Automatic Estimation of Severity of Parkinson's Disease Based on Speech Rhythm Related Features

Dávid Sztahó[1,2], Miklós Gábriel Tulics[1], Klára Vicsi[1]
[1]Department of Telecommunication and Media Informatics
Budapest University of Technology and Economics
[2]Hungarian Academy of Sciences
Budapest, Hungary

István Valálik
Department of Neurosurgery
St. John's Hospital
Budapest, Hungary

Abstract — **Diseases, such as Parkinson, impairs cognitive processes of patients, through which speech is also affected. In this paper, we propose a method for Parkinson's disease severity level estimation based on speech rhythm related features extracted from running speech (read texts and monologue) uttered by Hungarian Parkinson patients and healthy control population. Classification and regression models are built using various machine-learning methods for both linguistic types separately. Separate and joint decisions were made for the different text types. The final prediction was obtained by fusing the separate estimations for each speaker. Test trials were run in order to investigate, if age is a relevant feature for the machine learning tasks. It was found that the investigated features are useful and highly relevant for the automatic diagnosis of Parkinson's disease based on the classification and regression performances. The best results were obtained using support vector machine (and regression) with 84.62% accuracy for binary classification and 0.735 Spearman correlation for Parkinson severity level estimation measured on the Hoehn-Yahr scale.**

Keywords— Parkinson, speech, signal processing, support vector regression, neural networks, medical informatics

I. Introduction

Biomarkers are getting more and more popular in the field of medicine. Biomarkers can indicate diseases before more severe symptoms may arise [1]. Identification of these early signs may help patients to be treated in an earlier state of the disease. Automatic self-diagnostic tools may further improve the fast identification of the diseases.

Parkinson's disease (PD) is one of the most frequent neurologic diseases, its incidence rate is about 120/100 000 per year. In Hungary about 18-20 000 patients are affected. Because age is a risk factor, in an aging society it will occur even more frequently. The main symptoms of the disease are tremor, slowness and rigidity. 70% of the patients have tremor as the first sign [2]. Additionally, in case of PD patients, doctors also experienced cognitive impairments.

One of the biomarkers is the impairment of cognitive processes, therefore it is also relevant from the viewpoint of cognitive infocommunication [3, 4]. This impaired functioning affects many life functions, such as communication, speech. Researches show that speech can also contain useful information about PD [5, 6, 7]. Almost 89% of the patients have some kind of speech impairment, even as an early sign [5]. Informal descriptions state that speech has decreased loudness, improved vocal tremor and breathy voice production [8]. These facts indicate that speech may be utilized for automatic detection of PD. Speech can also be utilized as an interface for open-domain CogInfoCom systems [24, 25].

Several studies tried to detect PD based on speech using machine-learning algorithms. The first researches used sustained vowels [9]. These samples can contain useful information, because the improper muscle control can cause improper production of vowels. This type of speech is easy to use, it is robust; therefore, it is common in clinical practice in any field of research [9]. However, sustained vowel examination gives information about the operation of the vocal cord only, and about a steady state of the articulation organs. The focus of this paper is the investigation of speech rhythm related features. Therefore, running speech is needed to get information about the continuous process of speech production.

Various features have been investigated in the speech signal: harmonic-to-noise ratio [10, 11, 12, 13], jitter, shimmer [11], mel-spectral frequency coefficients (MFCC) [13, 14], ratio of voiced-to-unvoiced sounds, intelligibility, prosodic features [19], nonlinear voice features [18].

Decision support tools can be used in order to help building easy-to-use supplementary diagnostic software. These tools help to differentiate patients into healthy and PD categories, and may even try to predict to severity of the disease. Many of these [6, 11, 12, 13, 15] use Bayesian classifier, random forest, k-nearest neighbors, support vector machines and neural networks.

In [9] patients were classified in a binary decision task (healthy and PD) using Spanish, German, Czech speech samples. Based on acoustic features calculated at the beginning and the end of the voiced speech segments, different results are obtained for each language. This suggests that PD may affect speech in a language-dependent way.

Sustained vowels were also used for predicting the severity of the disease. Jain and Shetty [15] used a reduced UPDRS (Unified Parkinson's Rating Scale) range to predict the severity level of PD. They also included gender and age as a prediction variable. Age, as a risk factor of PD seems to be a good idea to increase the accuracy of automatic prediction, but it may decrease the chance to identify severe cases in a younger age. However, it could be a useful information.

