Automatic Detection of Parkinson's Disease Using Noise Measures of Speech

E.A. Belalcazar-Bolaños[1*], J.R. Orozco-Arroyave[1,3], J.D. Arias-Londoño[2], J.F. Vargas-Bonilla[1] and E. Nöth [3]
[1]Department of Electronics and Telecommunications Engineering, Universidad de Antioquia, Colombia.
[2]Department of Systems Engineering, Universidad de Antioquia, Colombia.
[3]Universität Erlangen-Nürnberg, Erlangen, Germany.
[*]Corresponding author: elkyn.belalcazar@udea.edu.co

[1]

***Abstract*—Parkinson's disease (PD) is a neurodegenerative disorder that is characterized by the loss of dopaminergic neurons in the mid brain. It is demonstrated that about 90% of the people with PD also develop speech impairments, exhibiting symptoms such as monotonic speech, low pitch intensity, inappropriate pauses, imprecision in consonants and problems in prosody; although they are already identify problems, only 3% to 4% of the patients receive speech therapy. The research community has addressed the problem of the automatic detection of PD by means of noise measures; however, in such works only the phonation of the English vowel /a/ has been considered. In this paper, the five Spanish vowels uttered by 50 people with PD and 50 healthy controls (HC) are evaluated automatically considering a set of four noise measures: Harmonics to Noise Ratio (HNR), Normalized Noise Energy (NNE), Cepstral HNR (CHNR) and Glottal to Noise Excitation Ratio (GNE). The decision on whether a speech recording is from a person with PD or from a HC is taken by a K nearest neighbors (k-NN) classifier, finding an accuracy of 66.57% when only the vowel /i/ is considered.**

Keywords: **Noise measures, k-nearest neighbor, Parkinson's disease, Spanish vowels.**

I. Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders with a prevalence rate exceeding $100/100.000$ [1]. PD is characterized by the loss of dopaminergic neurons in the mid brain and its main symptoms of PD are tremor, rigidity and other movement disorders. It is demonstrated that about 90% of the people with Parkinson's disease (PPD) also develop speech impairments [2], however only from 3% to 4% of the patients receive speech therapy [3]. Given that age is the single most important factor for PD and the fact that older population is growing, these figures could further increase in the not too distant future [4].

Different voice tests have been introduced to extract the symptoms of dysphonia. For example in sustained phonation approach [5], the test subject is introduced to pronounce a sentence constructed from representative linguistic units. Although this kind of tests are useful to assess dysphonia of the patient, their usefulness for evaluating the severity of the PD is still unclear. Different efforts to analyze the influence of the disease in the speech of PD patients have emerged, in [6] and [7] the authors study changes in the low frequency spectra in order to characterize possible displacements of the velum due to the lack of control of this limb. According to the results, low frequency region gives important information able to characterize speech impairments in PPD. On the other hand, different noise measures that quantify the increase of aeroacoustic noise due to excessive turbulence because of the incomplete vocal fold closure have been used along with different nonlinear dynamics features. In [8] the harmonic-to-noise ratio (HNR) and glottal to noise excitation ratio (GNE) are combined with eleven Mel-frequeny Cepstral Coefficients (MFCC) and different complexity measures such as correlation dimension, recurrence period density entropy, detrended fluctuation analysis, among others. With a set with 33 patients and 10 healthy controls, the authors report accuracies of up to 97.1% when a subset that includes noise measures, MFCC and nonlinear dynamics features is considered.

Although there are works tackling the problem of the automatic classification of speech signals uttered by PPD, there are few reported experiments considering the five Spanish vowels and even the performance of systems only considering noise measures is not evaluated yet.

In this paper we propose the use of only noise measures for the automatic classification of speech from PPD. The set of features includes harmonics to noise ratio (HNR), normalized noise energy (NNE), cepstral HNR (CHNR) and glottal to noise excitation ratio (GNE).

