Analysis of Speech from People with Parkinson's Disease through Nonlinear Dynamics

Juan Rafael Orozco-Arroyave[1,2], Julián David Arias-Londoño[1], Jesús Francisco Vargas-Bonilla[1], and Elmar Nöth[2]

[1] Universidad de Antioquia, Medellín, Colombia
[2] Friedrich-Alexander Universität, Erlangen - Nürnberg, Germany

Abstract. Different characterization approaches, including nonlinear dynamics (NLD), have been addressed for the automatic detection of PD; however, the obtained discrimination capability when only NLD features are considered has not been evaluated yet.

This paper evaluates the discrimination capability of a set with ten different NLD features in the task of automatic classification of speech signals from people with Parkinson's disease (PPD) and a control set (CS). The experiments presented in this paper are performed considering the five Spanish vowels uttered by 20 PPD and 20 people from the CS.

According the results, it is possible to achieve accuracy rates of up to 76,81% considering only utterances from the vowel /i/. When features calculated from the five Spanish vowels are combined, the performance of the system is not improved, indicating that the inclusion of more NLD features to the system does not guarantee better performance.

Keywords: Nonlinear dynamics, complexity measures, Parkinson's disease, speech signals.

1 Introduction

PD is a neurodegenerative disorder that results from the progressive death of dopaminergic cells in the substantia nigra, a region of the mid-brain. About 89% of PPD commonly develop speech impairments affecting different aspects such as respiration, phonation, articulation and prosody [1]. Speech impairments in PPD are related to the vocal fold bowing and incomplete vocal fold closure [2], besides the vocal production is a highly nonlinear dynamical system, thus the changes caused by impairments in the movement of different muscles, tissues and organs which are involved in the voice production process, such as those suffered by PPD, can be modeled using NLD analysis [3], [4].

NLD techniques have been applied for both the automatic assessment of pathological speech signals and the automatic evaluation of speech from PPD. In [5] the authors include four NLD features along with other 13 acoustic measures for the automatic detection of PD. The set of NLD features includes correlation dimension (D_2), Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA) and Pitch Period Entropy (PPE). According to their results, it is possible to achieve classification rates of up to 91.4%. Additionally, in [6] the

T. Drugman and T. Dutoit (Eds.): NOLISP 2013, LNAI 7911, pp. 112–119, 2013.

evolution of the PD through the time is studied using a set of features composed by different dysphonia measures (including some from NLD) and analyze their correlation with the evolution of the patients according to the Unified Parkinson Disease Rating Scale (UPDRS) [7] in a period of six months. The authors stated that UPDRS scale can be mapped with a precision of up to 6 points.

Despite the interest of the scientific community to apply NLD for the automatic assessment of speech from PPD, the discrimination capability of NLD features is not clear yet because they have been combined with other features such as acoustics and noise measures. In this paper, different state of the art NLD features are implemented and their discrimination capability is objectively evaluated on the automatic classification of speech signals from PPD and CS. The set of features considered in this study includes a total of 10 measures which have been used for the automatic detection of different speech disorders such as hypernasality [8] and dysphonia [5], [9]. The features are: correlation dimension, largest Lyapunov exponent, Lempel-Ziv complexity, Hurst exponent, RPDE, DFA, approximate entropy, approximate entropy with Gaussian kernel, sample entropy, sample entropy with Gaussian kernel.

The paper is organized as follows: section 2 presents a brief description of the methods that are applied in this work, in the section 3 the details of the performed experiments is given, section 4 shows the obtained results and finally, the section 5 provides the conclusions that are derived from the presented work.

2 Methodology

The general methodology that is applied in this work is depicted in figure 1. The signal is first preprocessed by means of its division into frames. After, the characterization is performed. In this case, only NLD features have been considered for this stage. With the aim of eliminate possible redundancy in the information provided by all NLD a features selection stage is required. Finally, the decision about whether a speech signal comes from a person with PD or a CS is taken through an automatic classification strategy. In this work, this step is performed using support vector machines (SVM). In the following subsections, more details of each part of the methodology will be provided.

