Early Detection of Parkinson's disease through Voice

Vikas
School of VLSI Design & Embedded Systems
National Institute of Technology
Kurukshetra ,India
vikas.rohilla11@gmail.com

R.K. Sharma
School of VLSI Design & Embedded Systems
National Institute of Technology
Kurukshetra, India
mail2drrks@gmail.com

***Abstract*— this paper present different methods of diagnosis of Parkinson's disease through voice in early stages. The aim of this study is to provide a simple, fast, cheaper method of detection of Parkinson's disease for the patients as there is no cure for this disease and the available therapies which provide some relief to them are too much costlier for this disease. Voice has different parameters that tell meaningful variations between normal people and Parkinson's patient like pitch, jitter, formant, shimmer, Mel-frequency Cepstral Coefficient (MFCC) and glottal pulse were founded out and analyzed.**

***Keywords*—MFCC, pitch, jitter, formant, shimmer, Parkinson's disease, glottal pulse, diagnosis of disease.**

I. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder that affects motor actions in human body. It is caused due to the deficiency of brain chemical named as 'Dopamine'. Major symptoms are tremors in hand, arm, legs and face, swallowing in body, speech problems, muscles rigidity, and movement problems. Though this disease occurs in aged persons yet in some persons its onset is quite early. Its onset occurs in a wide age range (45 years to 85 years). Speech problem is also a common symptom seen in Parkinson's disease patients. According to Metter and Hanson this symptom of PD is related to 'Hypo-kinetic Dysarthria' (HD) [1],[2]. Around 70-80% of PD affected patients have speech problem because of HD that comes in the area of phonetics science. Speech problems occur due to muscle contraction and because of this vocal cords do not work smoothly, resulting into voice disturbance. Higher mean pitch and more variations in formants are visible in PD affected patients. Fluctuations that result in shaking voice can be measured with the help of jitter. This is because laryngeal muscles can't be in stable position for more time [3].

MFCC coefficients and voice parameters like pitch, formants, jitter and shimmer of normal and PD affected persons have been computed using PRAAT and MATLAB based coding. The glottal pulse has also been extracted from the collected voice samples using APRAAT.

II. DATABASE AND PARAMETERS

A. Voice Database

For analysis of voice a database of 21 persons (especially five vowels a, e, i, o, u) has been recorded using Sony IC Recorder (ICD-UX513F). As each person was asked to pronounce the vowels loudly two or three times. Thus the total voice samples in the study were 105. In the available database 14 were of PD affected persons and rest were of normal healthy persons. The male persons whose voices were recorded have an age between 50 to 74 years and for females it was 45 to 85 years.

B. Mel-frequency Cepstral Coefficients

Mel-frequency Cepstral Coefficients are very important and popular parameter of voice. MFCC are most acoustic and robust in nature, therefore popularly used in speech recognition systems [4]. They are derived from a type of cepstral representation of voice signal. In Mel-cepstral, frequency bands are spaced equally on Mel-scale which calculates human auditory system's response more closely than normal cepstral (in which frequency bands are linearly spaced).

Basic steps to find out the MFCC coefficients are as shown in Fig.1. Any voice signal can be converted to Mel-frequency signal using (1).

$$Mel(f) = 2595\log_{10}\left(1+\frac{f}{100}\right) \qquad (1)$$

As shown in Fig.1, an audio signal is passed through Pre-emphasis block. The purpose of this block is to compensate the high- frequency part that was suppressed during the sound production mechanism of a person. Pre-emphasized signal splits into smaller frames of 15 – 40 ms duration. A voice signal shows large variations in its amplitude if it is of larger duration and therefore smaller signals are preferred for frame blocking. After splitting voice signal into frames, window function (hamming window) is multiplied with bunch of frames. Then, signal is passed through FFT (Fast Fourier Transform) block to convert signal into frequency domain.

Mel-filter bank block converts signal frequency into Mel-frequency using (1). Logarithm is used for channel

ISBN No.:978-1-4799-4949-6-@ 2014 IEEE

normalization. Finally, Discrete Cosine Transform (DCT) of logarithm block output done that de-correlates overlapped energies of Mel-filter bank. Output of DCT block provides MFCC values.

