

Degree of Parkinson's Disease Severity Estimation Based on Speech Signal Processing

Zoltan Galaz, Zdenek Mzourek, Jiri Mekyska, Zdenek Smekal, Tomas Kiska, Irena Rektorova, Juan Rafael Orozco-Arroyave, Khalid Daoudi

▶ **To cite this version:**

Zoltan Galaz, Zdenek Mzourek, Jiri Mekyska, Zdenek Smekal, Tomas Kiska, et al.. Degree of Parkinson's Disease Severity Estimation Based on Speech Signal Processing. IEEE 39th International Conference on Telecommunications and Signal Processing, Jun 2016, Vienna, Austria. hal-01328198

HAL Id: hal-01328198
https://inria.hal.science/hal-01328198

Submitted on 7 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Degree of Parkinson's Disease Severity Estimation Based on Speech Signal Processing

Zoltan Galaz*, Zdenek Mzourek*, Jiri Mekyska*, Zdenek Smekal*, Tomas Kiska*
Irena Rektorova†, Juan Rafael Orozco-Arroyave‡§ and Khalid Daoudi¶
*Signal Processing Laboratory, Department of Telecommunications
Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic
†First Department of Neurology
St. Anne's University Hospital, Pekarska 53, 65691 Brno, Czech Republic
‡Grupo de Investigación en Telecomunicaciones Aplicadas
Facultad de Ingeniería, Universidad de Antioquia, Medellin, Colombia
§Pattern Recogntion Lab, Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany
¶INRIa Bordeaux Sud-Ouest (GEOSTAT team)
200 av. de la vieille tour., 33405 Talence, France

***Abstract*—This paper deals with Parkinson's disease (PD) severity estimation according to the Unified Parkinson's Disease Rating Scale: motor subscale (UPDRS III), which quantifies the hallmark symptoms of PD, using an acoustic analysis of speech signals. Experimental dataset comprised 42 speech tasks acquired from 50 PD patients (UPDRS III ranged from 6 to 92). It was divided into subsets: words, sentences, reading text, monologue and diadochokinetic tasks. We performed a parametrization of the whole corpus and these groups separately using a wide range of conventional and novel speech features. We used guided regularized random forest algorithm to select features with maximum clinical information and performed random forests regression to estimate PD severity. According to significant correlations between true UPDRS III scores and scores predicted by the proposed methodology it was shown that information extracted through variety of speech tasks can be used to estimate the degree of PD severity.**

***Keywords*—hypokinetic dysarthria; Parkinson's disease; regression; severity estimation; speech processing.**

I. Introduction

Parkinson's disease (PD) is a chronic idiopathic disorder with the unknown aetiology characterized by the progressive loss of dopaminergic neurons in *substancia nigra pars compacta* [1]. Besides its hallmark motor symptoms, patients with PD often develop a multi-modal disruption of motor speech realization referred to as hypokinetic dysarthria (HD) [2]. According to the previous studies, HD affects the area of phonation, articulation, prosody, speech fluency and faciokinesis [3]–[5].

In summary, the following speech disorders associated with HD in PD have been observed: increased acoustic noise [1], reduced voice intensity [6], harsh breathy voice quality [7], increased voice nasality [8], reduced variability of pitch and loudness combined with speech rate abnormalities [9], imprecise consonant articulation [10], unintentional introduction of pauses [11], rapid repetition of words or syllables [11], sudden deceleration or acceleration in speech [12].

Recently, researchers have focused on the prediction of clinical rating scales evaluating severity of PD and its progression [7], [13]–[16]. This study aims to follow the trend and propose a model that can discriminate healthy and disordered speech by assessment of hypokinetic dysarthria with a special focus on estimation of PD severity using the acoustic analysis of speech signals. Degree of severity of PD is estimated according to the Unified Parkinson's Disease Rating Scale, motor subscale: UPDRS III (evaluation of motor function [17]).

The rest of this paper is organized as follows. Section II presents dataset and methodology respectively. This section provides a description of speech features, statistical analysis with selection of feature subset with the best discrimination power and mapping features to the participant's UPDRS III score in order to estimate the degree of PD severity. Experimental results are discussed in section III, and finally in section IV some conclusions are provided.

