2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340177

# Using the recurrence plots as indicators for the recognition of Parkinson's disease through phonemes assessment

Vasileios Skaramagkas[1,2,*], Anastasia Pentari[2,*], Dimitrios I. Fotiadis[3], Fellow, IEEE, and Manolis Tsiknakis[1,2]

***Abstract*— Parkinson's disease (PD) is considered to be the second most common neurodegenerative disease which affects the patients' life throughout the years. As a consequence, its early diagnosis is of major importance for the improvement of life quality, implying that the severe symptoms can be delayed through appropriate clinical intervention and treatment. Among the most important premature symptoms of PD are the voice impairments of articulation, phonation and prosody. The objective of this study is to investigate whether the voice's dynamic behavior can be used as possible indicator for PD. Thus in this work, we employ the recurrence plots (RPs) which derive from the analysis of the three modulated vowels /a/, /e/ and /o/, which belong to the PC-GITA dataset, and are fed as input images to a 3-channel Convolutional Neural Network-based (CNN) architecture, which, finally, differentiates the 50 PD patients from 50 healthy subjects. The experimental results obtained provide evidence that the RP-based approach is a promising tool for the recognition of PD patients through the analysis of voice recordings, with a classification accuracy achieved equal to $87\%$.**

Parkinson's disease, recurrence plots, CNNs, modulated vowels, speech analysis

## I. Introduction

Parkinson's disease (PD) affects in an increasing rate the aging population. PD patients suffer from motor and non-motor impairments in a multisymptomatic manner. Specifically, among the most important symptoms are the tremor, rigidity, bradykinesia as well as speech impairments such as dysarthria [1]. PD affects speech at the early stage of the disease and $89\%$ of the patients develop speech disorders [2]. These speech disturbances comprise reduced voice intensity and prosody as well as imprecise articulation, which is characterized by narrower pitch range, longer pauses along with tremor, harsh and breathy voice quality [1]. As a consequence, the evaluation of speech is considered critical and can provide a possible indicator for the early detection of PD.

The study of phonation and articulation is mainly based on the analysis of vowels, whilst the choice of words, sentences and monologue have been proved more appropriate for the evaluation of prosody [2]. As the clinical evaluation of classifying parkinsonian voice depends on subjective criteria and is based on the clinician's perception, recently the study of the patient's voice has been approached through mathematical and computerized methods [1]. The most common approaches exploit the four main speech aspects, i.e., the phonatory, the articulatory, the prosodic and the linguistic, which can be used as voice biomarkers.

Speech analysis has gained the researchers' interest the last decades, as speech is the most complex and important human motor skill [3]. The acoustic characteristics such as the fundamental frequency and the voice variability provide cues to the listener about the speaker's personality and identity. However, PD patients have a reduced voice variability, which makes their voice more mono loudness. As a consequence, the idea behind this study is whether the parkinsonian voice can be characterized and classified by approaches which exploit the voice variability. More specifically, we aim to investigate if PD patients' voice can be modelled by dynamic-based approaches and be differentiated from a healthy subject's voice.

Considering the speech as a dynamical system, we exploit two basic properties: (i) the fact that in a dynamical system similar states are repeated and (ii) that these states evolve in a similar manner [4]. Accordingly, a mathematical approach, which studies a system's dynamic behavior, is the recurrence plots (RPs) which study the dynamic interrelations among time-delayed copies of a time series and provide a binary matrix analogous to the correlation. In this work, we apply the RPs theory and construct matrices which represent the dynamic nature of the voice of PD patients as well as that of healthy subjects. Additionally, these matrices are analyzed by a deep learning architecture which is based on a 3-channel Convolutional Neural Network (CNN) model, thus leading to the classification of these two groups. To the best of our knowledge this is the first approach of PD analysis through a dynamic-based method which is further introduced to a deep learning model towards PD diagnosis.

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 945175 (Project: CARDIOCARE). This paper reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

*The first two authors contributed equally to this work.

