Automatic Detection of Parkinson's Disease: An Experimental Analysis of Common Speech Production Tasks Used for Diagnosis

Anna Pompili[1(✉)], Alberto Abad[1], Paolo Romano[1], Isabel P. Martins[2,4], Rita Cardoso[4,5,6], Helena Santos[4,5,6], Joana Carvalho[4,5,6], Isabel Guimarães[3,4,5], and Joaquim J. Ferreira[4,5,6]

[1] INESC-ID/IST, Lisbon, Portugal
{anna,alberto.abad}@l2f.inesc-id.pt
[2] Laboratório de Estudos de Linguagem, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
[3] Department of Speech Therapy, Escola Superior de Saúde do Alcoitão, SCML, Estoril, Portugal
[4] Instituto de Medicina Molecular, Lisbon, Portugal
[5] Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
[6] CNS - Campus Neurológico Sénior, Torres Vedras, Portugal

Abstract. Parkinson's disease (PD) is the second most common neurodegenerative disorder of mid-to-late life after Alzheimer's disease. During the progression of the disease, most individuals with PD report impairments in speech due to deficits in phonation, articulation, prosody, and fluency. In the literature, several studies perform the automatic classification of speech of people with PD considering various types of acoustic information extracted from different speech tasks. Nevertheless, it is unclear which tasks are more important for an automatic classification of the disease. In this work, we compare the discriminant capabilities of eight verbal tasks designed to capture the major symptoms affecting speech. To this end, we introduce a new database of Portuguese speakers consisting of 65 healthy control and 75 PD subjects. For each task, an automatic classifier is built using feature sets and modeling approaches in compliance with the current state of the art. Experimental results permit to identify reading aloud prosodic sentences and story-telling tasks as the most useful for the automatic detection of PD.

Keywords: Parkinson's disease · Phonation · Articulation · Prosody

1 Introduction

Parkinson's disease (PD) is a progressive degenerative disorder of the central nervous system characterized by motor and non-motor symptoms. The cardinal motor signs of PD include the characteristic clinical picture of resting tremor, rigidity, bradykinesia, and impairment of postural reflexes, while non-motor

K. Ekštein and V. Matoušek (Eds.): TSD 2017, LNAI 10415, pp. 411–419, 2017.
DOI: 10.1007/978-3-319-64206-2_46

symptoms include cognitive disorders, sleep and sensory abnormalities. These symptoms slowly worsen during the disease with a nonlinear progression. Motor symptoms of PD influence also the speech production of language. Dysarthria is typically observed in PD patients, it is characterized by weakness, paralysis, or lack of coordination in the motor-speech system, affecting respiration, phonation, articulation and prosody. The main deficits of PD speech are: loss of intensity, monotony of pitch and loudness, reduced stress, inappropriate silences, short rushes of speech, variable rate, imprecise consonant articulation and harsh and breathy voice. The standard method to evaluate and rate the neurological state of Parkinson's patients is based on the revised version, provided by the Movement Disorders Society, of the unified Parkinson's disease rating scale (MDS-UPDRS) [1]. The motor part of the MDS-UPDRS (Sect. 3) addresses speech evaluating volume, prosody, clarity and repetition of syllables. There are several speaking tasks that could be used to evaluate the extent of speech and voice disorders in PD. The most traditional of them are the sustained vowel phonation, rapid syllable repetition (diadochokinesis), and variable reading of short sentences, longer passages or freely spoken spontaneous speech [2].

Recently, the automatic detection of Parkinson's disease through speech has gained the interest of the scientific community. Current state of the art includes plenty of works targeting either the discrimination of PD patients from healthy controls (HC) [3–5] or the evaluation of the correlation among different speech characteristics and the severity of the disease [6–8]. These works differ on many aspects: on the set of features considered, on the speech tasks used for the analysis, and on the statistical approach used in the characterization of the problem. Nevertheless, there are few studies that specifically focus on comparing common speech production tasks typically used for diagnosis in terms of their utility for automatic PD discrimination [3].

