International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

EXTRACTION OF PHONATION MEASURES AFFECTED BY PARKINSON DISEASE

Savitha S. Upadhya[1] , Dr. A. N. Cheeran[2]

[1] *Ph D Research Scholar. V.J. T. I., Matunga,Mumbai,(India)*

[2] *Prof in Electrical Engineering Department, V. J. T. I., Matunga, Mumbai, (India)*

ABSTRACT

Parkinson's disease (PD) is a chronic neuro-degenerative disorder caused by the damage or loss of dopamine producing cells in the substantia nigra of the brain. This deficiency in the dopamine affects the neurotransmitter systems which are responsible for a variety of motor and non-motor deficits. The motor and non-motor deficits involve areas such as speech, EEG, gait, mood, sensation, thinking and behavior. The various speech subsystems affected by PD includes phonation, articulation, respiration and prosody. These PD affected speech subsystems are typically assessed using several acoustic measurements. The measures of phonation influenced by PD include fundamental frequency, jitter, shimmer, NHR and HNR ratios and voice onset time. This paper provides a systematic review of different algorithms used for the extraction of these phonation measures affected by Parkinson disease which may be helpful in diagnosing and or in classifying the severity of the disease.

Keywords—*Jitter ,phonation measures,pitch,shimmer, voice onset time.*

1. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative disorders which affect persons above the age of 65 years [1]. This disease is caused because of the damage of the dopamine producing cells in the substantia nigra of the brain. This dopamine deficiency affects the brain neurotransmitter which is responsible for regulating and coordinating the movement of the muscles which also includes those responsible for speech production too. Hence the intelligibility and quality of speech is reduced. The different PD related vocal impairments include dysarthia, dysphonia, hypophonia and monotone [2].

The different speech subsystems which are affected by Parkinson disease include respiration, phonation, articulation, and prosody [3]. The dysarthria features of PD are related to impairment in phonation, with the articulation being the second most affected speech subsystem. Patients suffering from PD can also have abnormalities related to all the factors of speech including monoloudness, monopitch, inaccurate articulation, variable speech rate, hoarseness, reduced stress, speech disfluencies, inappropriate silence, and others [4, 5]. Phonation refers to the vibration of the vocal cords to produce sound, articulation is the variation of the position and shape of articulators (tongue, lips, jaw) to produce sound, prosody is the variation in loudness, pitch, and timing associated with natural speech, and respiration is the action of the respiratory apparatus during exhalation, providing a continuous flow of air with sufficient volume and pressure to begin phonation [6]. The PD affected phonation measures include measurement of fundamental frequency or pitch, jitter - a perturbation measurement of fundamental frequency, shimmer - a perturbation measurement of amplitude and Noise to

Harmonic Ratio NHR - the ratio of amplitude of noise to tonal components in the speech. The other phonatory measure that can be affected by PD is voice onset time (VOT), defined as the duration of time from the release of a stop consonant to the onset of voicing for the following vowel. This paper provides an overview of different algorithms used for the extraction of phonation measures affected by Parkinson disease and an interpretation on the effect of Parkinson disease on these measures of phonation.

2. PHONATION MEASURES

Phonation represents the vibration of the vocal cords to produce sound. Vocal folds vibration during phonation creates pitch of the voice. The different features of phonation influenced by PD include Pitch frequency F_0, Jitter and its variants, Shimmer and its variants, NHR and HNR ratios and Voice onset time. For the measurements of these measures a speech task called sustained phonation is employed wherein the patient is asked to speak a vowel say /a/, /i/ or /u/ at a comfortable pitch and constant or stable loudness and as long as possible, at least 4 s [4].