[1]V. Skaramagkas and M. Tsiknakis are with the Dept. of Electrical and Computer Engineering, Hellenic Mediterranean University, GR-710 04 Heraklion, Crete, Greece vskaramag@ics.forth.gr

[2]V. Skaramagkas, A. Pentari and M. Tsiknakis are with the Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH), GR-700 13 Heraklion, Crete, Greece

[3]D. I. Fotiadis is with the Dept. of Biomedical Research, Institute of Molecular Biology and Biotechnology (FORTH), GR-451 10, Ioannina, Greece and the Dept. of Materials Science and Engineering, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR-451 10, Ioannina, Greece

## II. Related Work

People diagnosed with PD exhibit a diverse manifestation of heterogeneous symptoms which likely reflect different subtypes [5]. Speech tasks are among the most robust ways of evaluating PD and further, assessing the stage of PD to which a patient is. However, there are no widely accepted criteria and methodologies totally appropriate for the analysis of the speech recordings. Our study is based on the analysis of phonatory recordings and in more detail, on modulated vowels. The choice of the modulated vowels task was inspired by state-of-the-art studies, which aim to recognize the PD patients from short-duration tasks [1]. Regarding the analysis of phonemes, the most common approach is through feature extraction [1]. The evaluation of the PC-GITA database's sustained vowels proves that PD patients can be differentiated from healthy controls with a high accuracy [1]. On the other hand, spectrograms-based analysis, as proposed in [6], proved that the phonemes are a suitable manner of evaluating the PD patients' phonatory. However, the deep learning architectures have gained the researchers' interest recently, as they can provide more accurate detection of PD [7], [8]. The idea of RPs was inspired by the neuroscience field and specifically the study proposed in [9].

## III. Methodology

### A. The recurrence plots

Recurrence is a fundamental property of dynamical systems, which can be exploited to characterise the system's behaviour in phase space. A powerful tool for their visualisation and analysis called recurrence plot was introduced in the late 1980's [4]. In our application, suppose that a system is a time series $\mathbf{x} = [x_1, \ldots, x_N]$, where as $N$ we denote the number of samples which define the speech signal. Our purpose is to create $M$ 2-dimensional recurrences plots of equal size, one for each examined speech signal, with $M = 1, \ldots, 100$, i.e., equal to the number of the available subjects. In our mathematical implementation, we have to handle two limitations: (i) the speech signals are of long duration and (ii) the time series' RPs must have the same size, which implies that the input time series should have the same duration. As a consequence, regarding the first limitation we employed the mathematical approach denoted as "recurrence plot of recurrence plots" (RPofRPs) [10], whilst for the second limitation we downsampled the available time series, steps which will be analyzed next.

**Recurrence plot of recurrence plots:** As the time series are of long duration, first, we split each time series to $K$ segments of window length equal to $400$ samples, defined experimentally. Then, suppose that each segment exists in a phase space trajectory with embedding dimension equal to $m$. Each time series consists of $K$ segments, leading to a variety of different $m$ values for each segment. To overcome this inequality, we selected the minimum $m$ value among the different $m$'s as the most appropriate dimension of the phase space trajectory. The embedding parameter $m$ was estimated through the false nearest neighbor algorithm (fNN) [9].

After that, based on the RPofRPs method, we apply the short-term RP definition [10]. Analytically, each processed segment, of length $400$ samples, is a new time series $\mathbf{y} \in \mathrm{R}^{400\times 1}$, from which we extract a RP plot, by following the next procedure: Defining the copies of a time-series as:

$$Y(n) = [\mathbf{y}(n), \mathbf{y}(n+\tau), \ldots, \mathbf{y}(n+(m-1)\tau)], \qquad (1)$$

where $n$ denotes the segments' samples and $\tau$ the time delay, which usually equals to one.

Using the Euclidean distance, we define an unthresholded RP as follows: Suppose that we examine two distinct times, the $n_1$ and $n_2$, then the Euclidean distance $d_s$ is:

$$d_s(Y(n_1), Y(n_2)) = \sqrt{\sum_{m=0}^{m-1} (y(n_1 + m\tau) - y(n_2 + m\tau))^2}, \qquad (2)$$

and this is repeated for all the pairwise combinations of time $n_i$, with $i = 1, \ldots, 400$, leading to the RP of a segment out of the $K$'s. Then. the RPofRPs is defined as the summation of the matrix's elements, derived from the Euclidean distance between a pairwise combination of $(RP_i, RP_j)$, with $i, j = 1, \ldots, K$. To conclude, we take as a result the final RP plot, denoted as $\mathbf{R} \in \mathrm{R}^{K\times K}$.

However, the RP plots are binary matrices which are defined by thresholding the weighted $\mathbf{R}$ by a value denoted as $\varepsilon$. The parameter $\varepsilon$ defines if two state vectors are close together in the phase space trajectory. In our mathematical implementation, $\varepsilon$ was selected to be equal to $0.2$. As a consequence, after normalizing the final $\mathbf{R}$ recurrence plot values, by dividing the matrix with the maximum value, through the Heaviside function $\Theta(\cdot)$, if a $\mathbf{R}$ value is lower than $\varepsilon$ the function returns "1", otherwise "0".

