VOICED SPEECH CHARACTERISATION BASED ON EMPIRICAL MODE DECOMPOSITION

Mina Kemiha[1]

[1]Department of electronic, Jijel University, Algeria

ABSTRACT

Empirical Mode Decomposition (EMD) is a tool for the analysis of multi-component signals. The EMD algorithm decomposes adaptively a given oscillation modes namely the functions of intrinsic mode (IMFs) extracted from the signal itself signal. The analysis method is no need for a basic function fixed a priori as conventional analytical methods (eg Fourier transform and the wavelet transform). In this paper, the algorithm of empirical mode decomposition (EMD) is proposed as an alternative to estimate the vocal tract formants characterizing the vocal tract. The proposed method was tested on natural speech. LPC analysis of the first three functions intrinsic modes using the autocorrelation is calculated; a comparison was made between the LPC analysis of the first three vowel of MFIs studied and the LPC analysis of the speech signal.

KEYWORDS

Empirical mode decomposition, intrinsinc mode function, LPC analysis.

1. INTRODUCTION

Spectral analysis is one of the most common farming methods in signal processing, the class of so-called parametric methods expired enables better data spectral estimation was based model for determining the parameters of the latter. A parametric analysis method is analyzed by linear predictive coding LPC. LPC is defined as a method of encoding digital signal to analog years everything that has a particular value is provided by a linear function of the past values of the signal.

Recently, a new temporal signal decomposition method called Empirical Mode Decomposition (EMD), has been introduced by Huang et al. [1] for processing data from nonstationary and nonlinear processes. The analysis is adaptive in contrast to traditional methods such as wavelets where the basic functions are fixed. The EMD has received more attention in terms of applications [2]-[3], interpretation [4]-[5], and improvement [6]-[7]. The major advantage of the EMD is that the basic functions are derived from the signal itself.

The EMD is also used in speech analysis. In this paper, we combine EMD with linear prediction coding analysis (LPC); we exploit the characteristics of the empirical modes from the EMD to study a new approach.

The remainder of the paper is organized as follows. Empirical mode decomposition (EMD) algorithm is defined in Section 2. EMD combined with LPC in Section 3. Results based on real speech signals are presented in Section 4. Finally, conclusions are given in Section 5.

2. EMPIRICAL MODE DECOMPOSITION

The empirical mode decomposition has been proposed by Huang et al. as a new signal decomposition method for nonlinear and/or nonstationary signals [1]. Conventional signal analysis tools, such as Fourier or wavelet-based methods, require some predefined basis functions to represent a signal. Therefore, the EMD can be viewed as sub-band signal decomposition. The EMD relies on a fully data-driven mechanism that does not require any a priori known basis. The EMD decomposes a given signal into a collection of oscillatory modes, called intrinsic mode functions (IMFs). Each IMF can be viewed as a sub-band of the signal and represent fast to slow oscillations in the signal. The algorithm operates through the following steps:

1. Initialize the algorithm: $j=1$, initialize residue $r_0(t)=x(t)$ and fix the threshold δ
2. Extract local maxima and minima of $r_{j-1}(t)$
3. Compute the upper lower envelope $U_j(t)$, $L_j(t)$ by cubic spline interpolation of local maxima and minima, respectively
4. Compute the mean envelope $m_j(t)=\frac{(U_j(t)+L_j(t))}{2}$
5. Compute the jth component $h_j(t)=r_{j-1}(t)-m_j(t)$
6. $h_j(t)$ is processed as $r_{j-1}(t)$. Let $h_{j,0}(t)=h_j(t)$ and $m_{j,k}(t)$ $k=0,1,\ldots\ldots$ be the mean envelope of $h_{j,k}(t)$, then compute $h_{j,k}(t)=h_{j,k-1}(t)-m_{j,k-1}(t)$ until $SD(i)=\sum_{t=0}^{T}\frac{|h_{j,i-1}(t)-h_{j,i}(t)|^2}{(h_{j,i-1}(t))^2}$
7. Compute the jth IMF as $IMF_j(t)=h_{j,k}(t)$
8. Update the residue $r_j(t)=r_{j-1}(t)-IMF_j(t)$
9. Increase the sifting index j and repeat steps 2 to 8 until the number of local extrema in $r_j(t)$ is less than 3

The signal reconstruction process x(t), which involves combining the IMFs formed from the EMD and the residual

$$x(t)=\sum_{j=1}^{N} IMF_j(t)+r_N(t)$$

3. EMD COMBINED WITH LPC

To a better exploitation of IMFs, we operate an LPC analysis of the first tree IMFs of empirical mode decomposition of speech signal, in order to analyze the forming characterizing the speech signal [8].

