Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches

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

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

Abstract. In this paper we propose a methodology for the automatic detection of Parkinson's Disease (PD) by using several glottal flow measures including different time-frequency (TF) parameters and nonlinear behavior of the vocal folds. Additionally, the nonlinear behavior of the vocal tract is characterized using the residual wave. The proposed approach allows modeling phonation (glottal flow) and articulation (residual wave) properties of speech separately, which opens the possibility to address symptoms related to dysphonia and dysarthria in PD, independently. Speech recordings of the five Spanish vowels uttered by a total of 100 speakers (50 with PD and 50 Healthy Controls) are considered. The results indicate that the proposed approach allows the automatic discrimination of PD patients and healthy controls with accuracies of up to 78% when using the TF-based measures.

Keywords: dysarthria, nonlinear behavior, glottal flow, Parkinson's Disease, dysphonia, time-frequency.

1 Introduction

Parkinson's Disease (PD) is the most common neurodegenerative disorder in patients older than 65, it affects about 1.5 million of people in the United States of America, and the cost of their treatment will rise up to 50 million dollars in 2040 [1]. The speech symptoms of people with PD (PPD) include problems in respiration, phonation, articulation, and prosody [2]. Usually research related to PD is focused on measuring and identifying patterns in speech, using computational intelligence and pattern recognition techniques [3, 4], and it is showing a feasible detection. Such techniques also allow to model phonation, articulation, and prosody phenomena [5]. However, the phonation and articulation processes are clearly defined. When they are analyzed from speech recordings, the information from both processes are combined, making it difficult to conclude whether the results come from phonatory or articulatory impairments. This is the case in studies analyzing the nonlinear behavior directly from the speech signal [6, 7], where the authors did not conclude which impairment causes the nonlinear behavior in PD detection. In order to analyse only the phonation process by means of glottal closure patterns from the speech signal, it is necessary to apply techniques that separate the information contributed by the articulators and the glottis.

Abnormalities in the phonation process have been observed in the glottal closure pattern of PPD through laryngeal videoscopic examination [8]. It has revealed that the irregular glottal closure pattern is the most frequent symptom in PD speech, leading to a perceptual impression of breathy voice. Vocal fold bowing and slowed vibration are also observed [9]. These changes are caused by impairments in the movements of various muscles, tissues, and organs, which are involved in the voice production process [10], showing a highly nonlinear behavior. Laryngeal videoscopy is expensive and time consuming, thus the analysis of glottal patterns from speech signals is a good alternative to perform similar screenings. The glottal signal can be extracted from speech by means of Glottal Inverse Filtering (GIF) techniques [11]. GIF allows to estimate the glottal volume velocity waveform, i.e. the glottal flow from a speech signal, in which the effects of the vocal tract and lip radiation are cancelled; once the glottal flow is estimated, it is possible to reconstruct a residual wave by subtracting the glottal spectral components from the speech signal, and then phonation and articulation phenomena are considered separately.

This paper is focused on automatic detection of PD, considering glottal and residual flows with acoustic and nonlinear approaches. The aim of the work is to address two comparisons. First, the PD detection task by acoustic measures from the glottal flow and the nonlinear behavior produced by the glottal and residual flows are compared. Second, only nonlinear behavior is considered in order to analyze if the phonation process (glottal flow) or the articulation process (residual wave) provide more nonlinear information.

The rest of the paper is organized as follows. Section 2 describes the methodology, Section 3 provides details of experimental framework, Section 4 contains the results, and Section 5 comprises the conclusions derived from this work.

2 Methodology

Figure 1 displays the general stages of the proposed methodology. The recordings

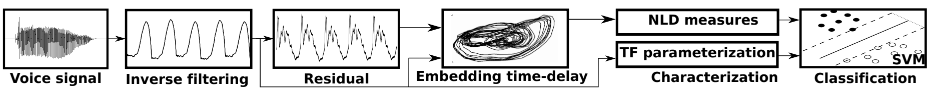

Fig. 1. General Methodology

are considered in time frames using Hamming windows with different lengths, one for time-frequency (TF) and other for nonlinear dynamics (NLD) measures. The length and time shift will be described when each feature set is introduced. The Parkinson's vs. Healthy decision is made using a Support Vector Machine (SVM). More details of the methodology are provided in the following subsections.