The rest of the paper is organized as follows. Section 2 presents the methodology that we are addressing to perform the experiments, section 3 provides details of the experimental framework with special attention to the classification and error estimation methodologies. Section 4 presents the obtained results and finally, in section 5 the conclusions derived from the work are provided.

II. Methodology

Figure 1 shows a schematic of the methodology used in this work. The most representative stages of the process will be explained in the following subsections.

A. Characterization

Speech recordings are preprocessed by means of a short temporal analysis using windows of 40ms length with an

[1] 978-1-4799-1121-9/13/$31.00 ©

Figure 1. Methodology

overlap of 20ms. After, each frame is characterized by means of the noise measures: HNR, CHNR, NNE and GNE.

1) Harmonics to Noise Ratio: This measure is based on the assumption that the acoustic wave of a sustained vowel consist of two components: a periodic component that is the same from cycle to cycle and an additive noise component that has a zero-mean amplitude distribution. The concatenation of the signal intervals gives rise to a new signal averaged over which the energy is estimated. Noise energy is calculated as the substraction between the energy of the original signal and the energy of the averaged signal. Finally, HNR is the relation between Harmonic structure energy, from the speech signal, and the additive noise due to different voice disorders. According to Yumoto, et al. [9] this ratio can be calculated using the following procedure:

1) The speech signal $x(n)$ is divided into N intervals whose length is the pitch and its average the sum of the intervals according to 1:

$$x_A(n) = \frac{1}{N}\sum_{i=1}^{N} x_i(n) \tag{1}$$

2) The harmonic component H is calculated from the smoothed speech signal $x_A(n)$ through the equation 2:

$$H = N\sum_{n=0}^{T} x_A^2(n) \tag{2}$$

Where T is the average size of the N estimated pitch periods:

3) The noise component N is estimated as the energy of the signal remaining after substracting the smoothed signal from the original signal x_A for each interval x_i. According to 3:

$$N = \sum_{i=1}^{N}\sum_{n=0}^{T}\{x_i(n) - x_A(n)\}^2 \tag{3}$$

4) Finally, the harmonic to noise ratio in dB is calculated as:

$$HNR_{dB} = 10log_{10}\left(\frac{H}{N}\right) \tag{4}$$

2) Cepstral Harmonics to Noise Ratio: The estimation of HNR in cepstral domain is based on the method proposed by Guus de Krom in [10]. The aim of this method is to consider all the spectral components of the speech, improving the precision in the estimation of the noise levels in a signal. In this process the speech signal $x(n)$ is divided into x_i intervals whose duration is the pitch period; such intervals are windowed and then the cepstrum $C_{\widehat{x}_i}$ are calculated. In the cepstral domain the harmonics are filtered (the process is also called liftering); then the Fourier transform is estimated and the result is the noise spectrum N_i. The harmonic spectrum is calculated as the subtraction between the logarithm of the spectrum of each frame $FFT(\widehat{x}_i)$ and the noise spectrum N_i. In order to make a more accurate estimate of the noise level, it is necessary to make a correction on its calculation, which involves finding the minimum between successive harmonics and the harmonic spectrum, resulting the vector mHi and subtracting the noise spectrum N_i. The result is the noise level $NOISE_i$. CHNR is the ratio between the absolute value of the Fourier transform of each frame $FFT(\widehat{x}_i)$ called $SIGNAL_i$, and the noise level $NOISE_i$, in decibels.

$$CHNR_{dB} = 10log_{10}\left(\frac{NOISE_i}{SIGNAL_i}\right) \tag{5}$$

3) Normalized Noise Energy: NNE is another feature based on noise measures from a speech signal and the method was proposed by Kasuya, et al. in [11]. It considers a windowed speech signal $x_i(n)$ in the m-th frame of vowel phonation of periodic components $s_i(n)$ and an additive noise component $w_i(n)$. It is represented by: $x_i(n) = s_i(n) + w_i(n)$.

Let $X_i(k)$,$S_i(k)$, and $W_i(k)$ be the discrete Fourier transform of $x_i(n)$, $s_i(n)$ and $w_i(n)$, respectively; then $X_i(k) = S_i(k) + W_i(k)$.