4. RESULTS AND DISCUSSION

To illustrate the effectiveness of the method we performed numerical simulations. The proposed approach has been tested on natural speech signals presented by three vowel / a /, / i /, / u /, and its performance in terms of accuracy has been compared to that of the LPC analysis of speech signal. The sampling rate of all speech signals used in the experiment is 11 kHz. The formants are defined as the ordered resonances of the vocal-tract, from the lowest to the highest. Figure 2 shows the different modes obtained from the empirical mode decomposition of the signal of the vowel / a /, presented in Figure 1 and the residue of the last algorithm step. And similarly for the vowels / i / and / u /, an EMD decomposition was performed. In our approach, we proceed to an LPC analysis of the IMFs represented in figure 3 and its comparison to results of the same analysis operated on speech signal. The results are depicted in figures 3, 4 and 5 for vowel / to / and / i / and / u /

respectively. We can see easily that each component has the same number of zero crossings as extrema and is symmetric with respect to zero line. We note that the first mode corresponds naturally to the highest frequency, and the last one corresponds to the lowest frequency, we compute an LPC analysis of the three first intrinsic mode functions using the autocorrelation method.

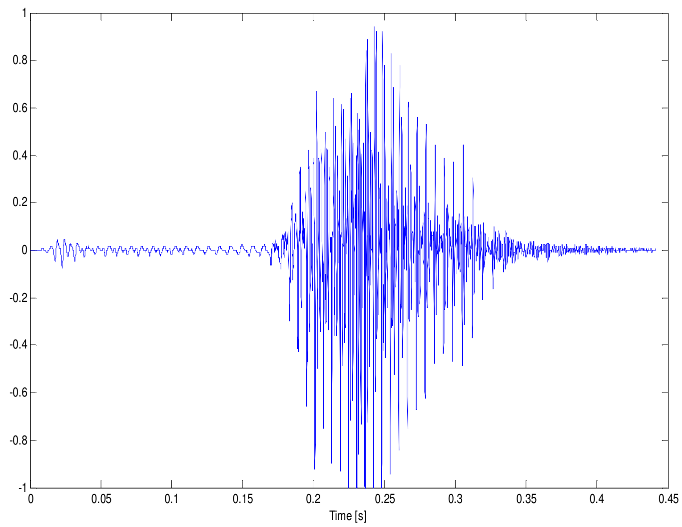

Figure 1. Waveform of the vowel /a/.

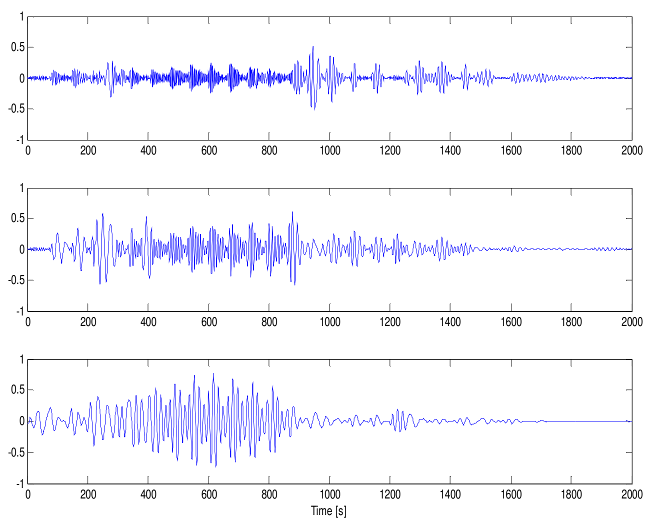

Figure 2. Waveform of the first tree IMFs of the vowel /a/.

