

Diagnosis and Tracking of Parkinson's Disease by using Automatically Extracted Acoustic Features

CJ Pérez^{1*}, Y Campos-Roca², L Naranjo³ and J Martín¹

¹Departamento de Matemáticas, Universidad de Extremadura, Cáceres (Spain)

²Departamento de Tecnologías de los Computadores y de las Comunicaciones, Universidad de Extremadura, Cáceres (Spain)

³Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, México DF (Mexico)

Abstract

A system that is capable of automatically discriminating healthy people from people with Parkinson's Disease (PD) from speech recordings is proposed. It is initially based on 27 features, extracted from recordings of sustained vowels. The number of characteristics has been further reduced by feature selection. The system has been tested by using a heterogeneous database, composed of 40 control subjects and 40 subjects with PD belonging to different severity stages of the disease and under prescribed treatment. Repeated measures per individual were averaged before being assigned to subject, avoiding the usual practice of considering measurements within the same subject as independent. The best overall accuracy obtained was 85.25%, with a sensitivity of 90.23% and a specificity of 80.28%. Additionally, a pilot experiment to track PD severity stages has been performed on 32 out of the 40 initial subjects with PD. To the authors' knowledge, this is the first speech-based experiment on automatic PD tracking by using the Hoehn and Yahr's scale (clinical metric mainly focused on postural instability). The results suggest that progression of voice impairment follows different developmental trajectories than postural instability, implying different degenerative mechanisms.

Keywords: Computer-aided diagnosis system; Longitudinal ordinal regression; Machine learning; Parkinson's disease; Speech recordings; Voice features

Introduction

People with Parkinson's Disease (PD) exhibit a chronic neurodegenerative disorder caused by the progressive degeneration and death of dopaminergic neurons, that play a key role in coordinating the movement at level of muscular tone.

Voice and speech, as dependent on laryngeal, respiratory and articulatory functions, are also affected in people with PD. Non-motor signs can also affect language, cognition and mood, which can impact on communication [1]. Vocal impairment is most likely one of the earliest signs of the disease [2]. Since the very early stages of PD, there are subtle abnormalities in speech that might not be perceptible to listeners, but they could be evaluated in an objective way by performing acoustic analyses on recorded speech signals.

Some authors found statistically significant differences of some acoustic features between healthy controls and subjects with PD [3-5]. Significance tests on prosodic parameters have also been used to evaluate the effectiveness of response to levodopa (L-dopa) and surgical treatment [6]. However, individually evaluating each feature is not enough to properly discriminate people with PD from healthy subjects. A classification-based approach must be performed to show the discrimination capability of a set containing the most relevant features.

Computer-Aided Diagnosis (CAD) systems seek to maximize the information that can be automatically extracted from medical tests (images, videos, voice,...) by quantitative calculations. In recent years numerous techniques have been developed to assess speech-related diseases. The monograph presented by [7] provides a current view of the state-of-the-art concerning automatic classification of speech signals for clinical purposes. This research line has a special development for the diagnosis and monitoring of PD. The Parkinson's Voice Initiative has played an important role in the spread of this topic.

Some authors have considered measures extracted from speech recordings to discriminate healthy people from those with PD [8-12]. In these investigations, speech samples are recorded to extract some specific characteristics of the signals and classify them by using different methods. These techniques have a great potential as biomarkers to provide early diagnosis of the disease [13]. Successfully addressing early diagnoses of people with PD is a key issue to improve the patients' quality of life. Undiagnosed and misdiagnosed subjects can also be benefited from these techniques. Note that it is estimated that 20% of people with PD remain undiagnosed [14], allowing the fast advance of the disease. Misdiagnosis is also an important problem since the symptoms can be confused with the ones of other neurodegenerative disorders.

As the disease progresses, voice impairment increases and most times the improper voice can be directly distinguished without the need of any computer aided analysis [4]. PD tracking has been previously performed by applying the Unified Parkinson's Disease Rating Scale (UPDRS, [15]), which reflects the presence and severity of symptoms. A revision of this scale was published in 2007 by the Movement Disorder Society (MDS), and it is known as the MDS-UPDRS [16]. Both scales require the patient's physical presence in a clinical center and the availability of expert clinical staff for long time. For many people with PD, visits to hospital are an additional complication. Advances in

***Corresponding author:** Carlos Javier Pérez Sánchez, Departamento de Matemáticas, Universidad de Extremadura, Cáceres (Spain), Tel: +34927257146; E-mail: carper@unex.es

