Journal of Engineering Science and Technology
Vol. 13, No. 9 (2018) 2792 - 2804

INVESTIGATION OF PITCH AND NOISE FEATURES EXTRACTED FROM VOICE SAMPLES OF HEALTHY AND PARKINSON AFFECTED PEOPLE USING STATISTICAL TESTS

SAVITHA S. UPADHYA*, A. N. CHEERAN

Department of Electrical Engineering, Veermata Jijabai Technological Institute, H.R. Mahajani Road, Matunga, Mumbai, Maharashtra 400019, India
*Corresponding Author: savitavjti@gmail.com

Abstract

Parkinson Disease (PD) is characterized by reduced muscular movements, which also affects the speech production muscles. This leads to the degradation of speech quality and intelligibility, which is captured by extracting voice parameters like pitch, amplitude and noise features. The statistical comparison of these parameters with those of healthy subjects is a first step, which helps in the diagnosis of the disease. In this paper, different pitch and noise features are investigated. The investigated pitch features are mean, maximum and minimum pitch. The noise features analysed are Harmonic to Noise Ratio (HNR), Noise to Harmonic Ratio (NHR) and Glottal to Noise Excitation (GNE) ratio. These features extracted from both the PD affected and healthy people are then investigated to determine whether a statistically significant difference exists between the two groups using statistical tests, which include the descriptive statistics, correlation analysis, box-plots, Shapiro-Wilk test and Mann-Whitney U tests. The results obtained from the statistical tests performed on the speech signal features extracted from the speech samples shows that a statistically significant difference between the two groups is observed in all the pitch features, NHR and HNR features but not in the GNE feature.

Keywords: Noise features, Parkinson disease, Pitch, Statistical tests.

1.Introduction

PD is caused because of damage in the brain cell which leads to a deficiency in the chemical dopamine. This lack of dopamine, in turn, leads to reduced muscular movements. Since the production of speech involves the movements of many auditory muscles, the speech of the people suffering from PD is also affected.

The speech of people with Parkinson's is characterized by reduced volume, varied volume, hoarseness, varied pitch and amplitude, etc. [1, 2]. Hence analysing the speech features might provide a useful cue for detection of PD. The diagnosis of PD is normally done using neurological tests and brain scans. Hence the trauma of going through scans and tests can be avoided if it is possible to detect the disease using speech signal features because the method to acquire the speech signal is non-invasive wherein the participants have to just speak over the microphone.

The production of speech involves three processes. The respiration process, which is related to the source of air includes the respiratory organs like lungs and bronchia. This is followed by phonation, which represents the vibration of the vocal cords to produce sound. The articulation process related to the modification of the position and shape of articulators like teeth, tongue, lips, jaw, etc. [3]. In examining phonation in PD speakers, the parameter generally used is the rate of vibration of cords called pitch or fundamental frequency, its perturbation called jitter, energy perturbation called shimmer, the amplitude of noise relative to tonal components in the speech called NHR and HNR ratios [4, 5].

Researchers have investigated various statistical tests on different features as a first step before performing further processing [1, 5-10]. Little et al. [6] used the two-sided Wilcoxon rank-sum test to determine the best classification performing a subset of features. Henriquez et al. [5] used the box plots to visualize a statistical difference in the speech features extracted between the healthy subjects and pathological subjects. Rusz et al. [1] also used the non-parametric two-sided Wilcoxon rank-sum test to compare individual speech features extracted. Bruckl et al. [7] adopted the statistical test-analyses of covariance as to compare the voice tremor features between the group of PD speakers and the healthy control group. The researchers found the tests to be significant at an alpha level of less than 0.05. Jun et al. [8] used the statistical t-test to examine the statistically significant difference for the detection of diabetes mellitus. Lopez-de-Ipina et al. [9] conducted the automatic dysphagia detection analysis, performed the statistical non-parametric Kruskal-Wallis test for automatic feature selection.

The objective of this paper is to extract the various pitch and noise features from the voice samples of healthy and persons with PD and provide a statistical comparison of these extracted parameters. Statistical tests are prerequisites to check whether the two groups differ from each other for the extracted parameters statistically. The results of these tests help us to find the optimized features, which can then be used in a classifier to discriminate between the two groups and predict PD. The different features extracted and analysed statistically in this investigation from the speech samples of healthy and PD affected people are Mean pitch frequency, Median pitch frequency, Standard deviation of the pitch frequency, Minimum pitch frequency and Maximum pitch frequency. The noise features analysed include NHR, HNR and GNE.

