Language Independent Assessment of Motor Impairments of Patients with Parkinson's Disease Using i-Vectors

N. Garcia[1(✉)], J.C. Vásquez-Correa[1,2], J.R. Orozco-Arroyave[1,2], N. Dehak[3], and E. Nöth[2]

[1] Faculty of Engineering, Universidad de Antioquia UdeA, Medellín, Colombia
{nicanor.garcia,jcamilo.vasquez,rafael.orozco}@udea.edu.co

[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
noeth@informatik.uni-erlangen.de

[3] Center for Language and Speech Processing, Jhons Hopkins University, Baltimore, USA
najim@jhu.edu

Abstract. Speech disorders are among the most common symptoms in patients with Parkinson's disease. In recent years, several studies have aimed to analyze speech signals to detect and to monitor the progression of the disease. Most studies have analyzed speakers of a single language, even in that scenario the problem remains open. In this study, a cross-language experiment is performed to evaluate the motor impairments of the patients in three different languages: Czech, German and Spanish. The i-vector approach is used for the evaluation due to its capability to model speaker traits. The cosine distance between the i-vector of a test speaker and a reference i-vector that represents either healthy controls or patients is computed. This distance is used to perform two analyses: classification between patients and healthy speakers, and the prediction of the neurological state of the patients according to the MDS-UPDRS score. Classification accuracies of up to 72% and Spearman's correlations of up to 0.41 are obtained between the cosine distance and the MDS-UPDRS score. This study is a step towards a language independent assessment of patients with neuro-degenerative disorders.

Keywords: Parkinson's disease · i-vectors · UPDRS score · Language independent assessment

1 Introduction

Parkinson's disease (PD) is a neuro-degenerative disorder which produces several motor and non-motor impairments. The motor symptoms include, among others tremor, rigidity, slowed movement, postural instability, lack of coordination and speech disorders [1]. Evaluating the condition of PD patients is difficult. Mobility

K. Ekštein and V. Matoušek (Eds.): TSD 2017, LNAI 10415, pp. 147–155, 2017.
DOI: 10.1007/978-3-319-64206-2_17

problems make attending medical appointments burdensome, while speech disorders may hinder the communication with the medical experts [2]. Currently, the assessment of the disease in the motor capabilities is evaluated with the third section of the Movement Disorder Society, Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [3]. This evaluation is subject to a clinical criterion and its intra- and inter-rater variability could be high. The diagnosis could be supported by computer aided systems, which could also improve the evaluation of the disease progression. On the other hand, only two of the 33 items of the MDS-UPDRS are related to the speech impairments of patients; however, speech disorders are among the most prevalent, and an early sign of further motor impairments [4]. In that way, speech signals could be used to assess the motor symptoms of PD patients.

There has been interest in the scientific community to develop computer aided tools to evaluate the condition of PD patients using information from speech. In the 2015 INTERSPEECH Computational Paralinguistics Challenge (ComPARE) the task of predicting the MDS-UPDRS score of PD patients from speech was addressed [5]. Speech recordings of 50 PD patients from the PC-GITA database [6] were considered for the train and development subsets. Recordings from eleven new patients were considered as the test set. All the speakers were native Spanish speakers. A Spearman's correlation coefficient of 0.39 was reported as baseline of the challenge. The winners of the challenge [7] grouped the speech tasks of each patient and used deep neural networks and Gaussian processes for the prediction, obtaining a correlation coefficient of up to 0.69. In [8] the authors classify the speech of PD patients vs. healthy controls (HC) speakers in different languages. Cross-language experiments were performed using data in three languages: Czech, German and Spanish. The reported accuracies range from 60% to 77%, depending on the languages used for train and test sets. In [9] a speaker model based on Gaussian mixture models-universal background models (GMM-UBM) was proposed to monitor the neurological state of PD patients. UBMs were trained with information from 61 PD and 50 HC speakers. Specific GMMs were adapted for seven PD patients recorded in three sessions. Then, the Bhattacharyya distance between the speaker models and the UBM was computed and correlated with the MDS-UPDRS score of the patients. A Pearson's correlation of up 0.60 was reported by the authors. In [10] the authors proposed a new regression method to track the progression of speech disorders. The method is based on a non-parametric learning strategy based on a probability distance measure between the speakers from the test and training sets. The authors consider data from 61 PD patients to predict the UPDRS score, obtaining a Pearson's correlation of up to 0.58.

