IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

EMD Based Speech Reconstruction for Different Assamese Dialect

Nisha Goswami

Dept. of Electrical Engineering,Girijananda Chowdhury Institute of Management and Technology, Guwahati, India

Corresponding Author: nisha.goswami75@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i6.457461 | Available online at: www.ijcseonline.org

Accepted: 10/Jun/2019, Published: 30/Jun/2019

Abstract—In this paper, reconstruction of speech signal is performed using a novel technique Empirical Mode Decomposition (EMD). EMD is mainly applicable for non-linear and non –stationary signals and therefore applicable for speech which is non linear and non-stationary. EMD is applied for finding the glottal source signal of speech which provides us the source information and then vocal tract filter is found out and original speech is reconstructed.
Speech samples are collected from different Assamese dialect and the experimental result derived establishes the effectiveness of the proposed method.

Keywords— Speech, Empirical Mode Decomposition (EMD), Intrinsic Mode Function (IMF), Fourier Transform (FT)

I. Introduction

In speech processing, glottal source and vocal tract separation is a crucial component. Speech is the result of many non-linearly interacting processes like almost all natural phenomena therefore any linear analysis has the potential risk of missing a great amount of information content [1]. In most of the data analysis method like Time-frequency or time –scale, data is assumed to be linear and stationary. Wavelet analysis work for non-stationary linear data. All the above methods show some windowing effect, low time resolution, low frequency resolution etc.[2]

The non-stationary attributes of a speech signal restricts direct application of conventional digital signal processing methods to a speech signal. Since EMD performs decomposition assuming the nonstationary nature of signal. In EMD technique frame by frame analysis is not required.
Speech is usually both non-linear and non –stationary. The empirical mode decomposition (EMD) is designed in such a way that it reduces non-stationary, multi component signals to a series of amplitude and frequency modulation (AM and FM) contributions. As a result it creates a bank of sub signals, termed intrinsic mode functions (IMF).The sum of all the IMFs produces the original signal**.** The last IMF or residual is of the lowest order [3]. The residual information provide us the information of source.

Some works have been reported which explores EMD for speech processing applications. In [4], the reported work illustrates a novel and effective method for suppressing residual noise from enhanced speech signals as a second-stage post-filtering technique using EMD**.** In [5], the authors show that the speech fundamental frequency can be captured in a single IMF. A new algorithm for pitch extraction based on the Ensemble Empirical Mode Decomposition (EEMD) is presented. In [6], EMD is used for extraction of long term structures in musical signals. Long-term musical structures provide information concerning rhythm, melody and the composition.

In this paper, speech synthesis applications using EMD is performed. EMD is applied for obtaining residual signal which provides us the information of source. After getting the source information, vocal tract filter response is determined and the original speech signal is reconstructed.
The experimental result derived establishes the effectiveness of the proposed method for different Assamese dialect.

II. Theoritical Consideration of EMD and its Usefulness

Empirical Mode Decomposition (EMD)-EMD- For processing of non-linear and non-stationary signals EMD is an adaptive tool. Based on the local behaviour of the signal, it segregates the constituent parts of the signal. It is able to analyse non-zero mean signals. Therefore no pre-processing is required, and is suitable to analyse the riding waves which may have no zero-crossings between two consecutive extrema [6]. Using EMD, signal can be decomposed into number of frequency modes called intrinsic mode function (IMF) which is similar to simple harmonic function.

Like simple harmonic function, an IMF represents a simple frequency mode. Mode frequency has amplitude and frequency as function of time. An IMF must satisfy two conditions. The first is that the number of zero crossing and number of extrema (sum of maxima and minima) must be differ by one or it should be equal. The second condition is that mean of the cubic splines must be equal to zero at all points [7]. The computational steps of the algorithm may be summarized as below [7].