Received July 07, 2016; **Accepted** September 07, 2016; **Published** September 14, 2016

Citation: Pérez CJ, Campos-Roca Y, Naranjo L, Martín J (2016) Diagnosis and Tracking of Parkinson's Disease by using Automatically Extracted Acoustic Features. J Alzheimers Dis Parkinsonism 6: 260. doi: [10.4172/2161-0460.1000260](https://doi.org/10.4172/2161-0460.1000260)

Copyright: © 2016 Pérez CJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

broadband telecommunication systems offer the possibility of remote monitoring [17]. Some approaches have addressed the problem of finding a statistical mapping between speech parameters and UPDRS scores, but all of them are based on only one multicenter study [18-22]. A simpler scale, which has wide utilization and acceptance, is Hoehn and Yahr's (H&Y [23]), but as a counterpart, it does not measure the symptom severity but the stage in an ordinal scale. There have been few attempts to correlate both metrics [24,25]. Up to our knowledge, no results have been presented on PD progression with H&Y scale by using voice recordings in a longitudinal study.

One important issue to deal with both in diagnosis and tracking applications is the choice of a proper statistical treatment. In this context, it has become usual in the literature to apply classification methods based on independent sample schemes to experiments including replicated recordings. This treatment of the data artificially increases the sample size and yields unreliable results [26-28]. Although some advances have been reported on this topic, there is a scientific challenge to improve and disseminate these techniques, so that they can be incorporated into protocols by neurological units. These procedures have the advantage over other traditional techniques of being non-invasive, objective, inexpensive and easy to self-administration so that they can be used in telediagnosis and telemonitoring contexts.

In this paper, a voice recording-based experiment on 80 subjects, 40 of them with PD, has been conducted. The aim was to study the potential of the automatically extracted voice features to discriminate between healthy controls and subjects with PD and to track PD progression by using the H&Y scale.

The outline of this paper is as follows. Section 2 presents the main information on participants, speech recordings, feature extraction and statistical data analysis. In Section 3, the statistical results are presented for both diagnosis and tracking PD progression. Section 4 presents a discussion on the obtained results as well as the advantages and limitations of the approach. Finally, Section 5 shows the conclusion.

Materials and Methods

In this section, information on participants, speech recordings, feature extraction and statistical data analysis is presented.

Participants

Firstly, 80 subjects older than 50 years were considered to perform voice recordings and to follow a survey protocol. Half of them were healthy: 22 men (55%) and 18 women (45%), and the other were subjects with PD: 27 men (67.5%) and 13 women (32.5%). The mean age ($\pm$ standard deviation) was 66.38 ± 8.38 years for the control group and 69.58 ± 7.82 years for the subjects with PD. None of the people in the control group has history of symptoms related to PD or any other kind of movement disorder syndrome. The research protocol was approved by the Bioethical Committee of the University of Extremadura and all the subjects signed an informed consent.

People with PD were diagnosed by expert neurologists and assigned to stages from 1 to 4, following the H&Y scale. The number of subjects in stages 1, 2, 3 and 4 was, respectively, 9, 13, 16 and 2 in 2014. All of them were under the effects of prescribed medication of levodopa. The time since the disease was detected ranged between 1 and 24 years, with mean 7.48 ± 5.30 years. One subject with PD was diagnosed when she was 34 years old, and at the moment of the survey she was 58 and in stage 4.

One year after the first application of the protocol, 32 out of the 40 subjects with PD were again applied the same protocol for tracking

disease progression. The number of subjects in stages 1, 2, 3 and 4 had changed to 6, 9, 14 and 3, respectively. The remaining 8 out of the 40 initial subjects with PD were no longer visiting the headquarters of *Asociación Regional de Parkinson de Extremadura* (Mérida) and *Confederación Española de Personas con Discapacidad Física y Orgánica* (Cáceres), where the protocol was applied.

Speech recordings

The vocal task was the sustained phonation of /a/ vowel at comfortable pitch and loudness, as constant as possible. This phonation had to be kept for at least 5 seconds and on one breath. The task was repeated three times per individual, and they were considered as repeated measurements. This task has been considered due to the good results provided in the scientific literature [8,29] and to keep the speech exercise as simple as possible, avoiding overtiring the participants, especially in the case of patients with advanced PD stages.

