© Springer-Verlag http://www.springer.de/comp/lncs/ index.html; published in: Text, Speech and Dialogue (TSD 2015), Proceedings; LNAI 9302, Springer, Cham, [...] , 2015, pp. 80–87

Automatic Detection of Parkinson's Disease in Reverberant Environments

Juan Rafael Orozco-Arroyave[1,2], Tino Haderlein[2], and Elmar Nöth[2]

[1] Faculty of Engineering, Universidad de Antioquia UdeA, Medellín, Colombia
[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

Abstract. Automatic classification of speakers with Parkinson's disease (PD) and healthy controls (HC) is performed considering a method for the characterization of the speech signals which is based on the estimation of the energy content of the unvoiced frames. The method is tested with recordings of three languages: Spanish, German, and Czech. Additionally, the signals are affected by two different reverberant scenarios in order to validate the robustness of the proposed method. The obtained results range from 85% to 99% of accuracy depending on the speech task, the spoken language, and the recording scenario. The method shows to be accurate and robust even when the signals are reverberated. This work is a step forward to the development of methods to assess the speech of PD patients without requiring special acoustic conditions.

Keywords: Parkinson's disease, reverberant evironments, unvoiced frames, multi-language

1 Introduction

Parkinson's disease (PD) is a neurological disorder that results from the death of dopaminergic cells in the substantia nigra of the midbrain [1]. It is the second most prevalent neurological disorder and affects about 2% of people older than 65 [2]. According to the Royal College of Physicians in London, PD patients should have access to a set of services and therapies including specialized nursing care, physiotherapy, and speech and language therapy, among others [3]. It is estimated that about 90% of people with PD develop speech impairments; however, only 3% to 4% of them receive speech therapy [4]. The symptoms observed in the speech of PD patients include reduced loudness, a monopitch and monoloudness kind of speech, breathy voice, and imprecise articulation, among others [4]. In addition to the aforementioned problems in the speech of PD patients, they develop also motor impairments that reduce their motion capabilities. The research community has shown interest in developing systems for the telemonitoring of people with PD from speech [5–7]. However, the performance of such systems in real-life conditions, i.e. in non-controlled noise and in reverberant environments, is still an unanswered question. The motion problems developed by PD patients make difficult to perform their recording in places different to their house or their room in a hospital. Thus, it is necessary to develop computational tools able to perform the analysis of the speech recordings even if such records are captured in reverberant environments or in non-controlled acoustic conditions. This paper presents a method to perform the automatic classification of speakers with PD and HC from speech recordings that are altered by two different reverberant scenarios. The method is tested with recordings of three databases including people speaking three different languages (Spanish, German, and Czech). The speech tasks evaluated include isolated sentences and the rapid repetition of the syllables /pataka/, which is also called diadochokinetic (DDK) evaluation.

The rest of the paper is organized as follows. Section 2 presents the details of the experimental setup. In Section 3 the obtained results are presented, and in Section 4 the conclusions derived from this study are provided.

2 Experimental Setup

The speech recordings are affected with two different reverberant scenarios. The first one considers a reverberant room that is characterized using a microphone situated at 60 cm in front of the speaker. The second one considers the impulse response obtained from several different angles and distances with respect to the speaker's position, and two different reverberation times. The original recordings, i.e. without any reverberation procedure, are also considered. The unvoiced frames of the speech recordings are segmented automatically using the software Praat [8]. Voiced frames are not considered in this study because in previous experiments we have shown that unvoiced frames are more discriminant than voiced ones [9]. The energy content of each unvoiced frame is measured considering 12 mel-frequency cepstral coefficients (MFCCs) and 25 energy bands distributed according to the Bark scale. The automatic classification of speakers with PD and HC is performed using a support vector machine with soft margin. Figure 1 summarizes the process introduced in this paper. The stages of the process are detailed in the following subsections.

