ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Empirical Mode Decomposition: A way for finding Pitch (Stuttered speech signal)

N Raju*, P Neelamegam.

SASTRA University, Thanjavur, 613401, India

ABSTRACT

In this paper we suggest a novel method for finding the pitch period in a stuttered speech signal taken from the UCLASS database in this we have applied Empirical Mode Decomposition (EMD) for finding the pitch period which is calculated using the Intrinsic Mode Function1(IMF1) and is been compared with the pitch period extracted using the standard PRAAT tool. The suggested method is also used to identify the frame as voiced speech or unvoiced speech. Speech signal being a non stationary signal and the EMD method used fits well for the non stationary signal and its validity is been compared with the standard methods applicable for stationary signals.

Keywords: PRAAT, Pitch, EMD, Speech, Stationary signals

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Speech processing which deals with the processing of speech signal is considered as a non stationary signal. It requires extracting the features from the speech signal for doing analysis. The various methods used for extracting the feature are done considering the speech signal to be stationary for short time which actually is not the case. There are various methods to deal with the non stationary signal which many researchers are working with like wavelet transform, STFT (Short Time Fourier transform), etc.

A novel method of analyzing the non stationary signal is been developed by N.E. Huang for adaptively representing non stationary signals as sums of zero-mean we can think of representing these signals in terms of amplitude and frequency modulated (AM–FM) components in which the component forms the basis function. This method Empirical Mode Decomposition (EMD) gives a gateway for analyzing the non stationary signal and it has been proved in the past.

EMD is designed primarily for obtaining representations of signals which are oscillatory, possibly non stationary or generated by a nonlinear system in this method of Decomposing a complicated set of data into a finite number of Intrinsic Mode Functions (IMF), that admit well behaved Hilbert Transforms is done.

This paper uses the EMD method for extracting the features which can be further considered for doing speech recognition or to identify the voiced speech and unvoiced speech for classification of speech sounds [1-13].

METHODOLOGY

The method used for extracting the pitch is done by using EMD. The following steps are used for determining the basis function of decomposition[1]. The basis function is called as Intrinsic Mode Function (IMF).

1. Identify all extrema of *x(t)*.
2. Interpolate the local maxima to form an upper envelope *u(x)*.
3. Interpolate the local minima to form an lower envelope *l(x)*.
4. Calculate the mean envelope: *m(t)=[u(x)+l(x)]/2*.
5. Extract the mean from the signal: *h(t)=x(t)-m(t)*
6. Check whether *h(t)* satisfies the IMF condition.

   YES: *h(t)* is an IMF, stop sifting.
   NO: let *x(t)=h(t)*, keep sifting.

Where x(t) is the speech signal recorded with 8KHz sampling rate. The recorded signal was then converted to the frame with the frame length of 180 samples and then the empirical decomposition was applied to it. Initially as the start of the emd function we assume the whole x(t) as the basis function which is considered as IMF 1. Figure 1 shows the plot of the recorded speech signal saying the word “one”.

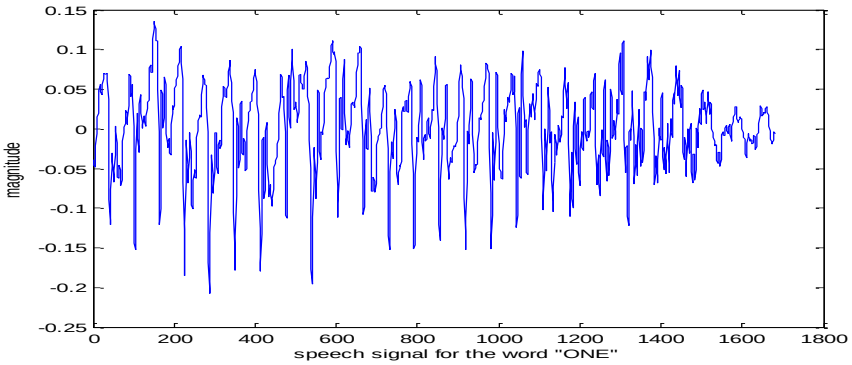


Figure 1- shows the plot of the recorded speech signal saying the word “ONE”