2.1 Glottal inverse filtering and residual wave estimation

The *Iterative and/or Adaptive Inverse Filtering (IAIF) [12]* is used to estimate the glottal flow. This method estimates the contribution of the glottal excitation on the speech spectrum with a low-order linear prediction (LP) model that is computed with a two-stage procedure. The vocal tract is then estimated using either conventional LP or discrete all-pole modeling (DAP). This method is based on an iterative refinement of both the vocal tract and the glottal components. As a result of the IAIF process, the glottal waveform $g(n)$ is obtained from the speech signal $s(n)$. Additionally, the residual waveform $r(n)$ can be estimated from $s(n)$ by subtracting the glottal log-spectral components.

2.2 Characterization

Time-frequency (TF) glottal flow parameterization

Glottal flows are preprocessed using windows with 200 ms length with an overlap of 50%, and each parameter is calculated for every glottal closure instant (GCI). The GCI is located using residual excitation and a mean-based signal algorithm [13]. Typically, time-domain features are estimated regarding the critical time instants, such as glottal opening and the glottal closure phases in the glottal flow pulse. From these critical instants, five time-domain features are obtained: **Open Quotient** (OQ), which is the ratio of the duration of the opening phase and the duration of the glottal cycle. **Closing Quotient** (CQ), defined as the ratio of closing phase duration and the glottal cycle duration. **Speed Quotient** (SQ), expressed as the ratio of opening and closing phase duration. **Amplitude Quotient** (AQ), which is defined as the ratio of the maximum of the glottal flow and the minimum of its derivative. Finally, **normalized AQ** with respect to the glottal period (NAQ). Figure 2 illustrates the described features extracted from the glottal flow. Besides, considering the spectrum of the glottal flow, some features are introduced: **H1H2**, defined as $H1 - H2$, where $H1$ and $H2$ are the first two harmonics of the glottal flow signal, and the **Harmonic Richness Factor** (HRF) is calculated as the ratio of the sum of the harmonics amplitude and the amplitude of the fundamental frequency.

Nonlinear Dynamics (NLD) measures

Glottal and residual flows are preprocessed by means of a short-time analysis using windows of 55 ms length with an overlap of 50%, where the glottal inverse filtering process has been applied previously. Before estimation of the nonlinear features, an embedding attractor has to be reconstructed from each flow. The

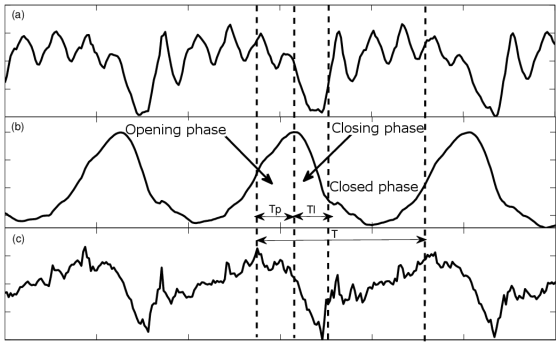

Fig. 2. Features extracted from the glottal flow. (a) Three periods of the voice signal, (b) the glottal flow, (c) derivative of the glottal flow signal.