Speaker models inspired by speaker verification and identification systems have shown promising results in evaluating PD from speech. The most recent breakthrough in speaker verification is the i-vector approach [11]. This strategy has also proven to be effective in many other speech tasks. Specially, it has shown excellent results in language identification [12,13]. Recently, i-vectors were used to identify the native language of a speaker from recordings in a second

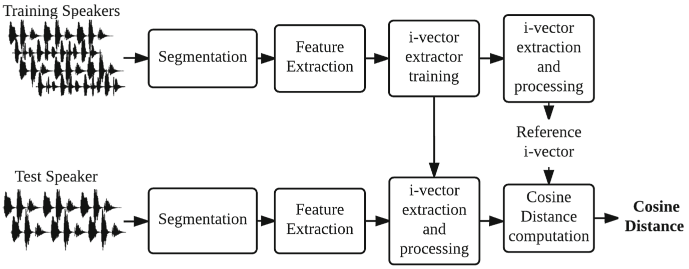
Training Speakers
Segmentation
Feature Extraction
i-vector extractor training
i-vector extraction and processing
Reference i-vector
Test Speaker
Segmentation
Feature Extraction
i-vector extraction and processing
Cosine Distance computation
Cosine Distance

Fig. 1. General methodology followed in this study.

language [14]. According to the reviewed literature, a language independent strategy to evaluate the PD condition from speech has not been enough addressed. Developing a computer aided system that can evaluate PD from speech in different languages would be a major step towards an unified objective assessment of the disease. Additionally, the use of i-vectors has been successfully applied to model speaker traits in multiple languages, which indicates that it could also be used to evaluate the PD condition from speech in a cross-language approach. In this study, a strategy based on i-vectors is used to assess the PD condition from speech in three languages: Spanish, German and Czech. Cross-language experiments are performed, i.e., train the models with utterances from one language, and test with the speech recordings from the other ones. The proposed approach is tested in two scenarios: (1) classification of PD vs. HC subjects, and (2) the prediction of the MDS-UPDRS score of the patients. Different i-vector extractors are trained with features related to specific dimensions of speech, e.g., phonation, articulation and prosody with the aim of evaluating the information provided by each dimension to represent the PD condition of the patients.

2 Methods

The methodology proposed in this study comprises four steps: (1) several feature sets are computed to analyze different speech dimension from speech, (2) a subset of speakers are used to train an i-vector extractor, (3) the i-vectors of speech signals are extracted, and (4) the cosine distance between a reference i-vector and the speaker i-vector is computed. This process is summarized in Fig. 1.

2.1 Feature Extraction

Four feature sets were considered in this study to model the speech impairments of PD patients. The first set comprises the Mel-Frequency Cepstral Coefficients (MFCCs), which are the classical features used to train i-vectors. 19 MFCCs and the log-energy extracted from 30 ms windows with time-shift of 15 ms were used to form a 20-dimensional feature vector. Non-speech frames were discarded using an energy-threshold voice activity detector (VAD). The other feature sets are formed with descriptors to assess the articulation, phonation, and prosody dimensions of speech. To evaluate articulation, the energy content in 22 Bark bands (BBE) in the voiced/unvoiced and unvoiced/voiced transitions were considered, as in [15]. The features considered to evaluate phonation and articulation in voiced segments are: the log-energy, the fundamental frequency (F_0), first and second formants (F_1 and F_2) and their first and second derivatives. Additionally, perturbation features such as Jitter and Shimmer are also included. These descriptors form a 14-dimensional feature vector. These features were computed from voiced segments using 30 ms long analysis frames with a time-shift of 5 ms. To evaluate prosody we followed the approach introduced in [16]: The log-F_0 and the log-energy contours within analysis frames were approximated using Lagrange polynomials of order $P = 5$. Analysis frames of 200 ms with time-shift of 50 ms were used as in [13]. A 13-dimensional feature vector is formed concatenating the six coefficients computed from the log-F_0 and the log-energy contours, along with the number of voiced frames in the utterance.