The speech data were recorded using a portable computer with an external sound card (TASCAM US322) and a headband microphone (AKG 520) featuring a cardioid pattern, positioned at approximately 8 cm from the lips. The digital recording was performed at a sampling rate of 44.1 KHz and a resolution of 16 bits/sample by using Audacity software (release 2.0.5). This recording conditions were consistent across the whole data collection.

The voice signal, composed of a set of sounds generated by the vocal system, is transformed into an electrical signal by means of the microphone. Then, producing a sustained /a/ lasting some seconds gives rise to a signal from which acoustic parameters can be extracted. Figure 1 shows the waveforms and spectrograms for a healthy subject and a person with PD.

Feature extraction

This study is based on 27 acoustic features coming from 4 different characteristic families. Features within each family are highly correlated. The four families are:

- Four pitch local perturbation measures: relative jitter, absolute jitter, Relative Average Perturbation (RAP) and Pitch Perturbation Quotient (PPQ) [30].
- Five amplitude perturbation measures: local shimmer, dB shimmer, 3-point Amplitude Perturbation Quotient (APQ3), 5-point Amplitude Perturbation Quotient (APQ5) and 11-point Amplitude Perturbation Quotient (APQ11) [30].
- Five Harmonic-to-Noise Ratio (HNR) features corresponding to different frequency bandwidths: HNR05 (0-500 Hz), HNR15 (0-1500 Hz), HNR25 (0-2500 Hz), HNR35 (0-3500 Hz) and HNR38 (0-3800 Hz). HNR values for each frame are extracted by using the freely available VoiceSauce toolbox [31].
- Thirteen Mel Frequency Cepstral Coefficients (MFCCs): from MFCC0 to MFCC12 [32].

In people with PD, both weak laryngeal and diaphragm control can produce vocal tremor [33]. Vocal fold stiffness and bowing cause changes in vocal fold mass and tension. Weak diaphragm control can cause fatigue across a prolonged voice loading task. Tremulous voices can not properly keep the sustained phonation and therefore show unstable fundamental frequency (high jitter) and amplitude (excess shimmer).

The third family of features measures the relative level of noise

present in speech. In people with PD, weak laryngeal control leads to an incomplete glottal closure. As a consequence, excess noise appears due to unphonated air escaped through these leaks in the glottis. This leads to lower HNR values.

The use of MFCCs for voice analysis stems from the field of speech and speaker recognition. Although specific coefficients do not have a clear physical meaning, a general interpretation can be made: low coefficients are related to the speech spectral envelope, which depends on articulators position. It has become customary in many speech processing applications to use the first 13 MFCC coefficients (12 of them represent the spectral envelope shape plus an energy parameter) [34]. PD is known to affect also articulation, therefore these low MFCC coefficients are promising features to complement the previous measures (related to phonation impairments) for PD diagnosis and tracking [19]. In this work, standard deviations of MFCC parameters are used to model the variability of spectral envelope within a sustained vowel.

After extraction, for each voice feature, the three replications per individual were averaged, avoiding an artificial sample size increase. Matlab software Release 2013 has been used to implement the feature extraction procedures.

Statistical data analysis

Independence-based t-tests are used to analyze statistically significant differences between mean values of the acoustic features in people with PD and control subjects. Support Vector Machine (SVM) methods are considered to analyze the predictive ability of the acoustic variables to discriminate healthy subjects from people with PD [35]. These methods are supervised learning models used for classification tasks. Since classes may not be separable by a linear boundary, SVMs can efficiently perform a non-linear classification using what is called a kernel function. The performance of SVMs considerably depends on the kernel. Thus it is important to determine the best choice among different kernel functions. Technical details on how SVM methods work can be found in [36]. Concerning feature selection, a backward sequential approach has been used to search for the optimal feature subset.

The sign test is applied to analyze if the H&Y distribution has significantly changed from the first to the second year. Paired t-tests were considered to compare the mean values of the acoustic features between the two years. In order to relate the H&Y scale to the acoustic features, a generalized linear mixed model for longitudinal ordinal data is considered [37]. The variable selection is performed by using a penalized backward continuation ratio model [38,39].

In order to estimate the model performance on new individuals for both diagnosis and progression, 10-fold cross-validation schemes are considered [40]. This is usual to estimate how accurately a predictive model will perform in practice. Specifically, in 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size subsamples. A single subsample (out of the 10 defined) is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds are averaged to produce a single estimation. This process can be repeated many times and the results averaged. For diagnostic purposes, the 10-fold cross-validation is performed in a stratified way so that each subsample has the same number of subjects with PD and control subjects.