state-space reconstruction is based on the time-delay embedding theorem [10]. A set of eight NLD measures is calculated after the embedding process: **Correlation dimension (D_2)** measures the space dimensionality occupied by the points in the reconstructed attractor. It is implemented according to the Takens estimator method [10]. **Largest Lyapunov exponent (LLE)** is estimated as the average divergence rate of neighboring trajectories in the attractor, according to the Rosenstein method [10]. **Lempel-Ziv complexity (LZC)** is used for complexity estimation in time series, its implementation consists of finding the number of different "patterns" present in a reconstructed binary string sequence. **Hurst exponent (H)** estimates the long-term dependencies in a time series, defined as the relation between the variation rank (R) of the signal and its standard deviation S, $\frac{R}{S} = cT^H$, where c is a scaling constant, and T is the duration of the segment, the estimation following the rank scaling method [10]. Moreover, entropy measures based on the uncertainty of a random variable are considered. By taking into account that in practical terms the Kolmogorov-Sinai entropy can not be computed, different estimation methods are used. 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 [14]. The main drawback of A_E is its dependence on the signal length due to the self-comparison of points in the attractor. In order to overcome this problem, the **sample entropy** (E_S) is proposed. The only difference lies in the non-comparison of embedding vectors with themselves. Another modification of A_E is the **approximate entropy with Gaussian kernel** E_AGK. It exploits the fact that a Gaussian kernel function can be used to give greater weight to nearby points by replacing the Heaviside function. Finally, the same procedure of changing the distance measure can be applied to define the **sample entropy with Gaussian kernel** E_SGK.

2.3 Classification

An SVM classifier is trained using a radial basis Gaussian kernel with bandwidth σ. To achieve a more robust classifier, the number of support vectors is also optimized with respect to the accuracy obtained in the training process avoiding over-fitting, increasing the generalization ability of the classifier and exhibiting better and more stable results [15]. This classifier is considered here due to its validated success in similar studies that addressed the problem of the automatic detection of pathological speech signals [16].

3 Experimental setup

3.1 Corpus of speakers – PC-GITA

This database [17] contains speech recordings of 50 patients with PD and 50 healthy controls (HC) sampled at 44.1 kHz with 16-bit resolution. The speakers in this database are balanced by gender and age between the two subgroups. All of the patients were diagnosed by neurologist experts; the mean values of their evaluation according to the UPDRS-III and Hoehn & Yahr scales are 38.2 and 2.3, respectively. None of people in the HC group has history of symptoms related to PD or any other kind of movement disorder. The recordings consist of sustained phonation of the five Spanish vowels: /a/, /e/, /i/, /o/, and /u/. Every person repeated the five vowels three times, thus in total the database is composed of 150 recordings per vowel on each class, e.g., PD or HC, respectively.

3.2 Experiment

TF glottal flow parameters and nonlinear behavior from the glottal and residuals waves are considered. First, each one of the seven TF parameters described above are calculated for every glottal closure instant for every person. Second, eight NLD measures, also described above, are obtained from every frame of each one of the signals. Finally, by considering the fact that every measure has a dynamic representation, four functionals are estimated on each parameter per recording: mean value (m), standard deviation (std), kurtosis (k), and skewness (sk).

The classification is performed using a soft margin support vector machine (SVM) with margin parameter C and a Gaussian kernel with parameter γ. The parameters of the SVM are optimized using steps of powers of ten through a gridsearch with $10^{-1} < C < 10^4$ and $1 < \gamma < 10^3$, and the accuracy on the test data as a selection criterion. Note that the optimization criterion could lead to slightly optimistic accuracy estimates, but as there are only two parameters to be optimized, the bias effect should be minimal. The SVM is tested following a 10-fold cross-validation strategy. The folds were formed randomly but ensuring speaker independence and balance in age and gender per fold.

3.3 Results

The results are presented in terms of accuracy, sensitivity, and specificity. The area under the ROC curve (AUC) is also presented in order to give compact information regarding the general performance of the system. Table 1 shows the results obtained when each Spanish vowel is modeled using the TF parameters over the glottal flow (Glot. TF), and using nonlinear features extracted from the glottal and residual flows (Glot. NLD and Res. NLD, respectively). Additionally, the last 3 rows of the table contain the results when each measure was obtained considering the union of the five Spanish vowels.