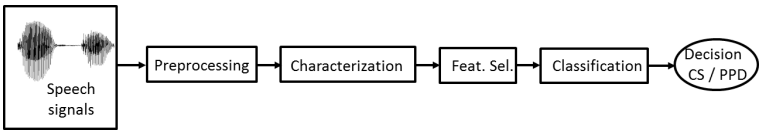

Fig. 1. General methodology

2.1 Nonlinear Dynamics Characterization

A set of ten NLD features is calculated to perform the automatic classification of speech signals from PPD and CS. The first step in the characterization process

is to embed the signal into the state space following the the time-delay embedding theorem originally proposed by Takens [10]. This theorem establishes that where there is a single sampled quantity of a dynamical system, it is possible to reconstruct a state space that is equivalent or diffeomorphic to the original state space that is unknown. Points in the state space form trajectories, and a set of trajectories from a time series is known as attractor [3]. After the embedding process, the NLD features are estimated as it is briefly described below.

Correlation Dimension (D_2): it is a measure of the space dimensionality occupied by the points in the reconstructed attractor. In this work, (D_2) is implemented according to the Takens estimator method [3]. The estimation requires the use of the correlation sum ($C(r)$), which is defined as: $C(r) = \sum_{i=1}^{N} C_i^m(r)$. Where,

$$C_i^m(r) = \frac{2}{N(N-1)} \sum_{j=i+1}^{N} \Theta\left(r - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|\right) \tag{1}$$

N is the number of points in the state space, Θ the Heaviside function and ($\|\cdot\|$) is a norm defined in any consistent metric space. D_2 is theoretically defined for an infinity amount of data ($N \to \infty$) and for small r, thus its general expression is written as:

$$D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{\partial \ln C\left(r, N\right)}{\partial \ln\left(r\right)} \tag{2}$$

Larges Lyapunov Exponent (LLE): this feature is estimated as the average divergence rate of neighbor trajectories in the attractor, according to the Ronsenstein method [3]. For this algorithm, once again the nearest neighbors to every point in the trajectories must be estimated. In this case, a neighbor must fulfill a temporal separation greater than the "period" of the time series, to be considered as a nearest neighbor. It is possible to state that the separation of points in a trajectory is according to the expression $d(t) = Ce^{\lambda_1 t}$, where λ_1 is the maximum Lyapunov exponent, $d(t)$ is the average divergence taken at the time t, and C is a normalization constant. Assuming that the $j - th$ pair of nearest neighbors approximately diverge at a rate of λ_1, it is possible to obtain the expression $ln(d_j(i)) = ln(C_j) + \lambda_1(i\Delta t)$, where λ_1 is the slope of the average line that appears when such expression is drawn on a logarithmic plane [3].

Lempel-Ziv Complexity (LZC): it is included for the estimation of the randomness of the voice signals. The method consists in finding the number of different "patterns" present in a given time series according to the algorithm presented in [11]. As the algorithm only considers binary strings; for the practical case, a value of 0 is assigned when the difference between two successive samples is negative, and 1 when such a difference is positive or null (see [11] for additional details).

Hurst Exponent (H): the possible long term dependencies in a time series can be estimated trough H. It is calculated following the rank scaling method [3]. Where the relation between the variation rank (R) of the signal, evaluated

in a segment, and its standard deviation S is given by $\frac{R}{S} = cT^H$, where c is a scaling constant, T is the duration of the segment and H is the Hurst exponent.

Entropy Measurements: in general, entropy is a measure of the uncertainty of a random variable. When there is a stochastic process with a set of independent but not identically distributed variables, the rate at which the joint entropy grows with the number of variables n is given by $H(X) = -\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i)$.