For speech signal , first identification of maxima and minima is done. Next step is the upper and lower envelope generation via cubic spline interpolation among all the maxima and minima respectively. For computing local mean series $m(t)$ point by point averaging of the two envelope is done.To obtain IMF, subtraction of $m(t)$ from the data i.e. $h(t) = x(t) - m(t)$ is performed and then properties of h(t) are checked to verify whether it Is a IMF or not. If If h is not a IMF, replace $x(t)$ with $h(t)$ and repeat the procedure. Again if h is a IMF, residue is evaluated as $m(t) = x(t) - h(t)$.The avove steps are repeated by shifting the residual signal, until at least two extrema remains. Therefore after performing EMD we obtain the glottal source information and with the help of this vocal tract of that person can be determined.

III. SYSTEM BLOCK DIAGRAM AND EXPERIMENTAL DETAILS

The complete system model is shown in Figure 1. Speech samples are taken for different Assamese dialects. There are mainly four dialect groups namely....

i) The eastern Assamese dialects spoken in the districts of Tinsukia, Dibrugarh, Lakhimpur, Sibsagar, Jorhat, Golaghat and Sonitpur.

ii) The central Assamese group of dialects spoken primarily in Nagaon and Morigaon districts and some parts of Sonitpur and Jorhat district also.

iii) The Kamrupi group of dialects are spoken in districts of Kamrup,Nalbari,Barpeta,Darrang,Kokrajhar and Bongaigoan

iv)The Goalparia group of Assamese dialects spoken primarily in Dhubri and Goalpara districts and in certain areas of Kokrajhar and Bongaigaon districts.

Firstly, the speech samples are collected for these four different dialects and the collected speech samples is presented to the EMD block to get the residual, which represents the glottal source information. After getting the residual , fft of the original speech $X(\omega)$ and residual $R(\omega)$ is taken. Vocal tract filter transfer function in frequency domain is obtained by

$$H(\omega) = X(\omega)/R(\omega)$$

Final reconstruction of speech can easily be done from the concept of source filter model. Now, the speech signal can be reconstructed with the help of source and vocal tract filter responses which appears similar to the original speech signal in both time and frequency domain which is explained in the next Section .

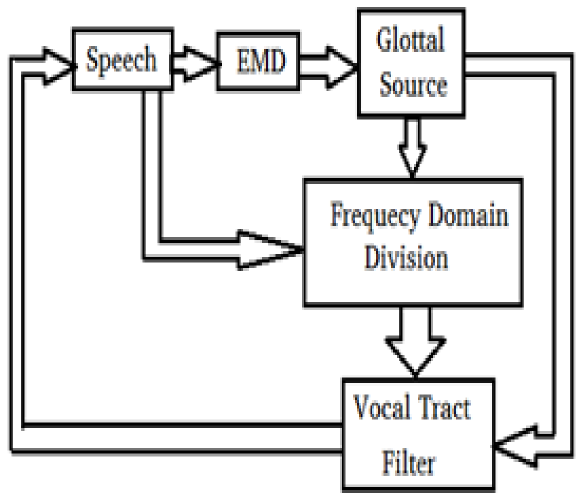
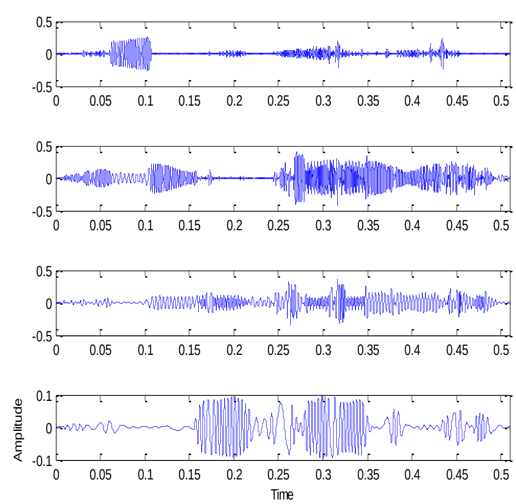


Fig 1: System Model

IV. RESULTS AND DISCUSSION

Firstly the sample taken from Kamrupi dialect. The EMD algorithm is applied to the speech signal. The speech signal along with their IMFs starting with IMF1,IMF2, IMF3, IMF4 upto IMF9 and the residual content are shown in Figure 2. Figure 3 represents the original and reconstructed speech signal.



