

Articulation and Empirical Mode Decomposition Features in Diadochokinetic Exercises for the Speech Assessment of Parkinson’s Disease Patients

J. C. Vásquez-Correa^{1,2}, C. D. Rios-Urrego², A. Rueda³, J. R. Orozco-Arroyave^{1,2}, S. Krishnan³, E. Nöth¹

¹Pattern Recognition Lab. Friedrich-Alexander Universität, Erlangen-Nürnberg, Germany

² Faculty of Engineering. Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

³ Department of Electrical and Computer Engineering, Ryerson University, Canada
{[juan.vasquez](mailto:juan.vasquez@fau.de)}@fau.de

Abstract. Speech impairments are one of the earliest manifestations in patients with Parkinson’s disease. Particularly, articulation impairments related to the capability of the speaker to move the limbs and muscles of the vocal tract have been observed in the patients. Articulation deficits have been evaluated in the patients mainly using diadochokinetic exercises, which consist in the rapid repetition of syllables like /pa-ta-ka/. This study considered different features to model several aspects of the diadochokinetic exercises, including the capacity to start/stop the vocal fold vibration, the speech rate, and the regularity of the diadochokinetic task. Articulation features are combined with others that result from an empirical mode decomposition procedure, which have been recently used to model dysphonia in Parkinson’s patients. The features are used to classify Parkinson’s patients and healthy speakers, and to predict the dysarthria severity of the participants according to a clinical scale. According to the results, articulation features are able to classify the presence of the disease with an accuracy up to 76%, and to predict the dysarthria level of the speakers with a Spearman’s correlation of up to 0.68.

Keywords: Parkinson’s disease · Articulation · Empirical mode decomposition · Dysarthria · Diadochokinetic exercises.

1 Introduction

Parkinson’s disease (PD) is a neurological disorder that alters the function of the basal ganglia in the midbrain, producing motor and non-motor deficits in the patients [1]. Motor symptoms include among others, bradykinesia, rigidity, resting tremor, and different speech impairments. Non-motor symptoms include depression, sleep disorders, impaired language, and others. Speech impairments

are an early and prominent manifestation that can contribute primarily to the diagnosis of PD [2]. The main symptoms of the speech of PD patients are grouped and called hypokinetic dysarthria. They include monopitch, reduced stress, imprecise consonants, and reduced loudness. Several studies in the literature have described the speech impairments of PD patients in terms of phonation, articulation, and prosody [3, 4]. Particularly, articulation impairments are related to the modification of position, stress, and shape of several limbs and muscles to produce speech. One of the first observed articulation symptoms was the imprecise production of stop consonants such as /p/, /t/, /k/, /b/, /d/, and /g/ [5, 6]. Other symptoms include reduced duration of voiced segments and transitions, and increased voice onset time (VOT) [6, 7]. These symptoms have been considered among the most important for the assessment of PD from speech [5, 6, 4].

Articulation analysis have been evaluated mainly using diadochokinetic exercises (DDK), which consist in the rapid repetition of syllables like /pa-ta-ka/. The execution of these exercises requires the continuous movement of different articulators such as lips, tongue and velum. Several researchers have addressed the task of modeling articulatory deficits considering DDK tasks. In [8], the authors modeled six different articulatory deficits in PD patients using DDK exercises: vowel quality, coordination of laryngeal and supra-laryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors reported an accuracy of 88% discriminating between PD patients and HC speakers, using a support vector machine (SVM) classifier. Another articulation model was proposed in [9], where the energy content in the transitions from unvoiced to voiced (onset) and from voiced to unvoiced (offset) segments was considered. The authors classified PD patients and HC speakers with speech recordings in three different languages (Spanish, German, and Czech) and reported accuracies in a range between 80% and 94% depending on the language. In [10] the authors proposed an articulation model based on temporal and spectral features extracted from the VOT segments from DDK exercises. The temporal features included the VOT duration, the VOT ratio, the vowel variability quotient, and the articulation rate. The spectral features considered 13 Mel frequency cepstral coefficients (MFCCs) extracted from the VOT segments. The authors considered a SVM classifier, and reported an accuracy of up to 92.2%. Recently, in [11], the authors proposed an articulatory model based on forced Gaussian mixture models (GMMs) to get time-stamps for the different phonetic structures that appear in an utterance. The different phonemes were segmented and grouped to train separate GMMs for each phoneme unit. The proposed method allowed to compare the features of each phonetic unit between PD and HC subjects independently. The classification was performed based on a threshold of the difference between the posterior probabilities from the models created for HC subjects and PD patients. The authors reported accuracies in a range from 81.0% to 94%.