ISSN: 0975-8585

The recorded speech is then converted into the frames with 180 samples each. The frame considered here is a non overlapping one. Once the speech signal is been converted into frames, then the individual frame is decomposed into IMF. The procedure for finding the IMF is done by first finding the maxima and the minima for the signal x(t). Then the maxima and the minima points is been interpolated using cubic spline method and the point wise mean is determined.

$$m(t)=[u(x)+l(x)]/2 \qquad \text{----(1)}$$

Equation 1 gives the calculation of the mean for u(x) upper envelope and the l(x) lower envelope. The extracted point wise mean is then subtracted from the initial IMF1 and the resultant signal is then assigned to IMF1. This process is done repeatedly until the stopping criteria is been met the stopping criteria [2] used here is the iteration goes on till the m (t) goes less than 0.05. When this stopping criteria is met then the signal x (t) is subtracted with IMF1 which becomes the IMF2 and the same procedure is followed until the whole signal x (t) is been decomposed and the maxima and the minima comes down to one each. The figure 2 shows the output of the EMD function [1-4] for the first frame of the speech signal .

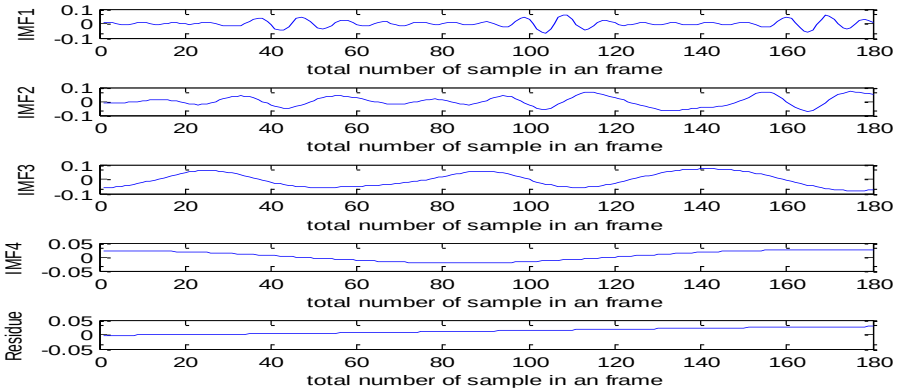


Figure 2: Output of the EMD function for frame1

From the figure 2 we can see that the signal is being decomposed into many IMF, forming the basis function for the signal x(t) . Thus the IMF gives the information with respect to the composition of the speech signal x(t) which can be used for determining the pitch for the individual frame. The figure 3 shows the waveform of the recorded speech for the digit 'ONE' done using the PRAAT tool. Upon doing the pitch analysis using PRAAT tool which is plotted in the figure4 shows the pitch[7] extraction for each frame with the frame length of 100 and the recorded time is of 0.21 seconds.

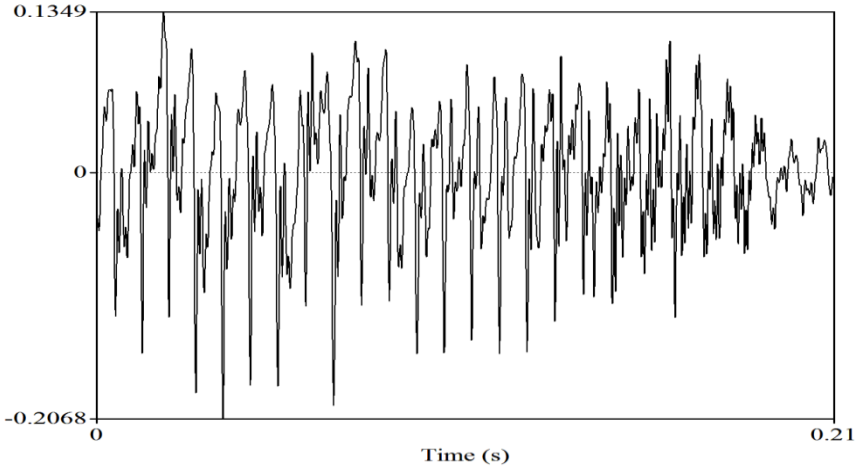


Figure 3- Waveform plot of the recorded speech signal for the word "ONE" using PRAAT tool.