In this paper, Section 2 gives a brief description of the speech features extracted. Section 3 explains the various statistical tests carried out in this investigation. Section 4 gives the experimentation details and the results obtained. Finally, the paper is concluded in Section 5.

2. Speech Feature Extraction

Sounds of speech are broadly classified as voiced and unvoiced speech. The vibration of vocal cords produces voiced sounds. The rate at which, the vocal cords vibrate is referred to as “pitch”. This pitch depends on the length, tension and mass of the vocal cords. Periodicity in the time domain speech waveform or harmonicity in the speech spectrum are cues to detect the pitch period. The quasi-periodic time structure of voiced speech can be captured by using different pitch estimation algorithms like the autocorrelation method, autocorrelation with centre clipping method and the average magnitude difference function method [11]. The concept of short time windowing is used in all these methods since speech signal is wide sense stationary.

NHR and HNR are used to quantify noise in the speech signal, which is caused mainly as a result of incomplete vocal cords closure. The HNR ratio is derived from the autocorrelation computed for each window of the speech signal. The maximum lag l_{max} corresponds to the sample that provides the maximum of the autocorrelation (with the exception of zero lag). Conceptually, for a signal without noise, the autocorrelation at the instant $R(l_{max})$ should be 1 [12, 13]. Then HNR is defined as in Eq. (1)

$$HNR(dB) = 10\log_{10}\left[\frac{R_{xx}(l_{\max})}{(1-R_{xx}(l_{\max}))}\right] \tag{1}$$

The GNE ratio aims to quantify the extent of noise in the speech signal. The opening and closing of vocal cords result in synchronous excitation whereas incomplete vocal cords closure leads to asynchronous excitation and hence turbulent noise. The correlation between the Hilbert envelopes of uniformly distributed frequency channels is used to compute the GNE. As proposed by Michaelis et al. [14], Fig. 1 shows the block diagram of the computation of GNE and the dotted portion is the computation of the Hilbert envelope. Detection of glottal pulses is achieved by inverse filtering the downsampled signal. The Hilbert envelopes of different frequency bands that use a stated bandwidth are then computed. The correlations between pairs of the frequency bands are computed and their maximum value is noted. The average of the resulting vector obtained gives the value of GNE. In our work, the GNE was calculated for three different bandwidths of the envelopes (1000 Hz, 2000 Hz, and 3000 Hz). For 1000 Hz bandwidth, 51 envelopes were calculated with center frequencies from 500 Hz to 4500 Hz in steps of 80 Hz. For 2000 Hz bandwidth, 31 envelopes were calculated with center frequencies from 1000 to 4000 Hz, in steps of 100 Hz. For 3000 Hz bandwidth, 21 envelopes were calculated with center frequencies from 1500 to 3500 Hz in 100 Hz steps.

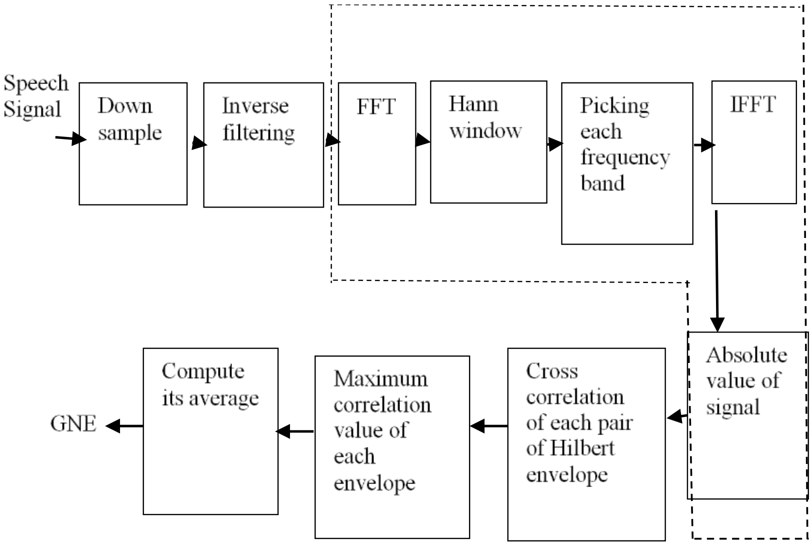


Fig. 1. Computation of GNE.

