



Progress report

Detecting Parkinson's disease using Vocal Data from Patients

Applied Probability and Statistics for Engineers

INDU 6310

Submitted to: Dr. Amar Sabih

Group members:

Abdulah Alhoothy: 40075668

Muhammad Haris: 40150933

Qamar Bani-Melhem: 40109566

Sourav Patne: 40137125

Yaoxin Li: 40092820

**Concordia University
Winter 2020**

Introduction

A person who has Parkinson disease or needs a diagnosis of Parkinson disease must pass through various stages of tests. These tests are designed by specialist doctors with their teams; to better understand the severity level of Parkinson's disease. Our study aims to discriminate healthy people from people with Parkinson's disease (PD) by detecting dysphonia. Each Patient out of 31 passed through an average of 6 iterations of voice tests. The diagnosis stage or data collection stage involves the application of pre-designed and verified measurement methods to all the speech signals.

Research has shown that approximately 90% of PWP exhibit some form of vocal impairment [3, 4]. Vocal impairment may also be one of the earliest indicators for the onset of the illness [5], and the measurement of voice is noninvasive and simple to administer. Thus, voice measurement to detect and track the progression of symptoms of PD has drawn significant attention [6, 7]. The Phonations were recorded by a Multi-Dimensional Voice Program (MDVP) is the premier software tool for quantitative acoustic assessment of voice quality of a patient under observation in the laboratory, it has a capability of calculating more than 22 parameters on a single vocalization input by the patient voice signal [8].

Dataset

In this report, we are introducing the Parkinson disease (PD) and the type of parameters that had been tested on the patients. We use real data from source [1], and analyze the data with different ways. We studied 31 patients, 23 of them with PD. Then we select the highly 10 uncorrelated parameters and start working on them. The period since diagnosis with PD is from 0 to 28 years, and the ages of the subjects are from 46 to 85 [8]. Multiple phonation tests were taken by the subjects (Averages of six phonations were recorded from each subject) ranging from one to 36 seconds in length [8].

The data chosen for the project had several parameters:

- MDVP (FO): Fundamental frequency (Fo) is the vibratory rate of the vocal folds. It can be measured in hertz or cycle per second (CPS). Average fundamental frequency during a conversation for males ranges from 100 to 150 Hz, whereas for females it ranges from 180 to 250 Hz.
- MDVP(FHI): maximum FO.
- MDVP(FLO): minimum FO.
- MDVP (Jitter %): Jitter is a measure of frequency instability. A normal voice has a small amount of instability during sustained vowel production Normal instabilities are influences by tissue and muscle properties. It is measured in %.
- MDVP (Jitter abs): Absolute jitter.
- MDVP (RAP): Relative measure of the pitch disturbance.
- MDVP (PPQ): Pitch perturbation quotient.
- MDVP (Shimmer): Shimmer is a measure of amplitude instability.
- MDVP (Shimmer db): Shimmer in db.
- Shimmer (APQ 3-5): Six measures of variation in amplitude perturbation quotient (APQ).
- (NHR): Noise-to-harmonics Ratio.
- (DFA): Signal fractal scaling exponent.
- Spread 1-2: Two nonlinear measures of fundamental frequency variation.
- (RPDE): recurrence period density entropy.
- (DFA): detrended fluctuation analysis.
- (PPE): Pitch period entropy.

Methodology

As we mentioned, we have a Dataset of 31 people with 195 records, 23 of them with Parkinson disease. The status attribute shows if the person is an illness with 1 or not with 0. For each patient, we studied 10 out of 24 attributes and did some analysis on them. In addition to the Introduction and Problem description we did some of the Data analysis and Statistical analysis as shown in the following steps:

1. We plotted a histogram for each attribute (column) and analyzed the distribution for the data.
2. We calculate the cross-correlation for the 24 attributes, then we calculated the correlation for every two attributes that had a high positive correlation with the target (0, 1). Let's say, spread 1 and ppe their correlation is equal to 0.96 so it's more than 0.65; we know that they are both correlated with each other. Now, we are correlating them with the target (status). Then, we will drop whichever has a weaker correlation with the target, in this case, is ppe (0.53).
3. We reduced the attributes to 10 according to our results, and this will be explained in details in the final report.
4. We picked a sample of 150 rows for each attribute, then we calculated the SD, variance, mean, standard error and confidence interval of 95% for each one of them.
5. We choose two attributes and apply the hypothesis test on them with the target (status) and analyze the results depending on p-value and alpha.
6. We compare the range of each attribute with status, to see if there is a clear distinction between values for patients has PD and healthy ones.

As long as we are learning more probability principles during our course, we will apply it to our data to complete the final report.

Tables and Figures

The following table shows the list of subjects with sex, age, Parkinson's stage and the number of years since diagnosis.