ISSN: 0975-8585

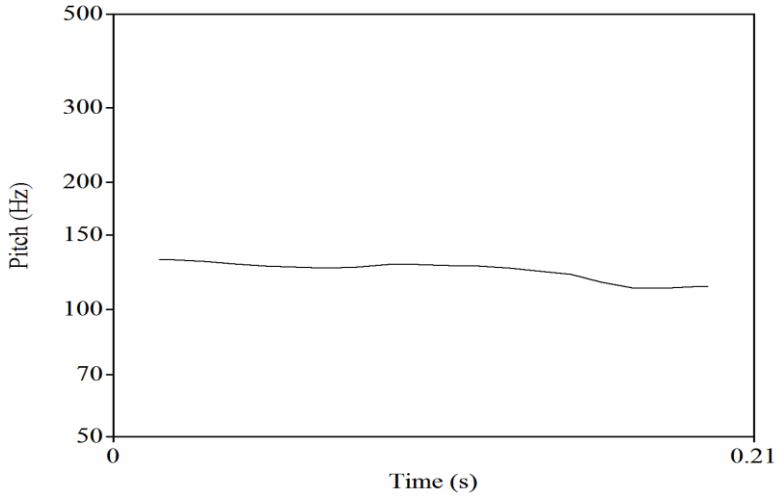
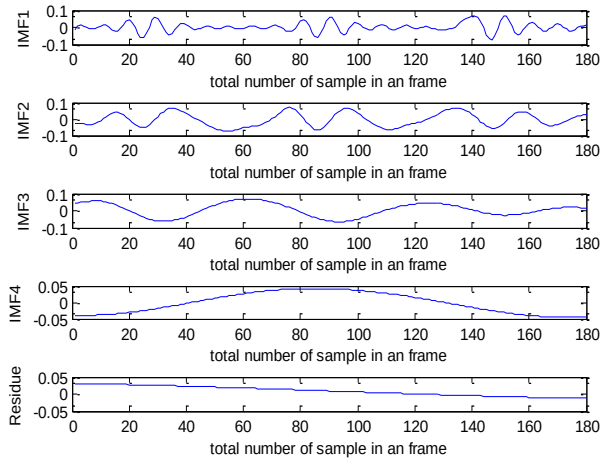
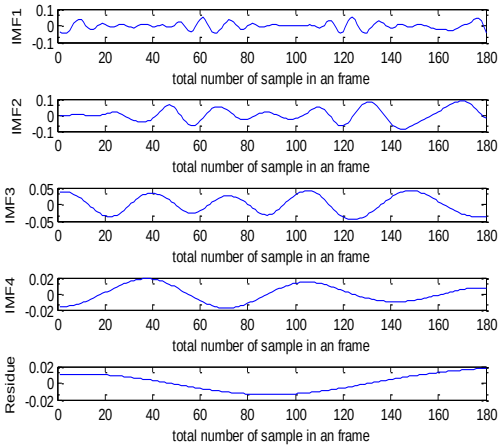
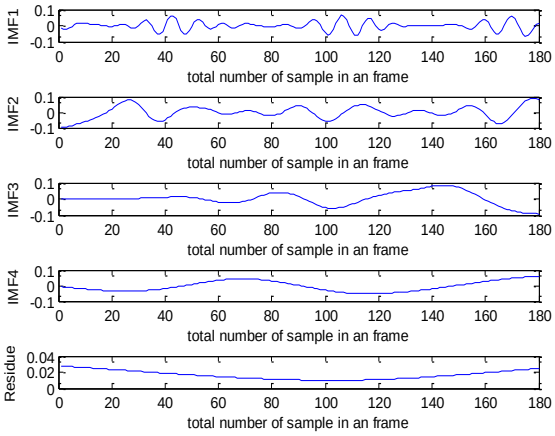