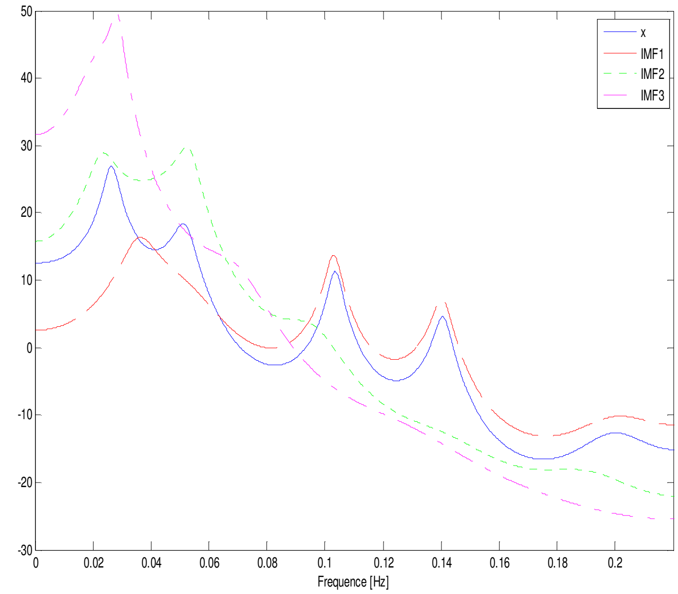

Figure 3. Comparison of the LPC analysis of the vowel / a / and the LPC analysis of the tree first IMFs

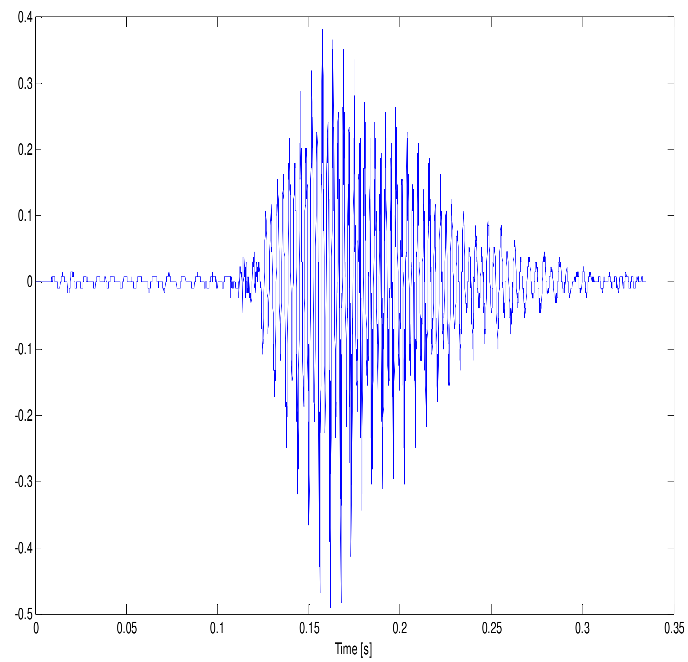

Figure 4. Waveform of the vowel /i/.

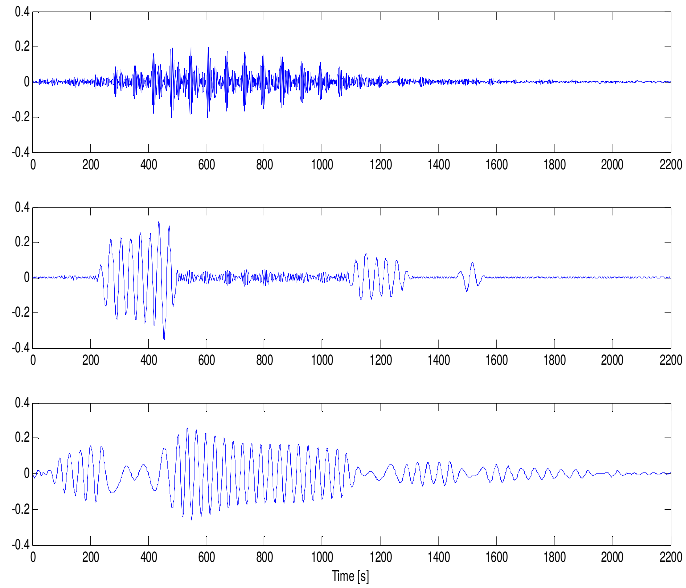

Figure 5. Waveform of the first tree IMFs of the vowel /i/.