II. Materials and methods

A. Speech corpus

A grand total of 100 Spanish native speakers from Colombia were studied. 50 of them suffer from PD (25 men/25 women; mean age 61.14 $\pm$ 9.61 years; mean disease duration 10.72 $\pm$ 9.25 years; UDPRS III score 36.74 $\pm$ 18.74; UDPRS IV score 2.29 $\pm$ 0.76) and the second half are gender and age matched controls, (25 men/25 women; mean age 60.90 $\pm$ 9.47 years) [18]. Each speaker performed 42 speech tasks including 24 isolated words, 10 sentences, one reading text, one monologue, and the rapid repetition of the syllables /pa-ta-ka/, /pa-ka-ta/ and /pe-ta-ka/.

This work was supported by the Czech Ministry of Health (15-28038A) and the following projects: SIX (CZ.1.05/2.1.00/03.0072), LO1401 and COST IC1206. For the research, infrastructure of the SIX Center was used. The authors would like to thank Fundalizanza Parkinson Colombia for their valuable contribution during the recording of the patients.

B. Speech features extraction

Building a regression model for the purpose of this study consisted of several stages: processing the original speech signals to extract distinctive, clinically useful properties (feature extraction stage), selecting a parsimonious, information-rich subsets of features (feature selection), and mapping the final features to the clinical outcome we aim to associate the speech signal with (feature mapping). For the purpose of speech feature extraction, Praat acoustic analysis software [19] and Neurological Disorder Analysis Tool [4] (NDAT) written in MATLAB and developed at the Brno University of Technology were used.

The feature extraction involved application of widely used speech signal processing algorithms. To objectively and automatically characterize clinically useful properties of the speech signals we computed signal to noise ratio derived from the discrete time wavelet transform (SNR DTWT), harmonics-to-noise ratio (HNR), noise-to-harmonics ratio (NHR), normalised noise energy (NNE), energy ratio (ER), detrended fluctuation analysis (DFA) and glottal-to-noise excitation ratio (GNE). We also used features based on the theory of empirical mode decomposition (EMD) to decompose the speech signal into intrinsic mode functions (IMF) and calculated SNR and NSR from the first few IMFs. Additionally we computed features based on fundamental frequency and amplitude perturbations. We applied features like, jitter (local, absolute), pitch perturbation quotient (PPQ), pitch period entropy (PPE) and glottis quotient (GQ) to track deviations in rhythmicity. We also quantified amplitude deviations using shimmer (local) and amplitude perturbation quotients (APQ3, APQ5). We also investigated the number of voice breaks (NVB) and a degree of voice breaks (DVB) to determine possible hesitation. Next, we calculated short-time energy (STE), low short-time energy ratio (LSTE), zero-crossing rate (ZCR), high zero-crossing rate (HZCR) and median frequency of power spectrum (MFPS), fraction of locally unvoiced frames (FLUF).

Next, we computed several statistical functionals [5]. In total, we extracted 715 features. Further description of the features can be found in our recent articles [4], [5], [20]. The exact features selected for each group of speech tasks separately and for whole corpus are shown in Table I.

TABLE I. COMPUTED SPEECH FEATURES

W tasks	S tasks	R tasks	M tasks	D tasks	A tasks
HNR	HNR	HNR	HNR	CP	HNR
HNR	NHR	NHR	NHR	RATE	NHR
MPSD	MPSD	MPSD	MPSD	jitt.	DFA
ZCR	ZCR	ER	ER	shimm.	GNE
F_0	F_0	ZCR	ZCR	VPL	NNE
PPE	PPE	F_0	F_0	OPL	F_0
STE	STE	PPE	PPE	SPL	jitt.
TKEO	TKEO	STE	STE	V+OPL	shimm.
jitt.	jitt.	TKEO	TKEO	O+SPL	GQ
shimm.	shimm.	NVB	NVB		ZCR
NNE	NVB	DVB	DVB		STE
SNR	DVB	jitt.	jitt.		MPSD
GNE	NNE	shimm.	shimm.		
IMF	SNR	NNE	NNE		
RDPE	GNE	HZCR	HZCR		
GQ	IMF	LSTE	LSTE		
DFA	FLUF	SF	SF		
	SF	FLUF	FLUF		
		GNE	GNE		

[1] W – words; S – sentences; R – read text; M – monologue; D – DDK; A – all; CP – DDK cycle periods; RATE – DDK rate; VPL – DDK voicing part lengths; OPL – DDK offset part lengths; SPL – DDK silence part lengths.

C. Feature selection

The objective of feature selection is to select a compact subset of features without loss of predictive information. Many different feature selection methods exist. They are generally divided into following categories: filters, wrappers, and embedded methods. Wrapper methods search for best feature subset for a given classifier, however, wrappers are often computationally very expensive. Embedded methods select feature subset using the information obtained from a classifier. Therefore the feature subset selected by embedded methods can compete with feature subsets selected by wrappers [21] and often can be computationally less demanding.