3. Statistical Tests

Once the features are extracted the next step is to explore the data using statistical tests. The different statistical tests performed in this investigation includes the descriptive statistics, which defines the metrics of central tendency, box plots, which illustrates the data distribution pictographically, correlation analysis, which signifies the dependencies of the features, Normality check test the Shapiro-Wilk test and Mann-Whitney U test to find whether a statistically significant difference exists between PD and healthy subjects.

3.1. Descriptive statistics and box plots

The first step in data analysis is the exploration of some statistical properties of the data and producing plots to get a feel for the data structure. To characterize data and probability distributions, descriptive statistics are normally used [15]. This gives the central tendency measures like mean, median, dispersion measures like variance and standard deviation. Symmetricity measures like skewness and peakedness parameters like kurtosis.

Another way to characterize a distribution or sample is by plotting the box plots. Box plot provides a pictographic representation of the following statistics: minimum, 25th percentile (lower quartile), median (50th percentile), the 75th percentile (upper quartile) and maximum. The lines extending from each end of the box plots are called whiskers, which shows the range of the data [16].

3.2. Correlation analysis

The correlation describes the degree of relationship between two variables. The strength of association between two random variables *A* and *B* can be estimated using correlation coefficients. The Pearson correlation coefficient is defined as in Eq. (2) [16].

$$R(A,B)_{Pearson} = \frac{Cov(A,B)}{\sqrt{Var(A).Var(b)}} = \frac{\sum_{i=1}^{N}(a_i - \mu_A)(b_i - \mu_B)}{\sqrt{\sum_{i=1}^{N}(a_i - \mu_A)^2 \sum_{i=1}^{N}(b_i - \mu_B)^2}} \quad (2)$$

where N is the number of realizations of the random variables A and B.

The correlation coefficient lies in the numeric range, (-1, 1), and the relationship between A and B is interpreted using (a) the sign, which represents the direction of the relationship, and (b) the magnitude. A negative sign signifies that the increase in the values of one variable leads to the decrease in the values of the other. The magnitude of the correlation coefficient signifies the strength of the statistical relationship [17]. The advantage of using the Pearson correlation is its simplicity and it evaluates the linear relationship between two continuous variables.

3.3. Shapiro Wilk test

Many of the statistical procedures like t-tests, analysis of variance, regression and other parametric tests assume that the data follow a normal distribution. According to the central limit theorem, (a) if the data are approximately normal then their sampling distribution will be normal; (b) the sampling distribution tends to be normal in large data [16, 18]. Normality can be found out by using normal plots or by performing significance tests that is, comparing the sample distribution to a normal one. The plots commonly used to visualize normality are the frequency distribution plots like histogram, boxplot and Quantile-Quantile Plot (Q-Q Plot) [16, 17]. The tests commonly used to assess the normality are Kolmogorov-Smirnov test, Chi-square test, Shapiro-Wilk test, Anderson-Darling test, Anscombe-Glynn Kurtosis test and the Jarque-Bera test [16, 18]. These tests compare the scores in the sample to a normally distributed set of scores with the same mean and standard deviation. The null hypothesis is that “sample distribution is normal”. In this investigation, the Shapiro-Wilk test, which is based on the correlation between the data and the corresponding normal scores is used. This Shapiro Wilk test gives a ‘p’ value and if this ‘p’ value is greater than the significant alpha level chosen (generally taken to be 0.05), then the null hypothesis that ‘the observed data follows a normal distribution’ is not rejected.

3.4. Mann Whitney U test

To determine whether any two groups of data are statistically different, various parametric and nonparametric statistical tests are used. Parametric tests like t-tests and the analysis of variance assume that the data is normally distributed. Non-parametric tests like Chi-square tests, the Fisher Exact Probability test, the Mann-Whitney test, the Wilcoxon Signed-Rank test, the Kruskal-Wallis test and the Friedman test do not follow such assumption. Hence, non-parametric tests are also called as “distribution-free" tests [19]. In this investigation, the Mann Whitney U test is used, which determines whether the median of a variable in one group is significantly different from the median of that variable in another group. This test does not require the distribution to have any particular shape. This test gives a “U-crit” (U-Critical) value and if this “U-crit” value is greater than U the hypothesis is defined in such a way, that the test is significant at the 0.05 level. This implies that there exists a significant statistical difference between the two data groups [20].