2.1 Fundamental Frequency-Pitch (Fo)

Speech sounds can be classified into two general categories, voiced and unvoiced sounds. When air is inhaled onto the lungs, no sound is produced. When air is exhaled from the lungs the tension in the vocal cords are so adjusted that they tend to vibrate, resulting in quasi periodic pulses of air which are then frequency modulated by the articulators to produce sounds called voiced sounds like the vowels /a/, /i/, /u/, etc. This rate of vibration of the vocal cords is called fundamental frequency pitch F_0. The pitch is a function of mass of the vocal folds, tension in the vocal cords and air pressure from the lungs. Various algorithms used for extracting pitch can be categorised as time-domain based tracking which looks for periodicity in time, frequency domain based tracking which looks for harmonics of fundamental frequency or both time-frequency domain based tracking jointly. Time domain pitch estimators identifies the quasi periodic time pattern of the speech signal or wavering of high and low amplitudes, or discontinuity points. Here the speech signal is examined over a short time window and short time analysis is performed. The time domain based pitch detection algorithms include Autocorrelation method, Autocorrelation method with center clipping and Average magnitude difference function (AMDF) method. In frequency domain, the pitch is determined by operating on a short-time frame of speech samples, and spectrally transforming them by performing Fourier transformation to obtain the periodicity information in the signal. Peaks in the spectrum at the fundamental and its harmonics represent periodicity. In cepstral analysis of speech the vocal tract parameters are separated from the source parameters [7].

2.1.1 Time Domain Pitch Detection Algorithms

Time domain based pitch detection algorithms include Autocorrelation method, Autocorrelation method with center clipping and Average magnitude difference function (AMDF) method. In these techniques the speech signal is examined frame by frame by performing short time analysis. All these techniques exploit the quasi periodic behavior of the speech signal in time domain.

2.1.1.1 Autocorrelation Method

This method is based on detecting the highest value of the autocorrelation function in the region of interest. For given discrete signal $x(n)$, the autocorrelation function is generally defined as in (1)

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

$$R(k) = \frac{1}{M} \sum_{n=0}^{M-1-k} x(n)x(n+k) \qquad 0 \le k \le M_o \tag{1}$$

where M is number of samples in a frame and M_o is the number of autocorrelation values to be computed. For the detection of pitch, if $x(n)$ is assumed to be a periodic sequence i.e. $x(n) = x(n+P)$ for all n, then its autocorrelation function will also be periodic with the same period i.e. $R(k) = R(k+P)$. Conversely, the periodicities in the autocorrelation function signify periodicity in the signal. Fig. 1 shows the time domain representation of a sustained vowel speech segment as a function of time and its corresponding autocorrelation function.