For the case of a state space, it can be partitioned into hypercubes of content ϵ^m and observed at time intervals δ, defining the Kolmogorov-Sinai entropy as:

$$H_{KS} = -\lim_{\substack{\delta\to\infty \\ \varepsilon\to 0 \\ n\to\infty}} \frac{1}{n\delta} \sum_{k_1,\ldots,k_n} p(k_1,\ldots,k_n) \log p(k_1,\ldots,k_n) \tag{3}$$

where $p(k_1,\ldots,k_n)$ is the joint probability that the state of the system is in the hypercube k_1 at the time $t = \delta$, k_2 at $t = 2\delta$, etc. For stationary processes, it can be shown that $H_{KS} = \lim_{\delta\to 0} \lim_{\varepsilon\to 0} \lim_{n\to\infty} (H_{n+1} - H_n)$.

In practical terms it is not possible to compute the equation 3 for $n \to \infty$, thus different estimation methods have been proposed in the literature. One of them is the *Approximate entropy* (A_E), which is designed for measuring the average conditional information generated by diverging points on a trajectory in the state space [12]. For fixed m and r, A_E is estimated as:

$$A_E(m,r) = \lim_{N\to\infty} \left[\Phi^{m+1}(r) - \Phi^m(r)\right] \tag{4}$$

where $\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)$, and $C_i^m(r)$ was defined in equation 1.

The main drawback of A_E is its dependence to the signal length due to the self comparison of points in the attractor. In order to overcome this problem, the *sample entropy* (S_E) is proposed as:

$$S_E(m,r) = \lim_{N\to\infty} \left(-\ln \frac{\Gamma^{m+1}(r)}{\Gamma^m(r)}\right) \tag{5}$$

The only difference between Γ in the equation 5 and Φ in the equation 4 is that the first does not evaluate the comparison of embedding vectors with themselves.

Another modification of A_E is the *approximate entropy with Gaussian kernel* A_EGK. It exploits the fact that Gaussian kernel function can be used to give greater weight to nearby points by replacing the Heaviside function by [13].

$$d_G(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\left(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_1\right)}{10r^2},\right) \tag{6}$$

The same procedure of changing the distance measure can be applied to define the *sample entropy with Gaussian kernel* S_EGK.

On the other hand, considering that the voice signal has two components, deterministic and stochastic, in [14] was proposed to analyzed the deterministic

component by means of the *recurrence period density entropy* ($RPDE$), considering a hypersphere of radius $r > 0$, containing a embedded data point $\boldsymbol{x}(t_i)$. The time $t_r = t_j - t_i$ is the recurrence time, where t_j is the instant at which the trajectory first returned to the same hypersphere. If $R(t)$ is the normalized histogram of the recurrence times estimated for all embedded points into a reconstructed attractor, the $RPDE$ can be defined as in the equation 7.

$$RPDE = \frac{-\sum_{i=1}^{t_{\max}} R(i) \ln R(i)}{\ln t_{\max}} \tag{7}$$

where t_{max} is the maximum recurrence time in the attractor. Besides, the stochastic component of the voice signals can be analyzed by means of the *detrended fluctuation analysis* (DFA) to estimate the scaling exponent α in non-stationary time series as is indicated in [14].

2.2 Feature Selection and Classification

In characterization stages large amounts of information are produced. Such information is represented in high dimensionality spaces which most of the times have redundant information. The reduction of the dimensionality of such spaces and the elimination of redundant information is performed applying the Sequential Floating Features Selection (SFFS) algorithm. It is such that finds the best subset of features of the original set through the inclusion and exclusion of features. In this procedure after each forward step, a number of backward steps are applied as long as the resulting subsets are better, in the sense of the accuracy, than the previous one [15]. The decision about whether a voice recording is from PPD or CS is taken by a SVM with Gaussian kernel which parameters are optimized in the training process [16]

3 Experimental Setup

3.1 Corpus of Speakers

Speech recordings from 20 PPD and 20 people from the CS are considered (10 women and 10 men). The ages of the men patients ranged from 56 to 70 (mean 62.9 $\pm$ 6.39) and the ages of the women patients ranged from 57 to 75 (mean 64.6 $\pm$ 5.62). For the case of the healthy people, the ages of men ranged from 51 to 68 (mean 62.6 $\pm$ 5.48) and the ages of the women ranged from 57 to 75 (mean 64.8 $\pm$ 5.65). All of the PPD have been diagnosed by neurologist experts and none of the people in the CS 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 60 recordings per vowel on each class. This database is built by *Universidad de Antioquia* in Medellín, Colombia.