For hypothesis tests, results are considered statistically significant

when p-values are lower than 0.05. No correction for multiple testing has been applied. IBM SPSS 19 and R software release 3.1.2 have been used for statistical analyses.

Results

The results for the acoustic measure-based experiments on PD detection and PD progression are presented in the following two subsections.

CAD system

All 27 acoustic features provide statistically significant differences between people with PD and healthy subjects. However, evaluating each feature individually is not enough to properly discriminate people with PD from healthy people. A classification-based approach is performed here.

Two hundred repetitions of 10-fold stratified cross-validation were used to estimate the classifier performance. The performance of SVM with backward sequential feature selection is evaluated based on four different kernel functions, namely, linear, quadratic, radial basis function (RBF) and multilayer perceptron (MLP). Classifier performance was measured using overall accuracy, sensitivity and specificity. The results are shown in Table 1.

The best overall accuracy (85.25%) is achieved with the RBF kernel and the 95% Confidence Interval (CI) is 85.07-85.43. Sensitivity of 90.23% (with 95% CI: 89.91-90.55) and specificity of 80.28% (with 95% CI: 80.12-80.43) are achieved.

Tracking PD progression

The H&Y stages observed in 2014 and 2015 are presented in a double entry table (Table 2).

Seven out of thirty two patients increased their disease stage in one year. Specifically, two patients changed from stage 1 to stage 2, four patients changed from stage 2 to 3, and one patient changed from stage 3 to 4. The sign test shows that this change in the H & Y distribution from the first to the second year is not statistically significant (p-value=0.125).

Paired t-tests were considered to compare the mean values of the acoustic features between the two years. Only one out of 27 acoustic variables provided statistically significant differences between mean values. Specifically, MFCC5 provides a p-value of 0.032.

The previous results show, separately, that there have not been too many changes in one year. Now, a longitudinal ordinal regression method is applied to relate the H&Y scale and the acoustic features in these two years. Two hundred repetitions of 10-fold cross-validation were used to estimate the model performance. The application of the longitudinal ordinal regression model with the considered variable selection provides an accuracy rate of 11.96% of correct classification for both stages simultaneously, and an accuracy rate of 31.68% for correct classification of only one H&Y stage (2014 or 2015). This clearly

	Accuracy (%)	Sensitivity (%)	Specificity(%)
RBF	85.25	90.23	80.28
Linear	84.29	84.75	83.83
MLP	81.51	84.10	79.93
Quadratic	79.50	85.12	74.38

Table 1: Overall accuracy, sensitivity and specificity from the CAD experiment.

