Phonological i-vectors to detect Parkinson's Disease

N. Garcia-Ospina[1], T. Arias-Vergara[1,2*], J. C. Vásquez-Correa[1,2*], J. R. Orozco-Arroyave[1,2], M. Cernak[3], and E. Nöth[2]

[1]Faculty of Engineering, University of Antioquia UdeA, Medellín, Colombia.
[2]Pattern Recognition Lab, University of Erlangen-Nrnberg, Erlangen, Germany.
[3]Logitech Europe S.A., Lausanne, Switzerland.
nicanor.garcia@udea.edu.co

Abstract. Speech disorders are common symptoms among Parkinson's Disease patients. These disorders affect the speech of patients in different aspects. Currently, there are few studies that consider the phonological dimension of Parkinson's speech. In this work we use a recently developed method to extract phonological features from speech signals. These features are based on the Sound Patterns of English phonological model. The extraction is performed using pre-trained Deep Neural Networks to infer the probabilities of phonological features from short-time acoustic features. An i-vector extractor is trained with these phonological features. We classify patients and healthy speakers and evaluate the dysarthria levels of the patients. This approach could be helpful to assess new specific speech aspects such as the movement of different articulators involved in the speech production process.

Key words: Parkinson's disease, phonological features, i-vectors

1 Introduction

Parkinson's disease (PD) is the second most common neuro-degenerative disorder worldwide after Alzheimer's [1]. PD patients suffer several motor and non-motor impairments. Among the motor symptoms, the most prominent are tremor, rigidity, slowed movement, postural instability, lack of coordination and different speech impairments. These symptoms limit the mobility and communication skills of patients, making it hard for them to attend appointments and therapy, and to adequately convey their symptoms to their physicians and caregivers [2]. Most PD patients develop hypokinetic dysarthria during the course of the disease, which include a group of speech disorders such as reduced loudness, monopitch, monoloudness, reduced stress, breathy, hoarse voice quality, and imprecise articulation. The disease severity is evaluated by neurologist experts following several tests. One of them is the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [3]. This is a perceptual scale used to assess motor and non-motor abilities of PD patients. In the third section of this scale, motor impairments are evaluated. As PD affects several aspects of

speech [4], it makes sense to model motor capabilities from speech considering different dimension such as phonation, articulation, prosody, and intelligibility [5, 6]. In recent years, the scientific community has been developing computer based aids to help physicians with the detection and evaluation of the disease. Different features extracted from the speech signal have been proposed to perform this automatic analysis on different dimensions of the affected speech: Phonation impairments in PD patients includes inadequate closing of the vocal fold and vocal fold bowing [7], which generates stability and periodicity problems in vocal fold vibration. Phonation in PD was automatically analyzed in [8], where features related to perturbation, noise content, and non-linear dynamics were used to evaluate whether the response of 14 PD patients to the Lee Silverman voice treatment is acceptable or unacceptable. The authors considered only information from sustained vowels, and reported an accuracy close to 90% when discriminating between acceptable vs. unacceptable utterances. The articulation problems are mainly related with reduced amplitude and velocity of the articulator movements [9], generating a reduced articulatory capability in PD patients to produce vowels [10] and to produce continuous speech. In [5] the authors modeled six different articulatory deficits in PD analyzing a diadochookinetic speech task uttered by 24 Czech native speakers, and reported an accuracy of 88% discriminating between PD patients and HC speakers. Prosody refers to intonation, loudness, and rhythm during continuous speech. Prosodic problems in PD patients includes a decrease in loudness and low variations of pitch, which is related to the frequency of vocal fold vibration (F0) [11, 12]. Prosody features were computed in [13]. The authors consider voiced segments as speech unit to compute features based on the F0 contour, energy contour, duration, and pitch periods to classify PD patients and HC speakers, and to classify the patients according to their neurological state in a 3-class approach (low, intermediate, and severe) state. The authors reported an accuracy of up to 74 classifying PD patients and HC speakers, and of 37% for the 3-class problem. Phonology studies the sounds of a language, e.g., the pronunciation of words. Few studies have analyzed the speech production of PD patients in phonological terms, and they focus on evaluating phonology from a neurological point of view. This is in part due to the difficulty in reliably estimate phonological features. Recently, a method to reliably estimate phonological features was proposed in [14]. These phonological features could be used in the analysis of dysarthric speech to assess the movements and capacity of specific articulators and parts of the speech production system. This method was used in [15] to evaluate the voice quality of PD patients.