4. Implementation and Results

The data samples used in this study has speech samples of the vowel/a/of length 3 to 4 seconds collected from 40 healthy and 40 PD affected contributors. The data set had both male and female participants in both the groups with their ages in the range of 50 to 78 years. The speech signal was recorded using a transcend voice recorder at a sampling frequency of 8 kHz, with a resolution of 16 bits. Praat voice analysis software [21] is used to obtain the different pitch features, HNR (dB) and NHR ratios. For computing the pitch features, the window length was set to 240 samples and Hamming window was used. The speech signal is divided into frames by windowing and the pitch frequency is computed for each window. These pitch frequencies obtained with every frame is then averaged and is taken or called as the mean pitch frequency feature of the speech signal sample analysed. The median of the pitch frequencies obtained from every analysed speech signal sample is the median pitch frequency feature. Similarly, the features standard deviation pitch, minimum and maximum pitch are computed for each and every analysed speech sample.

The GNE is implemented using MATLAB code. “Real Statistics using Excel” [12] is then used to perform the statistical tests on the features extracted.

The parameters used in the implementation of GNE are

- 30 window length ms (240 samples)
- Window overlap: 50% (120 samples)

The GNE was calculated for three different bandwidths of the envelopes (1000 Hz, 2000 Hz, and 3000 Hz). For 1000 Hz bandwidth, 51 envelopes were calculated with center frequencies from 500 Hz to 4500 Hz in steps of 80 Hz. For 2000 Hz bandwidth 31 envelopes were calculated with center frequencies from 1000 to 4000 Hz in steps of 100 Hz. For 3000 Hz bandwidth, 21 envelopes were calculated with center frequencies from 1500 to 3500 Hz in 100 Hz steps. The average of all the three GNE was taken in the analysis.

The statistical tests performed on these extracted features are descriptive statistics, box plots, correlation analysis-Pearson correlation coefficient, normality test-Shapiro Wilk test and non-parametric test-Mann Whitney U test.

4.1. Results of pitch analysis

Tables 1 and 2 give the descriptive statistics of the pitch features extracted namely the mean pitch, median pitch, standard deviation pitch, minimum pitch and maximum pitch for healthy and PD subjects.

Statistical significant difference between the two groups is observed in the metrics of central tendency for all the features of the pitch. From the descriptive statistics, it is observed that the mean of the mean pitch frequency F_0 is decreased and the mean of the standard deviation of the pitch is increased in persons suffering from PD. Since the mean pitch frequency is gender biased, a gender-dependent descriptive statistics test for the pitch features is performed. Here the dataset is divided into four groups namely healthy-male, PD-male, healthy-female and PD-female. Table 3 shows the results of the descriptive statistics obtained for the gender-dependent test.

From Table 3, it is again observed that the mean of the mean pitch frequency is reduced in PD subjects and also the mean of the standard deviation of the pitch is increased in PD subjects irrespective of the gender.

Figure 2 shows the box plot of the pitch features extracted for normal (healthy) and PD groups.

Figure 2 has two subplots. The upper plot shows the distribution of various features of pitch amongst the subjects. The box plots are shown in pairs. The left member of the pair relates to data from normal subjects labelled as ‘Nor’ in the plot and right member of the pair is of subjects suffering from PD labelled as ‘PD’. The lower plot of Fig. 2 is the SD of the pitch of normal and PD subjects. From Fig. 2 it is observed, that the variation of F_0, i.e., F_0 (SD) is large in people with PD when compared with healthy subjects. Also, non-overlapping notches in the box plots of the two groups for the pitch features extracted signifies that there exists a statistically significant difference between the two groups.

Table 1. Descriptive statistics of pitch features for healthy subjects.