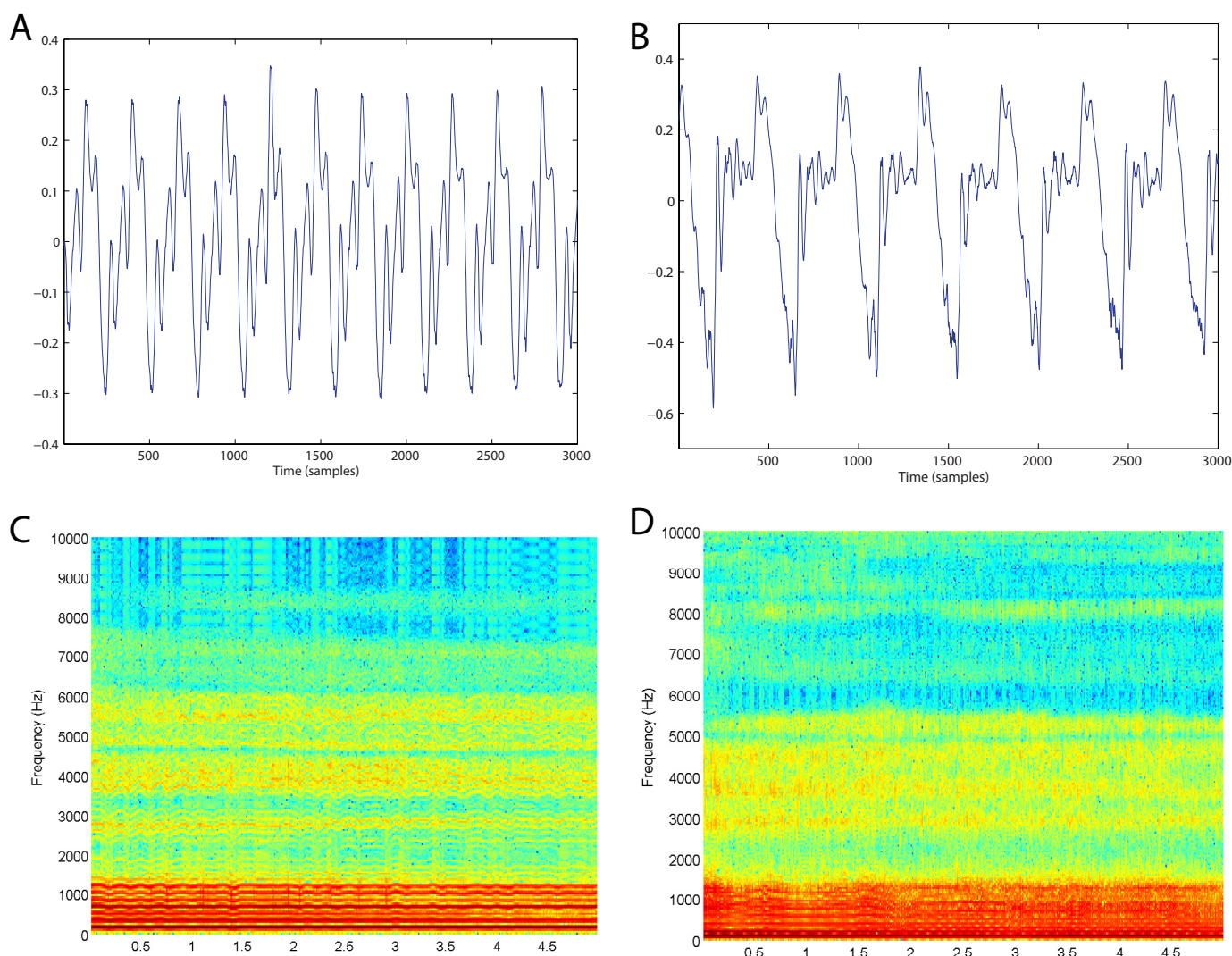


Figure 1: Waveforms, (a) and (b), and spectrograms, (c) and (d), for two voice recordings of control and PD subjects, respectively.

shows that the model is not accurately predicting the progression of the disease by using the acoustic features.

In the next section it is discussed about the advantages and limitations of acoustic feature-based approaches to aid in the diagnosis and tracking progression of PD.

Discussion

There is some evidence suggesting that speech disorders may precede other symptoms of PD. Acoustic analysis on recorded speech signals can help to detect subtle abnormalities in speech that might not be perceptible to listeners in early stages of the disease. This is mainly due to the fact that the muscles controlling voice and speech are affected by PD. This has opened an interesting research line to identify the differences between voices of healthy individuals and individuals suffering from PD, by developing, implementing, and testing feature extraction as well as pattern recognition algorithms.

Some authors have considered measures extracted from speech

recordings to discriminate healthy people from those with PD. In this context, it has become usual to conduct experiments with replicated recordings over the healthy subjects and the subjects with PD [8,9,11]. A common point in most of these approaches is that the classification methods are based on independent sample schemes instead of on repeated measurement frameworks. Note that each subject has several related measures which are not independent. Using the related measures as independent artificially increases the sample size, leading to inflated classification success proportions. [12,27] analyzed this problem and proposed how to handle recording replications.

A possibility to handle recording replications is to summarize by a representing value each feature extracted from all voice recordings belonging to the same individual. By using the database in [8], a best accuracy rate of 88.25% has been obtained in the discrimination of people with PD from healthy subjects by averaging features from the same individuals [12]. [8] obtained 91.4% by considering the voice recordings as independent and used the equivalent to 195 experimental units (195 voice recordings coming from only 32 individuals including

patients with PD and healthy controls). [11] compare their proposals with the ones from fifteen previously published papers that use the data in [8]. These authors report an accuracy rate of 100%, with the same speakers contributing large number of recordings and using them as if they came from different individuals.