Beside sustained vowels, Orozco-Arroyave and his colleges applied read sentences, words and short texts in order to

978-1-5386-1264-4/17/$31.00 ©2017 IEEE

distinguish PD patients from healthy control [17]. They used prosodic features along with characterization of voiced and unvoiced segments separately. Classification tests were carried out using three languages in mono- and cross-language ways. From 85% to 99% accuracies were found in the case of monolingual and from 60% to 99% accuracies were obtained in the case of cross-lingual experiments. It shows that there is a need for monolingual classification tests.

In this paper, we focus on speech rhythm related features. How these parameters are distorted in speech in case of such a cognitive decline as Parkinson's disease, and how these changes can be detected automatically. In [23] differences of rhythm related features are found using read text. We propose a method for classification of PD and estimating PD severity level using running speech (read texts and monologues) uttered by Hungarian Parkinson patients and healthy control population. A speech database has been recorded in Hungarian containing various speech tasks. Classification and regression models are built using various machine-learning methods, such as support vector machines and (deep) neural networks for each linguistic types separately. The final prediction was obtained by fusing the separate estimations for each speaker. Male and female differences were also investigated.

II. Database

For the experiments, a Hungarian speech database was recorded that contains speech samples from 51 PD patients. The audio footage took place at two health institutes in Budapest: Virányos Clinic and Semmelweis University. A part of the patients has subthalamic implants. In their case, the recordings were done before the DBS (deep brain stimulation) surgery took place. For healthy control (HC) population speech from 27 subjects were also recorded.

A. Patients

The severity of PD was labelled according to the Hoehn & Yahr scale (H-Y) [20]. The H-Y scale ranges from 1 to 5, where the values mean the followings: 1 - symptoms (like tremor) occur only unilateral; 2 - bilateral symptoms without walking difficulty; 3 - bilateral symptoms with minimal walking difficulty; 4 - bilateral symptoms with mild walking difficulty; 5 - bilateral symptoms with highly severe walking difficulty. The range is not linear and therefore we can't expect that patients with H-Y score 2 have twice as severe symptoms (also in speech impairment) as patients with H-Y score 1. This should be taken into consideration later on interpreting the results.

25 male speakers (mean H-Y score: 2.60(±0.84); mean age: 64.74(±10.48)) and 26 female speakers (mean H-Y score: 2.57(±0.96); mean age: 64.82(±9.32)) were recorded.

Beside the age and H-Y scores, additional data were recorded, such as taken medicines, smoking, illnesses that influence speech (such as asthma, cold and flu).

B. Healthy control

Speech of healthy control population, with same age distribution, has been also recorded using the same text material and recording environment. In order to balance the H-Y distribution of the database, 27 healthy subjects (14 male speaker with mean age of 67.14(±8.02) and 13 female speakers with mean age of 58.13(±6.27)) were recorded labelled with 0 H-Y score. The HC subjects had no known diseases and were not under any medical treatment.

C. Recording environment and text material

The recordings were made with external USB sound card (Terratec 6fire USB) with good quality A/D converter and low noise level (audio coding: PCM, sampling rate: 16 kHz, quantization: 16-bit). Two microphones were used: a clip-on condenser (Audio-Technika ATR3350) microphone and a close-field condenser microphone (Monacor ECM-100). Due to the impaired coordination ability of the patients, only the signal of the clip-on microphone was kept. The recordings were made in a quiet office environment (medical office).

From each subject (1) a read text ('The north wind and the sun') and (2) a monologue was recorded. The average duration of the read text is 45 seconds. The duration of the monologue was at least 1 minute. All speech samples were normalized to 0.99 peak level. The running speech samples were segmented on phoneme level using automatic speech recognizer [21].

III. Features

Running speech was characterized by different prosodic features in the case of read text and monologues. The list of features is described in Table I. The table also shows, whether the given feature is included in the feature vector for read text and monologue or not. As previously stated, Jain and Shetty [15] used the age as a feature for PD classification. Age is a risk factor, but may cause distortion. It may cause a younger patient to be detected with Parkinson with a lower likelihood. However, Parkinson is more likely to occur at a relatively older age, so it may worth to consider using it as an additional feature. The effect of age on the classification and regression results is examined in Section V.