Table 1. Performance considering TF and NLD measures

	Vowel	**Accuracy** (%)	**Sensitivity** (%)	**Specificity** (%)	**AUC**
Glot. TF	/a/	76 ± 8	73 ± 19	79 ± 14	$\boxed{0.81}$
Glot. NLD		73 ± 8	76 ± 10	70 ± 19	0.75
Res. NLD		77 ± 7	74 ± 12	81 ± 12	0.75
Glot. TF	/e/	77 ± 9	85 ± 14	69 ± 14	$\boxed{0.81}$
Glot. NLD		75 ± 10	78 ± 14	73 ± 21	0.76
Res. NLD		74 ± 8	73 ± 13	75 ± 21	0.72
Glot. TF	/i/	72 ± 11	73 ± 15	71 ± 28	$\boxed{0.76}$
Glot. NLD		72 ± 10	71 ± 19	73 ± 20	0.71
Res. NLD		72 ± 6	73 ± 16	71 ± 20	0.74
Glot. TF	/o/	75 ± 8	67 ± 13	84 ± 16	$\boxed{0.78}$
Glot. NLD		73 ± 7	74 ± 10	71 ± 17	0.71
Res. NLD		71 ± 12	67 ± 15	75 ± 20	0.70
Glot. TF	/u/	75 ± 8	69 ± 17	82 ± 8	$\boxed{0.77}$
Glot. NLD		73 ± 5	75 ± 12	71 ± 16	0.72
Res. NLD		75 ± 7	70 ± 17	81 ± 24	0.76
Glot. TF	Union	78 ± 10	78 ± 15	77 ± 19	$\boxed{0.81}$
Glot. NLD		75 ± 9	79 ± 16	71 ± 22	0.79
Res. NLD		75 ± 9	69 ± 12	81 ± 22	0.77

Glot. TF: Glottal Time-Frequency, Glot. NLD: Glottal Nonlinear Dynamics, Res. NLD: Residual Nonlinear Dynamics

Note that TF parameters achieve the best performance in most of the cases; only with the vowel /a/, the best performance is presented by NLD features from the residual flow with 76%. Although TF parameters present the best accuracy values, it can be noted that the sensitivity and specificity values are not strongly similar; thus, the ability to detect a person with PD or a HC is not the same. But when the vowels are considered jointly, the accuracy value achieves the best performance, and both sensitivity and specificity are sufficiently similar to detect pathological speakers or healthy persons.

Furthermore, when the AUC is considered, it can be noted that TF parameters achieved the best performance in all the vowels and with their union. Figure 3 shows the best ROC curves obtained when the vowels are considered jointly.

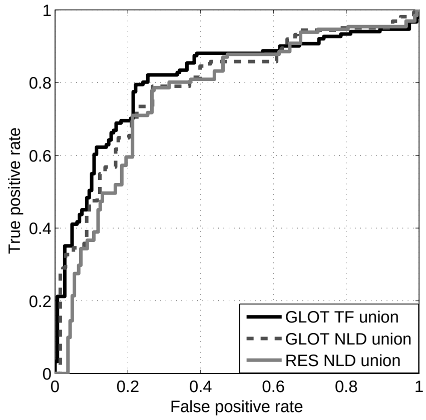

Fig. 3. ROC curves obtained when the vowels are merged

4 Conclusions

Problems in PPD related to vocal bowing and incomplete close of vocal folds are analyzed by means of the automatic separation of source, i.e., glottal flow, from the speech signal. Besides, when the source was estimated, it could obtain the residual wave, which gives information related to the articulation process. It could also be able to give some clues about the nonlinear behavior in the vocal tract, possibly due to the turbulent flow. One of the aims of the work was to determine which set of features offers more discriminatory capability to detect PD, either the TF parameters from the glottal flow, or the NLD measures estimated from the glottal and residual flows. In this sense, by means of accuracy and also with AUC measures of performance, the TF parameters are the best set in this task. It could be due to the ability of representing the glottal phases, describing in detail the phonatory process which is strongly involved in the speech impairments in PPD.