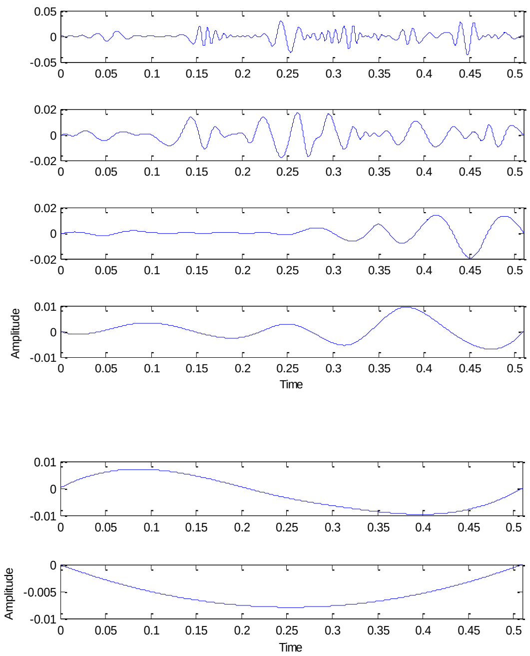


Fig 2: IMF 1 to IMF 9 and the residual signal

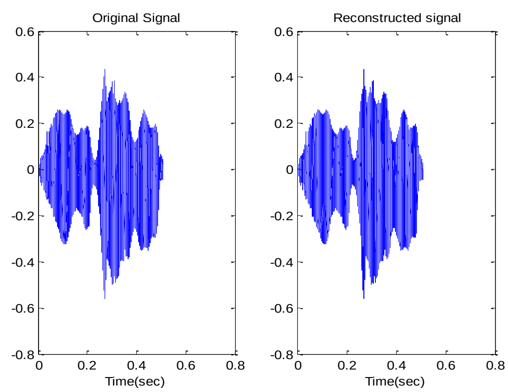


Fig 3: original and reconstructed speech signal

Secondly, The Goalparia group of Assamese dialect is taken as a sample of speech and EMD is applied on it.

The speech signal along with their IMFs starting with IMF1, IMF2, IMF3, IMF4 upto IMF11 and the residual content are shown in Figure 4. Figure 5 shows the original and the reconstructed speech signal.