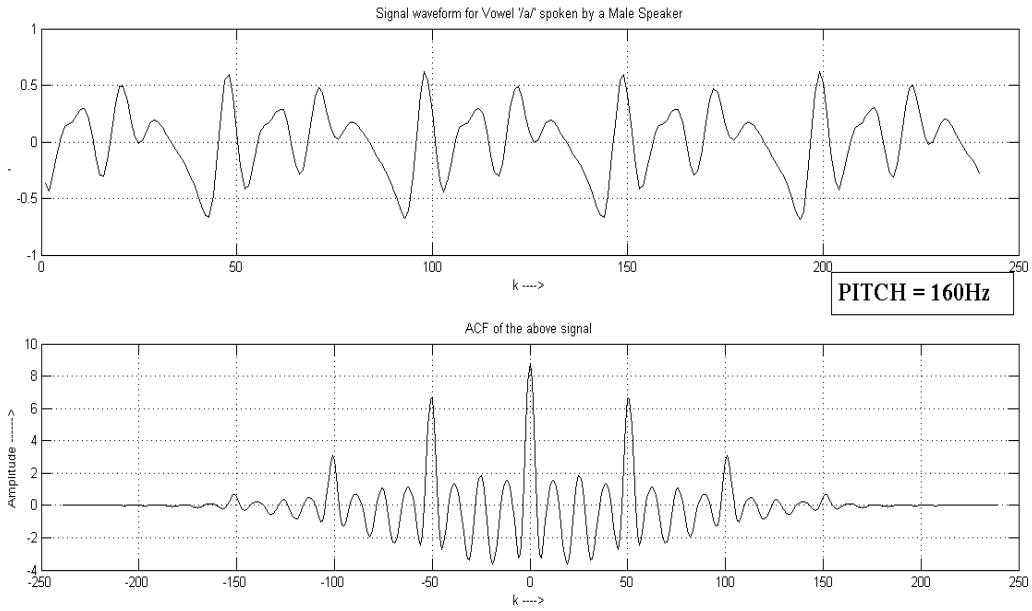

Fig.1 Autocorrelation of Sustained Vowel Segment

2.1.1.2 Autocorrelation Method with Center Clipping

The Autocorrelation method has a drawback that it has many peaks, as this method keeps much of the information in the speech signals. Most of these peaks are mainly because of the damped oscillation behavior of the vocal tract response which is responsible for the alteration in the periodic shape of the speech wave. Therefore in cases when the autocorrelation peaks due to the vocal tract response are higher than those due to the periodicity of the glottis excitation, the procedure of picking the largest peak in the autocorrelation function will fail [8].

To avoid this drawback the speech signal is pre processed to make the periodicity more evident while suppressing other unwanted features of the signal. Here center-clipping technique is used in the pre-processing stage. This removes the effects of the vocal tract transfer function and fewer peaks will appear in the autocorrelation function as compared to many peaks in the common autocorrelation method [8]. This will help us to estimate the pitch more accurately. In this technique, the relation between the input signal x(n), and the center-clipped signal y(n) is given as in (2)

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

$$y(n) = clc[x(n)] = \begin{cases} (x(n) - C_L), x(n) \geq \\ 0 \quad\quad\quad\quad , |x(n)| < \\ (x(n) + C_L), x(n) \leq \end{cases} \quad (2)$$

where C_L is the clipping threshold.

Normally, C_L is taken to be 25- 30% of the maximum absolute value of the signal in the signal frame [8]. Non-linear operations on the speech signal such as center-clipping tend to flatten the signal spectrum. This results in the increase of the distinctiveness of the true period peaks in the autocorrelation function [6].

Fig.2. shows a sustained vowel segment; its center clipped version, and shows the comparison between the autocorrelation function calculated from original signal frame and center-clipped signal frame. It is observed that the autocorrelation function obtained after center-clipping the signal contains fewer peaks as compared to the autocorrelation function without center clipping.

