1	0.53	0.42	0.40	0.38	0.48	0.26
Sentence 2	0.47	0.55	0.51	0.43	0.54	0.34
Sentence 3	0.46	0.45	0.48	0.27	0.46	0.19
Sentence 4	0.52	0.47	0.48	0.48	0.56	0.40
Sentence 5	0.43	0.44	0.44	0.46	0.48	0.35
Sentence 6	0.50	0.48	0.52	0.42	0.52	0.36
Sentence 7	0.53	0.48	0.54	0.47	0.52	0.42
Sentence 8	0.42	0.50	0.49	0.46	0.55	0.42
Sentence 9	0.40	0.44	0.43	0.38	0.46	0.29
Sentence 10	0.43	0.46	0.46	0.40	0.49	0.25
DDK1	0.62	0.58	0.58	0.57	0.55	0.50
DDK2	0.55	0.53	0.58	0.50	0.54	0.53
DDK3	0.57	0.50	0.52	0.37	0.53	0.51
DDK4	0.54	0.50	0.49	0.51	0.55	0.42
DDK5	0.47	0.52	0.49	0.44	0.31	0.34
DDK6	0.55	0.51	0.52	0.39	0.55	0.41
Read text	0.57	0.45	0.55	0.42	0.48	0.28
Monologue	0.45	0.53	0.52	0.44	0.52	0.44
Average*	0.50	0.49	0.50	0.43	0.51	0.37
Sentences	0.36	0.56	0.57	0.58	0.61	0.58
DDKs	0.64	0.55	0.62	0.57	0.58	0.57
Sentences + DDKs + Read text + Monologue	0.40	0.56	0.58	0.58	0.54	0.58

total m-FDA scale assigned by each phoniatrician. The highest correlations in the previous experiments were systematically obtained with LASSO and the rbf-SVR regressors. Thus we decided to choose the rbf-SVR to perform these experiments. The speech tasks included in this experiment were monologue, read text, and repetition of /pa-ta-ka/ (DDK1). The results are displayed in Table 7 . Note that the results are relatively stable among phoniatricians. The highest correlations were obtained with phoniatrician 1, who was the most experienced from the group. Conversely, the lowest correlations were obtained with phoniatrician 3, who was the least experienced. It is interesting to see also that correlations between 0.62 and 0.72 were obtained with the repetition of /pa-ta-ka/. This behavior is similar to what was observed in the experiments with the total score of the m-FDA, where the most accurate predictions were obtained mostly with DDKs.

3.2. Prediction of individual aspects of the m-FDA scale

Besides the prediction of the total score of the m-FDA scale, we wanted to analyze the performance of the proposed approach to predict sub-scores of the total scale. These sub-scores include aspects related with phonation, articulation, and prosody. The distribution of the sub-scores is depicted in Fig. 3. Phonation features were considered to model aspects related with problems in the larynx and in the respiration capability of the patients (items 1, 2 and 7). Articulation features were designed to model problems to start/stop the vocal fold vibration in continuous speech along with abnormal movements of lips, velum, and tongue (items 3, 4, 5, 6, 8, 9, 10, and 11). Finally, prosody features were considered to model monotonicity and intelligibility aspects (items 12 and 13).

The results are displayed in Table 8 . LASSO and rbf-SVR regressors were used in this case. As there is little prosodic variation in sustained phonations, no results were reported for Vowel /a/ and for the combination of all speech tasks. Highest correlations were obtained with articulation features (between 0.38 and 0.53), which is consistent with the results reported above. Among the results with articulation, the DDKs were the highest (0.53), which is also consistent with the previous experiments. These results confirmed previous observations reported several years ago in (Logemann et al., 1978; Logemann & Fisher, 1981) where the authors showed that although the speech impairments developed by Parkinson's patients include deficits in phonation, articulation, prosody, and intelligibility, the most compelling are related with articulation. Monotonicity, which is another impairment observed in PD patients, was better modeled by prosody features in the read texts, which is also consistent with previous findings reported in the literature (Orozco-Arroyave, 2016).

Table 7
Correlations between predicted and original m-FDA scores assigned by each phoniatrician (inter rater reliability: 0.75).