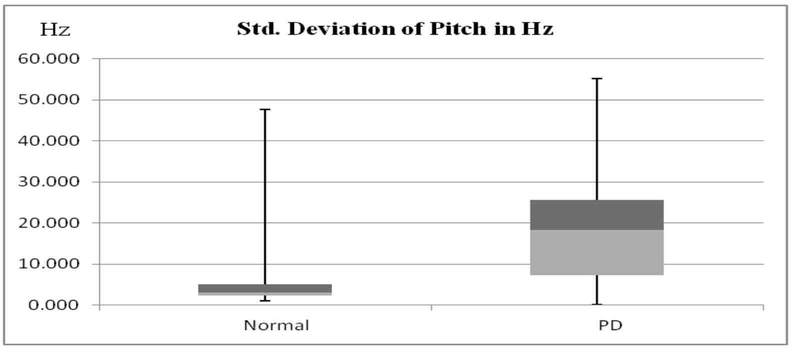
	Healthy subjects			
	Mean	**Median**	**Kurtosis**	**Skewness**
Median pitch (Hz)	190.9	191.68	4.83	1.27
Mean pitch (Hz)	**189.1**	185.75	5.32	1.42
SD pitch (Hz)	**7.46**	3.03	7.32	2.72
Min pitch (Hz)	156.4	161.62	1.69	0.77
Max pitch(Hz)	197.7	210.47	4.18	0.18

Table 2. Descriptive statistics of pitch features for PD subjects.

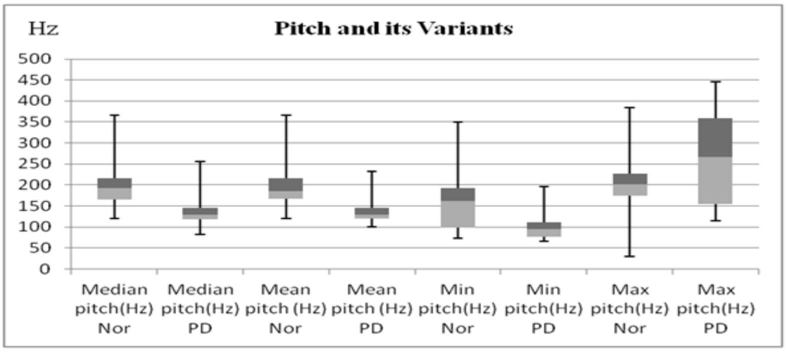
	Parkinson subjects			
	Mean	**Median**	**Kurtosis**	**Skewness**
Median pitch (Hz)	139.9	129.55	1.23	1.34
Mean pitch (Hz)	**140.9**	129.16	1.07	1.47
SD pitch (Hz)	**18.41**	18.26	0.07	0.70
Min pitch (Hz)	98.31	94.19	3.06	1.69
Max pitch (Hz)	262.4	267.48	-1.61	0.16

Table 3. Gender dependent descriptive statistics.

	Healthy male	**PD male**	**Healthy female**	**PD female**
Median pitch (Hz)	158.2	121.1	215.6	204.7
Mean pitch (Hz)	**158.3**	**122.4**	**213.1**	**187.9**
SD pitch (Hz)	**3.9**	**8.9**	**4.6**	**10.3**
Min pitch (Hz)	137.2	94.8	160.9	104.3
Max pitch(Hz)	181.7	202.1	216.5	277.6



a) Metrics of central tendency of pitch variants for normal and PD subjects



(b) Standard deviation of pitch for normal and PD subjects.

Fig. 2. Data distribution of pitch and its variants for normal (healthy) and PD subjects.

Many features can be highly correlated with other features because they measure very similar aspects of the signal. Therefore, calculation of the Pearson product-moment correlation coefficient was used to test for significant correlations among the pitch features. Using this test, the dimensionality of the feature set can be reduced. If the correlation value between two features is high, then only one feature can be used and the other can be discarded in future analysis. Table 4 presents the Pearson correlation coefficient between all pitch features.

Subsequently, from all highly correlated measures with a correlation coefficient of greater than 0.95 only one feature will be kept, which correlates with the greatest number of similar measurements and gains the most statistically significant differences between the PD and healthy groups. Table 4 bold entries indicate a high correlation between features. It is seen that the parameter mean pitch feature is highly correlated to the median pitch feature. Hence the median pitch feature can be discarded when features are passed through a classifier.

Tables 5 and 6 show the result of the Shapiro-Wilk Test carried out on healthy subjects and Parkinson subjects' data respectively to check, which of the pitch features follow a normal distribution.