Additionally, let $\left|\widehat{W}_i(k)\right|^2$ be an estimated of $|W_i(k)|^2$; then the NNE is defined as:

$$NNE_{dB} = 10log\left(\frac{\frac{1}{L}\sum_{k=N_L}^{N_H}\sum_{i=1}^{L}\left|\widehat{W}_i(k)\right|^2}{\frac{1}{L}\sum_{k=N_L}^{N_H}\sum_{i=1}^{L}|X_i(k)|^2}\right) \tag{6}$$

Where $N_L = [Nf_LT]$, $N_H = [Nf_HT]$ and, f_L and f_H are the low and high frequency respectively of the frequency band where the noise energy is evaluated, the brackets denote the greatest integer function, and T is the sampling period. Since the denominator in the eq. 6 can be directly computed from the Fourier transform of the input speech signal, the problem is to devise a methodology to obtain an estimate $\left|\widehat{W}_i(k)\right|^2$. Representing $S_m(k)$ and $W_m(k)$ in polar coordinates as:

$$S_i(k) = |S_i(k)|\, e^{j\theta(k)} \quad and \quad W_i(k) = |W_i(k)|\, e^{j\phi(k)} \tag{7}$$

it follows that,

$$|X_i(k)|^2 = |S_i(k)|^2 + |W_i(k)|^2 + 2\,|S_i(k)|\,|W_i(k)|\,cos[\theta(k) - \phi(k)]$$

Since $S_i(n)$ has been assumed to be a periodic component of the speech signal $x_i(n)$, $|S_i(k)|$ contributes to the harmonic structure of $|X_i(k)|$. Using a Hamming window with a relatively large value of M, $|S_i(k)|$ becomes small in the harmonic deep region. Thus an estimate $\left|\widehat{W}_i(k)\right|^2$ can be given in the deep region by:

$$\left|\widehat{W}_i(k)\right|^2 = |X_i(k)|^2 \quad , \quad k \in D_j \tag{8}$$

Where D_j is a set of k's corresponding to the i-th deep region.

4) Glottal to Noise Excitation Ratio: GNE was proposed by D. Michaelis in [12], this measure is well known for being more robust than other noise measures; property that is awarded primarily because its estimation process does not require the prior calculation of pitch periods, which gives advantage specially when working with high level of pathology, where the estimation of pitch is a very difficult problem. GNE measures the amount of excitation voice due to the vibration between the vocal folds versus the excitation noise caused by turbulence in the vocal tract.

The method starts by re-sampling the signal at $10Khz$, then it is necessary to find the glottal pulses of the voice, which can be achieved by using a linear prediction inverse filtering performed on $30ms$ interval of the signal. Subsequently, it is necessary to implement a series of bandpass filters using Hamming windows, whose number, location and bandwidth were set for an actual voice signal by J. Godino, et al. in [13]. Optimal values of bandwidth bandpass filter applied to seek voice disorders is $1000hz$, applied in bands that increase in steps of $300hz$. Finally, for each of the intervals filtered from the signal $x_i(n)$, whose duration is given by glottal pulses found in the process of inverse filtering, the Hilbert envelope and their respective cross-correlation sequences are calculated. The maximum of all sequences is the value of GNE.

B. Automatic features selection and classification

The selection of features is addressed through the application of principal components analysis (PCA). It is a statistical technique applied here to find out a low-dimensional representation of the original feature space, searching for directions with greater variance to project the data. Although, PCA is commonly used as a feature extraction method, it can be used to properly select a relevant subset of original features that better represent the studied process [14]. In this sense, given a set of features $(\boldsymbol{\xi}_k : k = 1, \ldots, p)$ corresponding to each column of the input data matrix $\mathbf{X}$, the relevance of each $\boldsymbol{\xi}_k$ can be analyzed for finding the resulting subspace $\mathbf{Y}$. More precisely, relevance of $\boldsymbol{\xi}_k$ can be identified looking at $\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_p \end{bmatrix}^\dagger$, where $\boldsymbol{\rho}$ is defined as $\boldsymbol{\rho} = \sum_{j=1}^{m} |\lambda_j \mathbf{v}_j|$. ($\lambda_j$ and $\mathbf{v}_j$ are the eigen-values and eigen-vectors of the initial matrix, respectively). Therefore, the main assumption is that the largest values of ρ_k point out to the best input attributes, since they exhibit higher overall correlations with principal components.