In this work, we propose to extract the phonological features with the previously mentioned method [14] and model them using the i-vector approach. These will be referred to as phonological i-vectors. The proposed model is tested in three scenarios: (1) the classification of PD patients vs. HC subjects, (2) the assessmento of the neurological state of the patients following the MDS-UPDRS-III scale, and (3) the assessent of the dysartrhia level of the patients followin a modified version of the Frenchay Dysartrhia assessment (m-FDA) scale.

2 Methods

2.1 Phonological Features

Phonological features were extracted using the deep learning approach from [14]. This process involves the following steps: (1) the speech signal is segmented into short-time frames, (2) 13 MFCCs and their derivatives, are computed for every frame of the speech signal, and (3) a set of 15 pre-trained DNNs infers the phonological posteriors from the acoustic feature vector. These posteriors are concatenated into a phonological feature vector z_t. The process is summarized in Figure 1, where X is the set of acoustic features and Z is the set of phonological features. A total of 15 phonological features are computed. Table 1 indicates a brief description of each feature.

Table 1. List of phonological features

Feature	Brief description
Vocalic	Refers to the vocal folds vibration without constriction in the vocal tract.
Consonantal	Indicates sounds where there is an obstruction of the vocal tract.
High	The body of the tongue is above its neutral position.
Back	The body of the tongue is retracted from its neutral position.
Low	The body of the tongue is below its neutral position.
Anterior	Indicates an obstruction located in front of the palato-alveolar region of the mouth.
Coronal	The blade of the tongue is raised from its neutral position.
Round	Refers to narrowed lips.
Rising	Differentiates diphthongs from monophthongs.
Tense	Indicates stressed vowels.
Voice	Indicates voiced sounds.
Continuant	Differentiates plosives from non-plosives.
Nasal	Indicates a lowered velum, where the air to escape through the nose.
Strident	Refers to sounds with more energy in high frequency components.
Silence	Tells that there is no speech in the frame.

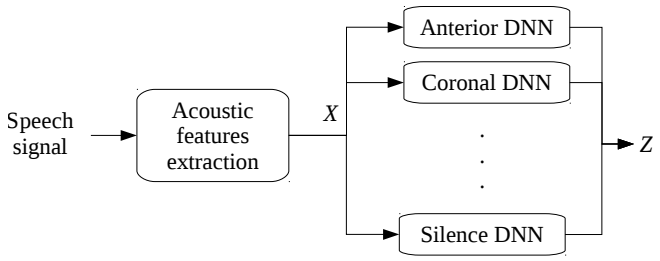

Fig. 1. Phonological feature extraction process (adapted from [14])

2.2 i-vector extraction

In the i-vector approach, factor analysis is used to define a new low-dimensional space known as the total variability space. Initially, for speaker verification applications, this space had the aim of modeling the speaker and the channel variability [16]. In pathological speech analysis applications, the speaker variability carries the information about the disorders in speech due to the disease.

The extracted i-vectors are processed in five steps: (1) the i-vectors computed from utterances of the same speaker are averaged to obtain one i-vector per speaker, (2) a whitening is process is applied by subtracting the mean of the training i-ivectors and performing a Principal Component Analysis [17]. No further processing such as PLDA is applied to the i-vectors as all the speech signals used in this study were recorded in similar acoustic conditions.

2.3 Classification methods

Cosine distance threshold The score computed from a test signal is compared with respect to a threshold θ. The score used in the i-vector approach is the average cosine distance. The cosine distance is used to compare two i-vectors. It considers only the "angle" between two i-vectors. In this case, the average cosine distance to a reference set of i-vectors is computed according to Equation 1.

$$\text{score}(w_{\text{test},j}) = \frac{1}{N}\sum_{i=1}^{N} C_i \frac{w_{\text{test},j} \cdot w_{\text{ref},i}}{||w_{\text{test},j}||||w_{\text{ref},i}||} \tag{1}$$

where C_i is the condition label of the reference i-vector: 1 for HC and -1 for PD patients. The larger the distance to the HC means the more affected the speech. With this is in mind, the condition is if $\text{score}(w_{\text{test},j}) > \theta$ it is considered from a PD patient. Also taking the previous argument into consideration, the threshold was set at $\theta = 0$. No parameters needed to be optimized for this method.

Support Vector Machines (SVMs) The goal of a SVM is to discriminate data points by using a separating hyperplane which maximizes the margin between two classes. A soft margin Support Vector Machine (SVM) with Gaussian kernel is used to classify PD vs. HC subjects. Two hyper-parameters need to be optimized in this classifier: the margin cost C and the bandwidth of the Gaussian kernel γ. Details of this optimization are given in Section 4.