It is observed from the Shapiro-Wilk test that none of the pitch features extracted followed a normal distribution. Hence the "distribution-free" Non-parametric test the Mann Whitney U test is carried out to test whether it is possible to obtain a statistical difference between the PD and healthy group using the pitch features, which are not normally distributed. The Mann Whitney U test showed the U-critical value to be 19039.1 and U value to be 16121.5. As U-critical (U-crit) is greater than U, the test is significant at the 0.05 level. This shows that there exists a significant statistical difference between the two groups for the pitch features extracted.

Table 4. Pearson correlation coefficients between all the pitch features.

	Median pitch (Hz)	Mean pitch (Hz)	SD pitch (Hz)	Min pitch (Hz)
Mean pitch (Hz)	**0.99**			
SD pitch (Hz)	-0.07	-0.11		
Min pitch(Hz)	0.68	0.73	-0.49	
Max pitch(Hz)	0.07	0.09	0.56	-0.02

Table 5. Result of Shapiro-Wilk test of pitch and its variants on healthy subjects.

	Median pitch (Hz)	Mean pitch (Hz)	Standard deviation (Hz)	Min pitch (Hz)	Max pitch (Hz)
W	0.89	0.88	0.56	0.93	0.92
p-value	0.000628	0.000571	7.29E-10	0.014233	0.005142
Normal	no	no	no	no	no

Table 6. Result of Shapiro-Wilk test of pitch and its variants on PD subjects.

	Median pitch (Hz)	Mean pitch (Hz)	Standard deviation (Hz)	Min pitch (Hz)	Max pitch (Hz)
W	0.84	0.79	0.94	0.81	0.88
p-value	4.45E-05	3.03E-06	0.03864	1E-05	0.00039
Normal	no	no	no	no	no

4.2. Results of noise feature analysis

Tables 7 and 8 give the descriptive statistics of the extracted noise features: NHR, HNR (dB) and GNE for healthy and PD subjects respectively.

Tables 7 and 8 show that the values of the metrics of central tendencies are different for the two groups for HNR and NHR features signifying a statistically significant difference between the two groups for these features but a marginal difference is observed in GNE between the PD and healthy groups. Also, from Fig. 3 (box plots) it is seen that the data distribution boxes do not overlap for NHR and HNR (dB) features but overlaps for GNE feature, which is interpreted as the two groups namely the PD and healthy groups cannot be differentiated by means of GNE feature. Table 8 shows the results of the correlation analysis conducted on the noise features.

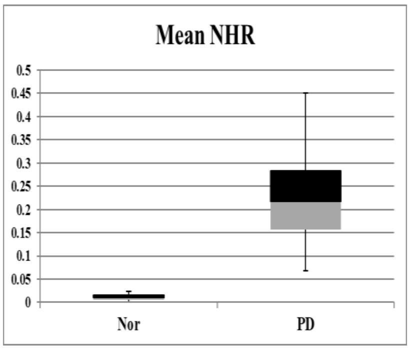
From Table 9, the analysis also signifies that the noise features are not correlated with one another. Hence, no feature can be discarded in the future analysis.

Table 7. Descriptive statistics of noise features for healthy subjects.

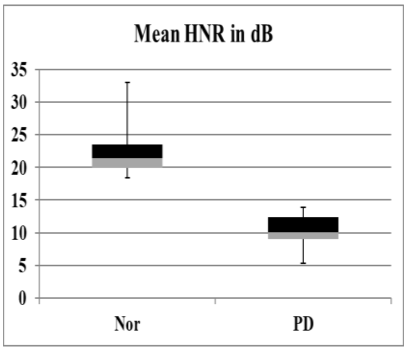
	Parkinson subjects					
	Mean	**Median**	**SD**	**Variance**	**Kurtosis**	**Skewness**
NHR	0.011	0.011	0.006	3.8E-05	-0.81	0.15
HNR in dB	22.48	21.48	3.57	12.75	2.15	1.54
GNE	0.85	0.85	0.02	0.0008	0.93	0.53

Table 8. Descriptive statistics of noise features for PD subjects.

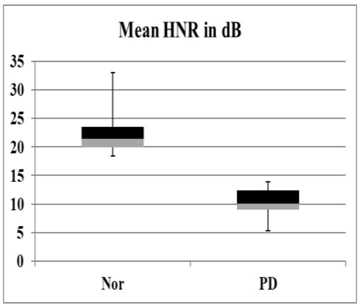
	Parkinson subjects					
	Mean	**Median**	**SD**	**Variance**	**Kurtosis**	**Skewness**
NHR	0.22	0.22	0.11	0.011	-0.46	0.45
HNR in dB	10.22	10.09	2.42	5.89	-0.69	-0.35
GNE	0.86	0.86	0.03	0.0009	0.43	0.09



(a) Mean of NHR parameter.