The decision of whether a voice recording is from PPD or HC is taken with a K nearest neighbor (K-nn) classifier. Considering that the aim of this work is to analyze the discrimination capability of the described features, this classifier is chosen because of its simplicity allows us to focus on the analysis of the considered features and not on the classifier.

III. Experimental framework

A. Database

The data for this study consists of speech recordings from 50 PPD and 50 HC sampled at $44.100hz$ with 16 quantization bits. All of the recordings were captured in a sound proof booth. The people that participated in the recording sessions are balanced by gender and age: the ages of the men patients ranged from 33 to 77 (mean 62.2 $\pm$ 11.2) and the ages of the women patients ranged from 44 to 75 (mean 60.1 $\pm$ 7.8). For the case of the healthy people, the ages of men ranged from 31 to 86 (mean 61.2 $\pm$ 11.3) and the ages of the women ranged from 43 to 76 (mean 60.73 $\pm$ 7.7). All of the PPD have been diagnosed by neurologist experts and none of the people in the HC group has history of symptoms related to Parkinson's disease or any other kind of movement disorder syndrome.

The recordings consist of sustained utterances of the five Spanish vowels, every person repeated three times the five vowels, thus in total the database is composed of 150 recordings per vowel on each class. This database was built by *Universidad de Antioquia* in Medellín, Colombia.

B. Experimental setup

The voice recordings were segmented and windowed using frames of $40ms$ with an overlap of $20ms$. The characterization of speech recordings is made considering the noise measures that were described above. Each measure is obtained from every frame of each voice signal and after that, four statistics are estimated per measure (mean value, standard deviation, kurtosis and Sweness). In this work we propose 2 realizations of the experiment: the first one consists of including each noise measure with its 4 statistics in order to analyze the discriminant capacity of each measure separately, and the second one consists of considering a total of 16 features (4 statistics of 4 nose measures) to represent each voice recording. Table I summarizes the set of features considered in the second realization of the experiment and the index assigned for each feature.

Table I
INDEX ALLOCATION FOR FEATURES

	GNE	HNR	CHNR	NNE
Mean	1	5	9	13
Std	2	6	10	14
Kurtosis	3	7	11	15
Skewness	4	8	12	16

The tests performed over the proposed system have been made following the strategy indicated in [15]. The 70% of the data is used for feature selection and for training the classifier and the remaining 30% is for testing; ten different subsets for training and testing are randomly formed and the process is repeated ten times in order to obtain confidence intervals for the estimation of the general performance of the proposed system.

IV. Results and discussion

Table II shows the results of the first experiment, where each measure is evaluated separately with its 4 statistics per vowel. The aim of this step is to analyze which features contribute significantly to the process of classification between PD and HC. It can be noticed that the fact to evaluate the measures individually provides very poor results, reporting accuracies of up to 62.29% when the NNE is considered in the vowel /a/.