The main goal of this work is to investigate the role of common speech production tasks used for the automatic detection of PD. To this end, we consider a new database of Portuguese speakers consisting of 65 healthy control (HC) and 75 PD subjects, each one of them performing 8 different speech tasks that are typically used in Speech and Language Therapy clinical evaluations. For each one of these tasks, we report individual automatic PD detection experiments based on a conventional machine learning approach. Feature extraction is based on a selection of the most representative measures typically considered in studies assessing how the symptoms of this disease affect the speech production. Thus, we are not interested in comparing the large amount of different acoustic measures and learning approaches that have emerged along the years, but rather in defining a feature set and classification strategy, based on the literature review, that can be suitable for assessing the different speech tasks. In the following, Sect. 2 reports on some recent works in the area of automatic detection of the disease. Section 3 describes the data and the speech tasks, while Sect. 4 explains the approach followed in this study. Results are presented and discussed in Sect. 5. Finally, the conclusions are provided in Sect. 6.

2 Related Work

In the last years there has been an increasing number of research works aiming at the automatic characterization and assessment of dysarthria in PD using speech. The primary focus of these studies is either discerning PD patients from HC or attempting to monitor the disease progression through the estimation of the UPDRS scale.

For instance, in Bocklet et al. [3], the authors investigate four different systems in order to assess acoustic, prosodic and voice-related features. The corpus used is composed of 88 German speaking subjects affected by PD and 88 HC subjects performing eight tasks. The proposed systems deal with different acoustic and prosodic features (MFCCs, F_0, energy, duration, pauses, jitter, and shimmer), and with the estimation of the parameters of the physical glottis. Another system performs feature extraction with openSMILE [9] considering the 1582 acoustic features of the INTERSPEECH 2010 Computational Paralinguistic Challenge (ComParE2010) baseline [10]. With the combination of all the speech tasks and the fusion of the four systems, the authors report a recognition result of 79% in discriminating HC subjects from PD patients.

In Bayestehtashk et al. [6], the authors investigated the automatic evaluation of the severity of PD from speech. The corpus used is composed of 168 English speaking patients at different stages of the disease. The recordings include three different tasks: sustained phonation of the English vowel /a/, diadochokinesis (DDK) evaluation and reading text. The classifiers are based on the ComParE2010 feature set. According to the results, the reading text and DDK tasks are the most effective to perform the evaluation of the extent of the disease. The authors reported a mean absolute error of 5.5 using the motor sub-scale of UPDRS that takes values in the [0, 108] range.

In Orozco-Arroyave et al. [5], the authors explore different acoustic measures on a set of recordings composed of the five vowels existing in the Spanish language. The corpus is composed of 50 subjects affected by PD and 50 healthy subjects, both groups are balanced by gender and age. The analysis includes several acoustic measures, among these the first two formants (F1 and F2), the pitch, the jitter, the shimmer, the vowel articulation index (VAI), the triangular vowel space area (tVSA), and three new measures based on the tVSA. Results, in agreement with previous studies, have shown that measures of the variability of the pitch are among the most important features. Also, combining articulation and phonation features led to an improvement in the results, achieving 81.3% of classification accuracy.

Finally, regarding PD for European Portuguese, Proença et al. [11] investigated acoustic and phonetic-prosodic characteristics of speech produced by PD patients while reading phonetically rich sentences and isolated words. The corpus is composed of 22 patients (12 females, 10 males) with different degrees of PD severity. Only vowels in continuous speech context were analyzed. First and second formant frequencies, vowel space area (VSA), VAI, MFCC, spectral and prosodic parameters were calculated for each speaker. Results have shown a centralization of vowel formant frequencies for PD speech, besides exploiting

acoustic, spectral and prosodic features for classifying PD speech have shown that dynamic features are of highest importance in this task.

3 Corpus Description

The FraLusoPark database [12] contains 140 European Portuguese speakers. The control group, composed of 65 healthy volunteers, is age-matched and sex-matched with the PD group, composed of 75 subjects. Patients were recorded twice, OFF medication (i.e.: at least 12 h after withdrawal of all anti-Parkinsonian drugs), and ON medication (i.e.: following at least 1 h after the administration of the usual medication).

Subjects were recorded in a quiet room, with a specialized speech recording equipment (Marantz PMD661 MKII recorder), using a unidirectional headset microphone sampled at 48 kHz with 16-bit resolution.

