Phonation and articulation analysis of Spanish vowels for automatic detection of Parkinson's disease

Juan Rafael Orozco-Arroyave[1,2], Elkyn Alexander Belalcázar-Bolaños[1], Julián David Arias-Londoño[1], Jesús Francisco Vargas-Bonilla[1], Tino Haderlein[2], and Elmar Nöth[2,3]

[1] Universidad de Antioquia, Medellín, Colombia
[2] Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany
[3] King Abdulaziz University, Jeddah, Saudi Arabia.

Abstract. Parkinson's disease (PD) is a chronic neurodegenerative disorder of the nervous central system and it can affect the communication skills of the patients. There is an interest in the research community to develop computer aided tools for the analysis of the speech of people with PD for detection and monitoring.
In this paper, three new acoustic measures for the simultaneous analysis of the phonation and articulation of patients with PD are presented. These new measures along with other classical articulation and perturbation features are objectively evaluated with a discriminant criterion. According to the results, the speech of people with PD can be detected with an accuracy of 81% when phonation and articulation features are combined.

Keywords: Parkinson's disease, phonation, articulation, acoustics, non-linear dynamics.

1 Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after the Alzheimer's disease [1] and about 1% of the people older than 65 suffer from this disease. About 90% of people with PD have disordered speech and such disorders are associated to motor impairments such as rigidity, bradykinesia, hypokinesia and tremor. Perceptually, speech and voice of people with PD are characterized by reduced loudness, monopitch, monoloudness, reduced stress, breathy, hoarse voice quality, imprecise articulation, among others [2]. All these symptoms are grouped and called hypokinetic dysarthria [3].

Voice problems are typically one of the first symptoms of PD, and while the disease is progressing, other speech problems appear affecting different speech characteristics, such as prosody, articulation, and fluency [2]. There are works focused on the automatic classification of speech of people with PD and people without any speech disorder or neurological disease, also called healthy controls

(HC) [4,5,6]; however, the real contribution of the features considered on those works remains unclear.

In order to provide an informative feedback during the speech therapy of people with PD, there exists the need for analyzing the speech of people with PD considering different characteristics. The Royal College of Physicians in the United Kingdom states that it is required to tackle the speech therapy of the patients considering respiratory exercises whose direct impact is in the loudness and phonation [7]. In [4] the authors analyze different acoustic measures to detect disordered speech in people with PD. The analysis includes different versions of jitter and shimmer along with several noise measures. A correlation analysis to decide which feature may be included in the classification stage is performed; however, there is not presented an analysis of the discriminant capability of each feature. Therefore, it is not clear which of the considered measures is more relevant for the classification process.

Recent works claim that it is required to consider phonation, articulation and prosody to fulfill a complete analysis of speech in patients with PD. In [8] the authors present an acoustic analysis of speech of people with PD considering phonation, articulation, and prosody. Phonation is evaluated with sustained phonation of vowels in order to have the vibration of the vocal folds for a long period of time. Articulation analysis is considered in two tasks: with a diadochokinetic (DDK) exercise which consists in the rapid repetition of the syllables /pa/-/ta/-/ka/, and using sustained phonation of vowels for the analysis of their first ($F1$) and the second ($F2$) spectral formant, respectively. In [5], the authors use acoustic and prosodic features along with features derived from a two-mass model of the vocal folds to evaluate the speech of 88 patients with PD and 88 HC. The accuracy rate reported in the paper is 90.5% when the prosodic features are considered. Articulation capability of people with PD is evaluated in [9], the authors calculate the vowel formants $F1$ and $F2$ and show that triangular vowel space area (tVSA) is lower in speech of people with PD than in HC.

This paper presents an analysis of several perturbation and articulation features of speech of people with PD and HC. The discriminant capability of the features is evaluated in different experiments considering a total of 300 recordings of the five Spanish vowels, 150 uttered by patients with PD and 150 by HC.

The paper is organized as follows: Section 2 includes details of the speech corpus, the characterization process and the experiments carried out. In Section 3, the results and the discussion are presented. Finally, the conclusions derived from this work are included in Section 4.

2 Experimental Setup

2.1 Corpus of Speakers

The speech corpus contains three repetitions of the five Spanish vowels uttered by 50 patients with PD and 50 HC (150 recordings per vowel on each group, PD and HC). The recordings were sampled at 44.1 KHz with a resolution of 16 bits. All of the speakers are balanced by gender and age, i.e. the age of the 25 male patients ranges from 33 to 77 years (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 men ranges from 31 to 86 years (mean 61.2 ± 11.3) and the age of the 25 women ranges from 43 to 76 years (mean 60.7 ± 7.7). All of the patients were diagnosed by neurologist experts; the values of their evaluation according to the UPDRS-III [10] and Hoehn & Yahr [11] scales are 36.7 ± 18.7 and 2.3 ± 0.8, respectively. None of the people in the HC group has a history of symptoms related to Parkinson's disease or any other kind of neurological disorder. Further details of this database can be found in [12].

