EMD Analysis of Speech Signal in Voiced Mode

Aïcha Bouzid[1], Noureddine Ellouze[2]

[1]Institute of Electronic and Communication of Sfax, Sfax, Tunisie
[2]National School of Engineers of Tunis, Tunis, Tunisie
N.Ellouze@enit.rnu.tn, bouzidacha@yahoo.fr

Abstract

Like almost all natural phenomena, speech is the result of many nonlinearly interacting processes; therefore any linear analysis has the potential risk of underestimating, or even missing, a great amount of information content. Recently the technique of Empirical Mode Decomposition (EMD) has been proposed as a new tool for the analysis for nonlinear and nonstationary data. We applied EMD analysis to decompose speech signal into intrinsic oscillatory modes. Besides, the LPC analysis of each mode provides an estimation of formants.

1. Introduction

Speech signal, as with many real-world signals, are nonstationary, making Fourier analysis unsatisfying since the frequency contents changes across the time. In time-frequency analysis, we analyse the frequency content across a small span of time and then move to another time position [1] and [2]. The major drawback of most time- frequency transforms is that the rectangular tiling of the time frequency plane does not match the shape of many signals.

On the other hand, basis decomposition techniques such as Fourier decomposition or the wavelet decomposition have also been used to analyse real world signals [3]. The main drawback of these approaches is that the basis functions are fixed, and do not necessarily match varying nature of signals.

In this paper, we use the empirical mode decomposition (EMD), first introduced by N. E. Huang and al. in 1998 [4]. This technique adaptively decomposes a signal into oscillating components. The different components match the signal itself very well. Because the approach is algorithmic, it does not allow expressing the different components in closed form. The EMD is in fact type of adaptive wavelet decomposition whose sub bands are built as needed to separate the different components of the signal.

EMD was applied to a number of real situations [5], [6] and [7], motivating us to consider work on naturally speech decomposition in order to delimit EMD limitations and possibilities.

The out line of the present paper is as follows. Firstly we introduce the new non linear decomposition technique known as empirical mode decomposition. Then we apply this technique to decompose a simple signal consisting of a sum of three pure frequencies. The second section presents results of this approach applied to speech signal decomposition. Computing the LPC analysis of different intrinsic mode functions provides measure of formant speaker. Last section concludes this work.

2. Empirical mode decomposition

The empirical mode decomposition is a signal processing technique proposed to extract all the oscillatory modes embedded in a signal without any requirement of stationarity or linearity of the data. The goal of this procedure is to decompose a time series into components with well defined instantaneous frequency by empirically identifying the physical time scales intrinsic to the data that is the time lapse between successive extrema [8].

Each characteristic oscillatory mode extracted, named Intrinsic Mode Function (IMF), and satisfies the following properties: an IMF is symmetric, has unique local frequency, and different IMFs do not exihibit the same frequency at the same time. In other words the IMFs are characterized by having the number of extrema and the number of zero crossings equal (or different at most by one), and the mean value between the upper and lower envelope equal to zero at any point.

The algorithm operates through six steps [4]:

1) Identification of all the extrema (maxima and minima) of the series x(t).
2) Generation of the upper and lower envelope via cubic spline interpolation among all the maxima and minima, respectively.
3) Point by point averaging of the two envelopes to compute a local mean series m(t).
4) Subtraction of m(t) from the data to obtain a IMF candidate d(t)=x(t)-m(t).
5) Check the properties of d(t):
    - If d is not a IMF (i.e it does not satisfy the previously defined properties), replace x(t) with d(t) and repeat the procedure from step 1
    - If d is a IMF, evaluate the residue m(t)=x(t)-d(t)

Repeat the procedure from step 1 to step 5 by sifting the residual signal.

The sifting process ends when the residue satisfies a predefined stopping criterion.

By construction, the number of extrema is decreased when going from one residual to the next (thus guaranteeing that the complete decomposition is achieved in a finite number of steps), and the corresponding spectral supports are expected to decrease accordingly. Selection of modes rather corresponds to an automatic and adaptative (signal dependent) time variant filtering [9] and [10].

At the end of the algorithm, we have:

$$x(t) = \sum_{i=1}^{n} d_i(t) + m_n(t) \tag{1}$$

where $m_n(t)$ is the residue and d_i is the intrinsic mode function at mode i that has the same numbers of zero crossing and extrema; and is symmetric with respect to the local mean.

Another way to explain how the empirical mode decomposition works is that it picks out the highest frequency oscillation that remains in the signal. Thus, locally, each IMF contains lower frequency oscillations than the one extracted just before. This property can be very useful to pick up frequency changes, since a change will appear even more clearly at the level of a IMF [5].

Figure 1 shows the starting point of signal decomposition and the IMF candidate obtained after little iteration.