This paper aims to evaluate the performance of three different articulation feature sets extracted from DDK exercises to model the speech deficits of PD

patients. The first set is based on the proposed in [9], which aims to model the difficulty of the patients to start/stop the vocal fold vibration. These features are based on modeling the transition between voiced and unvoiced segments in the speech signal. The second group of articulation features aims to model the speech rate and the regularity of the DDK exercises. Finally, the third feature set considers descriptors extracted from the empirical mode decomposition (EMD), which have been recently used to model dysphonia in PD patients [12, 13]. The combination of the articulation features to model the start/stop movement of the vocal folds with the features to model the regularity of the DDK exercises showed to be highly accurate to classify PD patients and HC subjects, and to predict the dysarthria level of the participants, according to a modified version of the Frenchay dysarthria assessment scale (m-FDA), which was proposed recently in [14].

2 Methods

2.1 Transition features

The first group of features aims to model the difficulties exhibited by PD patients to start/stop the movement of the vocal folds [9], and are based on the energy content in onset and offset segments. The border between voiced and unvoiced segments is detected based on the presence of the F_0 . Once the borders are detected, 40 ms of the signal are taken to the left and to the right, forming a segment with 80 ms length. The spectrum of the transitions is distributed into the first 17 critical bands according to the Bark scale, and the Bark band energies (BBE) are computed. 13 MFCCs and their first two derivatives are also computed in the transitions to complete the feature vector. Then, the mean, standard deviation, skewness, and kurtosis are computed for the features of consecutive transitions. Finally, the features obtained for onset and offset are concatenated, forming the final feature vector per utterance. Figure 1 shows the process to extract the articulation features.

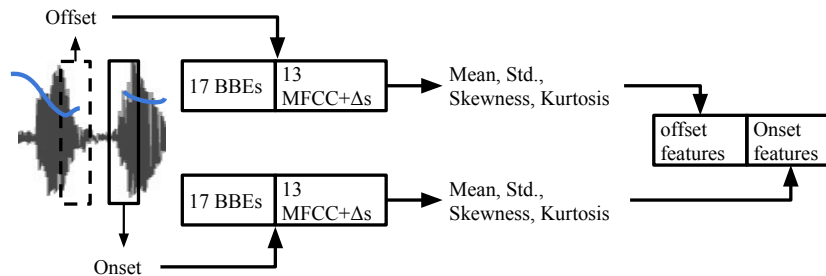


Fig. 1. Scheme for the extraction of articulation features

2.2 DDK regularity features

The second group of articulation features aims to model the regularity of the DDK exercises in terms of regularity, tone, rhythm, and duration. The feature set is formed with 55 features per utterance based on the speech rate, duration, and the F_0 contour when the subjects pronounce the DDK exercises. A detailed description of the features included in this set is shown in Table 1.

Table 1. Description of the features to model the regularity of the DDK exercises. **P**: pause duration, **V**: voiced duration, **U**: unvoiced duration

Num.	Feature	Description
1-4	F0-contour	Average, Standard deviation, Maximum, Minimum
5	Voiced rate	Number of voiced segments per second
6-9	Duration of Voiced	Average, Standard deviation, Maximum, Minimum
10-11	Duration of Unvoiced	Maximum, Minimum
12	Pause rate	Number of pauses per second
13-14	Duration of Pauses	Average, Standard deviation
15-18	F0 in first voiced segment	Average, Standard deviation, Skewness, Kurtosis
19-22	F0 in last voiced segment	Average, Standard deviation, Skewness, Kurtosis
23-24	Linear estimation of F0	Tilt, Mean square error (MSE)
25-30	Duration ratios	$P/(V+U)$, P/U , $U/(V+U)$, $V/(V+U)$, V/P , U/P
31	# voiced / # unvoiced	
32-55	Estimated F0 with a 5-degree Lagrange polynomial	Average, Standard deviation, skewness, and kurtosis

2.3 Empirical mode decomposition features

The third group of articulation features is based on the EMD computed on the DDK exercises. The EMD is a time-domain decomposition method commonly used in signal denoising. The method is based on the estimation of intrinsic mode functions (IMFs) through a sifting process in time-domain [15]. The original signal $s[n]$ can be reconstructed by adding all IMFs together, as shown in Equation 1. EMD-based features have been recently considered for the analysis of dysphonia in PD patients [12, 13].