	Phoniatrician 1	Phoniatrician 2	Phoniatrician 3	Median
Monologue	0.43	0.39	0.28	0.35
Read text	0.58	0.42	0.47	0.52
DDK1	0.72	0.62	0.62	0.67

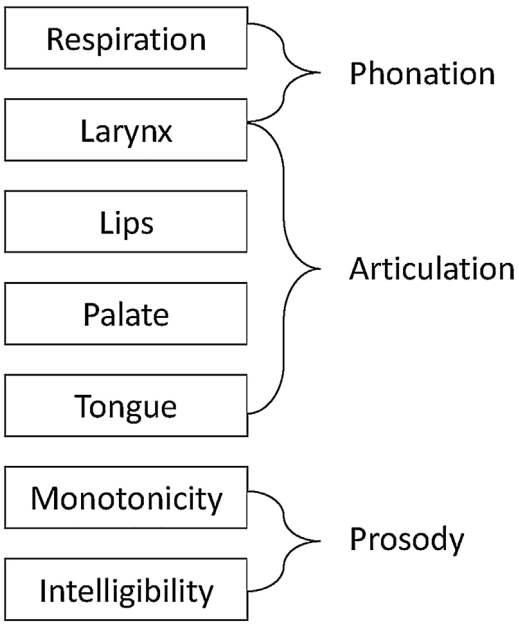

Fig. 3. Separation of the aspects in the m-FDA scale in terms of phonation, articulation, and prosody dimensions.

Table 8
Correlations between predicted and original m-FDA sub-scores according to the speech aspects depicted in Fig. 3. Sentences: combination of all sentences in the recording protocol, DDK1 to DDK6: repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/, respectively. DDKs: combination of the six DDK exercises.

	Phonation		Articulation		Prosody	
Task	LASSO	rbf-SVR	LASSO	rbf-SVR	LASSO	rbf-SVR
Vowel /a/	0.45	0.40	0.43	0.28	–	–
DDK1	0.49	0.36	0.49	0.39	0.32	0.03
DDK2	0.28	0.27	0.40	0.45	0.18	0.02
DDK3	0.06	0.25	0.53	0.42	0.18	0.03
DDK4	0.05	0.00	0.46	0.52	0.26	0.07
DDK5	0.27	0.05	0.50	0.43	0.11	0.19
DDK6	0.21	0.20	0.49	0.48	0.24	0.01
Read text	0.21	0.11	0.50	0.51	0.42	0.38
Monologue	0.17	0.08	0.38	0.48	0.22	0.16
DDKs	0.39	0.31	0.56	0.42	0.32	0.05
Sentences + DDK + Read text + Monologue + vowel /a/	0.49	0.27	0.53	0.40	–	–

3.3. Analysis using i-vectors

Results of predicting the total score of the m-FDA scale using i-vectors are shown in Table 9 . The analysis was performed for male and female speakers separately. Note that the correlation was higher for male speakers in the monologues, but higher for female speakers in read texts. Further research is required to understand the differences of the impact of PD in the speech of female and male subjects.

Table 9
Correlations between the predicted and original m-FDA scores using i-vectors. M&F: prediction for male and female speakers. M: prediction for male speakers. F: prediction for female speakers.

Speech task	M&F	M	F
All sentences	0.69	0.55	0.75
Read text	0.56	0.46	0.71
Monologue	0.69	0.68	0.58