2.1 Databases

Spanish: Recordings of the PC-GITA database [10] are considered. Seven speech tasks including six isolated sentences, and rapid repetitions of /pataka/ are evaluated. The corpus contains recordings of 100 speakers (50 with PD and 50 HC). The speakers are balanced by gender and age. The age of the 25 male patients ranges from 33 to 77 (mean 62.2 ± 11.2), and the age of the 25 female patients ranges from 44 to 75 years (mean 60.1 ± 7.8). For the case of HC, the age of the 25 male ranges from 31 to 86 (mean 61.2 ± 11.3), and the age of the 25 female ranges from 43 to 76 years (mean 60.7 ± 7.7). The recording sessions were performed in a sound-proof booth at Clínica Noel in Medellín, Colombia, using a dynamic omni-directional microphone and a professional audio card. The

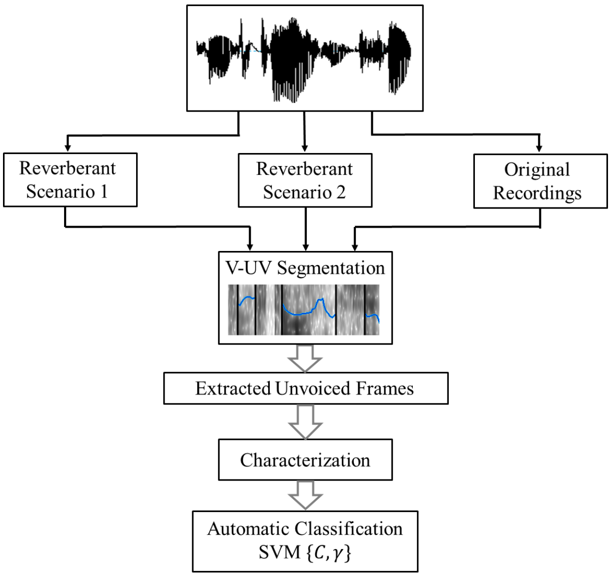


Fig. 1. Stages of the experimental setup

recordings were sampled at 44.1 kHz with a resolution of 16 bits. All of the patients were evaluated and diagnosed by a neurologist expert. The mean values of their neurological evaluation according to the unified Parkinson's disease rating scale (UPDRS-III) and Hoehn & Yahr scale [11] are 36.7 ± 18.7 and 2.3 ± 0.8, respectively. Further details of this database can be found in [10].

German: This corpus consists of 176 native German speakers (88 PD patients and 88 HC). The set of patients includes 88 people (47 male and 41 female). The age of the male patients ranges from 44 to 82 (mean 66.7 ± 8.4), while the age of the female patients ranges from 42 to 84 years (mean 66.2 ± 9.7). The HC group contains 88 speakers (44 male and 44 female). The age of the male ranges from 26 to 83 (mean 63.8 ± 12.7), and the age of the female is from 54 to 79 years (mean 62.6 ± 15.2). The participants were recorded at the Knappschaftskrankenhaus of Bochum in Germany. The sampling frequency of the recordings is 16 kHz with a resolution of 16 bits. The speakers read five isolated sentences and performed the DDK evaluation. The mean values of their neurological evaluation according to the UPDRS-III and Hoehn & Yahr scales are 22.7 ± 10.9 and 2.4 ± 0.6, respectively. Further details of this database can be found in [12].

Czech: A total of 33 native Czech speakers were recorded (19 PD patients and 14 HC). All of the participants of this database are male. The age of the patients ranges from 41 to 60 years (mean 61 ± 12). The age of the healthy group ranges from 36 to 80 years (mean 61.8±13.3). The patients were newly diagnosed with PD, and none of them had been medicated before or during the recording session. The participants were recorded in the General University Hospital in Prague, Czech Republic. The speech tasks considered in this paper include the

DDK evaluation and three isolated sentences. The signals were sampled at 48 kHz with a resolution of 16 bits. The mean values of the neurological evaluations according to the UPDRS-III and Hoehn & Yahr scales are 17.9 ± 7.4 and 2.2 ± 0.5, respectively. Further details of this database can be found in [13].