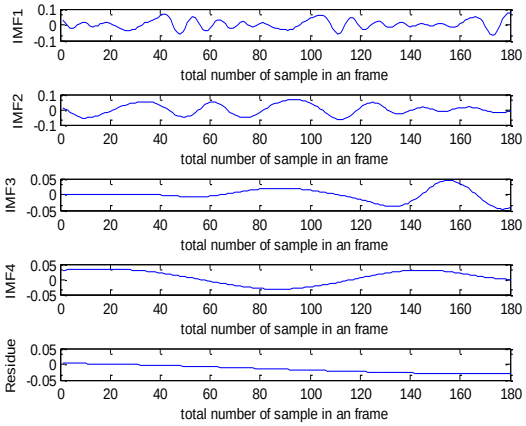
Figure 4- Pitch analysis plot for the recorded speech signal

The pitch analysis done for the recorded speech signal is then compared with the pitch analysis[8-9] done using EMD method. The figure 5 shows the plot of the IMF basis function which gives us the impression that upon doing the EMD analysis the IMF1 basis function can give the information of the pitch related feature in the frames of the speech signal.









ISSN: 0975-8585

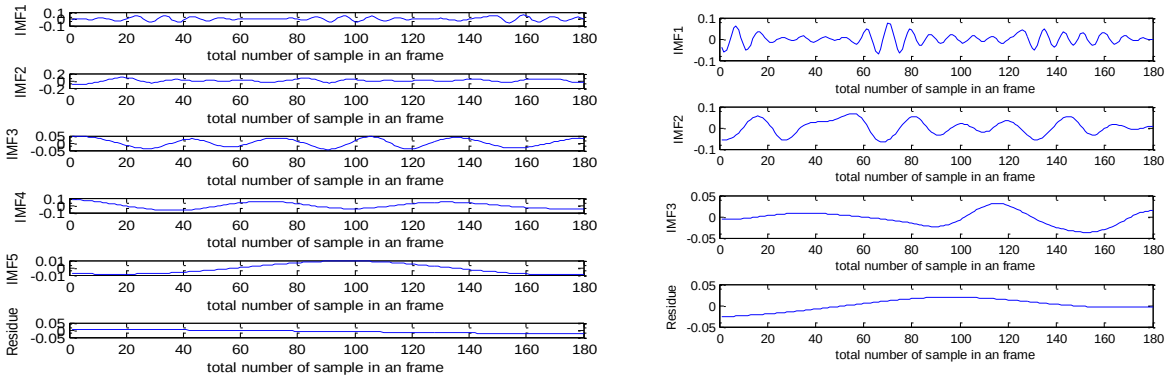


Figure 5- EMD for the speech frames

From the figure 5 we can see that the linear and the nonlinear cycles are easily separated by EMD methods in the same way as the linear methods like Fourier and Wavelet[5][11-13] do. The EMD method gives the better decomposition as compared to the later one which can be confirmed by figure 5.

As an experimental basis we took the speech signal from University College London's Archive of Stuttered Speech (UCLASS) Funded by The Wellcome Trust. The database which has the collection of shuttered speech recorded from various subject. Figure 6 is the plot of the waveform from the file taken from the database.