2.2 Characterization

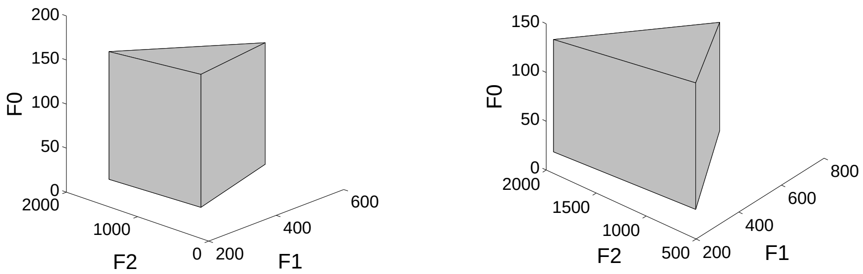
Acoustic Measures: Articulation The acoustic feature set includes measures that are calculated in the F1-F2 plane including a new set of features that are introduced in this paper. The classical acoustic features considered here are: The first two formants of each Spanish vowel calculated using a standard Linear Predictive Coding (LPC) filter with 12 coefficients: ($F1_a$, $F2_a$, $F1_e$, $F2_e$, etc.), the pitch value of each vowel: ($F0_a$, $F0_e$, $F0_i$, $F0_o$, $F0_u$), the vowel articulation index (VAI) [9] and the triangular vowel space area ($tVSA$) [9]. $F0$ value and its variability in time and amplitude measured through jitter and shimmer respectively, are features traditionally included in the analysis of speech of people with PD. With the aim of considering phonation and articulation information, in this paper we propose several measures: (1) the *vocal prism* whose base is the $tVSA$ and whose altitude is given by the variability of the pitch value estimated on the vowels /a/, /i/ and /u/, (2) the *vocal pentagon*, whose vertexes are the values of the F1 and F2 for the five Spanish vowels, and (3) the *vocal polyhedron* whose base is formed by the vocal pentagon and whose edges are given by the pitch variability obtained from the five Spanish vowels.

From this set of measures, different vocal features are derived: the volume of the vocal prism ($vPrism$), the area of the vocal pentagon ($aPenta$), the volume of the vocal polyhedron ($vPoly$) and the set of five measures that correspond to the coordinates of the centroid of each vertex in the *vocal pentagon*; they are named $CentPenta[Fk_j]$, where $k \in \{1, 2\}$ indicates the axes (F1 horizontal and F2 vertical) and $j \in \{a, e, i, o, u\}$. Figure 1 is included to give a graphical idea of the information that can be provided by the features proposed here. Note that the behavior of $tVSA$ and $aPenta$ is similar, e.g. the phonations of people with PD have a smaller area in the F1-F2 plane than HC, indicating that the articulation capability of the patients is reduced compared to HC. Note also that $aPenta$ includes information of the five Spanish vowels while $tVSA$ only can provide information about three.

$vPrism$ is the other measure introduced in this paper, its base is the vocal triangle and its altitude is the pitch variability; this measure can give information about articulation and phonation simultaneously. The measure is illustrated in Figure 2.

(a) $tVSA$ (b) $aPenta$

Fig. 1. Triangular vowel space area and vocal pentagon for PD and HC.



(a) HC (b) PD

Fig. 2. Vocal prism.

Perturbation Measures: Phonation Temporal and amplitude variation of the pitch, also known as jitter ($Jitt$) and shimmer ($Shimm$) respectively, are included along with another perturbation measure which is called Correlation Dimension CD [13].

$Jitt$ and $Shimm$ are measures commonly used to evaluate speech of people with PD; however they assume near periodicity in the speech signal, thus they could not be reliable in cases where the severity of the disease disables the estimation of the pitch [14]. In this work, the CD is included in the set of perturbation features due to its capability to describe periodicity in a speech signal. According to [14], the vibrations of the vocal folds can be represented as trajectories in the state space with time evolution, and such trajectories can give qualitative information about the phenomenon, thus periodic vibrations produce closed trajectories, whereas aperiodic vibrations produce irregular trajectories. With CD it is possible to quantify how aperiodic a speech signal is, so this measure can be considered as a perturbation feature, with the advantage of being able to give information even in cases where the estimation of other perturbation measures is not possible, i.e. when the disease is advanced [13]. See [15], for more details about the estimation of the CD.

2.3 Experiments

The recordings are windowed each 40 ms with an overlap of 20 ms. The features are calculated for each frame, forming a feature vector for each measure. Afterward, four statistics are calculated for each feature: mean value (m), standard deviation (std), kurtosis (k) and skewness (sk).