Participants were required to perform several speech production tasks with an increasing complexity in a fixed order: (1) three repetitions of the sustained phonation of the vowel /a/, (2) two repetitions of the maximum phonation time (vowel /a/ sustained as long as possible), (3) oral diadochokinesia (repetition of the pseudo-word *pataka* at a fast rate for 30 s.), (4–5) reading aloud of 10 words and 10 sentences, (6) reading aloud of a short text ("The North Wind and the Sun"), (7) storytelling speech guided by visual stimuli, and (8) reading aloud of a set of sentences with specific prosodic properties.

The total duration of the recordings is approximately 6 h and 31 min for the control group, and 7 h and 30 min for the PD group. Demographics data of the corpus are presented in Table 1.

Table 1. Demographic and clinical data for patient and control groups.

	PD patients		Controls	
	M	F	M	F
Gender	38	37	34	31
Age	64.6 ± 11.9	66.9 ± 8.5	62.4 ± 12.4	66.6 ± 14.4
Years diagnosed	6.7 ± 4.5	10.8 ± 5.6	–	–
MDS-UPDRS-III	32.1 ± 12.9	38.3 ± 14.5	–	–

4 Methodology

In this work, we use the database described in the previous section to conduct an analysis of the performance of automatic PD classification for each one of the 8 speech production tasks. For the purpose of this study, only recordings ON medication were considered. Data was manually preprocessed in order to remove the therapist's speech. Additionally, each spontaneous intervention introduced by the subject, that was not directly related with the task, was removed as well.

After that, recordings were down-sampled to 16 kHz. The selected feature set, described in detail in the next Sect. 4.1, has been extracted with the openSMILE toolkit [9]. The selected model is a Random Forest classifier as implemented in the WEKA toolkit [13]. This implementation relies on bootstrap aggregating, also known as bagging, a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. Bagging reduces variance and helps to avoid over-fitting. A stratified k-fold cross validation per speaker strategy is used for training and evaluation of each speech task separately, with k being equal to 5. Thus, we ensure that the train and the test sets at each iteration do not contain the same speakers. Also, the percentage of speakers of each class is balanced in the two data sets at each iteration.

4.1 Features Selected for PD Detection

Motor symptoms of PD affect also the motor-speech system, influencing the production of language at various dimensions: phonation, articulation, and prosody. Phonation problems are related with vocal fold bowing and incomplete closing of vocal folds [14,15]; articulation deficits are manifested as reduced amplitude and velocity of the articulatory movements in the lips, tongue and jaw [16]; prosody impairments comprise changes in loudness, pitch, and timing, which overall contribute to the resulting intelligibility of speech [17].

In the literature the most traditional measures used in examining phonation include measurement of F_0, jitter, shimmer, and Harmonics to Noise Ratio (HNR) [5,17]. Articulation is typically assessed considering differences in vocal tract resonances. The first and second formant frequencies and the vowel space area are frequently studied [11,18]. Prosodic analysis includes measurements of F_0, intensity, articulation rate, pause, and rhythm [3,17,19].

In this study we consider some of the measures that are repeatedly referred in the majority of the works examined. In particular, our custom set of features is reported in Table 2. First, these features are initially computed at the frame level, the so-called low-level descriptors, which are obtained based on a sub-set of the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) [20] and the MFCC pre-built configuration files. Then, in a second step, two functionals (mean and standard deviation) are applied in order to obtain a feature vector of constant length for the whole utterance. For some features, (F_0 and loudness), mean and standard deviation of the slope of rising/falling signal parts were also computed. Finally, we obtain a 114-dimensional feature vector composed of 78 MFCC based features and 36 GeMAPS based features. Notice that some other features also frequently mentioned in the literature (i.e.: the articulation rate, pause analysis, or VSA) were not considered in order to build a general purpose feature set, which could be suitable for each task under assessment.

Table 2. Description of the acoustic features based on 53 low-level descriptors plus 6 functionals.