2.2 i-Vectors

In this approach, factor analysis is used to define a new low-dimensional space known as the total variability space with the aim of modeling the speaker and the channel variability [11]. For applications related to pathological speech analysis, the speaker variability carries the information about the disorders in speech due to the disease. In the total variability space, an utterance is represented by a supervector M formed by concatenating the mean vectors of a GMM-UBM. The total variability space is expressed according to Eq. 1, where m is a speaker and channel independent supervector (the UBM), T is the total variability matrix and w corresponds to the i-vector.

$$\mathbf{M} = \mathbf{m} + \mathbf{Tw} \tag{1}$$

The i-vectors are processed in five steps: (1) i-vectors extracted from training speakers are normalized to zero mean and unit variance, i.e., Z-norm, (2) the normalized i-vectors computed from different speech tasks of a given speaker are averaged to obtain one i-vector per speaker, (3) the i-vectors of HC and PD

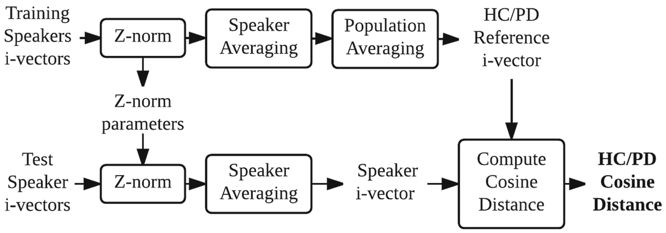

Fig. 2. i-vector processing.

speakers are averaged to obtain HC and PD reference i-vectors, respectively, (4) the i-vectors of a test speaker are normalized using the parameters from the training i-vectors, (5) the normalized i-vectors per utterace are averaged to obtain the speaker i-vector. Finally, the cosine distance between the HC/PD reference i-vectors and the speaker i-vectors is computed. The process is summarized in Fig. 2.

2.3 Cosine Distance

The cosine distance is used to compare two i-vectors w_1 and w_2. The distance is defined by Eq. 2. In this study, the i-vector of a single speaker is compared with a reference i-vector that represents the HC or PD population.

$$d_c(\mathbf{w}_1, \mathbf{w}_2) = 1 - \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{||\mathbf{w}_1||||\mathbf{w}_2||}. \tag{2}$$

2.4 Evaluation

The cosine distance between the test speaker i-vector and the reference i-vector is compared to a threshold to discriminate between PD patients and HC speakers. The development set is used to find the threshold that maximizes the accuracy. The prediction of the neurological state of a patient is evaluated using the Spearman's correlation coefficient between the real MDS-UPDRS score and the distance measure.

3 Data

Spanish- The PC-GITA database [6] is used in this study. It contains recordings of 50 PD patients and 50 healthy control (HC) speakers. All of them are native Colombian Spanish speakers. During the recordings, the participants were asked to perform different speech tasks including reading ten isolated sentences, and the repetition of /pa-ta-ka/, a diadochokinetic (DDK) exercise.

German- The German data contain recordings from 88 PD patients and 88 HC subjects. The speakers perform several speech tasks, including the repetition of /pa-ta-ka/, and reading five isolated sentences [17].

Czech- The Czech data are formed with recordings from 20 PD patients and 15 HC subjects. The patients were recorded at the time of diagnosis with PD, and none of them had been medicated before or during the recording session. The speech tasks performed by the speakers include the rapid repetition of /pa-ta-ka/, and several repetitions of a sentence [4].

4 Experiments and Results

Two speech tasks were analyzed independently in these experiments: the rapid repetition of /pa-ta-ka/, and read sentences. Data from the three languages are used in turn as training, development and test sets. All possible combinations are tested. The training data are used for several processes: (1) to train the UBM and the i-vector extractor, (2) to compute the normalization parameters, and (3) to obtain the HC and PD reference i-vectors. UBMs with different number of Gaussian components were trained in a range from $M = 2$ to $M = 2^9$ into powers of 2. The dimension of the i-vector dim_w was chosen following the relation $\text{dim}_w = \log_2(M) \cdot \text{dim}_f$, where M is the number of Gaussian components in the UBM and dim_f is the dimension of the feature vector.

Table 1. Accuracies (%) for the classification task.