2.2 Speech Tasks

The speech tasks uttered in Spanish are (1) Mi casa tiene tres cuartos, (2) Omar, que vive cerca, trajo miel, (3) Laura sube al tren que pasa, (4) Los libros nuevos no caben en la mesa de la oficina, (5) Rosita Niño, que pinta bien, donó sus cuadros ayer, and (6) Luisa Rey compra el colchón duro que tanto le gusta, (7) and the rapid repetition of /pataka/.

The speech tasks uttered in German are (1) Peter und Paul essen gerne Pudding, (2) Das Fest war sehr gut vorbereitet, (3) Seit seiner Hochzeit hat er sich sehr verändert, (4) Im Inhaltsverzeichnis stand nichts über Lindenblätentee, (5) Der Kerzenständer fiel gemeinsam mit der Blumenvase auf den Plattenspieler, and (6) the rapid repetition of /pataka/.

The speech tasks uttered in Czech are questions that differ in a couple of words among them. The set of questions is (1) Kolik máte teď u sebe asi peněz?, (2) Kolikpak máte teďka u sebe asi peněz?, (3) Kolikpak máte teďka u sebe asi tak peněz?, (4) and the rapid repetition of /pataka/. Unfortunately, we did not had access to sentences with more varied content.

2.3 Reverberation

Testing the robustness of a system for acoustic analysis with respect to different recording scenarios usually means collecting speech data in many rooms with different impulse responses. Additionally, the microphone(s) should be in different angles and distances from the speaking people who also have to be available in every location. Reverberating close-talking speech artificially with the help of pre-defined room impulse responses can reduce this effort. For this reason, the method introduced in [14] is applied here. The original audio samples from all three languages were converted to 16 kHz and 16 bit as a preprocessing step using SoX v14.3.1. In order to avoid too much clipping due to over-amplification, the volume of the Czech data is reduced to 0.98 of its original value. For the Spanish data, the factor 0.99 is used. For the German data is not necessary to apply such factor because in that case there were not clippings. Room impulse responses for reverberation were measured in a seminar room with the size 5.8 m × 5.9 m × 3.1 m. The microphone was at position (2.0 m, 5.2 m, 1.4 m). The reverberation time could be changed from $T_{60} = 250$ ms to $T_{60} = 400$ ms by removing sound absorbing carpets and sound absorbing curtains from the room. 12 impulse responses were measured for loudspeaker positions on three semi-circles in front of the microphone at distances 60 cm, 120 cm, and 240 cm following the method described in [15] (Fig. 2).

For *reverberant scenario 1*, the original close-talking speech data were convolved with the impulse response measured when the loudspeaker was at 60 cm distance right in front of the microphone. For *reverberant scenario 2*, the original data were divided into 12 parts; i.e., each part consisted of one twelfth of all recordings, as far as possible. Each of the parts was convolved with one of the 12 available impulse responses then.

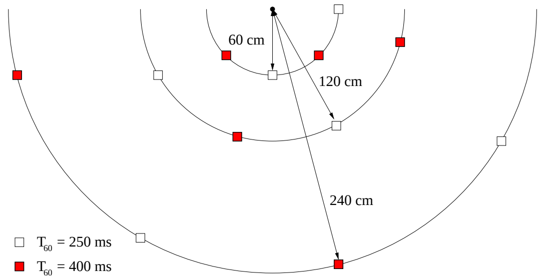


Fig. 2. Assumed speaker positions in the virtual recording rooms for artificially reverberated data; 12 room impulse responses from different positions and with two reverberation times (250 and 400 ms) were used.