In this paper we applied the embedded method: guided regularized random forest (GRRF) [21]. We used GRRF for its ability to deal with the node sparsity issue (ability to select the most useful features even if there is only a few instances for given node) that is a frequent case with high dimensional data. GRRF has 2 tuning parameters: λ and γ. However, according to authors it is sufficient to set first one to a fixed value and tune only the other one, which often leads to better results in terms of accuracy [21]. For detailed information about GRRF, see [21]. In this paper we used γ as the only parameter for GRRF.

D. Mapping features to UPDRS III

To obtain a preliminary insight into statistical properties of the selected features, Spearman's rank sum correlation coefficient (ρ) between the feature vectors and associated UPDRS III score was computed. Fig. 1 shows correlation graphs of the highest correlated feature for each speech task group giving a visual impression of the distribution of these feature values and their relationship to the severity of PD. All selected features significantly correlated with the speakers' diagnosis ($p < 0.001$).

Building a regression model requires forming of a functional relationship $y = f(x)$ to map the feature subspace into the clinical output we want to associate the speech signal with. For this purpose we used random forests regression algorithm in the classical supervised learning setup, therefore we used training data set to learn the final model to predict PD degree and evaluated this model using independent data set.

III. DISCUSSION

Firstly, we computed a variety of speech features, see Table I, to quantify HD. Next, we explored the data by computing ρ to quantify the association strength and relevance of speech features to the values of UPDRS III. We also plotted the correlation graphs of the highest correlated features, see

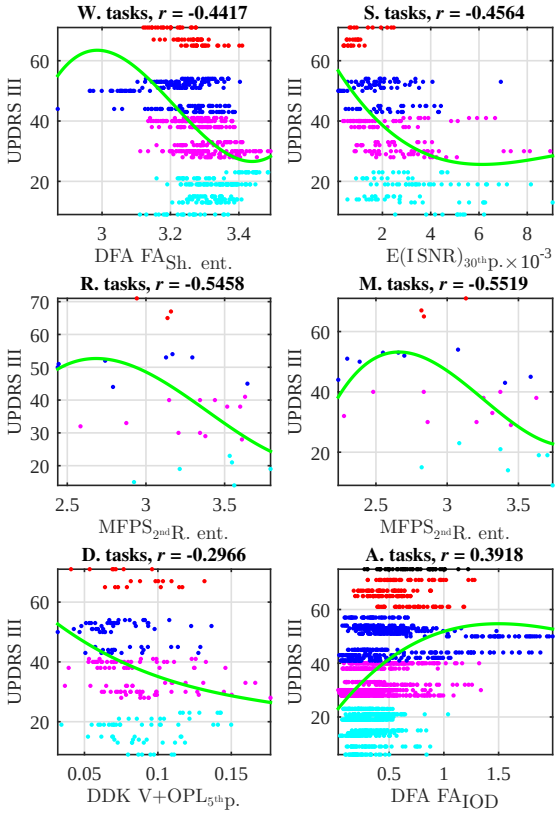


Fig. 1. Correlation graphs for all training speakers (excluding 5 outliers) from elected subsets of speech tasks: words (W.), sentences (S.), reading text (R.), monologue (M.), DDK (D.), whole corpus (A.). The speakers were divided linearly into 5 groups according their UPDRS III value. The green line represents a trend of PD severity for associated speech feature values (3^{rd} order polynomial fit curve). Figure notation: IOD – index of dispersion; Sh. ent – Shannon entropy; R. ent – Rényi entropy; p. – percentile; r – Spearman's correlation coefficient.

Fig. 1. All speech features in the figure significantly correlated with participants' UPDRS III value ($p < 0.001$).

Regression models predicting degree of PD severity was constructed considering 5 tree-based sub-models, one for each speech task group, and one for the whole corpus. The speech features selected for the speech tasks can be seen in Table I. Subsequently feature selection was applied using GRRF algorithm, see section II-C.