Train Lang.	Test Lang.	HC reference				PD reference			
		MFCCs	Art.	Phon.	Pros.	MFCCs	Art.	Phon.	Pros.
DDK									
Czech	German	47.4	47.4	52.6	53.7	47.4	47.4	52.6	53.7
	Spanish	43.0	48.0	44.0	50.0	43.0	48.0	44.0	50.0
German	Czech	58.8	47.1	**67.6**	58.8	58.8	47.1	**67.6**	58.8
	Spanish	50.0	48.0	60.0	50.0	50.0	48.0	60.0	50.0
Spanish	Czech	**61.8**	47.1	58.8	**61.8**	**61.8**	47.1	58.8	**61.8**
	German	52.6	40.6	50.9	53.7	52.6	40.6	50.9	53.7
Sentences									
Czech	German	55.1	54.0	57.4	53.4	55.1	54.0	57.4	53.4
	Spanish	60.0	50.0	50.0	50.0	60.0	50.0	50.0	50.0
German	Czech	**72.2**	58.3	44.4	55.6	63.9	58.3	47.2	55.6
	Spanish	**68.0**	55.0	50.0	50.0	55.0	55.0	50.0	50.0
Spanish	Czech	50.0	**72.2**	50.0	**63.9**	50.0	**72.2**	50.0	**63.9**
	German	**60.2**	39.8	52.8	57.4	**60.2**	39.8	52.8	57.4

Table 1 shows the results for the classification of PD vs. HC speakers. For the DDK speech task, only the test in Czech language shows accuracies higher than 65%. This could be explained due to Czech patients being diagnosed at the time of the recording and being in an earlier state of the disease than the patients from the other two languages. Accuracies below 50% could be explained by the fact that ranges of the cosine distance are likely to be different in the development and test sets. On the other hand, for the sentences, the i-vectors extracted with MFCCs and articulation-based features show the best results in most of cases. The 50% results when Spanish is used for test can be explained by the mismatch of cosine distance ranges in the development and test sets. For this case, the threshold could be set so all speakers in the test dataset are classified either as PD or HC. The similar results found using the HC and PD reference i-vectors may be due to both vectors being antiparallel.

Table 2 shows the results for the prediction of the MDS-UPDRS. For this case, articulation-based features provide the best result when evaluating the DDK speech task. Phonation features show good results in some cases, specially when the Spanish language is used for test, but show poor results when testing on Czech. This maybe due to the fact that Spanish is a more voiced language than the other two languages, but further experimentation is required. Correlations with the MDS-UPDRS score of up to 0.4 were achieved. Slightly higher correlations were obtained using the DDK speech task due to the fact that such a task is language independent, i.e., the speakers in the corpora uttered the same sounds. Good results were also obtained when analyzing isolated sentences, which is a language dependent speech task. This is encouraging and indicates that other speech tasks could also be analyzed in a cross-language setting.

Table 2. Spearman's correlation for the prediction task.

Train Lang.	Test Lang.	HC reference				PD reference			
		MFCCs	Art.	Phon.	Pros.	MFCCs	Art.	Phon.	Pros.
DDK									
Czech	German	−0.14	−0.25	0.14	−0.06	0.14	0.25	−0.14	0.05
	Spanish	**0.32**	0.23	0.20	−0.13	−**0.32**	−0.29	−0.17	0.17
German	Czech	**0.26**	0.38	−0.09	0.11	−**0.30**	−0.38	0.16	−0.25
	Spanish	0.04	0.20	0.24	0.09	0.25	−0.21	−**0.39**	0.31
Spanish	Czech	−0.32	**0.41**	−0.19	−0.48	0.45	−0.25	0.12	0.16
	German	0.15	−0.14	0.11	−0.24	−0.17	0.14	−0.11	0.14
Sentences									
Czech	German	0.06	0.11	−0.14	0.15	−0.06	−0.11	0.14	−0.15
	Spanish	0.16	−0.15	−0.01	0.12	−0.15	0.15	0.01	−0.12
German	Czech	0.18	−0.11	−0.30	−0.02	−0.21	0.13	0.32	0.07
	Spanish	0.26	0.08	0.37	0.02	−0.27	−0.08	−**0.37**	−0.02
Spanish	Czech	−0.02	0.29	0.05	**0.36**	0.05	−0.29	−0.05	−**0.36**
	German	0.10	0.04	0.11	0.04	−0.10	−0.04	−0.19	−0.04

For comparison, language dependent results using the same Spanish database and the same i-vector methodology can be found in [18].