IV. Classification and Regression Methods

Two main detection tasks were performed on the database. Binary classification (PD or healthy) decisions were made using k-nearest neighbors (k-NN, k=10), support vector machines (c-SVC with linear (C=1) and radial basis kernel function (C=8, γ=0.01)) and artificial feed-forward neural network (ANN with neurons in hidden layers: 10, 10) and deep neural network with rectifier activation function (DNN with neurons in hidden layers: 20, 20). In addition, estimation of H-Y score was done using simple linear regression, support vector regression (epsilon-SVR with linear (C=4) and radial basis kernel functions (C=4, γ=0.05)) feed-forward neural networks (ANN with of neurons in hidden layers: 10, 10) and deep neural networks with rectifier activation function (DNN with neurons in hidden layers: 20, 20) as regression methods. The optimal parameters that are marked at each the classifier and regression method are obtained in a grid search. RapidMiner Studio 7.5 was used for machine-learning tests.

TABLE I. LIST OF USED FEATURES. LAST COLUMN INDICATES, IF IT IS INCLUDED IN FEATURE VECTOR OF READ TEXT AND/OR MONOLOGUE

Feature	*Description*	*# of features*	*read text / monologue*
ΔC and ΔV	standard deviation of interval durations of consonant and vowels segments	2	+/+
speech duration	ratio of total duration of speech segments (without pauses) to the total duration of recording	1	+/+
%C and %V	ratio of the total consonant and vowel segment durations to the total speech duration	2	+/+
articulation rate	ratio of the uttered phonemes to the total speech duration	1	+/+
phoneme duration	mean and standard deviation of duration of phonemes /a/ and /e/	4	+/+
normalized Pairwise Variability Index (nPVIc and nPVIv)	mean of the consecutive interval duration differences for consonants and vowels separately normalized to the total duration of the segments	2	+/+
raw Pairwise Variability Index (rPVIc and rPVIv)	mean of the consecutive interval duration differences for consonants and vowels separately	2	+/+
pause ratio	ratio of the total duration of pause segments to the total duration of recording	1	+/+
total duration	total duration of recording	1	-/+
total pause duration	total duration of silence and pause segments	1	-/+
total vowel duration	ratio of total duration of vowel segments to the total duration of the recording	1	-/+
age	age of the speaker	1	+/+

In addition to the results of the separate decisions, a joint final decision was calculated for each subject. In the case of the binary classification, the separate decisions are merged using majority voting. If the two classifiers showed different results (healthy and PD), the final decision for the subject was 'healthy'. This facilitated that a healthy person to not to be detected as PD. In the case of H-Y score estimation (regression), the final decision was made by averaging the separate predicted H-Y scores.

Due to the limited size of the database leave-one-out cross-validation (LOOCV) was performed, by iterating through all of the samples in the database, leaving that sample out for testing and using all others for training. Although this will result in different models for each iteration step, we can use all samples for training and testing and the test samples are not included in none of the training sets. This trade-off is the cost for a better generalization ability.

V. RESULTS

A. Binary classification

The performance of binary classification task is evaluated by measuring accuracy, precision and recall. The results are shown in Table II. The accuracy results in Table II are comparable to the highest results reported in [17] in the case of prosodic features. There, the best reported monolingual accuracy is 83.9% (with the German database). The highest overall accuracy in Table II is obtained with c-SVC using rbf kernel function (read text: 83.56%, monologue: 85.11%, joint decision: 84.62%). Male samples are classified significantly more