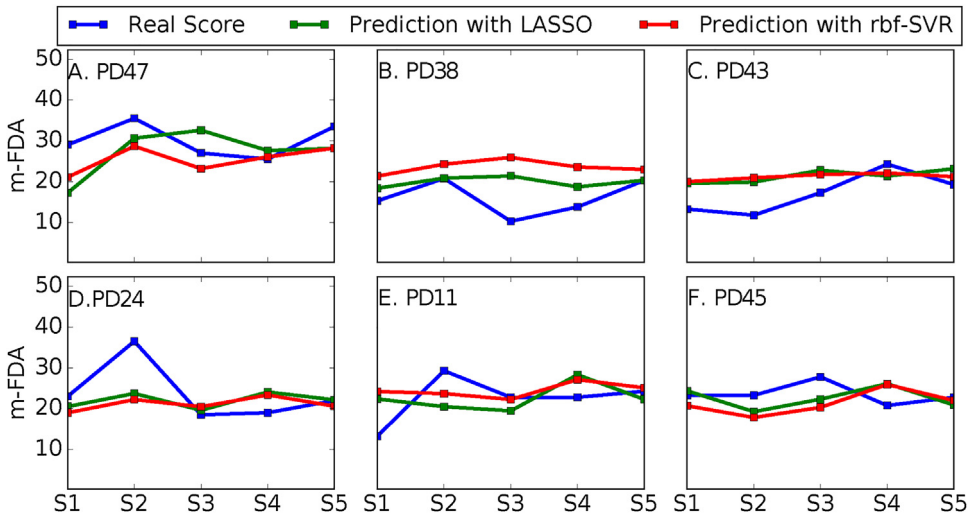

Fig. 4. Prediction of the m-FDA scores for the six patients recorded in the five sessions between 2012 and 2016.

3.4. Longitudinal analysis

Six of the 68 patients were recorded in five sessions collected between 2012 and 2016. A longitudinal analysis was performed with the recordings of those patients with the aim of evaluating the accuracy of the proposed approach to monitor the progression of speech impairments in PD patients. Table 1 includes detailed information of the patients considered for the longitudinal analysis. The combination of phonation, articulation, and prosody features was used for these experiments. Features were computed upon utterances of the six DDK exercises because these speech tasks provided the most accurate predictions in most of the previous experiments. The total score of the m-FDA scale is predicted with the rbf-SVR and LASSO regressors. Fig. 4 displays the prediction of the m-FDA scores. The curves indicate that it is possible to monitor the dysarthria level of PD patients by modeling DDK tasks with phonation, articulation, and prosody features.

We think that these results can be improved by using speaker-specific models like those based on GMM-UBM or i-vectors. We recently proposed a model to predict MDS-UPDRS-III using GMM-UBM models (Arias-Vergara, Vásquez-Correa, Orozco-Arroyave, & Nöth, 2018; Arias-Vergara, Vásquez-Correa, Orozco-Arroyave, Vargas-Bonilla, & Nöth, 2016). As the results were promising, we continued recording patients longitudinally and for future work we will consider such approaches to model the progression of speech impairments in patients with PD.

4. Conclusion

A modified version of the Frenchay Dysarthria Assessment scale (m-FDA) was introduced in this study as a specific and sensitive tool to assess speech impairments of people with PD. The scale is distributed into seven aspects (respiration, lips, velum, larynx, tongue, intelligibility, and monotonicity) which are evaluated in thirteen items. The evaluation protocol consists of different speech tasks including sustained phonation of vowel /a/, diadochokinetic exercises (i.e., repetition of syllables like /pa-ta-ka/), reading of a text with 36 words, and a monologue. The m-FDA scale can be administered based on speech recordings, i.e., the patient does not need to visit the clinician.

This paper evaluated the m-FDA scale to quantify the dysarthria level of PD patients. Additionally, a methodology to automatically predict the dysarthria level of PD patients considering speech recordings and the resulting m-FDA scores was proposed. Features of phonation, articulation, prosody, and intelligibility are extracted to model different aspects considered in the proposed scale. Articulation features were the most accurate among the evaluated speech dimensions, with Spearman's correlations of up to 0.63 between the real m-FDA scores (assigned by the clinician) and the predicted ones. The comparison among different speech tasks indicated that the DDK exercises and read texts are the most accurate to predict and monitor the dysarthria level of PD patients. The highest correlations among all of the experiments were obtained with the i-vectors approach ($\rho = 0.69$). It seems like the low-dimensional speaker representation provided by this approach gives reliable information related to the dysarthria level of the speakers.

The results reported here suggest that the proposed scale and the i-vectors approach should be considered in clinical practice as a tool to quantitatively evaluate and monitor the dysarthria level of PD patients. Further experiments are necessary to evaluate the suitability of the proposed approach in dysarthric speech signals with other origins like Huntington's disease and Ataxia.

Besides the prediction of the m-FDA scores, longitudinal analyses were performed in a subset of speakers who participated in five recording sessions collected between 2012 and 2016. The results indicated that it is possible to monitor the dysarthria level of PD patients over time; however, it is necessary to include more speakers to obtain more conclusive results. The outcomes of this study motivate us to continue working on these topics to develop robust and accurate methods for the automatic and unobtrusive

monitoring of speech impairments of PD patients.