$$s[n] = \sum_{i=1}^N \text{IMF}_i[n] \quad (1)$$

The process to extract the EMD-based features considered in this study is as follows: the vocalic and plosive segments from the speech signal are segmented using the envelope of the Hilbert transform of the signal. The inflection points of the envelope are used to segment the plosive sounds from the DDK exercises and the vowels (see Figure 2). Once the plosives and the vowels are segmented, 12 descriptors are computed for each segment based on the IMFs decomposition.

The computed descriptors include the number of IMFs obtained from the decomposition, the occupied bandwidth (OBW) of the first 10 IMFs, and the adaptive SNR estimated based on the cross-correlation between the IMFs and the original speech signal [12, 13]. Four statistical functionals (mean, standard deviation, skewness, and kurtosis) are computed for the 12 descriptors extracted from the plosives and vocalic sounds from an utterance, forming two 48-dimensional feature vectors (one for plosives and one for vowels). Finally, the features obtained for plosives and vowels are concatenated, forming the final feature vector per utterance.

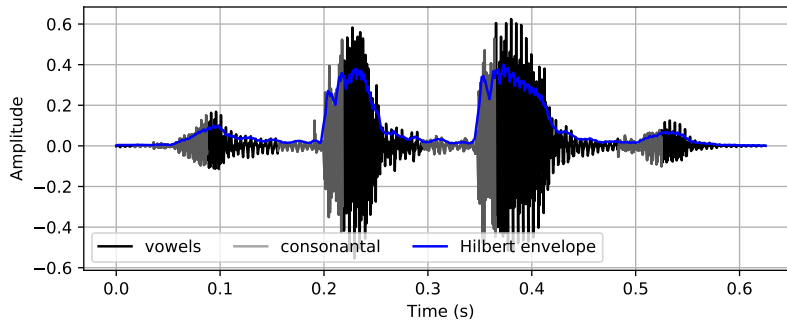


Fig. 2. Segmentation of the DDK exercises using the envelope of the Hilbert transform

2.4 Evaluation

The capability of the articulation and EMD-based feature sets to classify PD patients and HC subjects was evaluated with a SVM classifier with a Gaussian kernel. The complexity hyperparameter C and the bandwidth of the kernel γ were optimized in a grid-search with selection criterion based on the accuracy obtained in the train set, where $C \in \{10^{-5}, 10^{-4}, \dots, 10^3\}$ and $\gamma \in \{10^{-5}, 10^{-4}, \dots, 10^3\}$. In addition, the feature sets were used to predict the dysarthria level of the participants according to the m-FDA scale. The prediction was performed with a support vector regression (SVR) with a Gaussian kernel and an ε -insensitive loss function. The hyperparameter C , γ , and ε of the regressor were optimized in a grid-search with selection criterion based on the Spearman's correlation ρ obtained in the train set, where $C \in \{10^{-5}, 10^{-4}, \dots, 10^3\}$, $\gamma \in \{10^{-5}, 10^{-4}, \dots, 10^3\}$, and $\varepsilon \in \{10^{-4}, 10^{-3}, \dots, 10^1\}$. A Leave-one-out cross-validation strategy was performed for all classification and regression experiments.

3 Data

3.1 m-FDA scale

The evaluation of PD patients according to the MDS-UPDRS-III scale is suitable to assess general motor impairments of PD patients; however, the deterioration of the communication skills is not properly evaluated because such a scale only considers speech impairments in one of its items. We recently introduced the m-FDA scale, which is focused on speech impairments showed by PD patients and can be administered based on speech recordings [14]. The scale includes several aspects of speech: respiration, lips movement, palate/velum movement, larynx, tongue, monotonicity, and intelligibility. It has a total of 13 items and each of them ranges from 0 (normal or completely healthy) to 4 (very impaired), thus the range of the total score is from 0 to 52. The labeling process of the recordings was performed by three phoniatricians who agreed in the first ten speakers. Afterwards, each phoniatrician evaluated the remaining recordings independently. The inter-rater reliability among the labelers is 0.75, which was computed as the average Spearman's correlation between all pairs of labelers.