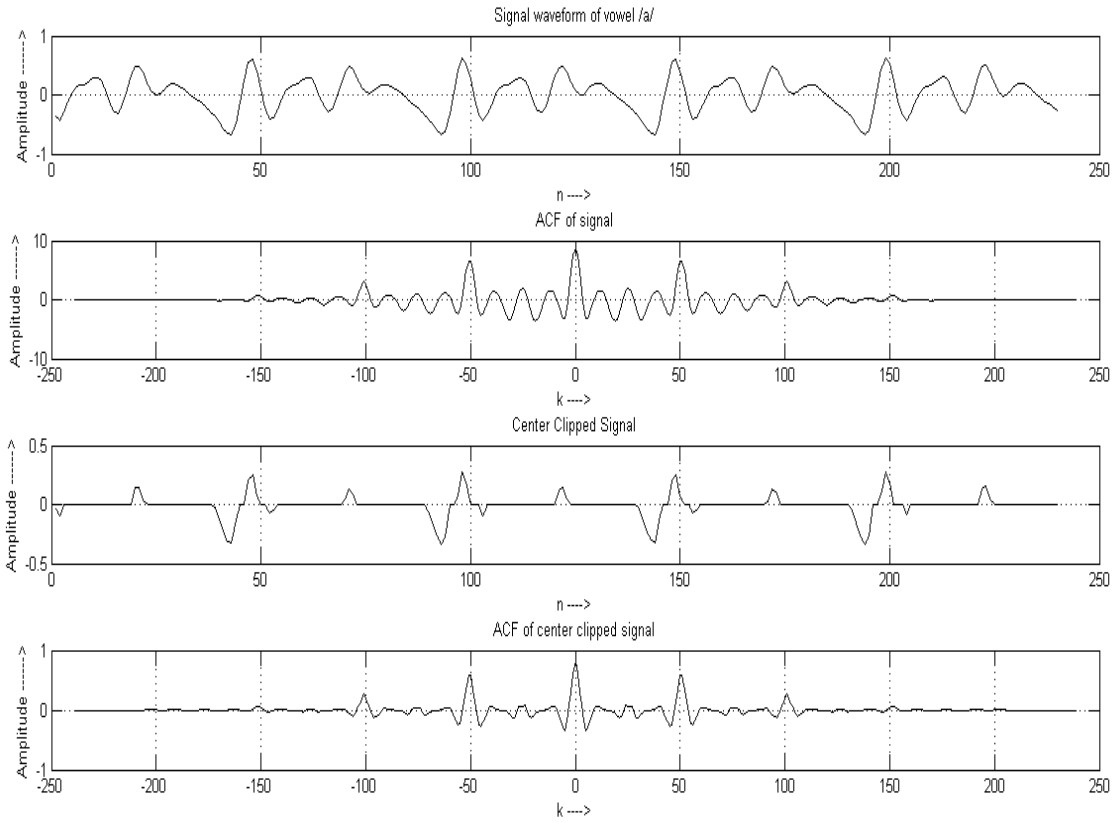

Fig.2 Sustained Vowel Segment - Autocorrelation Function and its Center-Clipped Version

2.1.1.3 Average Magnitude Difference Function Method (AMDF)

The average magnitude difference function (AMDF) is another type of time domain analysis similar to autocorrelation. Here a difference signal is obtained between the original and delayed speech, instead of correlating the input speech at different time delays. Then the absolute magnitude of the difference signal is taken at each delay value. For the frame of *M* samples, the short-time difference function AMDF is defined as in (3).

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

$$D_x(k) = \frac{1}{M} \sum_{n=0}^{M-1-k} | x(n) - x(n+k) | \qquad 0 \le k \le M_o \tag{3}$$

where *x(n)* are speech samples of analysis frame, *x(n+k)* are the samples time shifted on *k* samples and M is the length of the frame. The difference function is expected to have a strong local minimum if the lag *k* is equal to or very close to the fundamental period.

The advantage of average magnitude difference function method of pitch detection is simple implementation and low computational cost. This is because the no multiplications are required for AMDF calculations. Hence it is generally used in real-time applications. For each value of delay, the average absolute difference is taken over a window of *M* samples. The lags are taken from 16 to 160 samples. The value of the lag where minimum AMDF occurs is then identified as the pitch period [8].

In this extractor the accuracy is limited due to the effects of the spectral envelope of the vocal tract which cannot be separated from the glottis response completely. Fig 3 shows the AMDF function of a voiced speech segment.