The discrimination analysis is performed in two stages. The first is the linear Bayesian classification of recordings from PD and HC speakers. With the aim of obtain a sub set with the most discriminative measures, after the first classification stage, only the features with the minimum accuracy of 61% (after several tests with different numbers this was the accuracy that offered the best compromise between classification and number of features), are included in the second classification stage which consists of the classification with a soft-margin support vector machine (SVM) with Gaussian kernel [16]. This classifier is used because of its extensive usage in the state of the art for the automatic classification of pathological and healthy voices [4,5,15]. The parameters of the SVM are optimized using a 10-fold cross validation strategy.

3 Results and Discussion

The set with articulation features that exceed the threshold of accuracy includes $std[F1_o]$, $std[F0_u]$, $centPenta[F2_u]$, $std[Vprism]$, $m[F2_u]$, $std[F0_a]$, $m[VAI]$, $std[F1_i]$, $std[F1_a]$, $std[F0_i]$, $std[F0_e]$ and $std[F1_u]$. Four of the twelve features correspond to measures of the variability of the pitch (different statistics of $Jitt$, $Shim$, and CD), which confirms previous results reported in [9] evaluating utterances of German speakers. Note also that two of the measures introduced in this paper are in the selected group: $centPenta[F2_u]$ and $std[Vprism]$. These features can give information about phonation and articulation processes simultaneously, and have similar performance to classical acoustic features, such as vocal formants and VAI.

In addition, the perturbation features are tested on each Spanish vowel and the same criteria (accuracy 61%) is applied to decide whether a feature is included in the feature set of the second stage of classification.

The second stage of classification is performed with a soft-margin SVM with margin parameter C and Gaussian kernel with parameter γ, both optimized in a grid-search following a cross-validation strategy. Articulation and phonation features are combined in the same representation space, and the performance of the system is improved from 74.0% to 81.3% with a total of 73 features (phonation + articulation). Figure 3 shows the accuracy obtained when more features are added to the classification process incrementally.

The results obtained with each set of features (phonation and articulation), as well as with the union of all features, are presented in Table 1. The performance is presented in terms of accuracy, sensitivity, specificity and the area under the ROC curve (AUC).

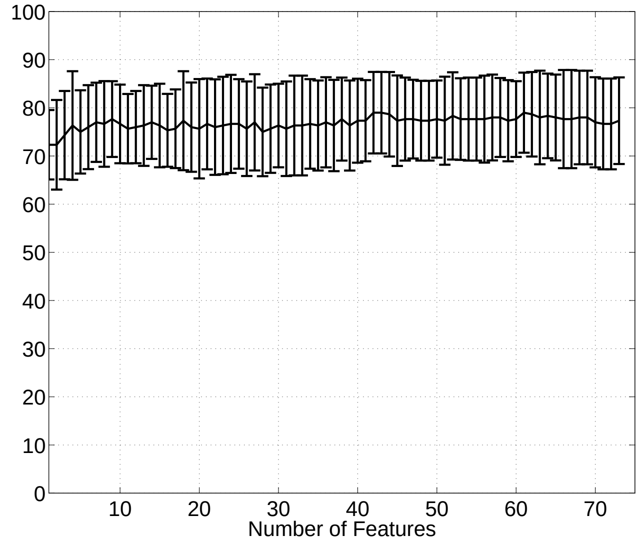


Fig. 3. Performance with the phonation and articulation features.

	Accuracy(%)	Sensitivity(%)	Specificity(%)	AUC
Articulation	79.3 ± 5.8	80.2 ± 10	79.3 ± 12.4	0.76
Phonation	74.0 ± 9.6	73.1 ± 9.9	73.8 ± 14.1	0.74
Union	81.3 ± 5.5	82.5 ± 9.9	80.9 ± 4.8	0.85

Table 1. Classification results with SVM.

The performance of each feature set can be observed in Figure 4 in a more compact way. Note the improvement in the accuracy when phonation and articulation features are combined.

4 Conclusions

New features for the simultaneous analysis of phonation and articulation in speech are presented. The features were evaluated with discriminative criteria in two stages. The first stage was performed with a linear Bayesian classifier and the second with a soft-margin SVM. Only the features that exceeded 61% of accuracy in the first stage were included for the second stage of classification. After the first classification stage, the set of features that remained for the second stage includes different standard features such as the variability of F0 and the vocal formants F1 and F2. Additionally, two of the features introduced in this paper remained for the second stage; one is $centPenta[F2_u]$ which is based on measurements of the second formant of the vowel /u/, and the other one is