5 Conclusion

In this work we address the task of cross-language evaluation of Parkinson's Disease speech using the i-vector approach. Data in Czech, German and Spanish were used. One of the languages is used for train, while the other two were used for parameter selection and test. All possible combinations were considered. Two reference i-vectors were created. These reference i-vectors represent the population of HC speakers or PD patients. Then, the cosine distance between one of these reference i-vectors and the i-vector of a test speaker was computed. This distance was used in two experiments: to classify PD patients and HC speakers, and to assess the prediction of the neurological state of the patients. Results are promising, with classification rates around 70% when using MFCCs and articulation features. Similar classification results were obtained using both reference i-vectors. In many cases, a positive correlation between the labels and the cosine distance to the HC reference i-vector was found. This means that the more affected the speech, the larger the difference to healthy speakers. A similar reasoning can be followed for the negative correlations when comparing test speakers with respect to the PD reference i-vector, i.e., the more affected the speech, the lower the difference to the PD speakers. Future work includes evaluating the use of techniques that can eliminate the variability of language in the i-vector space with the aim of improving the results and obtain a language independent method to evaluate the condition of patients with neurodegenerative disorders.

Acknowledgments. Thanks to CODI from University of Antioquia by the grant Numbers 2015-7683 and PRV16-2-01 and to COLCIENCIAS by the grant Number 111556933858.

References

1. Ahmed, A.M., et al.: Motor symptoms in Parkinson's disease: a unified framework. Neurosci. Biobehav. Rev. **68**, 727–740 (2016)
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 J. Biomed. Health Inf. **19**(6), 1862–1872 (2015)
3. Goetz, C.G., et al.: Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. Mov. Disord. **23**(15), 2129–2170 (2008)
4. Rusz, J., et al.: Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. J. Acoust. Soc. Am. **134**(3), 2171–2181 (2013)
5. Schuller, B., et al.: The INTERSPEECH 2015 computational paralinguistics challenge: nativeness, Parkinson's & eating condition. In: Proceedings of the 16th INTERSPEECH, pp. 478–482 (2015)
6. Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonil, 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, pp. 342–347 (2014)
7. Grósz, T., Róbert, B.-F., Gábor, G., Tóth, L.: Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks. In: Proceedings of the 16th INTERSPEECH, pp. 919–923 (2015)
8. Orozco-Arroyave, J.R., et al.: Automatic detection of Parkinson's disease from words uttered in three different languages. J. Acoust. Soc. Am. **139**(1), 481–500 (2016)

9. Arias-Vergara, T., Vasquez-Correa, J.C., Orozco-Arroyave, J.R., Vargas-Bonilla, J.F., Noth, E.: Parkinson's disease progression assessment from speech using GMM-UBM. In: Proceedings of the 17th INTERSPEECH, pp. 1933–1937 (2016)
10. Tu, M., Berisha, V., Liss, J.: Objective assessment of pathological speech using distribution regression. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2017)
11. Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. **19**(4), 788–798 (2011)
12. Dehak, N., Torres-Carrasquillo, P.A., Reynolds, D., Dehak, R.: Language recognition via i-vectors and dimensionality reduction. In: Proceedings of the 12th INTERSPEECH, pp. 857–860 (2011)
13. Martínez, D., Burget, L., Ferrer, L., Scheffer, N.: iVector-based prosodic system for language identification. In: Proceedings of the 37th ICASSP, pp. 4861–4864, March 2012
14. Senoussaoui, M., Cardinal, P., Dehak, N., Koerich, A.L.: Native language detection using the i-vector framework. In: Proceedings of the 17th INTERSPEECH, pp. 2398–2402 (2016)
15. Orozco-Arroyave, J.R., et al.: Towards an automatic monitoring of the neurological state of Parkinson's patients from speech. In: Proceedings of the 41st ICASSP, pp. 6490–6494 (2016)
16. Dehak, N., Dumouchel, P., Kenny, P.: Modeling prosodic features with joint factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. **15**(7), 2095–2103 (2007)
17. Skodda, S., Visser, W., Schlegel, U.: Vowel articulation in Parkinson's disease. J. Voice **25**(4), 467–472 (2012)
18. Garcia, N., Orozco-Arroyave, J.R., D'Haro, L.F., Dehak, N., Nöth, E.: Evaluation of the neurological state of people with Parkinson's disease using i-vectors. In: Proceedings of the 18th INTERSPEECH (2017, in Press)