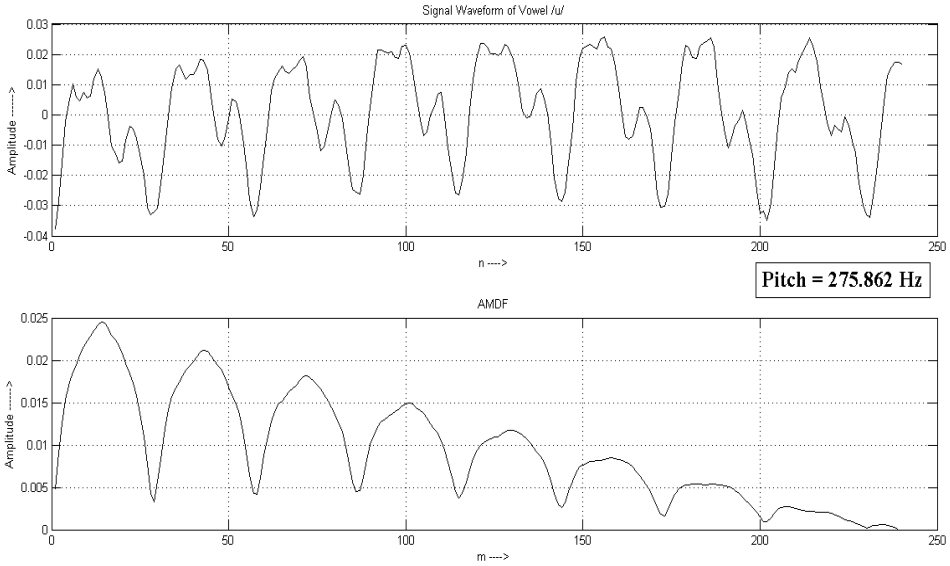

Fig 3. AMDF function of a voiced speech segment

2.2 Jitter Measurement

Jitter is a measure of variation of range of voice. It represents the variation of fundamental frequency from one cycle to the next. These measures are usually extracted by using the autocorrelation method for determining the fundamental frequency and identifying each cycle of vibration of the vocal cords. The other measurements related to jitter which also helps in measuring the perturbation are jitter local or Jitter absolute, jitter RAP (relative average perturbation)., jitter PPQ (period perturbation quotient), jitter relative etc. [5]

Jitter (absolute) is the variation of fundamental frequency (pitch) from one cycle to the other. It is the average absolute difference between successive periods, expressed as (4)

$$Jitter(abs) = \frac{1}{M-1} \sum_{i=1}^{M-1} |T_i - T_{i+1}| \tag{4}$$

where Ti are the computed pitch period lengths and M is the number of pitch periods extracted. [9].

Jitter relative is the average absolute difference between successive periods, divided by the average period. It is expressed in percentage as in (5):

$$Jitter(rel) = \frac{\frac{1}{M-1}\sum_{i=1}^{M-1}|T_i - T_{i+1}|}{\frac{1}{M}\sum_{i=1}^{M}T_i} \quad (5)$$

Jitter RAP is defined as the average absolute difference between a pitch period and the average of it and its two neighbors divided by the average period.

Jitter (PPQ5) is the five-point Period Perturbation Quotient, defined as the average absolute difference between a pitch period and the average of it and its four closest neighbors, divided by the average period [9].

All these variants of jitter can be extracted by determining the fundamental frequency and identifying every cycle using autocorrelation method.

2.3 Shimmer Measurement

The shimmer and its variants are measures of perturbation of amplitude. They are derived by measuring the maximum value of the amplitude of the signal within each vocal cycle.

Shimmer (dB) or shimmer local is expressed as the peak to peak amplitude variations in decibels, i.e. the average absolute base-10 logarithm of the difference between the amplitudes of successive periods, multiplied by 20 as in (6):

$$Shimmer(dB) = \frac{1}{M-1}\sum_{i=1}^{M-1}\left|20\log\left(\frac{V_{i+1}}{V_i}\right)\right| \quad (6)$$

where Vi are the peak-to-peak amplitude extracted and M is the number of fundamental frequency periods extracted[9].

Shimmer (relative) or Shimmer DDA is defined as the average absolute difference between the amplitudes of successive periods, divided by the average amplitude, expressed as a percentage as in (7):

$$Shimmer(rel) = \frac{\frac{1}{M-1}\sum_{i=1}^{M-1}|V_i - V_{i+1}|}{\frac{1}{M}\sum_{i=1}^{M}V_i} \quad (7)$$

Shimmer (APQ3) is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.

Shimmer (APQ5) is defined as the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude [5 9].

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

Shimmer (APQ11) is expressed as the 11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours, divided by the average amplitude [5 9].