(b) Mean of HNR parameter.



(c) Mean of GNE parameter.

Fig. 3. Data distribution of noise and its parameters for normal (healthy) and PD subjects.

Table 9. Pearson correlation coefficients between all the noise features.

	NHR	**HNR (dB)**
HNR (dB)	-0.87	
GNE	0.28	-0.11

The Shapiro Wilk test performed showed that the noise parameters follow a normal distribution in both healthy and Parkinson dataset except the Mean HNR (dB) parameter did not follow a normal distribution in healthy subjects. Similarly, the results of the Mann Whitney U test were not significant at the 0.05 level for GNE but was significant for NHR and HNR. This shows that a significant statistical difference between the two groups does not exist for the GNE parameter but exists for NHR and HNR features.

The pitch and the noise features extracted are then passed through a support vector machine classifier with the kernel function as a radial basis function. This model is tested using a 5-fold cross-validation method. The classification accuracy obtained using all the pitch and noise features extracted is 81.3% and using all the pitch and noise features without the GNE is 88.8%.

Hence overall from the different statistical tests conducted, the two groups namely the PD affected and healthy persons differed statistically for all the pitch features extracted as well as for NHR and HNR but failed to differ statistically for GNE. As GNE is a measure related to the vocal tract and HNR and NHR are related to phonation, the result obtained probably signifies that the disease affects the phonation parameters than the articulation parameters in early stages.

5. Conclusions

A statistical comparison is carried out to check whether it is possible to differentiate PD from healthy people using speech samples. The tests were carried out on various pitch and noise features extracted from the voice samples of the two groups of people. The descriptive statistics test shows that the pitch frequency F_0 is decreased and the deviation of the pitch from the mean pitch is increased in persons suffering from PD and in noise features mean of the NHR was higher in PD subjects whereas HNR was lower. But the significant difference was not observed for the GNE parameter. A similar interpretation is also visualized in the box plots. The correlation analysis conducted among the features showed that the mean pitch was highly correlated with the median pitch. The normality test: Shapiro-Wilk Test shows that none of the pitch features followed a normal distribution both for PD and healthy groups whereas among the noise features only the HNR (dB) did not follow a normal distribution. Hence the Mann Whitney U test is performed to determine whether there is a significant difference between the two groups of data. This test shows a significant statistical difference between the groups at a 5% significance level for all the pitch features, NHR and HNR but not for GNE. These features were then passed through a support vector machine classifier and a classification accuracy of 81.3% and 88.8% with and without the GNE feature respectively is obtained. Further from these variants of pitch and noise, optimum features can be selected by using dimensionality reduction techniques. These optimum features can then be used in classifiers like support vector machine classifier, artificial neural networks or deep neural networks to categorize the two groups.

Acknowledgement

We would sincerely like to thank the Parkinson Disease Movement Disorder Society (PDMDS) of India for allowing us to collect speech samples of participants from various PDMDS centers in Mumbai.

Nomenclatures

Cov (A, B)	Covariance of A and B
lmax	Maximum lag
log_{10}	Logarithmic to base 10
$R(A,B)_{Pearson}$	Pearson autocorrelation coefficient betweenvariables A and B
$R_{xx}(lmax)$	Autocorrelation of x at maximum lag
Var (A)	Variance of A
Var (B)	Variance of B

Greek Symbols

μ	Mean value

Abbreviations

Abs	Absolute
dB	Deci-Bels
FFT	Fast Fourier Transform
GNE	Glottal to Noise Excitation
HNR	Harmonic to Noise Ratio
IFFT	Inverse Fast Fourier Transform
NHR	Noise to Harmonic Ratio
PD	Parkinson Disease
PDMDS	Parkinson Disease Movement Disorder Society
Q-Q Plot	Quantile-Quantile Plot
SD	Standard Deviation
U-crit	U-Critical