The above procedure is repeated for all the time series, i.e., for the speech signals of each subjects, so as to give as input these binary plots to the deep architecture. However, as mentioned, the second limitation concerns the different lengths of the speech signals. To overcome this problem, we chose to undersample the time series to the minimum length of the whole number of the available time series.

### B. Multi-channel CNN-based architecture

The broad graphical depiction of our proposed model is presented in Fig. 1 and comprises of various components, including convolution, pooling, dropout, and dense layers. It consists of 3 inputs, each of them representing a modulated vowel. The selection of a multi-input network model provides to the user the freedom of editing the network parameters specifically for each of the chosen vowels as well as maintaining a better supervision of the effect of each vowel regarding the output of the model.

Each channel accepts a resized binary image of $48 \times 48$ as input and provides predicted probabilities as output. The architecture contains 4 (four) convolution (Conv2D) layers with $3\times3$ filters. The Rectified Linear Unit (ReLU) activation function is used in each Conv2D layer, which helps the model to avoid high vanishing gradient problems and learn

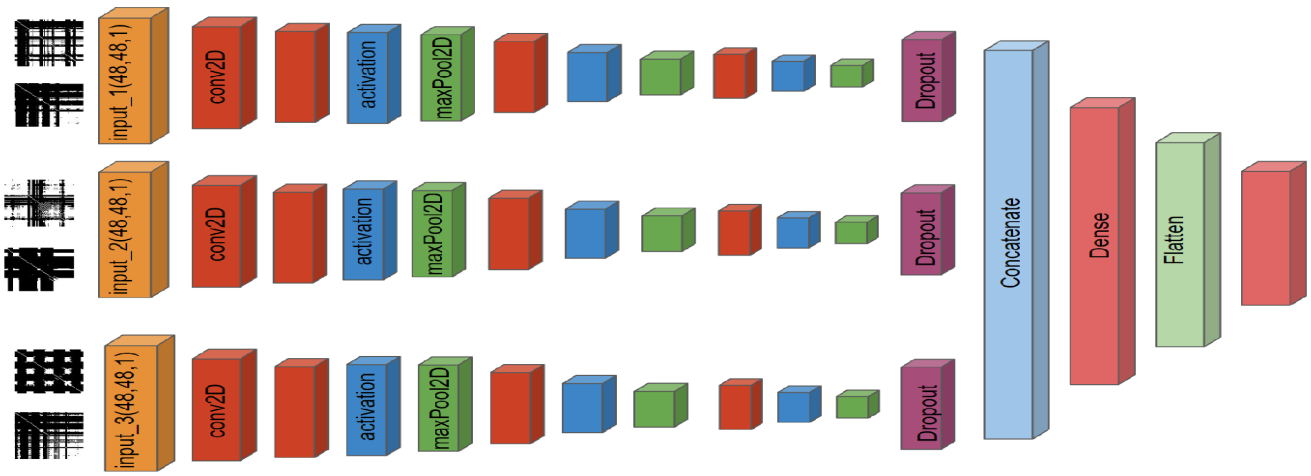


Fig. 1. Graphical representation of the multi-channel CNN-based architecture.

complex nonlinear functions while training. Moreover, 3 (three) max-pooling (MaxPool2D) layers are utilized with the kernel size of $2 \times 2$ after each Conv2D layer, except for the first one, to reduce the dimensions of resulting features maps from Conv2D layers and leave only high weighted features as output. Additionally, a dropout layer with a rate of .2 is utilized after the third MaxPool2D layer of each channel to avoid overfitting. Then, the outputs of the three channels are concatenated and served as a merged input for the remainder of the network model. Furthermore, one hidden layer, consisting of 10 neurons with a ReLU activation function in it, and one dense output layer, containing 1 neuron and a sigmoid activation function, are used to acquire probability for the two classes as an output of the model. Besides this, a flatten layer is utilized before the last dense layer to transform the multi-dimensional input to one-dimensional. A visual view of our proposed model is shown in Fig. 1 and hyper-parameters parameters of various layers are given in Table I. Finally, the final model consists of a total of 73,343 parameters.