3.2 Experiments

First, each Spanish vowel is considered separately. The recordings are preprocessed by means of its division into frames with $55ms$ of length with an overlap of 50%, according to [9]. After, NLD features are calculated for each frame and four statistics are calculated for each feature. The considered statistics are *mean value*, *standard deviation*, *skewness* and *kurtosis*, thus each recording will be represented by a total of 40 features (four statistics on ten features).

The validation of the system's performance is made by the division of the data into 70% for training and 30% for testing, following the methodology exposed in [17]. The 70% of the data are used for the feature selection and for training the classifier and the remaining 30% of the data are used for testing; the different subsets for training and testing are randomly formed ten times. For each pair of subsets (train and test), the classification process is repeated ten times, forming a total of 100 independent realizations of the experiment obtaining results with confidence intervals of the system's performance.

In order to look for better accuracies, the selected features per vowel are combined, collecting information from the five Spanish vowels. This combination is performed considering the same process that was described above. The features selection process is applied again and the decision about whether a speech recording is from PPD or CS is taken with a SVM. The results are presented according to [17], indicating accuracy rates, specificity and sensitivity. Specificity indicates the probability of a healthy register to be correctly detected and sensitivity is the probability of a pathological signal to be correctly classified.

4 Results

Table 1 shows the results obtained when each Spanish vowel is considered, and the last row of the table indicates the results obtained when all Spanish vowels are combined into the same space. According to the results, the best performance is obtained when the vowel /i/ is considered separately. This is an interesting result specially if it is considered that the production of the vowel utterances involved several muscles but among them the risorius muscles only participate in the phonation of vowels /e/ and /i/ [18]. Further experiments should be addressed to get a deeper understanding about the impact of PD in the movement of the muscles involved in speech production.

Although the combination of the five vowels can provide more information about the phenomena, the obtained results in such case are not as well as expected. It could happen because the combination of features from all vowels rises the dimensionality of the system and it can result in an increment of redundant information that does not contribute to the correct classification of the speech recordings. To present the results compactly, a detection error tradeoff (DET) curve is shown in figure 2. The line corresponding to the performance for vowel /i/ is separated from the others, indicating that such performance is the best among the Spanish vowels and even better than those obtained with the combination of features from the five vowels.

Table 1. Accuracy results obtained per vowel and with the combination of all vowels

Vowel	accuracy	sensitivity	specificity
/a/	72,49 ± 2,68	68,30 ± 7,08	76,67 ± 5,86
/e/	71,58 ± 5,29	69,50 ± 6,07	73,67 ± 8,72
/i/	**76,81 ± 3,77**	**76,61 ± 9,96**	**77,00 ± 7,62**
/o/	70,23 ± 3,71	69,99 ± 7,86	70,45 ± 7,83
/u/	73,36 ± 3,92	75,45 ± 3,85	71,28 ± 7,29
All vowels	74,03 ± 3,96	71,50 ± 8,92	76,06 ± 5,19

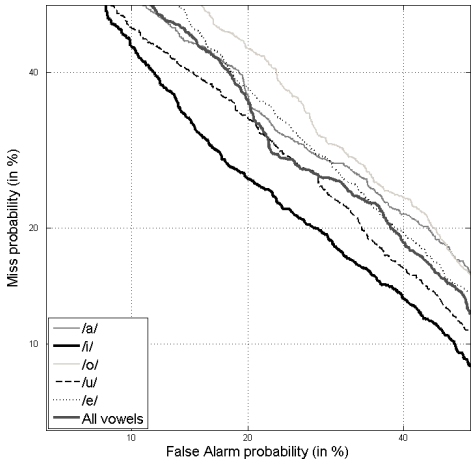

Fig. 2. DET curves for each vowel

5 Conclusions

Different state of the art NLD features have been evaluated in the automatic classification of speech recordings from PPD and CS. This work has considered experiments with the five Spanish vowels in order to state which of them provide better performance. According to our results, speech recordings from PPD and CS can be better classified considering NLD features by means of the evaluation of the vowel /i/. Additionally, with the voice samples evaluated, it is possible to state that the combination of features from the five vowels does not provides increases in the performance of the system.