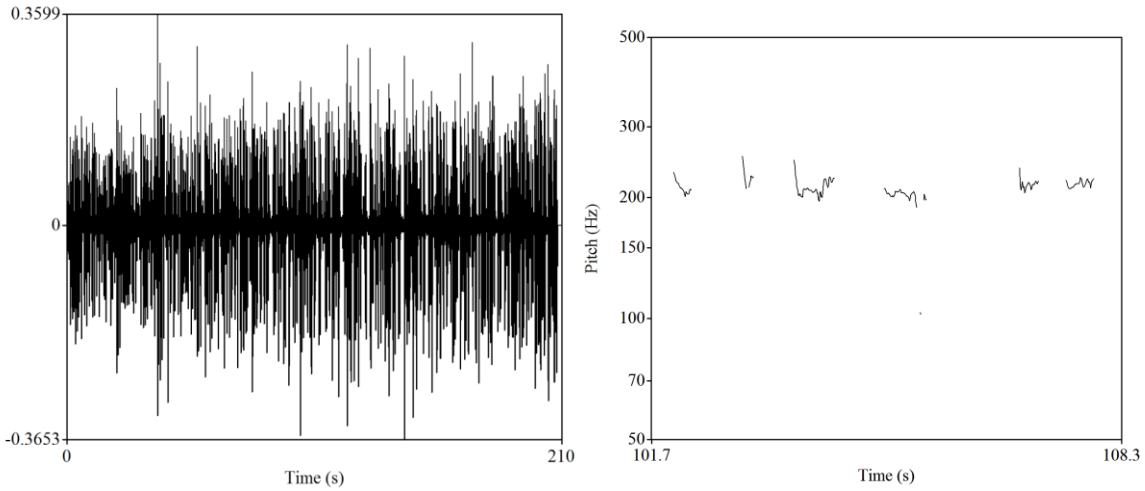


Figure 6- Plot for the wave file from UCLASS database and the pitch analysis using PRAAT tool.

The autocorrelation function taken for the waveform shown in the figure 1 is shown in figure 7 which shows that the method using EMD can also be used to detect the pitch for extracting the speech feature used for speech recognizing system. The speech signal is not been weighted by using any windowing technique[10]. Only the short time speech signal is considered. The figure 8 shows the pitch plot for the waveform F_0142_11y3m_1.wav file taken from the UCLASS[6] database .