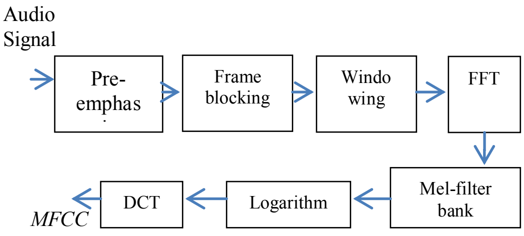

Fig. 1. Block diagram of MFCC

C. Formants

According to Gunnar Fant, Formants are the spectral peaks of the sound spectrum of voice [6]. Formants are the meaningful parameter of voice that can distinguish the vowels. This is due to the frequency components of voice. Normally we calculate first five formants only where 1st formant has lowest frequency among all. Mostly first two formants i.e. F1 and F2 we use to distinguish the vowels. We can calculate formants frequencies with the help of frequency components using different software like MATLAB, PRAAT. Alternatively spectrogram also tells us the meaningful frequency components and can be used to find out the formant values. Image of spectrogram is shown in Fig. 2 [7].

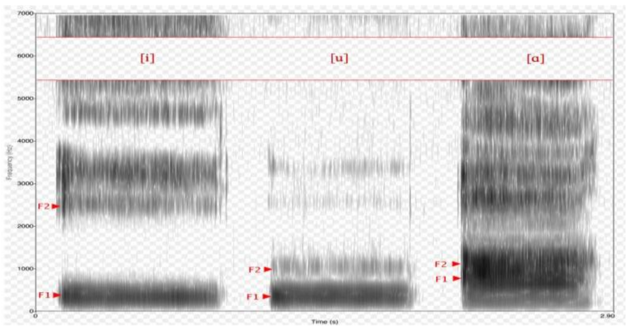

Fig. 2. Spectrogram showing formants f1 and f2 of vowel 'a, i, u'.

D. Pitch

Pitch depends upon the vocal folds vibrations, more vibrations of vocal folds will result in higher pitch values and vice versa. Pitch also depends on emotions, mood, swallowing on face etc. vocal folds contracts due to swallowing on face, neck (as in case of PD affected patients). Due to this more vibrations in vocal folds, that results in more vibrations in pitch values. Normally, pitch values in male voice is lower as compared to that of female voice due to larger length of vocal folds in males. [8].

E. Jitter and Shimmer

Jitter and Shimmer are measures of the cycle-to-cycle variations of fundamental frequency and amplitude, respectively, which have been largely used for the description of pathological voice quality [9].

Jitter (absolute) is the cycle-to-cycle variation of fundamental frequency, i.e. the average absolute difference between consecutive periods given by [10] is expressed as

$$Jitter\,(absolute) = \frac{1}{N-1}\sum_{i=1}^{N-1} \left|T_i - T_{i-1}\right| \quad (2)$$

Jitter (relative) is the average absolute difference between consecutive periods, divided by the average period. It is expressed in percentage given by [10] and is expressed as

$$Jitter\,(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | T_i - T_{i+1} |}{\frac{1}{N}\sum_{i=1}^{N} T_i} \quad (3)$$

In reference [10] shimmer (dB) is expressed as the variability of the peak-to-peak amplitude in decibels, i.e. the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20 is expressed as

$$Shimmer(dB) = \frac{1}{N-1}\sum_{i=1}^{N-1} | 20\log(A_{i+1} / A_i) | \quad (4)$$

Shimmer (relative) is defined as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, expressed as a percentage and is expressed as

$$Shimmer(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | A_i - A_{i+1} |}{\frac{1}{N}\sum_{i=1}^{N} A_i} \quad (5)$$

III. Results And Discussion

A. Analysis of Formants

Formants were calculated and analyzed for all database vowels using PRAAT. Here we found out first five formants of normal and PD affected person for vowel 'a' and then tracks were plotted against these calculated formant values for both PD affected person as well as normal person for vowel 'a'. In case of normal person first two formant value tracks are very similar to each other, as shown in Fig. 4. On the other hand, PD

ISBN No.:978-1-4799-4949-6-@ 2014 IEEE

affected person formant value tracks were not similar to each other as in case of normal person.

TABLE I. FORMANT FREQUENCIES OF NORMAL AND PD AFFECTED PERSONS FOR VOWEL 'A'.

S.No	Patient (P) / Normal (N)	Mean Formant Frequencies			
		F1	F2	F3	F4
1.	Normal 1 (N1)	388.925	2258.856	2768.9893	3441.589
2.	Normal 2 (N2)	464.609	2104.337	2890.4408	3409.625
3.	Patient 1 (P1)	955.117	2362.477	2701.3234	3663.396
4.	Patient 2 (P2)	455.994	1846.726	2688.6219	3877.603

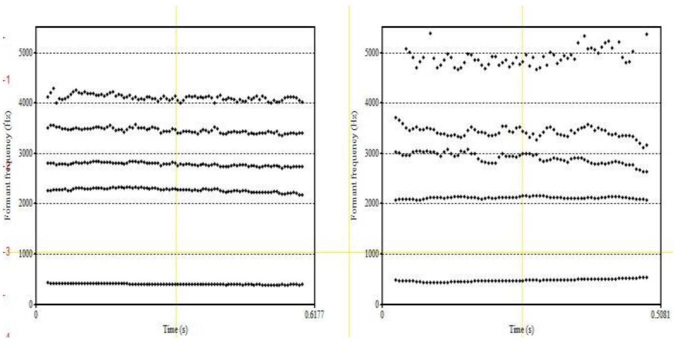

Fig. 3 First five formant values track of N1 and N2 for vowel 'a'.