TABLE I

HYPER-PARAMETERS OF THE PROPOSED ARCHITECTURE.

| Layer | Kernel size | Neurons | Activation | Dropout |
|---|---|---|---|---|
| Conv2D_1 (input)* | 3x3 | 8 | ReLU | - |
| Conv2D_2* | 3x3 | 10 | ReLU | - |
| MaxPool2D_1* | 2x2 | - | - | - |
| Conv2D_3* | 3x3 | 32 | ReLU | - |
| MaxPool2D_2* | 2x2 | - | - | - |
| Conv2D_4* | 3x3 | 64 | ReLU | - |
| MaxPool2D_3* | 2x2 | - | - | - |
| Dropout_1* | - | - | - | .2 |
| Concatenate | - | - | - | - |
| Dense_1 | - | 32 | ReLU | - |
| Flatten_1 | - | - | - | - |
| Dense_2 | - | 1 | Sigmoid | - |

* The layers of one out of the three total channels.

A visual view of our proposed model is shown in Fig. 1 and the hyper-parameters parameters of various layers are given in Table I.

## IV. EXPERIMENTAL EVALUATION

### A. Dataset description

The speech corpus used consists of 100 subjects, 50 healthy and 50 PD patients, and contains a variety of speech recordings in Spanish language. This database, known as PC-GITA, is balanced, as it contains equal number of healthy and PD subjects and moreover, each group consists of 25 men and 25 women. The age of male PD patients ranges from 33 - 77 whilst, for the women from 44 - 75. Regarding the healthy subjects, the age of male healthy controls ranges from 31 - 86 whilst, for the women from 43 - 76. The recordings are sampled in 44100 Hz using a dynamic omnidirectional microphone. The diagnosis of all the patients was made by neurologist experts and their evaluation was based on the UPDRS and H&Y scales. Moreover, the recordings were acquired when the patients were in ON-state, i.e., no more than 3 hours after their morning medication. Finally, none of the healthy control had symptoms associated to PD or any other neurological disease.

From the variety of the speech tasks, we employed the modulated vowels /a/, /e/, /i/, /o/ and /u/ to analyze, in terms of our study. Specifically, examining all the 3-modulated vowels combinations as well as the use of all 5 vowels, we concluded to the best 3-vowel combination based on the performance of our pipeline. It is worth to notice that, as modulated vowel task is denoted the evaluation of phonation from vowels recordings with a low to high manner.

### B. Comparative methods

The choice of the comparative methods was based on two criteria: (i) we selected the methods of the PC-GITA phonemes' analysis and (ii) among them, those which had achieved the highest accuracy results. Thus, we concluded to the [1], [6].

### C. Experimental setup and implementation

The implementation and experiments were conducted in a virtual environment based on Python version 3.9.7 that was installed on a personal computer with GTX GeForce 750 Ti GPU, Intel(R) Core(TM) i7-6700 CPU with 3.40 GHz clock speed, and 32 GB of RAM. For creating, training, and assessing the DL model, several frameworks and libraries are leveraged, including TensorFlow-GPU version 2.5.0 with the frontend of Keras-GPU. Utilizing a binary cross-entropy loss function and an Adam optimizer with an initial learning rate of $10^{-4}$, the loss of the model is computed and its weights are updated during training, respectively. In addition, the proposed model was trained on 32 minibatch sizes for 45 epochs, which took nearly 20 minutes to complete. In addition, NumPy is used for a variety of mathematical operations, such as reshaping and concatenation.

### D. Experimental Results

Regarding our experimental procedure, as time delay $\tau$ we selected the value 1, as this is proposed in [9]. We also examined different $\tau$ values, but the performance of our pipeline was not better. The embedding parameter $m$ was equal to 3. Finally, multiple thresholds were examined, to the range $[0.2 : 0.05 : 0.4]$ and the proposed results concern the parameter $\varepsilon = 0.2$.

For the procedure of the classification stage, we split the data into training and testing, with the number of the test data to be $15\%$ of the total number of examples. We evaluated the performance of our network model using the metrics of accuracy, precision, recall and specificity. Furthermore, we validated the models using a 10-fold cross-validation (10-CV). The experiments were repeated for 300 Monte-Carlo iterations.

In Table II, we present an overview of our experimental results as well as the results achieved in the comparative methods.