Funding

The work reported here was partially carried out during the 2016 Jelinek Memorial Summer Workshop on Speech and Language Technologies, which was supported by Johns Hopkins University via DARPA LORELEI Contract No HR0011-15-2-0027, and gifts from Microsoft, Amazon, Google, and Facebook. This work was partially supported by CODI from University of Antioquia (grants # PRG2015-7683 and PRV16-2-01). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 766287.

Acknowledgements

Thanks to the Center of language and speech processing (CLSP) from Johns Hopkins University, to Fundalianza Parkinson Colombia, and to the phoniatricians who scored the data.

Appendix A. Description of the speech exercises performed by the patients in this study.

Exercise	Description
Vowel /a/	Sustained phonation of the vowel /a/.
Sentence 1	Read aloud the sentence: Mi casa tiene tres cuartos. *Translation: My house has three rooms.*
Sentence 2	Read aloud the sentence: Omar, que vive cerca, trajo miel. *Translation: Omar, who lives near, brought honey.*
Sentence 3	Read the aloud sentence: Laura sube al tren que pasa. *Translation: Laura gets on the passing train.*
Sentence 4	Read aloud the sentence: Los libros nuevos no caben en la mesa de la oficina. *Translation: The new books do not fit in the office's table.*
Sentence 5	Read the aloud sentence: Rosita Niño, que pinta bien, donó sus cuadros ayer. *Translation: Rosita Niño, who paints well, donated her paintings yesterday.*
Sentence 6	Read aloud the sentence: Luisa Rey compra el colchón duro que tanto le gusta. *Translation: Luisa Rey buys the hard mattress that is so fond her.*
Sentence 7	Read aloud the sentence: Viste las noticias? Yo vi ganar la medalla de plata en pesas. Ese muchacho tiene mucha fuerza! *Translation: Did you see the news? I saw to win the silver medal in Weightlifting. That boy is very strong!*
Sentence 8	Read aloud the sentence: Juan se rompió una pierna cuando iba en la moto. *Translation: Juan broke his leg when he was driving his motorcycle.*
Sentence 9	read aloud sentence: Estoy muy triste, ayer vi morir a un amigo. *Translation: I am very sad, yesterday I saw a friend die.*
Sentence 10	Read aloud the sentence: Estoy muy preocupado, cada vez me es más dificil hablar. *Translation: I am very concerned, it is increasingly more difficult to talk.*
DDK1	rapid repetition of the syllables /pa-ta-ka/.
DDK2	rapid repetition of the syllables /pa-ka-ta/.
DDK3	rapid repetition of the syllables /pe-ta-ka/.
DDK4	rapid repetition of the syllables /pa/.
DDK5	rapid repetition of the syllables /ta/.
DDK6	rapid repetition of the syllables /ka/.

Read text	This task consists of reading a dialog between a doctor (D) and a patient (P). This text is phonetically balanced and contains almost all of the Colombian Spanish sounds. The dialog is as follows: P: Ayer fui al médico. D: Qué le pasa? Me preguntó. P: Yo le dije: Ay doctor! Donde pongo el dedo me duele. D: Tiene la uña rota? P: Sí. D: Pues ya sabemos qué es. Deje su cheque a la salida. *Translation* P: *Yesterday I went to the doctor.* D: *What happened to you? He asked me.* P: *I told him: ah doctor! Where I put my finger it pains me.* D: *Do you have the nail broken?* P: *Yes.* D: *Then we now know what is happening. Leave your check at the exit.*
Monologue	The patients were asked to speak about what they commonly do in a normal day, i.e., at what time they wake up, what kind of activities they do during the day, etc.

Appendix B. List of aspects, items, and speech exercises included in the m-FDA evaluations.

Aspect	Exercises and items
Respiration	1) Sustained vowel /a/ to assess the respiratory capability 2) DDK evaluations to assess the respiratory capability.
Lips	3) DDK evaluations to assess the strength of closing the lips. 4) Read text and monologue to assess general capability to control the lips.
Palate/Velum	5) Read text and monologue to assess nasal escape. 6) DDK evaluations to assess the velar movement.
Laryngeal	7) Sustained vowel /a/ to assess the phonatory capability. 8) Read text and monologue to assess the phonatory capability in continuous speech. 9) Read text and monologue to assess the effort to produce speech.
Tongue	10) DDK evaluations to assess the correctness and velocity to move the tongue. 11) Repetition of the syllable /ta/ to assess the correctness and velocity to move the tongue.
Intelligibility	12) Read text and monologue to assess intelligibility.
Monotonicity	13) Read text and monologue to assess monotonicity.