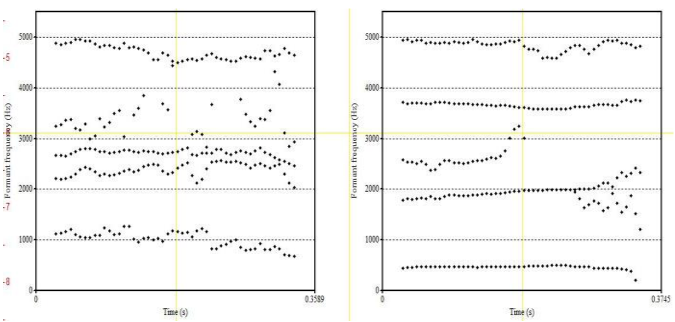

Fig.4 First five formant values track of P1 and P2 for vowel 'a'.

B. Analysis of Pitch

After analyzing pitch of normal and PD affected persons for all database voices, PD affected person shows higher pitch as compared to normal one in males but it is reverse in case of females. Pitch changes due to vocal cord vibrations more vibration results into higher pitch. PD affected persons have vibrations mostly in hands, neck, voice etc. that results in higher pitch. Some graph showing the differences between PD and normal person of same age group 45-75 years for vowel 'a', e, i, o, u in Fig. 6, 7, 8, 9, 10 respectively.

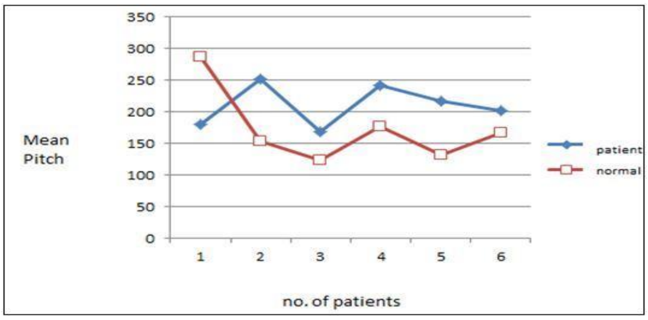

Fig. 5 Graph showing mean pitch for vowel 'a'.

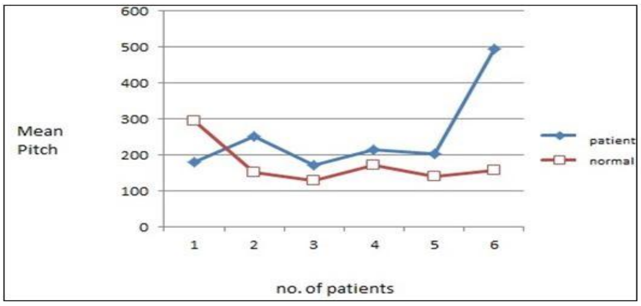

Fig. 6 Graph showing mean pitch for vowel 'e'

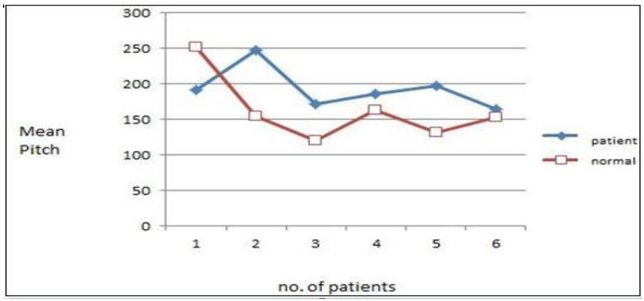

Fig. 7 Graph showing mean pitch for vowel 'i'