Other works in the state of the art report higher accuracy rates; however, such works consider NLD features combined with other acoustic measures such as Harmonics to Noise Ratio, jitter and shimmer. The results reported in this work allow to state which is the real contribution of each NLD feature for the automatic classification of speech signals from PPD and CS.

Acknowledgments. Juan Rafael Orozco Arroyave is under grants of "Convocatoria 528 para estudios de doctorado en Colombia 2011" financed by COLCIENCIAS. The authors give a special thanks to all of the patients and collaborators in the "Fundalianza Parkinson-Colombia" foundation, without their valuable support it would be impossible to address this research.

References

1. Ramig, L.O., Fox, C., Shimon, S.: Speech treatment for parkinson's disease. Expert Review Neurotherapeutics 8(2), 297–309 (2008)
2. Perez, K.S., Ramig, L.O., Smith, M.E., Dromery, C.: The parkinson larynx: tremor and videostroboscopic findings. Journal of Voice 10(4), 353–361 (1996)
3. Kantz, H., Schreiber, T.: Nonlinear time series analysis, 2nd edn. Cambridge University Press, Cambridge (2006)
4. Giovanni, A., Ouaknine, M., Guelfucci, R., Yu, T., Zanaret, M., Triglia, J.M.: Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds. Journal of Voice 13(4), 456–476 (1999)
5. Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O.: Suitability of dysphonia measurements for telemonitoring of parkinson's disease. IEEE Transactions on Bio-Medical Engineering 56(4), 1015–1022 (2009)
6. Tsanas, A., Little, M., McSharry, P., Ramig, L.: Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. IEEE Transactions on Biomedical Engineering 57(4), 884–893 (2010)
7. Kostek, B., Kaszuba, K., Zwan, P., Robowski, P., Slawek, J.: Automatic assessment of the motor state of the parkinson's disease patient-a case study. Diagnostic Pathology 7(1), 1–8 (2012)
8. Orozco-Arroyave, J., Arias-Londoño, J.D., Bonilla, J.V., Nöth, E.: Automatic detection of hypernasal speech signals using nonlinear and entropy measurements. In: Proceedings of the INTERSPEECH (2012)
9. Arias-Londoño, J., Godino-Llorente, J., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Domínguez, G.: Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. IEEE Transactions on Bio-medical Engineering 58(2), 370–379 (2011)
10. Takens, F.: Detecting strange attractors in turbulence. Dynamical Systems and Turbulence: Lecture Notes in Mathematics, vol. 898, pp. 366–381 (1981)
11. Kaspar, F., Shuster, H.G.: Easily calculable measure for complexity of spatiotemporal patterns. Physical Review A 36(2), 842–848 (1987)
12. Costa, M., Goldberger, A., Peng, C.: Multiscale entropy analysis of biological signals. Physical Review E 71, 1–18 (2005)
13. Xu, L.S., Wang, K.Q., Wang, L.: Gaussian kernel approximate entropy algorithm for analyzing irregularity of time series. In: Proceedings of the International Conference on Machine Learning and Cybernetics, pp. 5605–5608 (2005)
14. Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.E., Moroz, I.M.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical Engineering Online 6(23), 1–19 (2007)
15. Novovicova, J., Pudil, J.K.P.: Floating search methods in feature selection. Pattern Recognition Letters 15(11), 1119–1125 (1994)
16. Scholköpf, B., Smola, A.: Learning with Kernel. The MIT press (2002)
17. Sáenz-Lechón, N., Godino-Llorente, J., Osma-Ruiz, V., Gómez-Vilda, P.: Methodological issues in the development of automatic systems for voice pathology detection. Biomedical Signal Processing and Control 1, 120–128 (2006)
18. Phonetics, D.: Dissection of the speech production mechanism. Working Papers in Phonetics, UCLA (102), 1–89 (2002)