References

Ackermann, H., & Ziegler, W. (1991). Articulatory deficits in Parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry, 54*(12), 1093–1098.

Arias-Vergara, T., Vásquez-Correa, J. C., Orozco-Arroyave, J. R., Vargas-Bonilla, J. F., & Nöth, E. (2016). Parkinsons disease progression assessment from speech using GMM-UBM. *Annual conference of the international speech communication association (INTERSPEECH),* 1933–1937.

Arias-Vergara, T., Vásquez-Correa, J. C., Orozco-Arroyave, J. R., & Nöth, E. (2018). Speaker models for monitoring parkinsons disease progression considering different communication channels and acoustic conditions. *Speech Communication, 101*, 11–25.

Barnish, M. S., Whibley, D., Horton, S., Butterfint, Z. R., & Deane, K. H. (2016). Roles of cognitive status and intelligibility in everyday communication in people with parkinsons disease: A systematic review. *Journal of Parkinson's Disease, 6*(3), 453–462.

Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., & Rusz, J. (2011). Detection of persons with Parkinsons disease by acoustic, vocal, and prosodic analysis. *IEEE automatic speech recognition and understanding workshop (ASRU),* 478–483.

Bocklet, T., Steidl, S., Nöth, E., & Skodda, S. (2013). Automatic evaluation of Parkinson's speech – Acoustic, prosodic and voice related cues. *Annual conference of the international speech communication association (INTERSPEECH),* 1149–1153.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

De Letter, M., Santens, P., & Van Borsel, J. (2005). The effects of levodopa on word intelligibility in parkinson's disease. *Journal of Communication Disorders, 38*(3), 187–196.

Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P. (2011a]). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(4), 788–798.

Dehak, N., Torres-Carrasquillo, P. A., Reynolds, D., & Dehak, R. (2011b]). Language recognition via i-vectors and dimensionality reduction. *Annual conference of the*