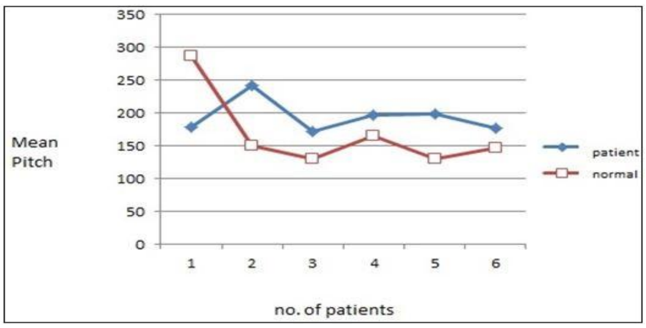

Fig. 8 Graph showing mean pitch for vowel ‘o’

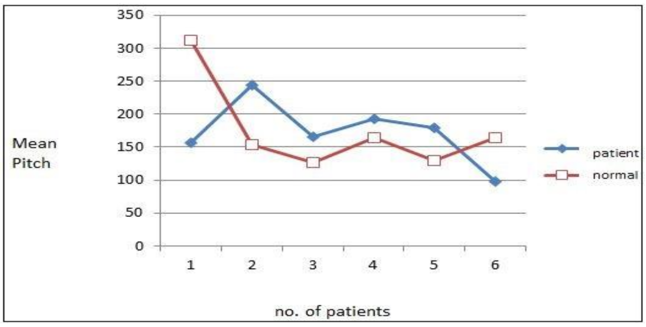

Fig. 9 Graph showing mean pitch for vowel ‘u’

C. Analysis of Glottal Pulse

Glottal pulse of normal and PD affected person for all voice samples were extracted using APRAAT, shape of glottal pulse in case of normal person is symmetric in nature as compared to PD affected person as shown in Fig. 11 and Fig. 12, for vowel a.

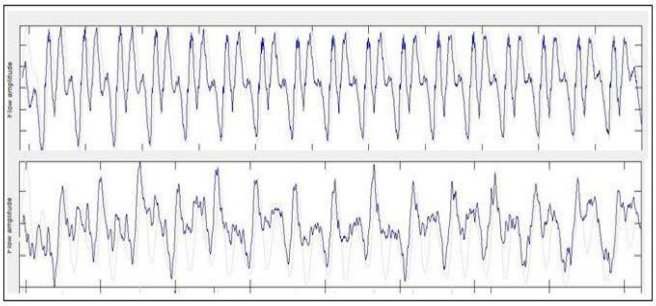

Fig. 10 Upper glottal pulse is for normal person and lower is for PD patient for vowel ‘a’

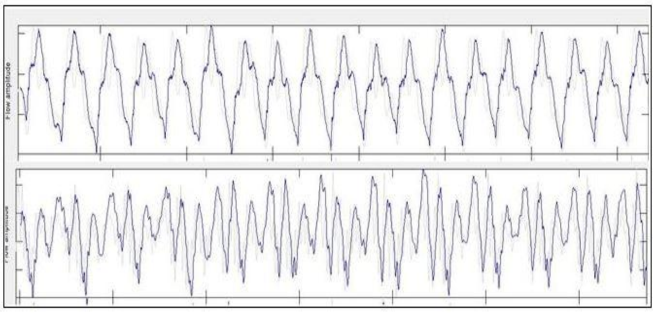

Fig. 11 Upper glottal pulse is for normal person and lower is for PD patient for vowel ‘e’

D. Analysis of MFCC

The most important parameter of voice i.e. MFCC was also calculated using MATLAB. Results are shown in Fig. 13 and Fig. 14, for PD patient and normal person respectively for vowel a.

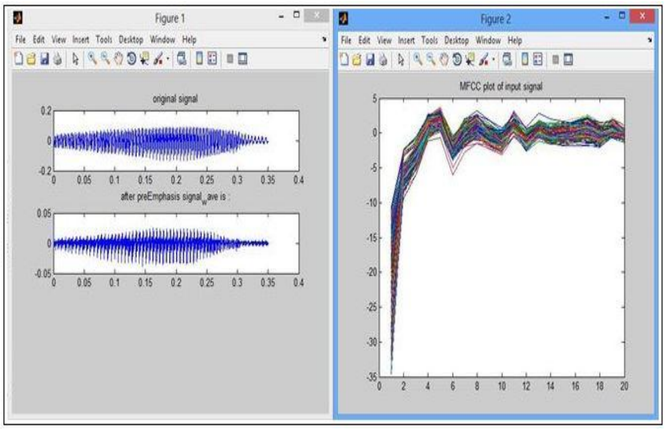

Fig. 12 MFCC values plot for PD patient of vowel ‘a’.