Next, we used RF regression algorithm (10-fold validation with 100 repetition) randomly permuting the data before splitting into training and testing subsets to find optimal model settings using *randomForest package* [22] written in R language and used these models to predict UPDRS III score of records from independent data set. The results are summarized in Table II and Fig. 2. The highest ρ was computed for the words task group ($\rho = 0.5709$) with the lowest RMSE $= 10.9520$. This model also explains the most of the variation in data (VE $= 63.56\,\%$) using a reasonable

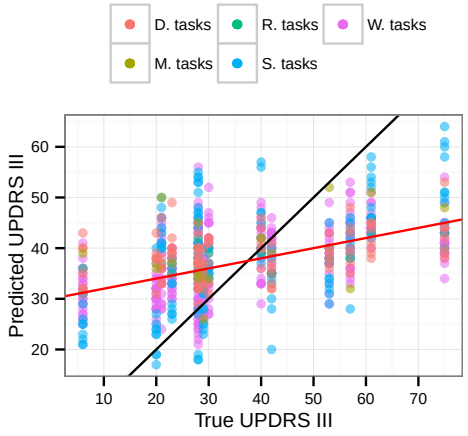


Fig. 2. Graph of predicted UPDRS III for the subsets of speech tasks: words (W.), sentences (S.), reading text (R.), monologue (M.), DDK (D.) computed from development data set as a function of true UPDRS III values. Black line visualizes ideal correlation ($\rho = 1$) between predicted and true UPDRS III values. Red line is fitted using least square method and visualizes Pearson correlation coefficient $\rho \doteq 0.502$.

amount of features (N $= 51$).

Finally, we used the regression models and predicted UPDRS III score of patients in test data set. ρ to estimate prediction accuracy. We considered two models (scenarios): a) regression model consisted of five submodels (built for speech task groups: words, sentences, reading the text, monologue and DDK tasks) - model M1; b) regression model built for the whole corpus - model M2. Resulting Spearman's correlation coefficients are: $\text{M1}_\rho = 0.1933$; $\text{M2}_\rho = -0.2500$. These results are discussed in section IV.

IV. Conclusion

In this paper we performed a complex acoustic analysis of speech in patients with PD in order to estimate a degree of the disease severity described by UPDRS III. The analysis was based on parametrization of several speech tasks described in [23]. The corpus was split into five subsets: words, sentences, reading text, monologue and diadochokinetic tasks. We also tried to analyse the whole corpus at once. Next, we performed feature extraction for all speech tasks subsets separately using the most suitable features described in Table I.

Consequently, we computed ρ to express the association of extracted speech features with the participants' clinical diagnosis (UPDRS III score) and plotted the correlation graphs of the features with the highest ρ achieved, see Fig. 1 to give a visual impression of the distribution of these feature values and their relationship to the severity of PD.

We used GRRF algorithm to select the feature subsets with the highest clinical relevance. Next, we built RF regression models from our reduced feature subsets. Description of the models we used to predict UPDRS III can be found in section III. Although we obtained ρ about 0.5 for speech tasks

TABLE II. PERFORMANCE OF REGRESSION MODELS

Speech tasks	ρ	RMSE	VE [%]	γ	N
W tasks	0.5709	10.9520	63.56	0.20	51
S tasks	0.4886	11.5227	59.66	0.20	50
R tasks	0.5014	19.3689	-13.98	0.05	59
M tasks	0.2136	18.5541	-4.60	0.20	25
D tasks	0.3222	17.3337	8.71	0.15	170

[1] W – words; S – sentences; R – read text; M – monologue; D – DDK; ρ – Spearman's rank sum correlation coefficient between predicted and true UPDRS III score for devel data set; RMSE – root-mean-square error; γ – GRRF parameter setting the feature subset size (lower γ results in larger feature subset); VE – variation explained; N – number of selected features.

groups individually, see Table II, when we used M1 regression model, resulting ρ was significantly lower ($M1_\rho = 0.1933$). This loss was probably caused by inappropriate speech task groups construction. From Fig. 1 it is evident that the number of observations in R. tasks group and M. tasks group is too small in comparison with the rest. Therefore the model (M1) built from these submodels (words, sentences, reading text, monologue, DDK) does not sufficiently represent subjects' speech degradation.

In contrast, M2 model achieved higher ρ compared to M1, which is probably caused by a selection of more optimal set of features during the feature extraction stage, see Table I. However, the regression models built for the speech tasks groups individually still outperforms both M1 and M2. Therefore, the results indicate that a selection of the speech features specific for a given speech task can in general increase prediction power of the regression model.

In this paper, we proved that PD severity estimation based on acoustic analysis of speech signal has a great potential in the field of Parkinson's disease analysis. Nevertheless, there is still space for deeper investigation. In our previous studies [5], [20], [24] of HD in PD we mainly focused on HD quantification and identification. In our future studies we will follow our recent research in the field of objective assessment of PD [25] and focus on increasing prediction accuracy of several scales developed to rate motor (freezing of gait) and non-motor (depression, sleeping disorders, cognitive impairment) symptoms of PD.

REFERENCES

[1] O. Hornykiewicz, "Biochemical aspects of parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.