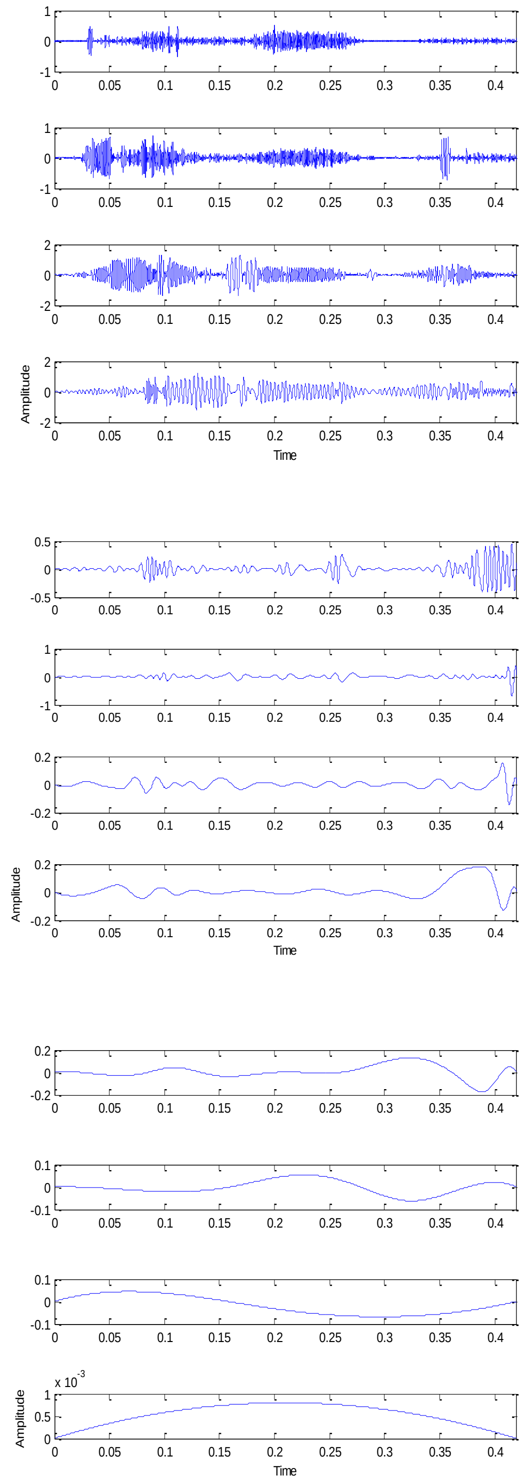


Fig 4: IMF 1 to IMF 11 and the residual signal

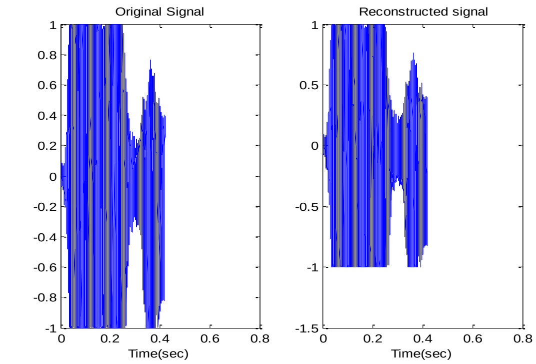


Fig 5: The original and the reconstructed speech

The eastern Assamese dialects and the central Assamese group of dialects are also taken and EMD based reconstruction algorithims applied to these samples and it is observed that it gives satisfactory performances in both the cases. Figure 6 represents the original and reconstructed speech signal of central Assamese dialect. Figure 7 shows for eastern Assamese dialect of the same.

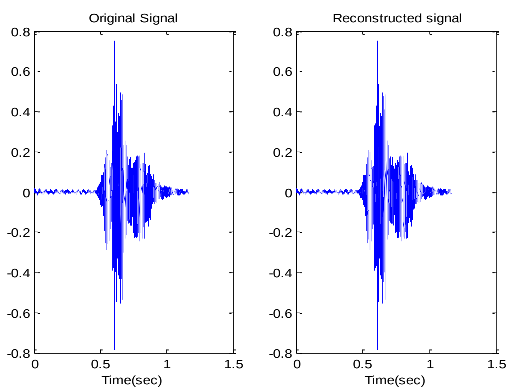


Fig 6: The original and the reconstructed speech

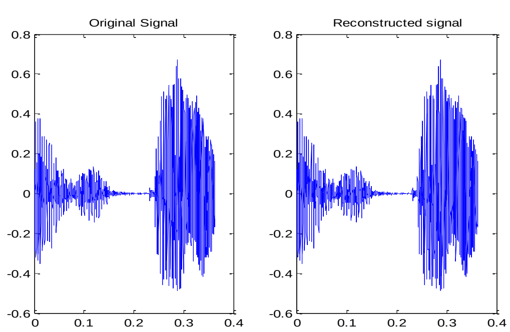


Fig 7: The original and the reconstructed speech

The similarity test between original and the reconstructed speech samples are done by the correlation method and the dissimilarity between original and the reconstructed speech samples for different Assamese dialect is shown in table1.

Table1: Similarity measure between original and reconstructed signal

Dialects	Speech samples	Dissimilarity measure
Kamrupi	Sample 1	1.68%
	Sample 2	0.31%
	Sample 3	1.91%
	Sample 4	1.38%
Goalparia	Sample 1	1.52%
	Sample 2	0.71%
	Sample 3	3.92%
	Sample 4	0.4%
Central	Sample 1	0.9%
	Sample 2	1.41%
	Sample 3	1.49%
	Sample 4	1.62%
Eastern	Sample 1	1.92%
	Sample 2	0.2%
	Sample 3	0.83%
	Sample 4	1.64%