2.4 Preprocessing and Characterization of Unvoiced Frames

In order to avoid possible bias introduced by the channel, i.e. microphone and sound card, mean cepstral subtraction is performed before the extraction of features from the recordings i.e., characterization. The energy content of the unvoiced frames is measured considering 25 Bark band energies (BBEs). 12 MFCCs are also calculated, as in [9]. Four low level descriptors are calculated over all feature vector, i.e. mean value, standard deviation, skewness, and kurtosis, forming a 148-dimensional feature vector per recording.

2.5 Classification

A soft margin support vector machine (SVM) with Gaussian kernel is considered to discriminate between PD speakers and HC. The complexity of the SVM (C) and the bandwidth of its kernel (γ) are optimized in a grid search with $10^{-1} < C < 10^4$ and $10^{-1} < \gamma < 10^3$. The optimization criterion is based on the accuracy on test, which could lead to slightly optimistic estimates, but considering that only two parameters are optimized, the bias should be minimal. The classifier is trained following a 10-fold cross validation strategy for the Spanish and German recordings. Each fold was chosen randomly but assuring the balance in age, gender, and the speaker independence. Due to the smaller number of recordings, leave-one-speaker-out (LOSO) cross-validation was used for the Czech data.

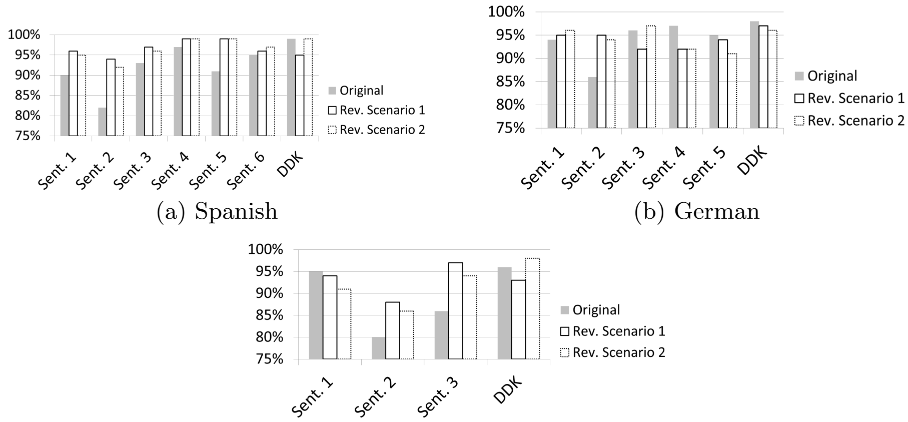
3 Results and Discussion

Table 1 shows the results obtained with the two reverberant scenarios and with the original recordings of the three databases. The results are presented in terms of accuracy, specificity, and sensitivity. Note that the highest accuracies on each database are obtained with the DDK evaluations. This result is in accordance with previous studies that highlighted such a task to be appropriate to assess PD speech [16]. Note also that there is an improvement in the accuracies when the signals are affected by the reverberant scenarios. This behavior can probably be explained by the reverberation process which is introducing information from voiced frames into the unvoiced regions, and thus the method is taking advantage of such "additional" information. The results reported in this paper indicate that the method works properly under reverberant conditions, so it could be used in environments where the acoustic conditions cannot be controlled.

Table 1. Results with recordings affected by two reverberant scenarios. Sent: Sentence. Rev: Reverberant. Acc: Accuracy (%), Spec: Specificity (%), Sens: Sensitivity (%)