TABLE II. CLASSIFICATION AND REGRESSION RESULTS

		Classification results								
		Read text			*Monologue*			*Joint decision*		
	Classifier	*Acc. [%]*	*Prec. [%]*	*Rec. [%]*	*Acc. [%]*	*Prec. [%]*	*Rec. [%]*	*Acc. [%]*	*Prec. [%]*	*Rec. [%]*
non-age	k-NN	75.34	74.85	76.61	80.85	82.43	68.88	76.92	75.61	78.00
	svm-rbf	80.82	88.33	74.07	82.98	84.47	72.58	83.33	87.20	76.80
	svm-linear	83.56	87.24	78.54	82.98	82.73	73.69	83.33	85.28	77.67
	ann	78.08	76.49	76.49	76.60	71.30	70.32	73.08	71.46	73.31
	dnn	79.45	79.62	75.28	75.53	69.88	68.46	71.79	69.46	70.59
age	k-NN	79.45	79.28	81.40	75.53	70.03	65.15	78.21	81.59	78.62
	svm-rbf	**83.56**	**89.66**	**77.78**	**85.11**	**88.44**	**75.18**	**84.62**	**88.04**	**78.65**
	svm-linear	80.82	81.92	76.37	80.85	77.04	74.41	80.77	79.17	77.45
	ann	79.45	78.03	77.58	81.91	77.86	79.57	75.64	74.06	76.14
	dnn	79.45	78.32	76.81	82.98	79.22	79.22	80.77	79.28	81.81
		Regression results								
		Read text			*Monologue*			*Joint decision*		
	Regression method	*RMSE*	*Pearson*	*Spearman*	*RMSE*	*Pearson*	*Spearman*	*RMSE*	*Pearson*	*Spearman*
non-age	linear regression	1.558	0.414	0.506	1.275	0.612	0.624	1.194	0.670	0.672
	svm-rbf	1.086	0.712	0.702	1.265	0.599	0.589	1.113	0.726	0.731
	svm-linear	1.621	0.431	0.500	1.273	0.616	0.621	1.207	0.663	0.664
	ann	1.429	0.582	0.605	1.607	0.453	0.424	1.345	0.592	0.572
	dnn	1.235	0.629	0.663	1.326	0.613	0.589	1.212	0.692	0.711
age	linear regression	1.581	0.481	0.596	1.293	0.601	0.618	1.150	0.699	0.715
	svr-rbf	**1.052**	**0.732**	**0.744**	**1.239**	**0.625**	**0.629**	**1.097**	**0.736**	**0.735**
	svr-linear	1.611	0.400	0.427	1.261	0.627	0.624	1.232	0.649	0.640
	ann	1.601	0.560	0.606	1.246	0.451	0.469	1.401	0.559	0.569
	dnn	1.323	0.579	0.615	1.351	0.586	0.593	1.250	0.671	0.697

precisely than females (Table III). Joint decision showed improvement only in some cases. Including age (as in [15]) in the feature vector showed some increase in the results (except in the 'svm-linear' case), but it is not significant.

B. Regression

The performance of the regression methods are evaluated by the RMSE value, the Spearman and Pearson correlations between the predicted and the original H-Y scores. The results are also displayed in Table II. The best results (lowest RMSE and highest correlations) are also achieved with support vector machines (regression in this case) using radial basis kernel function as in the classification tests. In this task, it clearly outperformed the other regression methods. Here, joint decision had more effect on the results. In all cases, the performance measures did improve. Age as a feature also resulted a slight improvement.

The results of svr-rbf method are depicted in Figs 1, 2 and 3 for read text, monologue and joint decision (age is included in the feature vector). The predicted H-Y scores are plotted in function of original H-Y scores for each speaker (linear trend lines are also displayed). Male and female samples are marked separately. As in the case of binary classification, the prediction also performed better on male than on female samples (Table III).

TABLE III. RESULTS FOR MALE AND FEMALE SETS SEPARATELY FOR BEST PERFORMING MACHINE LEATNING METHODS USING JOINT DECISION

Gender	**Classification**			**Regression**		
	Acc. [%]	***Prec. [%]***	***Rec. [%]***	***RMSE***	***Pearson***	***Spearman***
male	94.87	96.30	92.86	1.085	0.741	0.748
female	74.36	76.76	63.46	1.202	0.723	0.728

VI. CONCLUSION

In our study, we have two aims. On one hand, we investigate Parkinson's disease, how it affects the patient's speech, what speech characteristics are distorted and to what extent when the disease progresses. Our study contribute to a better understanding of cognitive processes operating in the formation of speech. On the other hand, with the automatic detection of early Parkinson's disease, we are going to create an IT tool that contribute to prevent further cognitive decline in patients.

In this paper, speech rhythm related features were used in order to distinguish patients with Parkinson's disease from healthy control subjects. Beside this binary classification task, the severity of the disease (measured with H-Y scores) was also predicted automatically using various regression methods. A new Hungarian speech database was introduced that contains speech from patients with Parkinson's disease. The rhythm related features included vowel and consonant timing derived measures, utterance and pause duration related parameters. For these measurements running speech was recorded: a read text and a monologue from each speaker. From the various machine-learning methods, support vector machines performed the best, in both classification and regression tasks.