3.2 Participants

Recordings of the PC-GITA database [16] are considered in this study. The data contain speech utterances from 50 PD and 50 HC Colombian Spanish native speakers balanced in age and gender. All patients were recorded in ON state, i.e., no more than three hours after their daily medication. The DDK exercises pronounced by the subjects included the rapid repetition of the syllables /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/. Additional information from the participants is shown in Table 2. The results from the statistical tests show that the data are gender, age, and education level balanced, and that there is a significant difference between the m-FDA scores assigned for PD patients and HC subjects.

Table 2. General information of the subjects. Time since diagnosis, age and education are given in years. ^a p-value calculated through chi-square test. ^b p-value calculated through t-test.

	PD patients	Healthy controls	Patients vs. controls
Gender [F/M]	25/25	25/25	p=1.00 ^a
Age [F/M]	60.7(7.3)/61.3(11.7)	61.4(7.1)/60.5(11.6)	p=0.98 ^b
Education level [F/M]	11.5(4.1)/10.9(4.5)	11.5(5.2)/10.6(4.4)	p=0.88 ^b
Total m-FDA score [F/M]	29.8(8.6)/28.2(9)	7.6(9.2)/5.1(7.3)	p<<0.005 ^b
Time since diagnosis [F/M]	12.6(11.5)/8.7(5.8)		
MDS-UPDRS-III [F/M]	37.6(14.0)/37.8(22.1)		

4 Experiments and Results

Two experiments are performed: (1) the classification of PD patients and HC subjects using the three different feature sets and their combination, and (2) the prediction of the dysarthria severity following the m-FDA scale. Table 3 shows the results for the first experiment.

The transition, regularity, and the EMD feature sets were considered separately and also their combination using an early fusion strategy. The classification was performed with the features computed from the six DDK exercises, and also with their combination. The results suggest that the combination of transition and regularity features are the most accurate to classify PD patients and HC subjects. A detailed evaluation of the importance of the different features have been performed in related research [17]. In addition, note that the combination of the six exercises slightly improved the accuracy w.r.t. the obtained with the individual exercises per feature set. When all of the exercises and feature sets are merged the classification accuracies decrease w.r.t. those obtained with the combination of transition and regularity features only. This result indicates that EMD features are not complementary to the other two feature sets. For the separate exercises, note that the repetition of /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/ produce the highest accuracies, which can be explained because the difficulties for the patients to perform those exercises, i.e., the patients have to move the lips, the tongue and the velum continuously, while in the other exercises the patients only have to move one articulator.

Table 3. Classification of PD patients and HC subjects using the transition (**Trans.**), regularity (**Reg.**), and EMD features computed upon DDK exercises. **Acc.:** Accuracy (%), **Prec.** Precision (%), **Rec.** Recall (%), **F1.** F1-score. **Fusion.** Fusion of the features of the six DDK exercises

Exercise	Trans.				Reg.				EMD				Trans.+Reg.				Trans.+Reg.+EMD			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
/pa-ta-ka/	74	81	64	0.74	69	72	62	0.69	70	72	66	0.70	75	82	64	0.75	71	72	68	0.71
/pa-ka-ta/	67	68	64	0.67	58	59	54	0.58	53	51	57	0.54	73	76	68	0.73	67	68	64	0.67
/pe-ta-ka/	70	70	70	0.71	64	65	60	0.64	53	53	54	0.53	72	72	72	0.72	66	69	58	0.66
/pa/	61	63	52	0.61	54	54	50	0.50	52	52	56	0.52	67	7	58	0.67	66	69	58	0.66
/ta/	60	61	59	0.60	70	75	60	0.69	67	68	64	0.67	58	59	54	0.58	67	68	64	0.67
/ka/	70	79	60	0.71	64	71	68	0.73	63	70	54	0.62	72	81	58	0.71	72	79	60	0.72
Fusion	75	77	73	0.75	63	67	52	0.63	71	77	60	0.71	76	77	74	0.76	68	73	58	0.68

The features from the fusion of the six DDK exercises are used to predict the dysarthria severity of the patients according to the m-FDA scale. The results are shown in Table 4 for the combination of transition and regularity features, and for the EMD features. Similarly to the classification experiments, the highest correlation is obtained with the combination of transition and regularity features ($\rho=0.6782$). The “strong” correlation obtained is statistically significant, and it is comparable to the obtained in related studies, where the same problem was addressed [14, 18].