To extract shimmer and its type's energy contours by computing the short time energy of the speech frame or computing the short time average magnitude of the speech frame can be employed. The short time energy of the speech signal provides a meaningful representation which reflects the variations in the amplitude. The short time energy is defined in (8) as

$$E = \sum_{k=0}^{n-1} x^2(k) \tag{8}$$

n = no of samples in each frame

$x(k)$ = Amplitude of each sample

One difficulty with short time energy is that it is very sensitive to large signal amplitudes. Since the computation uses square, large sample to sample variations in x are emphasized.[10]. A simple way to overcome this problem is to define an average magnitude function where the average of the absolute values of the samples are computed and is known as short time magnitude and is expressed as in (9)

$$M = \sum_{k=0}^{n-1} |x(k)| \tag{9}$$

n = no of samples in each frame

$x(k)$ = Amplitude of each sample

2.4 NHR and HNR Ratios

The Noise to harmonic ratio (NHR) and harmonics-to noise (HNR) ratios are obtained from the signal-to-noise estimates by computing the autocorrelation of each cycle. Noise-to-harmonics ratio (NHR) is the ratio of the amplitude of noise to tonal components. Harmonics to Noise ratio (HNR) is the ratio of the amplitude of tonal to noise components.

2.5 Voice Onset Time (VOT)

Voice-onset time (VOT) is a feature of stop consonants production. It is defined as the length of time that passes between the release of a stop consonant and the onset of voicing, the vibration of the vocal cords, or, periodicity. If the vibration of vocal folds for the following vowel begins before the stop release then the VOT is said to be negative, e.g.:/ga/, /ba/. If the vibration of vocal folds for the following vowel begins at the point of stop release then the VOT is said to be zero, e.g.:/ka/, /pa/. If the vibration of vocal folds for the following vowel begins after the stop release then the VOT is said to be positive, e.g.:/pha/, /kha/. Fig. 4 shows the different phases of stop consonant followed by a vowel.

VOT can be measured using spectrographic analysis [11]. Spectrogram is a two dimensional display in which the vertical dimension corresponds to frequency, horizontal dimensions corresponds to time and darkness of

pattern is proportional to the signal energy. For VOT measurement the patient is asked to repeat steady syllables like /pa/-/ta/-/ka/ constantly and as long as possible.