The second aim of this work was to analyze whether the nonlinearity in speech signals comes from the phonation or articulation process. The results show a similar behavior when NLD measures of glottal and residual flows are compared, thus it seems like phonation is not the only phenomenon in speech that is providing nonlinearities; there should be a nonlinear effect in the articulation process when a turbulent flow passes through the vocal tract. This work is our first approach to PD detection using nonlinear behavior of the glottal flow. For future work more nonlinear features will be considered to improve the accuracy and robustness of the models.

Acknowledgments. This work was financed by COLCIENCIAS, project № 111556933858.

References

1. Ramig, L., Fox, C., Sapir, S.: Speech treatment for Parkinson's disease. Expert Review of Neurotherapeutics **8**(2) (2008) 297–309
2. Ho, A., Jansek, R., Marigliani, C., Bradshaw, J., Gates, S.: Speech impairment in a large sample of patients with Parkinson's disease. Behavioral Neurology **11** (1998) 131–137
3. Tsanas, A., Little, M., McSharry, P., Spielman, J., Ramig., L.: Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Transactions on Biomedical Engineering **59**(5) (2012) 1264–71
4. 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 IEEE ASRU. (2011) 478–483
5. Orozco-Arroyave, J., Arias-Londoño, J., Vargas-Bonilla, J., Daqrouq, K., Skodda, S., Rusz, J., Hönig, F., Nöth, E.: Automatic detection of Parkinson's disease in continuous speech spoken in three different languages. Journal of The Acoustical Society of America **139**(1) (2016) 481–500
6. Orozco-Arroyave, J., Arias-Londoño, J., Vargas-Bonilla, J., Nöth, E.: Analysis of speech from people with Parkinson's disease through nonlinear dynamics. In: NOLISP 2013. Volume 7911 of Lecture Notes in Computer Science Springer. (2013) 112–119
7. Little, M., McSharry, P., Hunter, E., Spielman, J., Ramig, L.: Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Biomedical Engineering **56**(4) (2009) 1015–1022
8. Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M., Gunal, D.: Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurol Scand **117**(1) (2008) 26–34
9. Merati, A., Heman-Ackah, Y., Abaza, M., Altman, K., Sulica, L., Belamowicz, S.: Common movement disorders affecting the larynx: A report from the neurolaryngology committee of the AAO - HNS. Otolaryngology-Head and Neck Surgery **133** (2005) 654–665
10. Kantz, H., Schreiber, T.: Nonlinear time series analysis. 2 edn. Cambridge University Press (2004)
11. Walker, J., Murphy, P.: A review of glottal waveform analysis. In: WNSP 2007. Volume 4391 of Lecture Notes in Computer Science. (2007) 1–21
12. Alku, P., Svec, J., Vilkman, E., Sram, F.: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Communication **11** (1992) 109–118
13. Drugman, T., Thomas, M., Gudnason, J., Naylor, P., Dutoit, T.: Detection of glottal closure instants from speech signals: A quantitative review. IEEE Transactions on Audio, Speech, and Language Processing **20**(3) (2012) 994–1006
14. Kaspar, F., Schuster, H.: Easily calculable measure for complexity of spatiotemporal patterns. Phys. Rev. A Gen. Phys. **36**(2) (1987) 842–848
15. Schölkopf, B., Smola, A.: Learning with Kernels. The MIT Press (2002)
16. Maier, A., Haderlein, T., Eysholdt, U., Rosanowski, F., Batliner, A., Schuster, M., Nöth, E.: PEAKS - A system for the automatic evaluation of voice and speech disorders. Speech Communication **51**(5) (2009) 425–437
17. Orozco-Arroyave, J., Arias-Londoño, J., Vargas-Bonilla, J., Gonzáles-Rátiva, M., Nöth, E.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of the 9th LREC. (2014) 342–347