Subject code	Sex	Age	Stage (H&Y)	Years since diagnosis
S01	M	78	3.0	0
S34	F	79	2.5	¼
S44	M	67	1.5	1
S20	M	70	3.0	1
S24	M	73	2.5	1
S26	F	53	2.0	1½
S08	F	48	2.0	2
S39	M	64	2.0	2
S33	M	68	2.0	3
S32	M	50	1.0	4
S02	M	60	2.0	4
S22	M	60	1.5	4½
S37	M	76	1.0	5
S21	F	81	1.5	5
S04	M	70	2.5	5½
S19	M	73	1.0	7
S35	F	85	4.0	7
S05	F	72	3.0	8

S18	M	61	2.5	11
S16	M	62	2.5	14
S27	M	72	2.5	15
S25	M	74	3.0	23
S06	F	63	2.5	28
S10 (healthy)	F	46	n/a	n/a
S07 (healthy)	F	48	n/a	n/a
S13 (healthy)	M	61	n/a	n/a
S43 (healthy)	M	62	n/a	n/a
S17 (healthy)	F	64	n/a	n/a
S42 (healthy)	F	66	n/a	n/a
S50 (healthy)	F	66	n/a	n/a
S49 (healthy)	M	69	n/a	n/a

Table 1: List of subjects with sex, age, Parkinson's stage and the number of years since diagnosis [8].

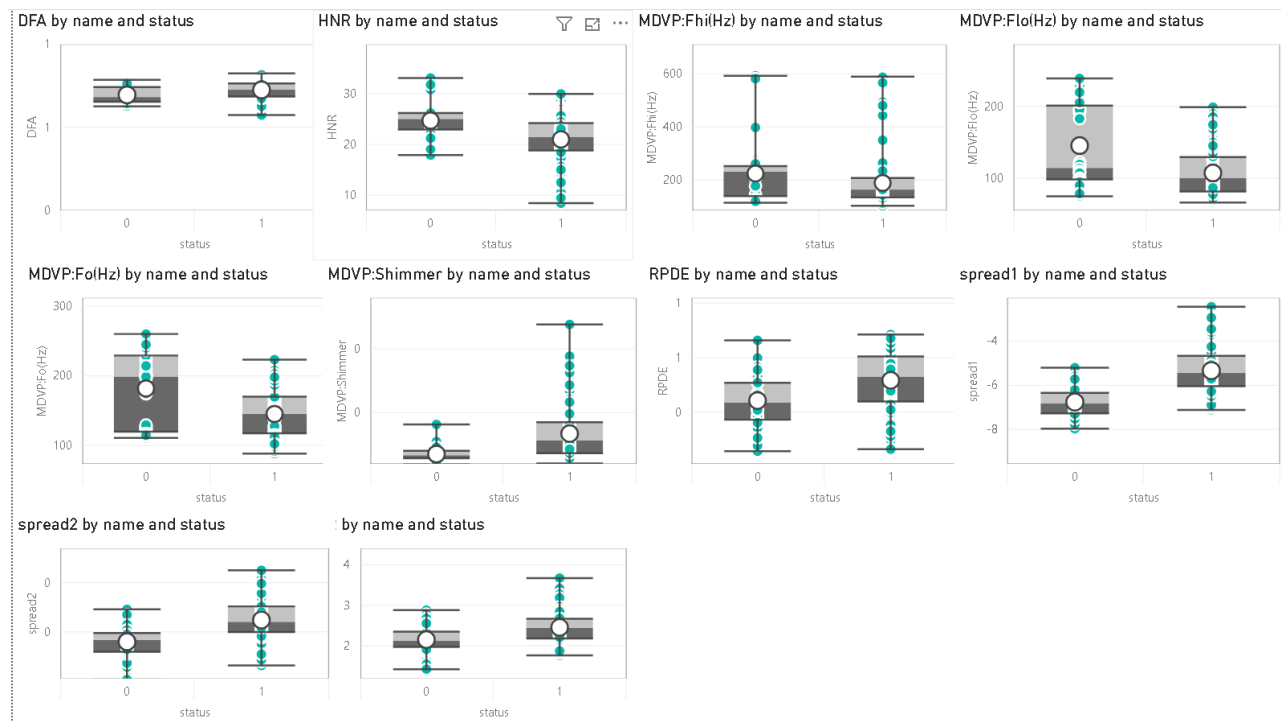


Figure 1: Comparing the range of each attribute by status to see if there is a clear distinction between values for patients with PD and healthy ones.

We will explain this figure and table in the final report.

References

- [1] Little, M.A., McSharry, P.E., Roberts, S.J. et al. Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMed Eng OnLine* 6, 23 (2007). <https://doi.org/10.1186/1475-925X-6-23>
- [2] A. E. Lang and A. M. Lozano, "Parkinson's disease - First of two parts," *New England Journal of Medicine*, vol. 339, pp. 1044-1053, 1998.
- [3] A. K. Ho, R. Iannsek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behavioural Neurology*, vol. 11, pp. 131-137, 1998.
- [4] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and Co-Occurrence of Vocal-Tract Dysfunctions in Speech of a Large Sample of Parkinson Patients," *Journal of Speech and Hearing Disorders*, vol. 43, pp. 47-57, 1978.
- [5] J. R. Duffy, *Motor speech disorders : substrates, differential diagnosis, and management*, 2nd ed. St. Louis, Mo.: Elsevier Mosby, 2005.
- [6] S. Sapir, J. L. Spielman, L. O. Ramig, B. H. Story, and C. Fox, "Effects of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals With Idiopathic Parkinson Disease: Acoustic and Perceptual Findings," *J Speech Lang Hear Res*, vol. 50, pp. 899-912, 2007.
- [7] D. A. Rahn, M. Chou, J. J. Jiang, and Y. Zhang, "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis," *Journal of Voice*, vol. 21, pp. 64-71, 2007.
- [8] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman and L. O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease," in *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015-1022, April 2009.