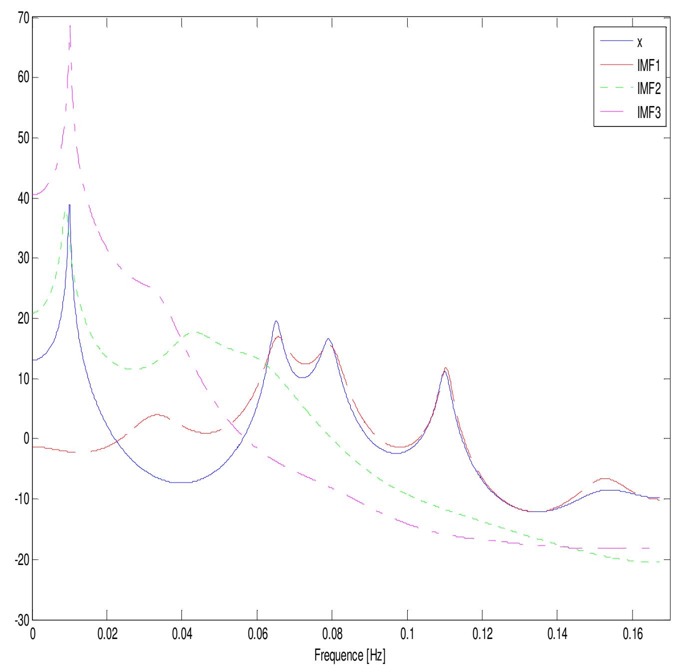

Figure 6. Comparison of the LPC analysis of the vowel / i / and the LPC analysis of the tree first IMFs

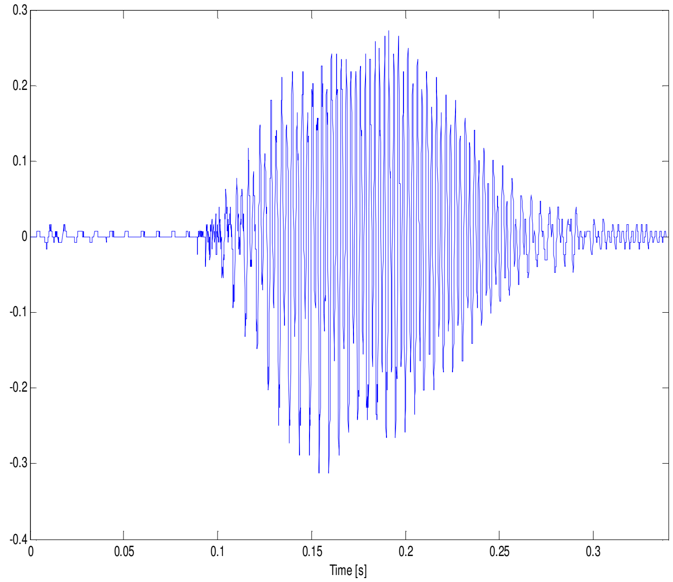

Figure 4. Waveform of the vowel /u/.

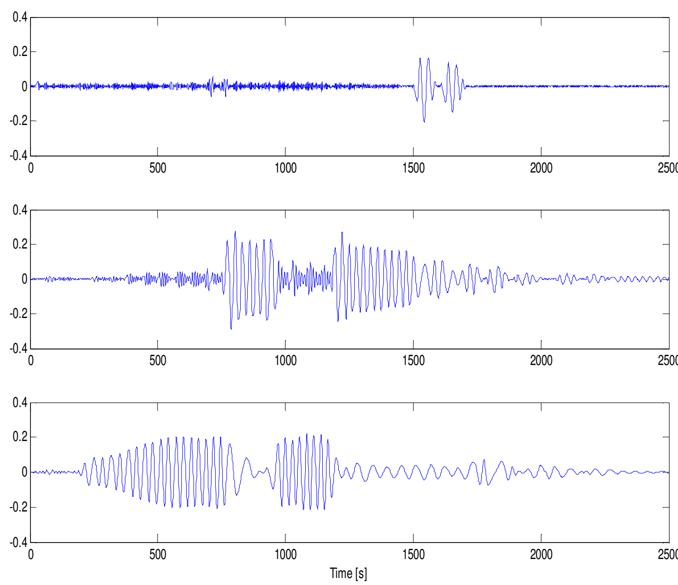

Figure 8. Waveform of the first tree IMFs of the vowel /u/.