international speech communication association (INTERSPEECH), 857–860.
Dimauro, G., Di-Nicola, V., Bevilacqua, V., Caivano, D., & Girardi, F. (2017). Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. *IEEE Access, 5*(November), 22199–22208.
Enderby, P. M., & Palmer, R. (2008). *FDA-2: Frenchay Dysarthria Assessment: Examiner's Manual. Pro-ed*.
Enderby, P. (1980). Frenchay dysarthria assessment. *British Journal of Disorders of Communication, 15*(3), 165–173.
Garcia, N., Orozco-Arroyave, J. R., D'Haro, L. F., Dehak, N., & Nöth, E. (2017). Evaluation of the neurological state of people with parkinsons disease using i-vectors. *Proceedings of INTERSPEECH,* 299–303.
Godino-Llorente, J., Gómez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering, 53*(10), 1943–1953.
Goetz, et al. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129–2170.
Google cloud (2018). *Google Speech-To-Text API.* https://cloud.google.com/speech-to-text/.
Haderlein, T. (2007). *Automatic evaluation of tracheoesophageal substitute voices, volume 25 of Studien zur Mustererkennung.* Berlin, Germany: Logos Verlag.
Hanson, D. G., Gerratt, B. R., & Ward, P. H. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. *The Laryngoscope, 94*(3), 348–353.
Hemmerling, D., Orozco-Arroyave, J. R., Skalski, A., Gajda, J., & Nöth, E. (2016). Automatic detection of parkinsons disease based on modulated vowels. *Annual conference of the international speech communication association (INTERSPEECH),* 1190–1194.
Hlavnicka, J., Cmejla, R., Tykalova, T., Sonka, K., Ruzicka, E., & Rusz, J. (2017). Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Nature Scientific Reports, 7*(12), 1–13.
Hornykiewicz, O. (1998). Biochemical aspects of Parkinson's disease. *Neurology, 51*(2 Suppl. 2), S2–S9.
Logemann, J. A., & Fisher, H. B. (1981). Vocal tract control in Parkinson's disease: Phonetic feature analysis of misarticulations. *Journal of Speech and Hearing Disorders, 46*(4), 348–452.
Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 43*(1), 47–57.
Miller, N., Allcock, L., Jones, D., Noble, E., Hildreth, A. J., & Burn, D. J. (2007). Prevalence and pattern of perceived intelligibility changes in parkinsons disease. *Journal of Neurology, Neurosurgery & Psychiatry, 78*(11), 1188–1190.
Nöth, E., Orozco-Arroyave, J. R., Vásquez-Correa, J. C., Bocklet, T., Hannik, J., Cernak, M., et al. (2016). *Remote Monitoring of Neurodegeneration through Speech. Tech. rep.* Johns Hopkins University.
Novotný, M., Rusz, J., Čmejla, R., & Růžička, E. (2014). Automatic evaluation of articulatory disorders in parkinson's disease. *IEEE/ACM Transactions on Audio, Speech and Language Processing, 22*(9), 1366–1378.
Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., et al. (2016a]). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *Journal of the Acoustical Society of America, 139*(1), 481–500.
Orozco-Arroyave, J. R., Belalcázar-Bolaños, E. A., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Skodda, S., Rusz, J., et al. (2015). Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases. *IEEE Journal of Biomedical and Health Informatics, 19*(6), 1820–1828.
Orozco-Arroyave, J. R., Vásquez-Correa, J. C., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Skodda, S., et al. (2016b]). Towards an automatic monitoring of the neurological state of the Parkinson's patients from speech. *International conference on acoustic, speech, and signal processing (ICASSP),* 6490–6494.
Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Gonzalez-Rátiva, M. C., & Nöth, E. (2014). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. *Language resources and evaluation conference (LREC),* 342–347.
Orozco-Arroyave, J. R. (2016). *Analysis of speech of people with Parkinson's disease* (1st ed.). Berlin, Germany: Logos-Verlag.
Patel, S., Parveen, S., & Anand, S. (2016). Prosodic changes in parkinson's disease. *The Journal of the Acoustical Society of America, 140*(4), 3442.
Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinsons disease. *The Journal of the Acoustical Society of America, 129*(1), 350–367.
Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., et al. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America, 134*(3), 2171–2181.
Schölkopf, B., & Smola, A. J. (2001). *Learning with kernels.* The MIT Press.
Skodda, S., Grönheit, W., & Schlegel, U. (2011a]). Gender-related patterns of dysprosody in Parkinson disease and correlation between speech variables and motor symptoms. *Journal of Voice, 25*(1), 76–82.
Skodda, S., Visser, W., & Schlegel, U. (2011b]). Vowel articulation in parkinson's disease. *Journal of Voice, 25*(4), 467–472.
Tsanas, A., Little, M. A., Fox, C., & Ramig, L. O. (2014). Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*(1), 181–190.
Vásquez-Correa, J. C., Orozco-Arroyave, J. R., & Nöth, E. (2016). Word accuracy and dynamic time warping to assess intelligibility deficits in patients with parkinsons disease. *XXI symposium on signal processing, images and artificial vision (STSIVA),* 1–5.
Zeißler, V., Adelhardt, J., Batliner, A., Frank, C., Nöth, E., Shi, R. P., et al. (2006). *The prosody module. SmartKom: Foundations of multimodal dialogue systems*139–152.

J.C. Vásquez-Correa is Electronics Engineer and Master in Computer Science from University of Antioquia (Medellín, Colombia). Currently he is PhD student from the same University at the Signals Processing Lab and from the Pattern Recognition Lab at University of Erlangen (Erlangen, Germany).

J.R. Orozco-Arroyave is Electronics Engineer and Master in Computer Science from University of Antioquia (Medellńn, Colombia). He is PhD in Computer Science from University of Erlangen (Erlangen, Germany) and currently he is Professor and the head of the Signals Processing Lab at University of Antioquia and Adjuct Researcher in the Pattern Recognition Lab at University of Erlangen (Erlangen, Germany).

T. Bocklet is Computer Scientist from the PattUniversity of Erlangen (Erlangen, Germany) and PhD from the same University. Currently he is Researcher and Product Architect at INTEL Corporation in Munich, Germany.

E. Nöth is PhD in Computer Science from University of Erlangen (Erlangen, Germany) and currently he is Professor and the head of the Speech Processing and Understanding group at the Pattern Recognition Lab at University of Erlangen (Erlangen, Germany).