Table II
PERFORMANCE MEASURES PER FEATURE

Vowel	Feature	Accuracy	Specificity	Sencitivity
/a/	**GNE**	0.568±0.039	0.578±0.042	0.564±0.052
	HNR	0.588±0.030	0.581±0.040	0.599±0.034
	CHNR	0.544±0.043	0.548±0.044	0.548±0.064
	NNE	**0.622±0.039**	**0.626±0.046**	**0.620±0.038**
/e/	**GNE**	0.598±0.031	0.619±0.060	0.596±0.058
	HNR	0.591±0.028	0.600±0.050	0.588±0.030
	CHNR	0.527±0.035	0.536±0.050	0.544±0.038
	NNE	0.592±0.034	0.593±0.029	0.593±0.046
/i/	**GNE**	0.585±0.037	0.583±0.040	0.592±0.052
	HNR	0.586±0.031	0.585±0.046	0.597±0.046
	CHNR	0.486±0.026	0.484±0.027	0.489±0.030
	NNE	0.565±0.024	0.572±0.031	0.561±0.033
/o/	**GNE**	0.611±0.023	0.608±0.024	0.616±0.040
	HNR	0.535±0.026	0.549±0.025	0.522±0.031
	CHNR	0.497±0.032	0.496±0.031	0.500±0.041
	NNE	0.550±0.041	0.553±0.047	0.551±0.043
/u/	**GNE**	0.527±0.029	0.528±0.039	0.531±0.033
	HNR	0.499±0.016	0.503±0.016	0.498±0.027
	CHNR	0.491±0.027	0.495±0.025	0.490±0.038
	NNE	0.606±0.045	0.618±0.062	0.605±0.053

In the second realization of the experiment, all features are considered in the same representation space. In order to eliminate redundancy and to reduce dimensionality, we have used an automatic feature selection process based on PCA that gives a subset of features that better represent the phenomena and also gives the order of features according to their contribution in terms of the cumulative variance. Table III indicates which of the features remain after the features selection process per vowel.

Table III
INDEXES OF SELECTED FEATURES

Vowel	Feature Index											
/a/	11	16	9	2	4	1	7	13	6	14	15	
/e/	9	4	16	2	13	12	5	15	8	6	14	
/i/	4	12	13	16	9	2	1	8	6	14		
/o/	12	4	13	7	16	1	14	6	9	2	5	15
/u/	4	5	16	10	12	13	15	7	2	1	6	14

Considering the order of the features given by the PCA process, we evaluate the accuracy in the classification incrementally, i.e. initially, only the first feature is considered, then the first two features, and so on. Figure 2 shows the increasing in the accuracy rate while more features are considered, it can be noticed that in all cases the system is not able to reach a stable zone of accuracy. It can also be noted that the best accuracy rate is obtained for the vowel /i/.

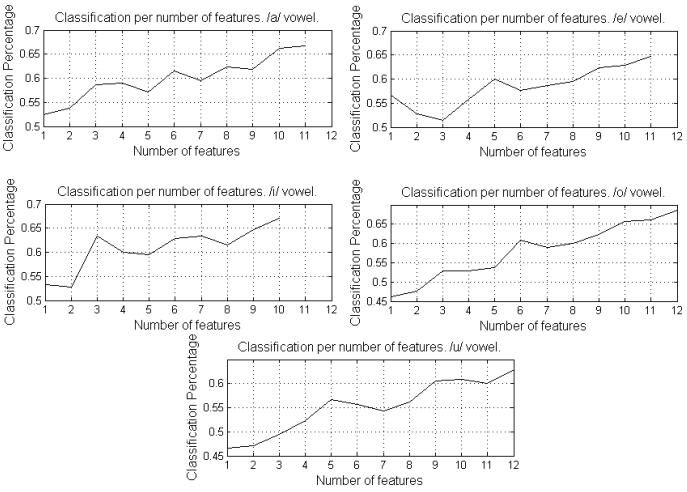

Figure 2. Success rate per vowel

Table IV indicates the results obtained per vowel in terms of accuracy, specificity and sensitivity. Note that the vowels /e/ and /i/ exhibit the best results while the worst performance is achieved with vowel /u/.

Table IV
PERFORMANCE MEASURES WITH ALL THE FEATURES

Vowel	Accuracy	Specificity	Sensitivity
/a/	0.652±0.021	0.670 ±0.039	0.650±0.057
/e/	**0.664±0.036**	**0.659±0.062**	**0.677±0.039**
/i/	**0.665±0.032**	**0.695±0.043**	**0.645±0.045**
/o/	0.652±0.029	0.690±0.064	0.633±0.032
/u/	0.619±0.034	0.634±0.048	0.612±0.042

V. Conclusion

Four noise measures are considered to characterize speech phonations from people with Parkinson's disease. The evaluations have been performed on the five Spanish vowels and according to the results, when such kind of measures are considered separately, they do not show significant contributions in the task of automatic classification of speech from PPD and HC.