ISSN: 0975-8585

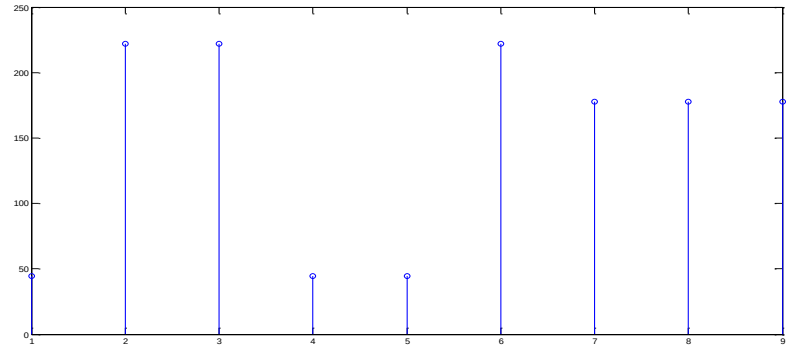


Figure 7 – Plot of pitch using EMD

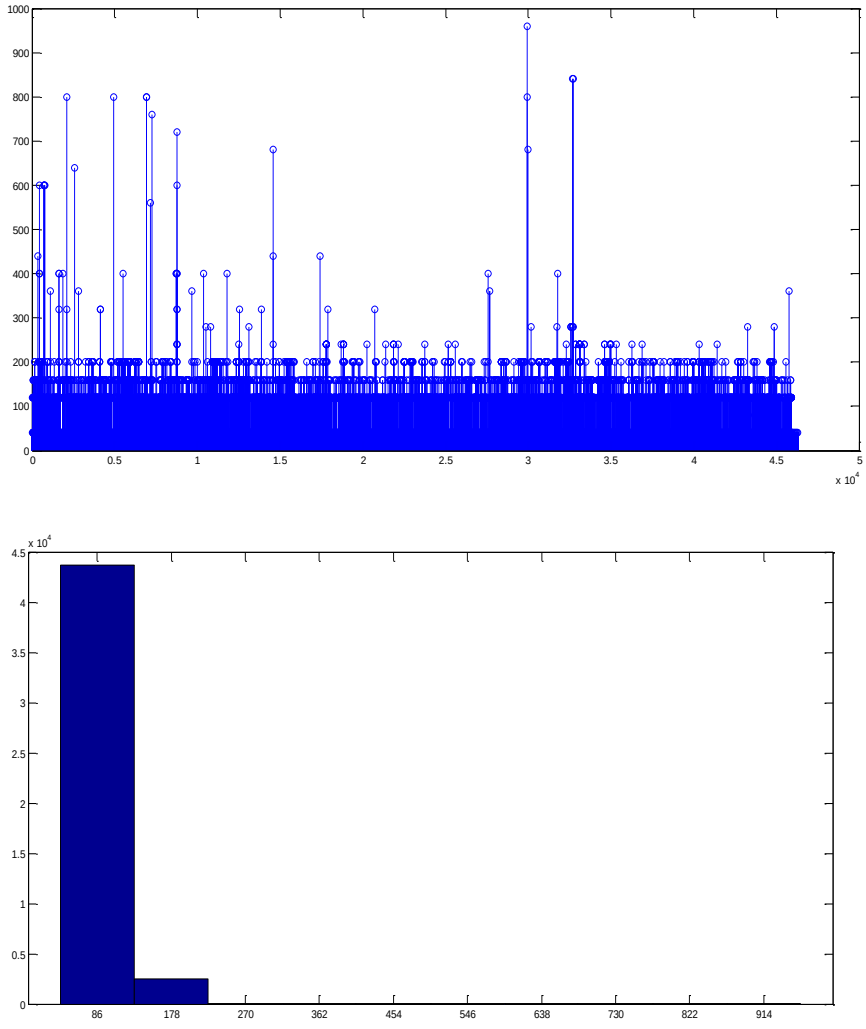


Figure 8- Plot for the autocorrelation for the wav file from UCLASS.

CONCLUSION

Thus we can see that the method using EMD process which is very much suitable for non stationary signal fits well for finding the pitch values for the speech signal same as the many conventional method explored in past and present. This method can be given a validation by using any speech recognizing system. Mostly the EMD process is used for enhancement of speech signal.

ISSN: 0975-8585

REFERENCES

[1] Huang N. E, Proc. Royal Soc. London A, 1998; 454: 903-995.
[2] Rilling G, Flandrin P,Gonçalves P, IEEE-EURASIP Workshop on Nonlinear Signal and Image ProcessingNSIP-03, 2003;
[3] Rilling G, Flandrin P, Gonçalves P, Lilly J M, ,Signal Processing Letters
[4] Huang N. E., , Proc. Royal Soc. London A, 2003; 459: 2317-2345
[5] Deering R , Kaiser J F, ICASSP 2005;
[6] Howell, P., Davis, S., Bartrip, J. ,Wormald, L. Stammering Research, 2004;1: 309-315.
[7] Ananthapadmanabha T V ,Yegnanarayana B,IEEE Trans. on Acoustics, Speech, and Signal Processing, 1975; 23: 562-570.
[8] Cheng Y M, Shaughnessy D O, IEEE Trans. on Acoustics, Speech, and Signal Processing, 1989; 37: 1805-1815.
[9] Strube H W, Journal of the Acoustical Society of America,1974; 56:1625-1629.
[10] Sondhi M M, IEEE Trans. on Audio and ElectroAcoustics, 1968; 16: 262266.
[11] Kumar, R.P., Balaji, V.S., Raju, N,Advances in Intelligent Systems and Computing,2016; 397: 603-613.
[12] Elamaran, V., Upadhyay, H.N., Raju, N., Narasimhan, K. ,International Journal of Pharmacy and Technology, 2015;7 (3): 9802-9810.
[13] Raju, N., Arjun, N., Manoj, S., Kabilan, K., Shivaprakaash, K,Journal of Artificial Intelligence, 2013;6 (2): 161-167.