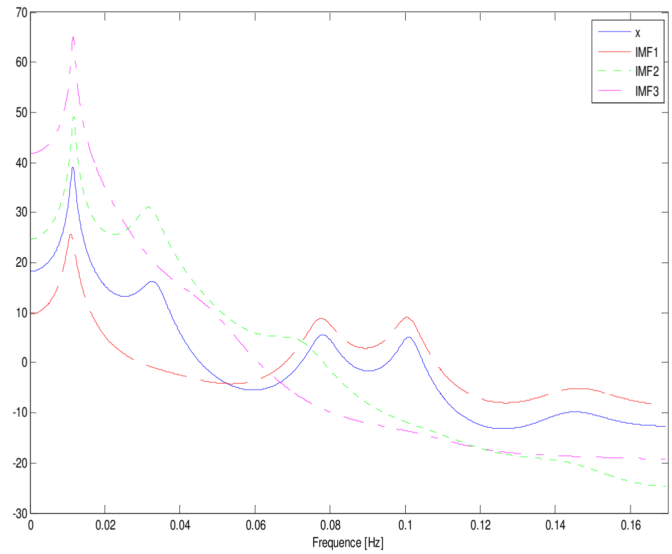

Figure 9. Comparison of the LPC analysis of the vowel / u / and the LPC analysis of the tree first IMFs

As shown by Figure 3, and figure 4 and figure 5, the first IMFs have not submitted the low frequency signals but only the high frequencies. These results can be interpreted by the frequency response of an equivalent filter while the collection of filters resulting equivalents of any of the LPC analysis of IMFs given vowel tends the estimation of formants of voice leading. Although EMD is a non-linear decomposition method, but the formants of the speech signal are preserved and properly evaluated as can be seen in table 1 and 2 and 3 which respectively presents a comparison between the values of formants obtained from the LPC analysis of the vowel / a /, and the vowel / i /, and the vowel / u / and the analysis LPC of their three first IMFs.

Table 1. Comparison between the value of formants obtained by LPC analysis of the vowel / a / and the LPC analysis of these first three IMFs.

	F1	F2	F3	F4
Speech /a/	0.002639	0.05116	0.1035	0.1406
IMF1	-	-	0.1028	0.1403
IMF2	-	0.05224	-	-
IMF3	0.02785	-	-	-

Table 2. Comparison between the value of formants obtained by LPC analysis of the vowel / i / and the LPC analysis of these first three IMFs.

	F1	F2	F3	F4
Speech /a/	0.01188	0.08082	0.1098	0.1331
IMF1	-	-	0.1098	0.1345
IMF2	-	0.07869	-	-
IMF3	0.01152	-	-	-

	F1	F2	F3	F4
Speech /a/	0.01469	0.09805	0.1456	0.1909
IMF1	-	-	0.1452	0.1899
IMF2	-	0.09506	-	-
IMF3	0.1678	-	-	-

5. CONCLUSIONS

In this work, we have proposed a new methodology to decompose a speech signal into different oscillatory modes and to extract the resonant frequencies of the vocal tract i.e. formants from the LPC analysis of different intrinsic mode functions called IMFs. The combination of EMD with the LPC method allows the extraction and proper evaluation of different formants characterize different signals studied.

REFERENCES

[1] N.E. Huang and al. “The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis”. Proc. Royal Society, 1998,454(1971):903–995.

[2] F. Salzenstein A.O. Boudraa, J.C. Cexus and L. Guillon. “ If estimation using empirical mode decomposition and nonlinear teager energy operator”. Proc. IEEE ISCCSP, Hammamet, 2004.,pp 45–48.

[3] A.O. Boudraa S. Benramdane, J.C. Cexus and J.A. Astolfi. “Transient turbulent pressure signal processing using empirical mode decomposition”. Proc. Physics in Signal and Image Processing, Mhoulouse, 2007.

[4] P. Flandrin, G. Rilling, and P. Goncalves. “Empirical mode decomposition as a filter bank”. IEEE Sig. Proc. Lett., 11(2):112–114, 2004.

[5] Z. Wu and N.E. Huang. “ A study of the characteristics of white noise using the empirical mode decomposition method”. Proc. Roy. Soc. London A, 2004,460:1597–1611.

[6] B. Weng and K.E. Barner. “Optimal and bidirectional optimal empirical mode decomposition”. Proc. IEEE ICASSPToulouse, 2007, 3:1501–1504.

[7] R. Deering and J.F. Kaiser. “The use of a masking signal to improve empirical mode decomposition”. Proc. IEEE Philadelphia, 2005.

[8] Ai’cha Bouzid and Noureddine Ellouze. “voiced speech analysis empirical mode decomposition”. Springer 2007.pp 213-220.