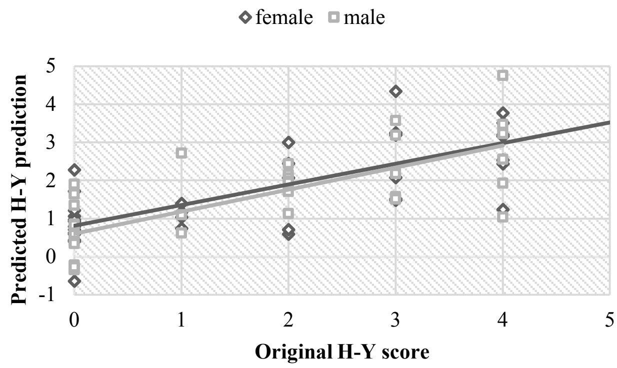

Fig. 1. Scatter-plot of the original and the predicted H-Y scores for *read text* in the case of svm-rbf regression method. Linear trend lines are shown for male and female speakres separately.

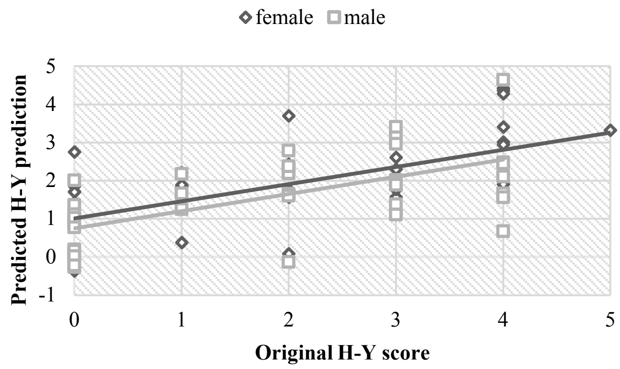

Fig. 2. Scatter-plot of the original and the predicted H-Y scores for *monologues* in the case of svm-rbf regression method. Linear trend lines are shown for male and female speakres separately.

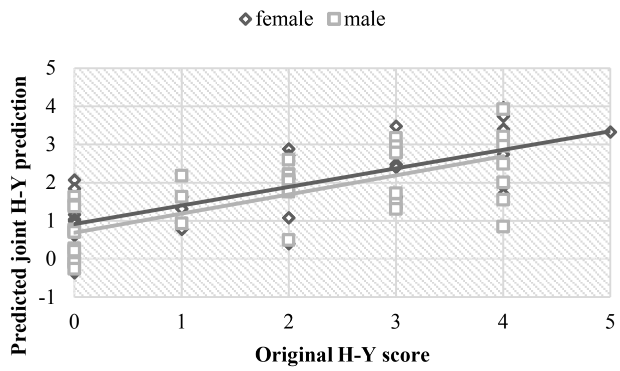

Fig. 3. Scatter-plot of the original and the predicted *joint* H-Y scores in the case of svm-rbf regression method. Linear trend lines are shown for male and female speakres separately.

Due to the speech style differences, separate and joint decisions were made for the two types of text material. In the case of binary classification task, joint decision did not always increased the overall performance. However, it did help in the case of regression. Test trials were run in order to investigate, if age is a relevant feature for the machine learning tasks. It was found that it slightly increased the overall performance. Although, age must be used carefully in a real world diagnostic system, this finding coincides with previous studies.

Gender differences were found in both binary classification and regression cases. In case of classification, males had significantly higher accuracy. This coincides with previous findings in [22].

Although, the investigated features are useful and highly relevant for the automatic diagnosis of Parkinson's disease based on the classification and regression performances, we could determine more information about the robustness and the generalization of the machine learning methods with increasing the number of speakers in the database.

Also, the machine learning models that are built can be utilized to create new artificial cognitive capability or even a complex cognitive sensor of the human - ICT combo that are applicable in cognitive infocommunication.

Although, the tremor symptoms of the patients decreased, subjectively perceived speech ability of patients with sub-thalamic implants did not change, when their implants was turned on and off. It is a remarkable experience in cognitive point of view. An examination is planned, in which the impact of the implant to motor functions of speech will be investigated in detail. Statistical tests will be carried out to reveal differences in the (presently used and newly introduced) measured acoustic features according to the state of the stimulus of the implant.

ACKNOWLEDGMENT

The research was partly funded by the Postdoctoral Fellowship Programme of the Hungarian Academy of Sciences (POSTDOC-77).

REFERENCES

[1] C. W. Turck. "Biomarkers for psychiatric disorders." Springer, 2008.

[2] M. Rajput, A. Rajput, and A. H. Rajput. "Epidemiology." In Handbook of Parkinson's Disease, 2007.

[3] P. Baranyi and A. Csapo, "Definition and synergies of cognitive infocommunications," in Acta Polytechnica Hungarica , vol. 9, no. 1, pp. 67–83, 2012.