2.4 Evaluation of the neurological state

The prediction of the neurological state and the dysarthria level are evaluated with the Spearman's correlation between the real score and the score given by equation 1. In this case, two different reference sets are used: the first includes i-vectors only from HC and the second is formed with i-vectors from PD patients only. The condition label is set $C_i = 1$ for all the i-vectors in the reference set.

3 Data

In this work the i-vector extractor was trained using a speech corpus collected for Speaker Verification. This corpus contains recordings from 103 young healthy native Colombian Spanish speakers. The speakers were asked to read aloud ten short utterances ten times each. The configuration with the lowest EER was selected for this experiment: an UBM with 64 Gaussians and 100-dimensional i-vectors. This i-vector configuration is used to extract i-vectors from the speech signals of PD patiens and age balanced Healthy Controls (HC) described in Section 4.

3.1 PC-GITA speech corpus

The PC-GITA speech corpus contains recordings of 50 PD patients (25 male and 25 female) and 50 healthy controls (HC), all of them native Colombian Spanish speakers. The recordings were captured in a sound-proof booth using professional audio equipment. The original sampling frequency was 44.1 kHz, but the recordings were down-sampled to 16 kHz for this study. During the recordings, the participants were asked to perform different speech tasks including ten read short sentences. All the patients were diagnosed by a neurologist expert and their neurological state was assessed according to the MDS-UPDRS [3]. Additional information of this corpus can be found in [18].

3.2 m-FDA scale

The evaluation of the neurological state 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 of the PD patients is only evaluated in one of its 33 items. A modified version of the FDA scale (m-FDA) based only on speech recordings was developed in [15, 6]. This modified scale includes several aspects of speech: respiration, lips movement, palate/velum movement, larynx, tongue, monotonicity, and intelligibility. The scale has a total of 13 items and each of them ranges from 0 (normal or completely healthy) to 4 (very impaired), thus the total score of the scale ranges from 0 to 52. The labeling process of the recordings of the PC-GITA database 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.

4 Experiments and results

The experimental methodology used in this work comprises the following steps: 1) The phonological features from the peech signals are extracted, 2) the phonological features from training signals are used to train an i-vector extractor, and 3) i-vectors are extracted from the features of test signals and are processed. This is summarized in figure 2.

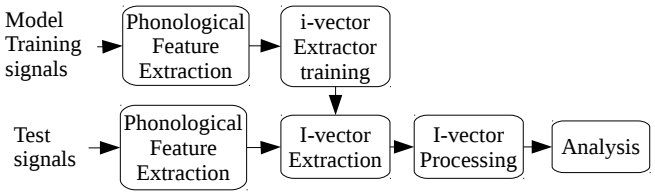

Fig. 2. General experimental methodology

4.1 Experimental setup and validation

For this study the data is randomly split for training and test as follows: 60% is used to perform training and development and the remaining 40% is used to test. Speaker independence is guaranteed between training and test sets. The optimization the hyper-parameters of the classifiers (if needed) is done by performing a 5-fold cross validation over the 60% training data. The best hyper-parameters found in the this process are then used to perform the test. To validate the results, this process is repeated ten times. The train and test sets are randomly chosen on each repetition of the experiment. The mean and standard deviation over the ten iterations are reported.

As a baseline, the phonological features are modeled with four functionals (mean, standard deviation, skewness, and kurtosis) computed from all the phonological features of a given speaker to form a 60-dimensional feature vector. A SVM classifier is trained on these features vectors [4].

4.2 Classification results

Table 2. PD vs. HC classification results

Method	**Accuracy** [%]	**Sensitivity** [%]	**Specificity** [%]	**F1-score**
Baseline	55.2 ± 3.1	49.0 ± 14.8	61.5 ± 14.3	0.51 ± 0.10
Threshold	77.5 ± 7.3	77.0 ± 12.1	78.0 ± 8.4	0.77 ± 0.08
i-vectors SVM	73.5 ± 8.2	64.0 ± 16.7	83.0 ± 8.7	0.70 ± 0.11

The results in Table 2 show that the threshold classification method has a better accuracy and better sensitivity. That indicates that is more capable of correctly classifying PD patients from HC. This would be more desirable in a clinical setting, where further tests could be used to discard a false positive.

4.3 Estimation of neurological state and dysarthria level

As mentioned in Section 2, two different reference i-vector sets are considered for this experiment. One comprises the i-vectors from the HC of the training

set and the other the PD speakers in the training set. The assessment of the neurological state and dysarthria level of a patient is evaluated by computing the Spearman's correlation coefficient between the MDS-UPDRS-III or m-FDA label and the average cosine distance.