References

1. Rusz, J.; Cmejla, R.; Ruzickova, H.; and Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson’s disease. *Journal Acoustical Society of America*. 129(1), 350-367.
2. Rusz, J.; Cmejla, R.; Ruzickova, H; Jiri, K.; Veronika, M.; Jana, P.; Jan, R.; and Ruzicka, E. (2011). Acoustic markers of speech degradation in early untreated parkinson’s disease. *Proceedings of the Forum Acusticum.* Aalborg, Denmark, 2725-2730.
3. Rusz, J. (2012). *Acoustic analysis of voice and speech disorders*. Ph.D. Thesis. Faculty of Electrical Engineering, Czech Technical University, Prague.
4. Tsanas, A.; Little, M.A.; McSharry, P.E.; Spielman, J.; and Ramig, L.O. (2012). Novel speech signal processing algorithms for high accuracy classification of parkinson’s disease. *IEEE Transactions on Biomedical Engineering*, 59(5), 1264-1271.
5. Henriquez, P; Alonso, J.B.; Ferrer, M.A.; Travieso, C.M.; Godino-Llorente, J.I.; and Diaz-de-Maria, F. (2009).Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Transactions on Audio, Speech and Language Processing*, 17(6), 1186-1195.
6. Little, M.A.; McSharry, P.E.; Hunter, E.J.; Spielman, J.; and Ramig, L.O. (2009). Suitability of dysphonia measurements for telemonitoring of parkinson's disease. *IEEE Transactions Biomed Engineering*, 56(4), 1015-1022.

7. Bruckl, M.; Ghio, A.; and Viallet, F. (2018). Measuremnt of tremor in the voices of speakers with parkinson's disease. *Procedia Computer Science*, 128, 47-54.
8. Jun, M.-H.; Kim, S.; Ku, B.; Cho, J.; Kim, K.; Yoo, H.-R; and Kim, J.U. (2018). Glucose-independent segmental phase angles from multi-frequency bioimpedance analysis to discriminate diabetes mellitus. *Scientific Reports*, 8, Article No. 648, 11 pages.
9. Lopez-de-Ipina, K.; Calvo, P.; Faundez-Zanuy, M.; Clave, P.; Nascimento, W.; de Lizarduy, U.M.; Alvarez, D.; Arreola, V.; Ortega, O.; Mekyska, J.; and Sanz-Cartagena, P. (2018). Automatic voice analysis for dysphagia detection. *Speech, Language and Hearing*, 21(2), 86-89.
10. Gillivan-Murphy, P.; Miller, N.; and Carding, P. (2018). Voice tremor in parkinson's disease: An acoustic study. *Journal of Voice: Official Journal of the Voice Foundation*. PMID: 29395332.
11. Upadhya, S.S.; and Cheeran, A.N. (2016). Extraction of phonation measures affected by parkinson disease. *International Journal of Electrical and Electronic Engineers*, 8(2), 70-78.
12. O'Shaughnessy, D. (2001). *Speech Communications: Human and machine* (2nd ed.). New Jersey: Wiley-IEEE Press.
13. Tsanas, A. (2012). *Accurate telemonitoring of parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning*. Ph.D. Thesis. University of Oxford, St. Cross College.
14. Michaelis, D.; Gramss, T.; and Strube, H.W. (1997). Glottal-to-noise excitation ratio-a new measure for describing pathological voices. *Journal Acustica Acta Acustica*, 83, 700-706.
15. Jaggi, S. (2014). Descriptive statistics and exploratory data analysis. Indian Agricultural Statistics Research Institute. *Descriptive Statistics and Data Exploration*, 18 pages.
16. Gupta, S.P. (2007). *Statistical methods*. New Delhi, India: Sultan Chand & Sons Pvt. Ltd.
17. Zaiontz, C. (2013-2015). Real statistics resource pack software. Retrieved May 1, 2016, from www.real-statistics.com.
18. Ghasemi, A.; and Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology & Metabolism*, 10(2), 486-489.
19. Sullivan, L. (2017). Non-parametric tests. Retrieved August 15, 2017 from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/.
20. DeCoster, J. (2006). Testing group differences using t-tests, anova, and nonparametric measures. Retrieved August 21, 2017 from http://www.stat-help.com/ANOVA%202006-01-11.pdf**.**
21. Boersma, P.; and Weenink, D. (2013). Praat: Doing phonetics by computer. Retrieved June 2, 2015 from http://www.praat.org/.