Descriptors	Functionals
Logarithmic F_0 (1), Loudness (1)	mean and stdev, mean and stdev of the slope of rising/falling signal parts (x6)
Jitter (1), Shimmer (1), Formant 1 bandwidth (1), Formant 1, 2, 3 frequency (3), amplitude (3), Harmonic to Noise Ratio (1), Harmonic difference: H1-H2 (1), H1-A3 (1), MFCC [1–12] (12), LOGenergy (1), First and second derivative of MFCC and Log-energy (26)	mean and stdev (x2)

4.2 Sentence-Level vs. Segmental Feature Extraction

On a first attempt, the recordings of every speech production task for each speaker have been processed as described previously to obtain a feature vector of 114 elements. We refer to this approach as sentence-level feature extraction. This strategy results in a single feature vector per speaker and task. In other words, cross-validation experiments for each task are limited to only 140 sample vectors, which will probably result in poorly trained models and less reliable results. Alternatively, in order to increase the number of samples, we have also performed a segmental feature extraction strategy. In this case, we obtain a feature vector as previously described for each audio sub-segment of fixed length equal to 4 s with a time shift of 2 s. This approach permits increasing the amount of training samples for the cross-validation experiments, besides extracting more detailed information of the speech productions.

5 Experimental Results and Analysis

Table 3 shows classification accuracy (%) results for each speech production task following the two feature extraction strategies described previously: sentence-level and segmental. As expected, the former approach led to poorer results, mostly motivated by the reduced number of training samples (only 112 in each task at each cross-validation iteration). However, we also believe that we are losing valuable information when applying the functionals to long speech segments as the ones corresponding to each speech production. On the other hand, the segmental feature extraction strategy leads to very remarkable improvements in terms of classification accuracy. In particular, the reading words task achieves a maximum of 40.6% relative improvement, followed by the reading sentences task with 31.5% relative improvement.

Overall, from these results we can observe that the reading prosodic sentences task achieved the best recognition accuracy (85.10%). In fact, this is the best

Table 3. Task-dependent recognition results on the 2-class detection task (PD vs. HC).

Feature extraction results - accuracy (%)		
Task	Sentence level	Segmental
Sustained vowel phonation (/a/)	55.00	58.14
Maximum phonation time (/a/)	60.00	75.65
Rapid syllables repetitions	60.71	73.28
Reading of word	54.29	76.35
Reading of sentences	62.14	81.74
Reading of text	65.00	79.86
Storytelling guided by visual stimuli	66.43	82.32
Reading of prosodic sentences	70.71	85.10

performing task also in the case of sentence-level feature extraction. This observation confirms the relevance of this task, which was carefully designed in order to explore language-general and language-specific details of PD dysprosody. The second most discriminant task in terms of automatic PD classification is the storytelling one (82.32%). As a matter of fact this task corresponds to the production of spontaneous speech, since the subject has to create a story based on temporal events represented in a picture. Although its overall duration is extremely variable and dependent on the speaker, this task definitely contains many important acoustic and prosodic information. This result is very encouraging for the development of tele-monitoring applications that may use spontaneous speech recorded over the telephone. The next most discriminant tasks are those consisting of reading short passages of text and sentences. Again, we believe that these productions are richer in terms of acoustic and prosodic information, which makes them more convenient for automatic PD detection in contrast to less informative rapid syllables repetitions or maximum phonation time of vowel /a/. In general, it is likely that more complex tasks will contain more linguistics phenomena, like for instance co-articulations, that may provide important cues for discrimination. Moreover, these more complex tasks consist generally of longer speech productions, which is expected to be beneficial for the segmental feature extraction approach. Nevertheless, we also note that both feature extraction strategies provide coherent results in terms of identifying the top-4 most significant speech production tasks. Finally, we observe that the sustained phonation of vowels /a/ is the task that achieved the worst results with the segmental approach by a large margin (58.14%). From a quick analysis we notice that this task is the one with shortest speech productions, resulting in less speech segments. Anyway, this result deserves a deeper analysis, since it is in contradiction with some previous observations for other languages.

6 Conclusions

In this work, we have analyzed the potential discriminative ability of a large set of common speech production tasks for the automatic detection of PD. For this purpose, we considered a database containing European Portuguese PD and HC speakers performing 8 tasks designed to assess speech disorders at various dimensions. For each task, automatic classification experiments have been conducted using a Random Forest classifier and a custom set of acoustic features carefully selected based on the study of the state of the art. The experimental results have shown that the most important production tasks are reading of prosodic sentences and storytelling, achieving a PD classification accuracy of 85.10% and 82.32%, respectively. Future work includes the analysis of conversational speech production tasks, besides the use of i-vector based classifiers, which is the current state of the art in speech recognition tasks such as speaker or language recognition.

Acknowledgments. This work was supported by Portuguese national funds through – Fundação para a Ciência e a Tecnologia (FCT), under Grants SFRH/BD/97187/2013 and Projects with reference UID/CEC/50021/2013 and CMUP-ERI/TIC/0033/2014.