In this paper, there are three recordings of sustained /a/ per subject and the instances match the subjects in the experiment, i.e., 80 subjects (40 of them with PD), so the sample size has not been artificially increased. [10,41] presented different approaches to avoid considering measurements within the same subject as independent. [10] proposed to aggregate information with central tendency and dispersion metrics with a study considering 20 healthy subjects and 20 subjects with PD. They obtained accuracy rates between 65.00% and 77.50% by using 26 voice features with several classification methods. [41] performed an experiment with the same sample size for the same purpose with a different methodology. They attained an accuracy rate of 81.08%. In our experiment the best accuracy rate is 85.25%. These results are remarkable, since they have been obtained by using a heterogeneous database, composed of people with PD belonging to four different stages of the H&Y scale (initial, two mild stages and advanced) and under prescribed treatment, so some effects of the disease were decreased.

This noninvasive low-cost tool can be considered as a diagnostic test in primary attention when the doctor has concerns about the medical condition. This would give another evidence to derive the patient to specialized neurological units, what may help in early diagnosis and with long-term undiagnosed patients. Note that it is estimated that 20% of people with PD remain undiagnosed [14].

However, CAD systems for helping in PD detection constitute only the first step. Many people with PD will eventually experience different degrees of vocal impairment as their condition advances. PD tracking has been previously performed by applying the UPDRS. Most approaches are based on only one multicenter study that tested the feasibility of a computer-based at-home testing device (AHTD) in 52 early-stage unmedicated people with PD over a period of six months [17]. Voice recordings of sustained /a/ from 42 out of 52 patients were considered by Tsanas and collaborators in several publications for tracking purposes [18-19,42-44]. Voice characteristics considered in [42] were used by other authors for the same purpose [20,22].

However, as in the classification problem, these approaches considered the features extracted from the recordings as independent. Since each patient has provided many replications during the six months, there is no longer independence. This artificial treatment of the data increases the sample size to the equivalent of 5,875 individuals, instead of the real 42 individuals with his/her replications at times 0, 3, and 6 months, and provides better estimated accuracies than they actually are. [26,28] propose alternative modelling methods to address repeated measurements in this context. However, the results show that the approaches provide higher estimations than the inter-rater variability for UPDRS, that is about 4-5 points.

There have been some tries to correlate the UPDRS scale with the H&Y scale [25,45]. There is an inconsistency at least one time [25] found statistically significant correlations (with moderate or low strength) between the items in section III of UPDRS scale and H&Y. This led to the possibility of using this scale instead of UPDRS to track PD progression. Up to now, no experiment based on acoustic measurements has been developed for this purpose in a longitudinal framework. In this paper, we have tracked the progression of 32 patients in one year. Seven of them increased their disease stage in this period. However, this change in the H&Y distribution from the first to the second year is not statistically significant. Besides, only one out of 27 acoustic variables provided statistically significant differences between mean values from 2014 to 2015. This reveals small changes in one year. The next question is how accurately the values of H&Y scale can be estimated by using the features extracted from the voice recordings. In this case, the acoustic features extracted with the longitudinal regression model considered do not accurately predict the progression of PD based on the H&Y scale (Table 3).

Although the results from this pilot study are limited by the sample size and period length, the findings might be explained by the fact that the H&Y scale is focused only on some specific motor deficits. In particular, H&Y scale considers unilateral versus bilateral disease and evaluates the presence or absence of postural instability, thereby leaving other aspects of motor deficit and non-motor symptoms unassessed. Further work on a larger-scale experiment (more subjects and over a longer period) should be focused on this H&Y scale and on a different clinical metric. [42] use total and motor UPDRS scores. The sum of specific items in the MDS-UPDRS is proposed here for future work, keeping the subject-based approach.

A great advantage of this methodology is that voice recordings can be obtained from remote locations, allowing to build a telemonitoring system [17]. Telemonitoring can be objective, simple, noninvasive and facilitates fast, frequent remote tracking of disease progression. For many people with PD, visits to hospital are an additional complication. On the other hand, it significantly alleviates the national health systems of excessive workload and the large associated costs of clinical human expertise. The implementation of the proposed approach in real-time through Android or iOS-based smartphones would provide an added value for all the implied parts. However, it is necessary more research before telemonitoring can be incorporated in neurological units for tracking PD progression.

Conclusion

Human voice is affected by PD, due to disorders of laryngeal, respiratory and articulatory functions. CAD systems based on acoustic features as the one proposed here can be used as a diagnostic test with a relatively high sensitivity and specificity. Therefore, voice features can be used as biomarkers in primary attention and they can help doctors to early diagnose. This is currently important for access to treatment of symptoms and is expected to be a key point associated with the

		Stage in 2015				Total
Stage in 2014	Stage 1	6	2	0	0	
	Stage 2	0	7	4	0	11
	Stage 3	0	0	10	1	11
	Stage 4	0	0	0	2	2
	Total	6	9	14	3	32

Table 2: Double entry table for H&Y stages in 2014 and 2015.

development of new treatments in the near future. Currently, no single definitive diagnostic test is available for PD, and accurate diagnosis has been a significant challenge, particularly among clinicians without particular expertise in movement disorders.