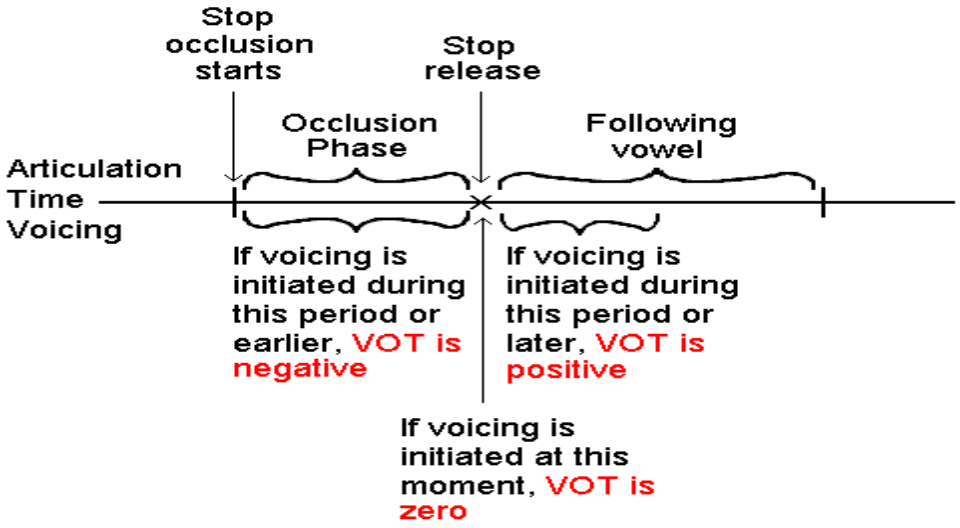

Fig. 4 Phases of stop consonant followed by a vowel

III. EFFECT OF PD ON PHONATION MEASURES-INTERPRETATION

The phonation measures affected by PD can be extracted by the algorithms explained in previous section. Since patients with PD suffer from impairment of vocal folds, their mean value of fundamental frequency Fo is expected to be higher when in comparison with healthy controls. Also the variation of F_0 i.e F_0-Standard Deviation is estimated to increase in sustained vowel prolongation. It can be observed that although considerable differences can be found between absolute and range values of F_0 in PD patients compared to Healthy persons these measures are not used, since they are affected by individual differences such as gender. Due to incorrect vocal fold vibrations, jitter and its variants as well as shimmer and its variants are estimated to be higher in PD patients. Hence shimmer and jitter can be used as measures to evaluate the instability of vocal cord vibrations. When compared with healthy subjects the NHR ratio is estimated to be higher and HNR ratio significantly smaller in PD patients. This is because PD patients suffer from incomplete vocal fold closure which produces small puffs of air during closure of vocal folds resulting in additional noise in the acoustic signal. The noise in speech can be also generated by turbulent airflow through the vocal fold. These ratios are used for assessing voice hoarseness. Also there would be deficits in producing normal voice onset time. Hence overall significant deficits in phonation features due to PD may be perceptually interpreted as poor voice quality and described in terms like hoarseness, harshness, breathy and rough voice, or even voice tremor.

IV. CONCLUSION

The PD affected phonation measures include fundamental frequency or pitch, jitter- a perturbation measurement of fundamental frequency, shimmer- a perturbation measurement of amplitude and HNR ratios -the amplitude of

tonal components relative to noise components in the speech. and voice onset time (VOT). For the extraction of these features, several algorithms namely autocorrelation method and its center clipping method and average magnitude difference method can be used to compute fundamental frequency and jitter. The accuracy of the pitch extraction can be improved by using the autocorrelation method based on center clipping technique. For extracting shimmer and its variants short time energy or short time magnitude algorithm can be used. The voice onset time can be obtained by spectrographic analysis.

REFERENCES

[1] G. Hornykiewicz O. "Biochemical aspects of Parkinson's disease". *Neurology,51:S2-9.* Review. PubMed PMID: 9711973.1998, Aug1998.

[2] Mohammad S Islam, Imtiaz Parvez, Hai Deng "Performance Comparison of Heterogeneous Classifiers for Detection of Parkinson's Disease Using Voice Disorder (Dysphonia)" IEEE 3rd International Conference on Informatics, Electronics & Vision 2014", Dhaka, Bangladesh. ISBN: 978-1-4799-5180-2/14/$31.00 ©2014 IEEE, 2014.

[3] Henriquez P, Alonso JB, Ferrer MA, Travieso CM, Godino-Llorente JI, Diaz-de-Maria F. "Characterization of healthy and pathological voice through measures based on nonlinear dynamics". IEEE Transactions on Audio Speech and Language Processing, vol 17, pp 1186-1195, 2009.

[4] Rusz J, Cmejla R, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease", *Journal Acoustic. Society of. America. 129 (1),* January 2011.

[5] Rusz J, Cmejla R "Acoustic markers of speech degradation in early untreated Parkinson's disease" Forum Acusticum 2011, pp 2725-2730, 27. June - 1. July, Aalborg, Denmark, 2011. European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

[6] Jan Rusz, "*Acoustic Analysis of Voice And Speech Disorders*," Ph.D. dissertation, Dept of Electrical Engineering and Information Technology, Czech Technical University, Prague, March 2012

[7] Doughlas O'Shaughnessey , *Speech Communication Human And Machine,* 2nd Edn, Universities Press India Limited, India 2001

[8] Savitha S Upadhya, "Pitch Detection in Time and Frequency domain" IEEE Xplore, Digital Library, ISBN: 978-1-4577-2077-2, DOI: 10.1109/ICCIT.2012.6398150.

[9] Mohammad Shahbakhi1, Danial Taheri Far, Ehsan Tahami , Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine "*J. Biomedical Science and Engineering, vol 7, pp 147-156, March 2014.*

[10] L. R.Rabiner and R. W.Schafer , *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

[11] http://ecconcord.ied.edu.hk/phonetics_and_phonology/wordpress/learning website/chapter_3_consonants_new.htm. Accessed on 25 April 2016