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have discussed about the novel EMD technique for reconstruction of speech signal for different Assamese dialect.Since EMD is an adaptive tool for processing of non-linear and non-stationary signals, it is very effective to be applied on speech signals.From experimental results, it can be seen that EMD is satisfactory method for decompositing speech signals into number of IMFs and obtaining the lowest frequency component, which is considered as the source signal in case of speech. Applying the concept of source-filter model, vocal tract filter transfer functiom is determined and original speech signal is reconstructed for four different Assamese dialects. The proposed approach next can be used for voice conversion application with the help of one person's source information and vocal tract filter of another person.

REFERENCES

[1] A. Bouzid, N. Ellouze, " EMD Vocal Tract Frequency Analysis of Speech Signal", 4th International Conference Sciences of Electronic, Technologies of Information and Telecommunications , PP TUNISIA

[2] S. S. Sheth ,"Extraction of Pitch from Speech Signals using Hilbert Huang Transform".

[3] B. M. Battista, C. Knapp, T. McGee, and V. Goebel, Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data, Geophysics , vol72, PP 29-37, 2007.

[4] T. Hasan and Md. K. Hasan, "Suppression of Residual Noise From Speech Signals Using Empirical Mode Decomposition", IEEE Signal Processing Letters, vol 16, no. 1, Jan 2009.
[5] G. Schlotthauer, M. E. Torres and H. L. Rufiner, "A New Algorithm For Instantaneous F0 Speech Extraction Based On Ensemble Empirical Mode Decomposition", 17th European Signal Processing Conference (EUSIPCO 2009), 2009.
[6] P. Heydarian and J. D. Reis , "Extraction Of Long-Term Structures In Musical Signals Using The Empirical Mode Decomposition", Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05), Madrid, Spain, 2005.
[7] G. Schlotthauer, M. E. Torres and H. L. Rufiner, "Voice Fundamental Frequency Extraction Algorithm Based on Ensemble Empirical Mode Decomposition and Entropies", Proc. of the World Congress on Medical Physics and Biomedical Engineering, Germany,2009
[8] J. W. Huang and B. Milkereit, "Empirical Mode Decomposition Based Instantaneous Spectral Analysis and its Applications to Heterogeneous Petrophysical Model Construction'', joint convention of CSPG, CSEG and CWLS, 2009.
[9] G. Rilling , P. Flandrin ,P. Goncalves, "Empirical Mode Decomposition as a Filter Bank", IEEE Signal Processing Letters, vol. 11, Issue 2, pp- 112 - 114, 2004.
[10] O. Douglas: "Speech communication Human and Machine", IEEE Press, 2nd Ed., , 2000
[11] L. R. Rabiner and and B. H. Juang: "Fundamentals of Speech Recognition", Prentice-Hall, 1993.
[12] N. Goswami, M. Sarma, K.K.Sarma, "Reconstruction of Speech Signal using Empirical Mode Decomposition Based Glottal Source Extraction", In Proceedings of 1st IEEE International Conference on Emerging Trends and Application in Computer Science, pp 27-32, Shillong, India, 2013.
[13] N. Goswami, M. Sarma and K.K.Sarma, "Empirical Mode Decomposition Based Reconstruction of Speech Signal in Noisy Environment", IEEE International Conference on Signal Processing and Integrated Network, Noida, India, 2014.
[14] N. Goswami, M. Sarma and K.K.Sarma,"Effective Speech Signal Reconstruction Technique using Empirical Mode Decomposition Under Various Conditions", Recent Trends in Intelligent and Emerging system Design, Springer, 2015.

Authors Profile

Ms.N.Goswami-received M.Tech degree in Electronics and Communication Technology from Gauhati University in the year 2013 and currently pursuing ph.D from Gauhati University. She is working as a Assistant Professor in the department of Electrical Engineering at Girijananda Chowdhury Institute of Management & Technology under Gauhati University since 2012. She has two IEEE publications and one Springer book chapter and attended numbers of national workshops.Area of interest includes speech technology, signal processing, Analog circuit design etc