	Spanish: Rev. Scenario 1			Rev. Scenario 2			Original Signals		
	Acc	Spec	Sens	Acc	Spec	Sens	Acc	Spec	Sens
Sent. 1	98 ± 4	98 ± 6	98 ± 6	94 ± 8	94 ± 14	94 ± 14	90 ± 9	90 ± 14	90 ± 14
Sent. 2	94 ± 7	92 ± 10	96 ± 8	91 ± 11	88 ± 17	94 ± 14	81 ± 7	84 ± 18	78 ± 15
Sent. 3	93 ± 10	100 ± 0	86 ± 19	96 ± 5	94 ± 10	98 ± 6	92 ± 13	94 ± 10	90 ± 19
Sent. 4	97 ± 7	100 ± 0	94 ± 14	99 ± 3	100 ± 0	98 ± 6	97 ± 5	98 ± 6	96 ± 8
Sent. 5	96 ± 8	100 ± 0	92 ± 17	96 ± 8	100 ± 0	92 ± 17	90 ± 9	92 ± 10	88 ± 14
Sent. 6	95 ± 9	94 ± 14	96 ± 8	97 ± 5	98 ± 6	96 ± 8	94 ± 7	94 ± 10	94 ± 14
DDK	96 ± 7	100 ± 0	92 ± 14	99 ± 3	98 ± 6	100 ± 0	$\boxed{99 \pm 3}$	100 ± 0	98 ± 6
	German: Rev. Scenario 1			Rev. Scenario 2			Original Signals		
Sent. 1	95 ± 6	97 ± 5	93 ± 9	96 ± 5	98 ± 5	94 ± 6	93 ± 5	92 ± 8	94 ± 9
Sent. 2	94 ± 4	94 ± 8	94 ± 9	93 ± 6	93 ± 8	93 ± 9	86 ± 6	84 ± 14	87 ± 14
Sent. 3	91 ± 9	93 ± 8	89 ± 14	96 ± 4	98 ± 5	94 ± 9	96 ± 5	95 ± 6	97 ± 8
Sent. 4	92 ± 4	94 ± 8	90 ± 10	93 ± 7	90 ± 13	97 ± 6	97 ± 6	96 ± 8	98 ± 7
Sent. 5	93 ± 2	90 ± 6	97 ± 6	90 ± 5	92 ± 8	89 ± 9	94 ± 5	98 ± 5	91 ± 11
DDK	97 ± 4	98 ± 7	97 ± 6	97 ± 5	97 ± 6	97 ± 6	$\boxed{98 \pm 3}$	99 ± 4	97 ± 6
	Czech: Rev. Scenario 1			Rev. Scenario 2			Original Signals		
Sent. 1	93 ± 17	94 ± 25	93 ± 21	90 ± 21	81 ± 41	99 ± 3	93 ± 18	89 ± 32	98 ± 10
Sent. 2	86 ± 23	99 ± 3	72 ± 46	85 ± 23	83 ± 38	88 ± 31	86 ± 23	79 ± 41	93 ± 19
Sent. 3	97 ± 10	100 ± 0	94 ± 20	93 ± 17	100 ± 0	87 ± 33	86 ± 23	88 ± 34	84 ± 37
DDK	93 ± 17	96 ± 13	90 ± 29	$\boxed{98 \pm 9}$	100 ± 0	95 ± 19	94 ± 16	99 ± 3	88 ± 31

In order to show the results more compactly, Figure 3 contains the values of the Area Under the receiver operating characteristic Curves (AUC) obtained with the speech tasks of the three databases in the three scenarios (two reverberant and the original recordings). The proposed method shows to be accurate and robust in reverberant environments. The results indicate that it is possible to discriminate between speakers with PD and HC with accuracies ranging from 85% to 99% considering recordings captured in non-controlled acoustic conditions. The results with several speech tasks were higher in the reverberated scenarios. Our hypothesis is that this behavior is explained by the introduction of suprasegmental information from the voiced frames into the unvoiced regions owing to the reverberation process. Further experiments modeling the voiced/unvoiced transitions could lead to validate this hypothesis.



(a) Spanish

(b) German

(c) Czech

Fig. 3. AUC values obtained with each speech task of the three languages. Sent: Sentence. Rev. Scenario: Reverberant Scenario.

4 Conclusions

A method to discriminate between speakers with PD and HC is presented in this paper. The experiments consider speech recordings that are affected by two reverberant scenarios. The method consists on the automatic segmentation and characterization of the unvoiced segments. Since speech recordings of three different languages are considered, the method seems to be language-independent. Additionally, the it shows to be robust against particular plosive sounds of the considered languages, i.e., in the repetition of /pataka/, the plosive sounds /p/ and /t/ are aspirated in German but not in Czech and Spanish. This work is a step forward to develop computational tools for the assessment of speech of PD patients with non-controlled acoustic conditions.