PD tracking is also an objective. The results considering predicting H&Y scale from acoustic features suggest that this clinical scale should be substituted for this purpose by another metric, such as the sum of specific items of the MDS-UPDRS scale.

The small sample size and the duration of the experiment is a limitation. There is a great difficulty to recruit patients for this kind of experiments. It is necessary more research on new patients, before these techniques can be incorporated into protocols by neurological units, possibly in a remote way. This would produce a definitive feedback to the scientific community allowing an iterative procedure that would lead to better results.

Acknowledgement

Thanks to the anonymous participants and to Carmen Bravo and Rosa María Muñoz for carrying out the voice recordings and providing information from the people with PD. We are grateful to the *Asociación Regional de Parkinson de Extremadura* and *Confederación Española de Personas con Discapacidad Física y Orgánica* for providing support in the experiment development. We also thank three anonymous referees for comments and suggestions which have greatly improved this paper.

This research has been supported by *Ministerio de Economía y Competitividad*, Spain (Project MTM2014-56949-C3-3-R), *Gobierno de Extremadura*, Spain (Projects GR15052 and GR15106), *UNAM-DGAPA-PAPIIT* (Project IA106416), Mexico, and *European Union* (European Regional Development Funds).

References

- Miller N, Revista de Logopedia, Foniatria y Audiología (2009) Communication changes in Parkinson's disease. Elsevier 29: 37-46.
- Duffy JR (2005) Motor speech disorders: Substrates, differential diagnosis and management. Elsevier.
- Rahn DA, Chou M, Jiang JJ, Zhang Y (2007) Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice* 21: 64-71.
- Skodda S, Rinsche H, Schlegel U (2009) Progression of dysprosody in Parkinson's disease over time- A longitudinal study. *Mov Disord* 24: 716-722.
- Shao J, MacCallum JK, Zhang Y, Sprecher A, Jiang JJ (2010) Acoustic analysis of the tremulous voice: Assessing the utility of the correlation dimension and perturbation parameters. *Journal of Communication Disorders* 43: 35-44.
- Azevedo LL, Reis C, Souza IS, Cardoso F (2013) Prosody and levodopa in Parkinson's disease. *Arquivos de Neuro-Psiquiatria* 71: 835-840.
- Baghai RL, Beet SW (2013) Automatic speech signal analysis for clinical diagnosis and assessment of speech disorders. Springer Briefs in Electrical and Computer Engineering - Speech Technologies, Springer.
- Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering* 56: 1015-1022.
- Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO (2012) Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering* 59: 1264-1271.
- Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgun F, et al. (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics* 17: 828-834.
- Hariharan M, Polat K, Sindhu R (2014) A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine* 113: 904-913.
- Pérez CJ, Naranjo L, Martín J, Campos-Campos-Roca Y (2014) A latent variable-based Bayesian regression to address recording replication in Parkinson's disease. 22nd European Signal Processing Conference: 1447-1451.
- Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *Journal of Acoustical Society of America* 129: 350-367.
- Schrag A, Ben-Shlomo Y, Quinn N (2002) How valid is the clinical diagnosis of Parkinson's disease in the community? *Journal of Neurology, Neurosurgery & Psychiatry* 73: 529-534.
- Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ (2002) Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders* 17: 867-876.
- Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, et al. (2007) Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format and clinimetric testing plan. *Mov Disord* 22: 41-47.
- Goetz CG, Stebbins GT, Wolff D, DeLeeuw W, Bronte-Stewart H, et al. (2009) Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device. *Movement Disorders* 24: 551-556.
- Tsanas A, Little MA, McSharry PE, Ramig LO (2010) Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Transactions Biomedical Engineering* 57: 884-893.
- Tsanas A, Little MA, McSharry PE, Ramig LO (2010) Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *The Royal Society Interface* 8: 842-855.
- Eskidere O, Ertac F, Haniçli C (2012) A comparison of regression methods for remote tracking of Parkinson's disease progression. *Expert Systems with Applications* 39: 5523-5528.
- Mohammadi P, Hatamlou A, Masdari M (2013) A comparative study on remote tracking of Parkinson's disease progression using data mining methods. *International Journal in Foundations of Computer Science and Technology* 3: 71-83.
- Castelli M, Vanneschi L, Silva S (2014) Prediction of the unified Parkinson's disease rating scale assessment using a genetic programming system with geometric semantic genetic operators. *Expert Systems with Applications* 41: 4608-4616.
- Hoehn MM, Yahr MD (1967) Parkinsonism: Onset, progression and mortality. *Neurology* 17: 427-442.
- Scanlon BK, Katzen HL, Levin BE, Singer C, Papapetropoulos S (2008) A formula for the conversion of UPDRS-III scores to Hoehn and Yahr stage. *Parkinsonism and Related Disorders* 14: 379-380.
- Tsanas A, Little MA, McSharry PE, Scanlon BK, Papapetropoulos S (2012) Statistical analysis and mapping of the unified Parkinson's disease rating scale to Hoehn and Yahr staging. *Parkinsonism and Related Disorders* 18: 697-699.
- Pérez CJ, Naranjo L, Martín J, Campos-Roca Y (2014) Bayesian regression models to address replications of recordings for parkinson's disease tracking. 21st International Conference on Computational Statistics C1530.
- Naranjo L, Pérez CJ, Campos-Roca Y, Martín J (2016a) Addressing voice recording replications for Parkinson's disease detection. *Expert Systems with Applications* 46: 286-292.
- Naranjo L, Pérez CJ, Martín J (2016b) Addressing voice recording replications for tracking Parkinson's disease progression. *Medical and Biological Engineering and Computing* Accepted for publication 1-9.
- Orozco-Arroyave JR, Arias-Londoño JD, Vargas-Bonilla JF, Gonzalez-Rátiva MC, Nöth E (2014) New spanish speech corpus database for the analysis of people suffering from parkinson's disease. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*.
- Baken J, Orlikoff RK (2000) Clinical measurement of speech and voice. Singular Thomson Learning.
- Shue YL, Keating P, Vicens C, Yu K (2010) Voicesauce: A program for voice analysis. *Energy: H1-A1*.
- Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing* 28: 357-366.