However, the success rate increases when the representation space is formed considering information from all the measures together.

Even though the considered features are not discriminant enough, it is worth to highlight that the proposed methodology allows to add other kind of features and then to increase the classification rates.

For future work more features, from different nature and domains must be considered in order to broaden the analysis capabilities to account possible phenomena in PD that may be have not been considered yet in the state of the art.

Acknowledgment

Juan Rafael Orozco Arroyave is under grants of "Convocatoria 528 para estudios de doctorado en Colombia, generación

del bicentenario, 2011" funded by COLCIENCIAS. The authors give a special thanks to all of the patients and collaborators in the *Fundalianza Parkinson-Colombia*, Without their valuable support it would be impossible to address this research. This work was granted by COLCIENCIAS, project # 111556933858.

REFERENCES

[1] A. S. von Campenhausen, B.Bornschein, R.Wick, K.Botzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger, , and R. Dodel, "Prevalence and incidence of parkinsons disease in europe," *Eur. Neuropsychopharmacol.*, vol. 15, pp. 473–490, 2005.

[2] A. Ho, R. Iansek, C. Marigliani, J. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with parkinson's disease," *Behavioral Neurology*, vol. 11, pp. 131–137, 1998.

[3] L. Ramig, C. Fox, and S. Shimon, "Speech treatment for parkinson's disease," *Expert Review Neurotherapeutics*, vol. 8, no. 2, pp. 297–309, 2008.

[4] S. D. Eeden, C. Tanner, A. L. Bernstein, R. Fross, A. Leim-peter, D. A. Bloch, , and L. Nelson, "Incidence of parkinsons disease: Variation by age, gender, and race/ethnicity," *Am. J. Epidem*, vol. 157, pp. 1015–1022, 2003.

[5] P. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. V. D. Heyning, M. Remacle, and V. Woisard, "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS), Eur Arch Otorhinolaryngol*, vol. 258, no. 7, pp. 77–82, 2001.

[6] P. Vijayalakshmi and M. Reddy, "Assessment of dysarthric speech and an analysis on velopharyngeal incompetence," in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2006, pp. 3759–3762.

[7] E. Belalcázar-Bolaños, J. Orozco-Arroyave, J. Vargas-Bonilla, J. Arias-Londoñno, C. Castellanos-Domínguez, and E. Nöth, "Low-frequency of speech for automatic detection of parkinson's disease," in *Lecture Notes in Computer Science*, vol. 7930, 2013, pp. 283–292.

[8] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[9] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics to noise ratio as hoarseness index of degree of hoarseness," *Journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1544–1549, 1982.

[10] G. de Krom, "Cepstrum based technique for determining a harmonics-to-noise ratio in speech signals," *Journal of Speech, Language and Hearing Research*, vol. 36, no. 2, pp. 254–266, 1993.

[11] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pahologic voice," *Journal of Acoustical Society of America*, vol. 80, no. 5, pp. 1329–1334, 1986.

[12] D. Michaelis, T. Gramss, and H. Strube, "Glottal to noise excitation ratio - a new measure for describing pathological voices," *Acustica/Acta acustica*, vol. 83, pp. 700–706, 1997.

[13] J. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, P. Gómez-Vilda, M. Blanco-Velasco, and F. Cruz-Roldán, "The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders," *Journal of Voice*, vol. 24, no. 1, 2010.

[14] G. Daza-Santacoloma, J. Arias-Londoño, J. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and C. G. Castellanos-Domínguez, "Dynamic feature extraction: an application to voice pathology detection," *Intelligent Automation and Soft Computing*, vol. 15, no. 4, pp. 665–680, 2009.

[15] N. Sáenz-Lechón, J. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, pp. 120–128, 2006.