Acknowledgments Juan Rafael Orozco-Arroyave is under grants of COLCIENCIAS through the program "Convocatoria Nº 528, generación del bicentenario 2011". This work was also financed by COLCIENCIAS through the project Nº 111556933858. The authors thank CODI, "estrategia de sostenibilidad 2014-2015 from Universidad de Antioquia" for the support to develop this work.

References

1. Hornykiewicz, O.: Biochemical aspects of Parkinson's disease. Neurology **51**(2) (1998) S2–S9
2. de Rijk, M.C., Launer, L.J., Berger, K., Breteler, M.M., Dartigues, J.F., Baldereschi, M., Fratiglioni, L., Lobo, A., Martinez-Lage, J., Trenkwalder, C., Hofman, A.: Prevalence of Parkinson's Disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology **54**(11 Suppl 5) (2000) S21–S23
3. Worth, P.: How to treat Parkinson's disease in 2013. Clinical Medicine **13**(1) (2013) 93–96
4. Ramig, L.O., Fox, C., Sapir, S.: Speech treatment for Parkinson's disease. Exp. Rev. Neurother. **8**(2) (2008) 297–309
5. Zicker, J.E., Tompkins, W.J., Rubow, R.T., Abbs, J.H.: A portable microprocessor-based biofeedback training device. IEEE Transactions on Biomededical Engineering **27**(9) (Sept 1980) 509–515
6. Wirebrand, M.: Real-time monitoring of voice characteristics using accelerometer and microphone measurements. Master's thesis, Linköpings universitet, Linköping, Sweden (2011)
7. Vásquez-Correa, J.C., Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Nöth, E.: New computer aided device for real time analysis of speech of people with Parkinson's disease. Rev. Fac. Ing. Universidad de Antioquia **1**(72) (2014) 87–103
8. Boersma, P., Weenink, D.: Praat, a system for doing phonetics by computer. Glot International **5**(9/10) (2001) 341–345
9. Orozco-Arroyave, J.R., Hönig, F., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Skodda, S., Rusz, J., Nöth, E.: Automatic detection of parkinson's disease from words uttered in three different languages. In: Proceedings of the 15th INTERSPEECH. (2014) 1573–1577
10. Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, 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 Language Resources and Evaluation Conference (LREC). (2014) 342–347
11. Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R.G., Moore, C.G., Wenning, G.K., Yahr, M.D., Seidl, L.: Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Movement Disorders **19**(9) (2004) 1020–1028
12. Skodda, S., Visser, W., Schlegel, U.: Vowel articulation in Parkinson's diease. J. of Voice **25**(4) (2011) 467–472 Erratum in J. of Voice. 2012 Mar;25(2):267-8.
13. Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., Ruzicka, E.: Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. Journal of the Acoustical Society of America **134**(3) (2013) 2171–2181
14. Haderlein, T., Nöth, E., Herbordt, W., Kellermann, W., Niemann, H.: Using Artificially Reverberated Training Data in Distant-Talking ASR. In V. Matoušek et al., ed.: Proc. Text, Speech and Dialogue; 8th International Conference, TSD 2005; Karlovy Vary, Czech Republic, 2005. Volume 3658 of Lecture Notes in Artificial Intelligence., Berlin, Heidelberg, Springer–Verlag (2005) 226–233
15. Herbordt, W.: Combination of robust adaptive beamforming with acoustic echo cancellation for acoustic human/machine interfaces. PhD thesis, University Erlangen-Nuremberg, Germany (January 2004)

16. Harel, B.T., Cannizzaro, M.S., Cohen, H., Reilly, N., Snyder, P.J.: Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment. Journal of Neurolinguistics **17** (2004) 439–453