References

1. Movement disorder society task force on rating scales for Parkinson's disease. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations (2003)
2. Goberman, A.M., Coelho, C.: Acoustic analysis of Parkinsonian speech I: speech characteristics and L-Dopa therapy. NeuroRehabilitation **17**(3), 237–246 (2002)
3. Bocklet, T., Steidl, S., Nöth, E., Skodda, S.: Automatic evaluation of Parkinson's speech-acoustic, prosodic and voice related cues. In: Interspeech, pp. 1149–1153 (2013)
4. Orozco-Arroyave, J.R., Hönig, F., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Skodda, S., Rusz, J., Nöth, E.: Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease. In: Interspeech, pp. 95–99 (2015)
5. Orozco-Arroyave, J.R., Belalcázar-Bolaños, E.A., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Haderlein, T., Nöth, E.: Phonation and articulation analysis of Spanish vowels for automatic detection of Parkinson's disease. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2014. LNCS, vol. 8655, pp. 374–381. Springer, Cham (2014). doi:10.1007/978-3-319-10816-2_45
6. Bayestehtashk, A., Asgari, M., Shafran, I., McNames, J.: Fully automated assessment of the severity of Parkinson's disease from speech. Comput. Speech Lang. **29**(1), 172–185 (2015)
7. Arias-Vergara, T., Vasquez-Correa, J., Orozco-Arroyave, J.R., Vargas-Bonilla, J.F., Nöth, E.: Parkinson's disease progression assessment from speech using GMM-UBM. In: Interspeech, pp. 1933–1937 (2016)
8. Orozco-Arroyave, J.R., Vasquez-Correa, J., Hönig, F., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Skodda, S., Rusz, J., Noth, E.: Towards an automatic monitoring of the neurological state of Parkinson's patients from speech. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6490–6494. IEEE (2016)

9. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the Munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM International Conference on Multimedia, MM 2010, pp. 1459–1462. ACM, New York (2010)
10. Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S.: The INTERSPEECH 2010 paralinguistic challenge. In: Interspeech (2010)
11. Proença, J., Veiga, A., Candeias, S., Perdigão, F.: Acoustic, phonetic and prosodic features of Parkinson's disease speech. In: STIL-IX Brazilian Symposium in Information and Human Language Technology, 2nd Brazilian Conference on Intelligent Systems, Brazil (2013)
12. Pinto, S., Cardoso, R., Sadat, J., Guimarães, I., Mercier, C., Santos, H., Atkinson-Clement, C., Carvalho, J., Welby, P., Oliveira, P., D'Imperio, M., Frota, S., Letanneux, A., Vigario, M., Cruz, M., Martins, I.P., Viallet, F., Ferreira, J.J.: Dysarthria in individuals with Parkinson's disease: a protocol for a binational, cross-sectional, case-controlled study in French and European Portuguese (FraLusoPark). BMJ Open **6**(11), e12885 (2016)
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. Newsl. **11**(1), 10–18 (2009)
14. Hanson, D.G., Gerratt, B.R., Ward, P.H.: Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope **94**(3), 348–353 (1984)
15. Perez, K.S., Ramig, L.O., Smith, M.E., Dromey, C.: The Parkinson larynx: tremor and videostroboscopic findings. J. Voice **10**(4), 354–361 (1996)
16. Skodda, S., Visser, W., Schlegel, U.: Vowel articulation in Parkinson's disease. J. Voice **25**(4), 467–472 (2011)
17. Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J. Acoust. Soc. Am. **129**(1), 350–367 (2011)
18. Vásquez-Correa, J., Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Nöth, E.: Design and implementation of an embedded system for real time analysis of speech from people with Parkinson's disease. In: Symposium of Signals, Images and Artificial Vision - 2013, STSIVA - 2013, pp. 1–5, September 2013
19. Skodda, S., Schlegel, U.: Speech rate and rhythm in Parkinson's disease. Mov. Disord. **23**(7), 985–992 (2008)
20. Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., Andr, E., Busso, C., Devillers, L.Y., Epps, J., Laukka, P., Narayanan, S.S., Truong, K.P.: The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Trans. Affect. Comput. **7**(2), 190–202 (2016)