33. Gómez-Vilda P (2013) Voice as vehicular tool to organic and neurological disease tracking: How far we may go? Plenary Session BIOSTEC.
34. Esposito A, Chile-Zanuy M, Edwards AM, Cordasco G, Drugman TH, et al. (2016) Recent advances in non-linear speech processing. *Springer* : 93-102.
35. Gil D, Magnus J (2009) Diagnosing Parkinson by using artificial neural networks and support vector machines. *Global Journal of Computer Science and Technology* 9: 63-71.
36. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, first edition.
37. Hadfield JD (2010) MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* 33: 1-22.
38. Archer KJ (2013) A penalized continuation ratio model for predicting an ordinal response, R Package glmpathcr.
39. Park MY, Hastie T (2007) L1 regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society* 69: 659-677.
40. Webb A (2002) Statistical pattern recognition. John Wiley and Sons.
41. Orozco-Arroyave JR, Arias-Londoño JD, Vargas-Bonilla JF, Noth E (2013) Analysis of speech from people with Parkinson's disease through non-linear dynamics, in: T. Drugman , T. Dutoit (Eds.), *Advances in Nonlinear Speech Processing*, volume LNAI 7911 of *Lecture Notes in Artificial Intelligence*. Subseries of *Lecture Notes in Computer Science*, Springer-Verlag 112-119.
42. Tsanas A, Little MA, McSharry PE, Ramig LO (2010) Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Dallas, US 594-597.
43. Tsanas A, Little MA, McSharry PE, Ramig LO (2011) Robust parsimonious selection of dysphonia measures for telemonitoring of Parkinson's disease symptom severity. In: *Models and analysis of vocal emissions for biomedical applications: 7th international workshop*. Manfredi C (Edtr) Firenze University Press, Firenze, Italy 169-172.
44. Tsanas A, Little MA, McSharry PE, Ramig LO (2013) Using the cellular mobile telephone network to remotely monitor Parkinson's disease symptom severity. *IEEE Transactions on Biomedical Engineering*.
45. Scanlon BK, Katzen HL, Levin BE, Singer C, Papapetropoulos S (2010) A revised formula for the conversion of UPDRS-III scores to Hoehn and Yahr stage. *Parkinsonism and Related Disorders* 16: 151-152.