[2] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *J. Behav. Neurol.*, vol. 11, no. 3, pp. 131–137, 1999.

[3] I. Eliasova, J. Mekyska, M. Kostalova, R. Marecek, Z. Smekal, and I. Rektorova, "Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease," *J. Neural Transm.*, vol. 120, no. 4, pp. 597–605, 2013.

[4] J. Mekyska, *et al.*, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94–111, 2015.

[5] Z. Galaz *et al.*, "Prosodic analysis of neutral, stress-modified and rhymed speech in patients with parkinson's disease," *Comput. Methods. Programs. Biomed.*, vol. 127, pp. 301 – 317, 2016.

[6] K. K. Baker, L. O. Ramig, E. S. Luschei, and M. E. Smith, "Thyroarytenoid muscle activity associated with hypophonia in Parkinson's disease and aging," *Neurology*, vol. 51, no. 6, pp. 1592–1598, 1998.

[7] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. R. Soc. Interface*, vol. 8, no. 59, pp. 842–855, 2010.

[8] K. A. Spencer and M. A. Rogers, "Speech motor programming in hypokinetic and ataxic dysarthria," *Brain Lang.*, vol. 94, no. 3, pp. 347–366, 2005.

[9] S. Skodda, H. Rinsche, and U. Schlegel, "Progression of dysprosody in Parkinson's disease over time–a longitudinal study," *Mov. Disord.*, vol. 24, pp. 716–722, 2009.

[10] N. Roy, S. L. Nissen, C. Dromey, and S. Sapir, "Articulatory Changes in Muscle Tension Dysphonia: Evidence of Vowel Space Expansion following Manual Circumlaryngeal Therapy," *Journal of Communication Disorders*, vol. 42, no. 2, pp. 124–135, 2009.

[11] R. Moretti, *et al.*, "Speech initiation hesitation following subthalamic nucleus stimulation in a patient with parkinson's disease," *Eur. Neurol.*, vol. 49, no. 4, pp. 251–253, 2003.

[12] M. Gentil, P. Pollak, and J. Perret, "Parkinsonian dysarthria," *Rev. Neurol.*, vol. 151, no. 2, pp. 105–112, 1995.

[13] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of Parkinson's Disease progression by noninvasive speech tests," *IEEE Trans. Bio-Med. Eng.*, vol. 57, no. 4, pp. 884–893, 2010.

[14] M. Asgari and I. Shafran, "Predicting severity of parkinson's disease from speech," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, Aug 2010, pp. 5201–5204.

[15] M. Castelli, L. Vanneschi, and S. Silva, "Prediction of the unified parkinson's disease rating scale assessment using a genetic programming system with geometric semantic genetic operators," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4608–4616, 2014.

[16] B. E. Sakar, C. Sakar, G. Serbes, and O. Kursun, "Determination of the optimal threshold value that can be discriminated by dysphonia measurements for unified Parkinson's disease rating scale," in *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on*, Nov 2015, pp. 1–4.

[17] S. Fahn and R. L. Elton, *UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale*. Recent developments in Parkinson's Disease. Macmillan, Florham Park, 1987.

[18] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.

[19] P. Boersma and D. Weenink, "Praat: Doing Phonetics by Computer (Version 5.3.10)," 2012. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[20] Z. Smekal *et al.*, "Analysis of phonation in patients with Parkinson's disease using empirical mode decomposition," in *2015 International Symposium on Signals, Circuits and Systems (ISSCS)*, July 2015, pp. 1–4.

[21] H. Deng and G. Runger, "Gene selection with guided regularized random forest," *Pattern Recognit.*, vol. 46, no. 12, pp. 3483–3489, 2013.

[22] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[23] S. Björn, S. Stefan, B. Anton, H. Simone, H. Florian, O.-A. Juan Rafael, N. Elmar, Z. Yue, and W. Felix, "The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson's & Eating Condition," in *Proceedings INTERSPEECH 2015*, ISCA, Ed., 2015.

[24] J. Mekyska *et al.*, *Recent Advances in Nonlinear Speech Processing*. Cham: Springer International Publishing, 2016, ch. Perceptual Features as Markers of Parkinson's Disease: The Issue of Clinical Interpretability, pp. 83–91.

[25] J. Mekyska, Z. Galaz, Z. Mzourek, Z. Smekal, and I. Rektorova, "Assessing progress of Parkinson's using acoustic analysis of phonation," in *2015 International Work Conference on Bioinspired Intelligence (IWOBI)*, June 2015, pp. 115–122.