[4] P. Baranyi, A. Csapo, and G. Sallai, "Cognitive Infocommunications (CogInfoCom),"Springer International, 2015.

[5] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth. "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," In LREC, 2014, pp. 342-347.

[6] 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." Behavioural neurology, 11(3), 1999, pp. 131-137.

[7] J.R. Orozco-Arroyave, F. Hönig, J.D. Arias-Londoño, J.F. Vargas-Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease" Proceedings of 16th INTERSPEECH, Dresden, Germany, 2015, pp. 95-99.

[8] G. An, D. G. Brizan, M. Ma, M. Morales, A. R. Syed, and A. Rosenberg. "Automatic Recognition of Unified Parkinson's Disease Rating from Speech with Acoustic, i-Vector and Phonotactic Features." In: INTERSPEECH. 2015, pp. 508-512.

[9] I. R. Titze and D. W. Martin. "Principles of voice production." The Journal of the Acoustical Society of America, 104(3), 1998. pp. 1148-1148.

[10] J. Kim, M. Nasir, R. Gupta, M. V. Segbroeck, D. Bone, M. Black, Z. I. Skordilis, Z. Yang, P. Georgiou, and S. Narayanan. "Automatic estimation of Parkinson's disease severity from diverse speech tasks." In Proceedings of Interspeech, September 2015.

[11] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun. "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings." IEEE Journal of Biomedical and Health Informatics, 17(4), 2013, pp. 828-834.

[12] Y. Alemami and L. Almazaydeh. "Detection of Parkinson Disease through Voice Signal Features. Journal of American Science," 10(10), 2014, pp. 44-47.

[13] A. Frid, E. J. Safra, H. Hazan, L. L. Lokey, D. Hilu, L. Manevitz, L. O. Ramig, and S. Sapir. "Computational diagnosis of Parkinson's Disease directly from natural speech using machine learning techniques." In: Software Science, Technology and Engineering (SWSTE), 2014 IEEE International Conference on. IEEE, 2014, pp. 50-53.

[14] T. Khan, J. Westin, and M. Dougherty. "Classification of speech intelligibility in Parkinson's disease." Biocybernetics and Biomedical Engineering, 34(1), 2014, pp. 35-45.

[15] S. Jain and S. Shetty. "Improving accuracy in noninvasive telemonitoring of progression of Parkinson's Disease using two-step predictive model." Electrical, Electronics, Computer Engineering and their Applications (EECEA), 2016 Third International Conference on. IEEE, 2016, pp. 104-109.

[16] Zhang, H. H., Yang, L., Liu, Y., Wang, P., Yin, J., Li, Y. and Yan, F. "Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples." Biomedical engineering online, 15(1), 122. 2016.

[17] Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S. and Nöth, E. "Automatic detection of Parkinson's disease in running speech spoken in three different languages." The Journal of the Acoustical Society of America, 139(1), 2016, pp. 481-500.

[18] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello and I. M. Moroz. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." BioMedical Engineering OnLine, 2007, pp. 6-23.

[19] T. Khan, J. Westin and M. Dougherty. "Classification of speech intelligibility in Parkinson's disease." Biocybernetics and Biomedical Engineering 34.1, 2014, pp. 35-45.

[20] M. Hoehn and M. Yahr. "Parkinsonism: onset, progression and mortality." Neurology, 17(5), 1967, pp. 427-427.

[21] B. Tarján, G. Sárosi, T. Fegyo and P. Mihajlik. "Improved recognition of Hungarian call center conversations." In The 7th International Conference on Speech Technology and Human-Computer Dialogue (SpeD 2013), pp. 1-6.

[22] D. Sztahó and K. Vicsi. "Estimating the Severity of Parkinson's Disease Using Voiced Ratio and Nonlinear Parameters." In International Conference on Statistical Language and Speech Processing, 2016, pp. 96-107.

[23] M. Pettorino, M. G. Busà and E. Pellegrino. (2016). "Speech Rhythm in Parkinson's Disease: A Study on Italian." In INTERSPEECH 2016, 2016, pp. 1958-1961.

[24] G. Wilcock and K. Jokinen. "Towards cloud-based speech interfaces for open-domain coginfocom systems." In, 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), 2013, pp. 413-418.

[25] P. Baranyi, A. Csapo and P. Varlaki. "An overview of research trends in coginfocom." In Intelligent Engineering Systems (INES), 2014 18th International Conference on Cognitive Infocommunications (CogInfoCom), 2014, pp. 181-186.