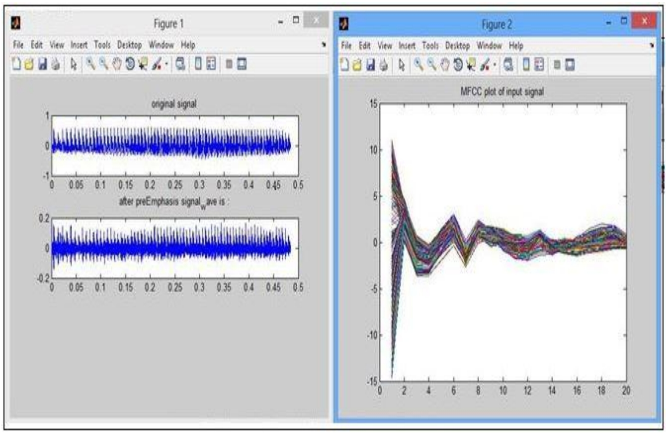

Fig. 13 MFCC values plot for normal person of vowel ‘a’.

E. Analysis of Jitter and Shimmer

After that jitter and shimmer was calculated and analyzed that was also more in case of PD patients than normal person of same age group from 45 to 75 years. Variations are shown in TABLE I and II for PD patient and normal person respectively.

TABLE II. JITTER AND SHIMMER %(LOCAL) VALUES OF PD PATIENT FOR VOWEL ‘E’.

S.No.	Jitter (% local)	Shimmer (%local)
1.	0.369	1.708
2.	1.381	12.166
3.	0.355	6.560
4.	0.467	5.480
5.	0.353	1.381

ISBN No.:978-1-4799-4949-6-@ 2014 IEEE

6.	0.688	21.345
7.	0.369	1.708

TABLE III. JITTER AND SHIMMER %(LOCAL) OF NORMAL PERSON FOR VOWEL 'E'.

S.No.	Jitter (% local)	Shimmer (%local)
1.	0.558	4.041
2.	0.179	1.833
3.	0.290	1.834
4.	0.361	4.327
5.	0.357	3.207
6.	1.891	4.717

IV. CONCLUSION

The voice parameter such as pitch, formants, jitters, shimmer, glottal pulse and MFCC have been investigated for both normal and PD affected males and females. The observation, on the basis of analysis done, can be summarized as under:-

- PD patient male have higher pitch as compared to normal person.
- Formants have more variations as compared to normal person.
- In case of normal person glottal pulse, there emerged a similar pattern whereas it is not in case of PD affected persons.
- Jitter and Shimmer values are more in PD patients than normal persons.

V. REFERENCES

[1]. Hartelius L and Svensson P, "Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey," Folia Phoniatr Logop 1994:46(1):9-17.

[2] Goberman A.M. and Mc Millan J, "Relative speech Timing in Parkinson's disease," commun Science Discord 2005,32:22-29

[3] Jiri Mekyska, Irena Rektorova and Zdenek Srnekal, "Selection of optimal Parameters for Automatic Analysis of speech Disorders in Parkinson's disease".

[4] Md. Sahidullah and Goutam Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," Science direct, Speech Communication 54 (2012) 543–565.

[5] Meinard Müller,"Information Retrieval for Music and Motion," Springer. p. 65. ISBN 978-3-540-74047-6

[6] Fant.G,"Acoustic Theory of Speech Production. Mouton & Co," The Hague, Netherlands, 1960.

[7] Titze and I.R., "Principles of Voice Production," Prentice Hall, ISBN 978-0-13-717893-3, 1994.

[8] Tonya Reiman, "The Human Voice - Pitch," http://www.bodylanguageuniversity.com.

[9] Mireia Farrús, Javier Hernando and Pascual Ejarque, "Jitter and Shimmer Measurements for Speaker Recognition".

[10] Raymond E. Slyh, W.Todd Nelson and Eric G. Hansen, "Analysis of mrate, Shimmer, Jitter and Fo counter features across stress and speaking style in the susas database."

[11] Ch. Ramaiah and Dr. V. Srinivasa Rao, " Speech samples recognition based on MFCC and vector Quantization", International Journal on Computer Science and Emerging Trends (IJCSET) Vol. 01, No.02, December 2012, 1-7.

[12] Jun Qi1, Dong Wang2, Yi Jiang1, Runsheng Liu1, "Auditory features based on Gammatone filters for Robust Speech Recognition."

[13] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi," Voice Recognition Algorithms using Mel-Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Volume 2, Issue 3, March 2010, ISSN 2151-9617.

[14] Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk," Speech Recognition using MFCC", International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand).

[15] Kishore Prahallad, "Speech Technology: A Practical Introduction,Topic: Spectrogram, Cepstrum and Mel-Frequency Analysis".

ISBN No.:978-1-4799-4949-6-@ 2014 IEEE