Table 3. Spearman's correlation

Label	HC Ref.	PD Ref.
MDS-UPDRS-III	0.646 ± 0.101	-0.649 ± 0.099
m-FDA	0.581 ± 0.063	-0.574 ± 0.067

The results in table 3 show that the phonological i-vectors average cosine distance is more correlated to the neurological state of the patient than to the phonological evaluation. The negative correlation found when using the PD reference set is consistent with the hypothesis that a more affected speech has a larger cosine distance to the reference i-vectors.

5 Conclusion

In this work we introduced the use of phonological i-vectors extracted from the speech of PD patients and age balanced HC to perform the classification of PD vs. HC and to estimate their neurological state and their dysarthria level. These i-vectors are extracted from phonological posteriors obtained using pre-trained DNNs.

The average cosine distance had better classification results than the SVM. One of the main advantages of this approach is that it requires less parameters to be optimized than other like those based on neural networks. Additionally, it can be used to assess the neurological state and dysarthria.

Future work includes modeling subsets of the phonological features with i-vectors to assess specific items in the m-FDA scale. This can help in obtaining interpretable results such that are suitable to guide the phoniatrician or clinician when defining the patient's therapy. Also, we want to test the language independence asssertion about the phonological features and perform similar analyses in different languages.

Acknowledgments

The work reported here was financed by CODI from University of Antioquia by grants Number 2015–7683. This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 766287.

References

1. Sveinbjornsdottir, S.: The clinical symptoms of Parkinson's disease. Journal of Neurochemistry (139) (jul 2016) 318–324
2. Stamford, J.A., Schmidt, P.N., Friedl, K.E.: What engineering technology could do for quality of life in Parkinson's disease: A review of current needs and opportunities. IEEE Journal of Biomedical and Health Informatics **19**(6) (nov 2015) 1862–1872
3. C. G. Goetz et al.: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Movement Disorders **23**(15) (2008) 2129–2170
4. Orozco-Arroyave, J.: Analysis of speech of people with Parkinson's disease. 1st edn. Logos-Verlag, Berlin, Germany (2016)
5. Novotnỳ, M., et al.: Automatic evaluation of articulatory disorders in Parkinson's disease. IEEE/ACM Trans. on Audio, Speech and Language Processing **22**(9) (2014) 1366–1378
6. Orozco-Arroyave, J.R., Vásquez-Correa, J.C., et al.: Neurospeech: An open-source software for Parkinson's speech analysis. Digital Signal Processing (In press) (2017)
7. Hanson, D.G., et al.: Cinegraphic observations of laryngeal function in Parkinson's disease. The Laryngoscope **94**(3) (1984) 348–353
8. Tsanas, A., Little, M.A., Fox, C., Ramig, L.O.: Objective automatic assessment of rehabilitative speech treatment in parkinson's disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering **22**(1) (2014) 181–190
9. Ackermann, H., Ziegler, W.: Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. Journal of Neurology, Neurosurgery & Psychiatry **54**(12) (1991) 1093–1098
10. Skodda, S., Visser, W., Schlegel, U.: Vowel articulation in parkinson's disease. Journal of Voice **25**(4) (2011) 467–472
11. Ho, A.K., et al.: Speech impairment in a large sample of patients with Parkinson's disease. Behavioural neurology **11**(3) (1999) 131–137
12. Darley, F.L., et al.: Differential diagnostic patterns of dysarthria. Journal of Speech, Language, and Hearing Research **12**(2) (1969) 246–269
13. Bocklet, T., et al.: Automatic Evaluation of Parkinson's Speech – Acoustic, Prosodic and Voice Related Cues. In: Annual Conference of the International Speech Communication Association. (2013) 1149–1153
14. Cernak, M., Potard, B., Garner, P.N.: Phonological vocoding using artificial neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (April 2015) 4844–4848
15. Cernak, M., Orozco-Arroyave, J., Rudzicz, F., Chirstensen, H., Vásquez-Correa, J., Nöth, E.: Characterisation of voice quality of Parkinsons disease using differential phonological posterior features. Computer Speech & Language **46** (2017) 96–208
16. Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-End Factor Analysis for Speaker Verification. IEEE Transactions on Audio, Speech, and Language Processing **19**(4) (may 2011) 788–798
17. Garcia-Romero, D., Espy-Wilson, C.: Analysis of i-vector Length Normalization in Speaker Recognition Systems. In: Proceedings of the 12th INTERSPEECH. (sep 2011)
18. Orozco, J.R., Arias, J.D., Vargas, J.F., González Rátiva, M.C., Nöth, E.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of the 9th LREC. (2014) 342–347