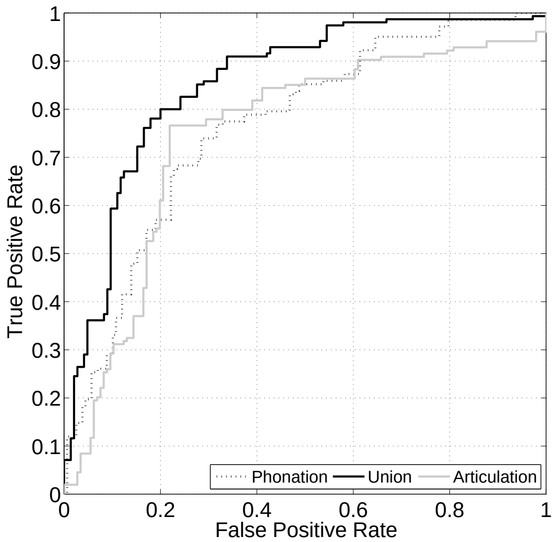


Fig. 4. ROC curves obtained with phonation and articulation features and their union.

$std[Vprism]$ which compiles information about the variability of F0, F1, and F2 measured on the corner vowels /a/, /i/, /u/.

The phonation analysis is performed considering standard features such as jitter, shimmer, and correlation dimension. According to the results, when phonation features are evaluated individually, jitter is the most discriminative; however, when these measures are considered in the same representation space, the accuracy is improved.

The accuracy obtained in the automatic classification of speech of people with PD and HC was improved from 74.0% to 81.3% when articulation and phonation features are combined.

Acknowledgments. Juan Rafael Orozco Arroyave is under grants of COLCIENCIAS through the program "Convocatoria № 528, generación del bicentenario 2011". Elkyn Belalcazar Bolaños is under grants of "Convocatoria 617, jovenes investigadores e innovadores 2013" funded by COLCIENCIAS. This work was also financed by COLCIENCIAS through the project № 111556933858.

References

1. de Rijk, M., et al: Prevalence of Parkinson's Disease in Europe: A collaborative study of population-based cohorts. Neurology **54** (2000) 21–23
2. Ramig, L., Fox, C., Sapir, S.: Speech Treatment for Parkinson's Disease. Expert Review Neurotherapeutics **8**(2) (2008) 297–309

3. Darley, F., Aronson, A., Brown, J.: Differential Diagnosis Patterns of Dysarthria. Motor Speech Disorders (1975)
4. Little, M.A., McSharry, P., Hunter, E., Spielman, J., Ramig, L.: Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Bio-medical Engineering **56**(4) (2009) 1015–1022
5. Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., Rusz, J.: Detection of Persons with Parkinson's Disease by Acoustic, Vocal and Prosodic Analysis. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). (2011) 478–483
6. Bayestehtashk, A., Asgari, M., Shafran, I., Mcnames, J.: Fully Automated Assessment of the Severity of Parkinson's Disease from Speech. Computer Speech & Language, 2014 1–14 To appear.
7. National Collaborating Centre for Chronic Conditions: Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care. London: Royal College of Physicians (2006)
8. Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E.: Quantitative Acoustic Measurements for Characterization of Speech and Voice Disorders in Early Untreated Parkinson's Disease. The Journal of the Acoustical Society of America **129**(1) (2011) 350–367
9. Skodda, S., Visser, W., Schlegel, U.: Vowel Articulation in Parkinson's Disease. Journal of Voice **25**(4) (2011) 467–472
10. Stebbing, G., Goetz, C.: Factor Structure of the Unified Parkinson's Disease Rating Scale: Motor Examination Section. Movement Disorders **13** (1998) 633–636
11. Hoehn, M.M., Yahr, M.D.: Parkinsonism: Onset, Progression, and Mortality. Neurology **17** (1967) 427–442
12. Orozco-Arroyave, J., Arias-Londoño, J., Vargas-Bonilla, J., González-Rátiva, M., 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. To appear.
13. Lee, V., Zhou, X., Rahn, D., Wang, E., Jiang, J.: Perturbation and Nonlinear Dynamic Analysis of Acoustic Phonatory Signal in Parkinsonian Patients Receiving Deep Brain Stimulation. Journal of Communication Disorders **41**(6) (2008) 485–500
14. Jiang, J., Zhang, Y., McGilligan, C.: Chaos in Voice: From Modeling to Measurement. Journal of Voice **20**(1) (2006) 2–17
15. Orozco-Arroyave, J.R., Vargas-Bonilla, J.F., Arias-Londoño, J.D., Murillo-Rendón, S., Castellanos-Domínguez, G., Garcés, J.: Nonlinear Dynamics for Hypernasality Detection in Spanish Vowels and Words. Cognitive Computation **5**(4) (2013) 448–457
16. Schölkopf, B., Smola, A.: Learning With Kernel. The MIT press (2002)