TABLE II
EXPERIMENTS USING DIFFERENT TRIADS OF VOWELS.

| Vowels | Accuracy | Precision | Recall | Specificity |
|---|---|---|---|---|
| /a/,/e/,/o/-proposed | **.87** | **.92** | .75 | **.91** |
| 5 sustained [1] | .84 | - | **.84** | .85 |
| modulated(male) [6] | .84 | - | - | - |
| modulated(female) [6] | .76 | - | - | - |

Notice that, all possible 3 modulated vowels combinations were examined concluding to the proposed combination with the highest classification accuracy.

## V. CONCLUSIONS

In this work we investigated whether RPs, an advanced technique of nonlinear data analysis, which describe the dynamic nature of a system can be applied on speech signals and further, whether they can successfully used as a possible indicator of Parkinson's disease. We have processed long duration time windows of voice signals through the recurrence plots theory, thus creating binary images which were given as inputs to a DL architecture. The results obtained prove that, the RPs constitute able and appropriate quantities for the classification of PD patients and healthy controls with high accuracy. It is worth to mention that, in terms of our experimental analysis, we also examined the traditional feature-based classification, i.e., we extracted the Recurrence Quantification Analysis features and applied a SVM-based classification. However, the resulted accuracy was not higher than our proposed.

Nevertheless, it is important to note that the performance of the model is highly dependent on the quality and size of the dataset it is trained on, so it can generalize well to new data. In the future, we aim to extend our analysis to the evaluation of more speech tasks, rather than the ones concerning the evaluation of the phonation, i.e., we will analyze tasks which examine the articulation and prosody. Moreover, the patients were in the ON-state of medication. Thus, we could also examine the OFF-state as the medication affects the motor PD symptoms, including the speech changes. Therefore, it is possible that the speech recordings taken from PD patients in the ON-state of medication may not accurately reflect the true severity of their speech changes. Finally, a possible extension of this work is to enhance our analysis by introducing the Mel-Spectrograms to our pipeline as they are widely used in the speech analysis field.

## REFERENCES


[1] N. D. Pah, M. A. Motin, D. K. Kumar, "Phonemes based detection of parkinson's disease for telehealth applications," in Sci Rep 12, 9687, 2022.

[2] J. R. Orozco-Arroyave, J. D. Arias-Londoño,J. F. Vargas-Bonilla, M. C. González-Rátiva, E. Nöth E., "New Speech Corpus Database for the Analysis of People Suffering From Parkinson's Disease," In Proc. Of the International Conference on Language Resources and Evaluation (lrec), pages 342-347, Reykjavik, Iceland. 2014.

[3] V. P. Martine, N. Xavier, M. Francis, P. Nathalie, "Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance," in Frontiers in Psychology, vol. 9, 2018.

[4] N. Marwan, M. C. Romano, M. Thiel, J. Kurths, "Recurrence plots for the analysis of complex systems," inPhysics Reports, vol. 438, Issues 5–6, pp. 237-329, 2007.

[5] A. Tsanas, S. Arora, "Data-Driven Subtyping of Parkinson's Using Acoustic Analysis of Sustained Vowels and Cluster Analysis: Findings in the Parkinson's Voice Initiative Study,", in SN COMPUT. SCI. 3, 232, 2022.

[6] D. Hemmerling, J. R. Orozco, A. Skalski, E. Noeth, "Automatic Detection of Parkinson's Disease Based on Modulated Vowels,", in Proc. Interspeech 2016, pp. 1190-1194, 2016.

[7] P. Varalakshmi, B. T. Priya, B. A. Rithiga, and R. Bhuvaneaswari, "Parkinson disease detection based on speech using various machine learning models and deep learning models," 2021 International Conference on System, Computation, Automation and Networking, ICSCAN 2021, pp. 1-6, 7 2021.

[8] B. Karan, S. S. Sahu, and K. Mahto, "Stacked auto-encoder based time- frequency features of speech signal for parkinson disease prediction," 2020 International Conference on Artificial Intelligence and Signal Processing, AISP 2020, pp. 1-4, 1 2020.

[9] A. Pentari, G. Tzagkarakis, P. Tsakalides, P. Simos, G. Bertsias, E. Kavroulakis, K. Marias, N. J. Simos, F. Papadaki, "Changes in resting-state functional connectivity in neuropsychiatric lupus: A dynamic approach based on recurrence quantification analysis," in Biomedical Signal Processing and Control, vol. 72(3):103285, Part A, 2022.

[10] Fukino, Miwa and Hirata, Yoshito and Aihara, Kazuyuki, "Coarse-graining time series data: Recurrence plot of recurrence plots and its application for music," in Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 26(2):023116,, 2016.