Table 4. Prediction of the m-FDA score of PD patients and HC subjects using transition (**Trans.**), regularity (**Reg.**) and EMD-based features computed on DDK exercises. ρ Spearman’s correlation coefficient.

Exercise	Trans.+Reg.		EMD		Trans.+Reg.+EMD	
	ρ	p-val	ρ	p-val	ρ	p-val
Fusion of exercises	0.678	$\ll 0.005$	0.500	$\ll 0.005$	0.629	$\ll 0.005$

5 Conclusion

This study evaluated and compared the performance of three different articulation features for the assessment of the speech of PD patients when they perform different DDK exercises. The first set of features are used to model the capabilities of the patients to start/stop the vibration of the vocal folds. The second feature set aimed to model the regularity of the movement of the articulators when they perform the exercises. Finally, the third feature set based on EMD features was included to evaluate the harmonic structure and the noise presence in the speech signal.

The results indicated that the combination of the first two feature sets are the most accurate to classify the presence of the disease, and also to predict the dysarthria level of the participants. The combination of different exercises seems to be more appropriate and accurate than considering DDK exercises separately. Additionally, the most complex exercises like /pa-ta-ka/ or /pa-ka-ta/, where the patients have to move more articulators showed to be more accurate than the exercises where the patients have to move only one articulator. Further research should consider an exhaustive analysis about the importance of individual articulation features for the assessment of the dysarthria severity of the patients.

Acknowledgments

The work reported here was financed by CODI from University of Antioquia by grant Number 2017–15530. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 766287.

References

1. O. Hornykiewicz. Biochemical aspects of Parkinson’s disease. *Neurology*, 51(2):S2–S9, 1998.
2. J. Rusz et al. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinsons disease. *Journal of the Acoustical Society of America*, 129(1):350–367, 2011.
3. J. R. Orozco-Arroyave, J. C. Vásquez-Correa, et al. Neurospeech: An open-source software for parkinson’s speech analysis. *Digital Signal Processing*, 77:207–221, 2018.

4. J. Hlavnicka, R. Cmejla, T. Tykalova, K. Sonka, E. Ruzicka, and J. Ruz. 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, 2017.
5. H. Ackermann and W. Ziegler. Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 54(12):1093–1098, 1991.
6. T. Tykalova, J. Ruz, J. Klempir, R. Cmejla, and E. Ruzicka. Distinct patterns of imprecise consonant articulation among parkinsons disease, progressive supranuclear palsy and multiple system atrophy. *Brain and language*, 165:1–9, 2017.
7. K. Forrest, G. Weismer, and G. S. Turner. Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America*, 85(6):2608–2622, 1989.
8. M. Novotný, J. Ruz, et al. Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM Trans. on Audio, Speech and Language Processing*, 22(9):1366–1378, 2014.
9. J. R. Orozco-Arroyave. *Analysis of speech of people with Parkinson's disease*, volume 41. Logos Verlag Berlin GmbH, 2016.
10. D. Montaña, Y. Campos-Roca, and C. J. Pérez. A diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of parkinsons disease. *Computer methods and programs in biomedicine*, 154:89–97, 2018.
11. L. Moro-Velazquez et al. A forced gaussians based methodology for the differential evaluation of parkinson's disease by means of speech processing. *Biomedical Signal Processing and Control*, 48:205–220, 2019.
12. A. Rueda and S. Krishnan. Feature analysis of dysphonia speech for monitoring parkinson's disease. In *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2308–2311. IEEE, 2017.
13. A. Rueda and S. Krishnan. Clustering parkinsons and age-related voice impairment signal features for unsupervised learning. *Advances in Data Science and Adaptive Analysis*, 10(02):1840007, 2018.
14. J. C. Vázquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth. Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease. *Journal of communication disorders*, 76:21–36, 2018.
15. N. E. Huang et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.
16. J. R. Orozco-Arroyave et al. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In *Language Resources and Evaluation Conference, (LREC)*, pages 342–347, 2014.
17. A. Rueda, J. C. Vázquez-Correa, Ríos-Uregu C. D., J. R. Orozco-Arroyave, S. Krishnan, and E. Nöth. Feature representation of pathophysiology of parkinsonian dysarthria. In *Proceedings of INTERSPEECH*, pages 1–5, 2019.
18. M. Cernak et al. Characterisation of voice quality of parkinsons disease using differential phonological posterior features. *Computer Speech & Language*, 46:196–208, 2017.