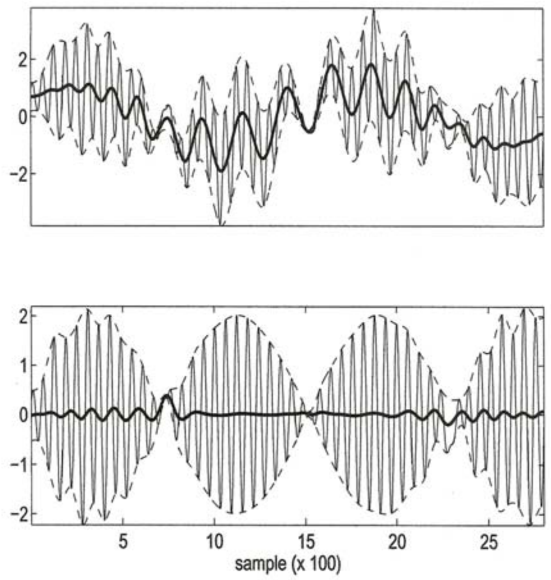

Figure 1: At the top: the original signal with upper and lower envelope. The thick line represents the point by point mean value of the envelopes. Below: the signal d after little iteration. The iteration continue until becomes IMF.

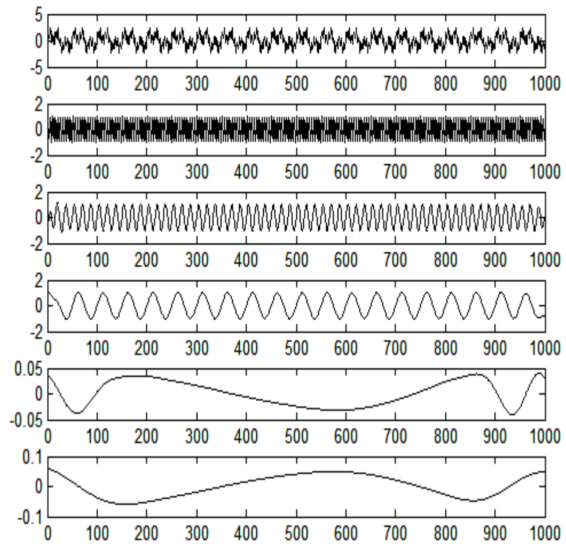

Figure 2: Decomposition of sum of 3 sinus signal of frequency 100Hz, 300 Hz and 1000 Hz by 5 first IMFs.

Figure 2, shows a signal which is the sum of three pure frequencies having the following frequencies: 100 Hz, 300 Hz and 1000 Hz, and the five IMFs followed by the residue. The signal is composed by 1000 samples with a sampling frequency of 20 kHz.

We can see 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 which corresponds naturally to the highest frequency shows clearly the 1 kHz frequency present in the signal. Consequently the second mode depicts 300Hz frequency and the third one corresponds to the lowest frequency i.e. 100Hz.

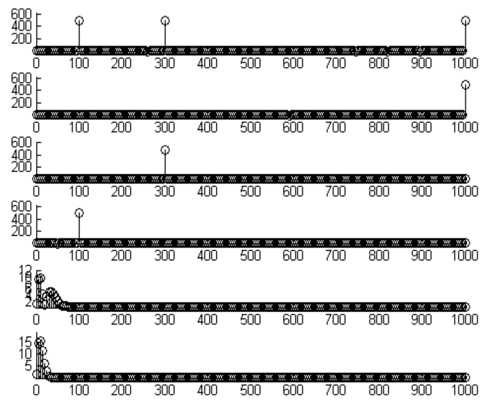

Figure 3: Spectral analysis of composite signal (frequency 100Hz, 300 Hz and 1000 Hz) and its 5 first IMFs.

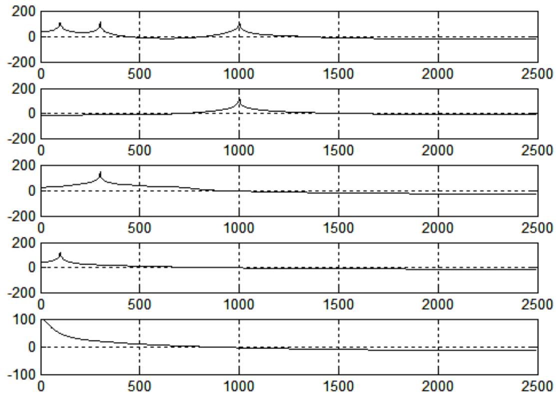

Figure 4: LPC analysis of composite signal and its IMFs

Fourier analysis of the composite signal and its IMFs as depicted in Figure 3, shows that the highest frequency is identified from the first IMF and the lowest one is given by the third IMF.

We compute also the LPC analysis of signal and the 3 IMFs using autocorrelation method. As expected, LPC analysis shows peaks at 1 kHz, 300 Hz and 100 Hz.

This analysis demonstrates once again the efficiency of the proposed method in decomposing the signal in spectral domain [12]. The proposed decomposition detects all frequencies constituting the signal separetly.

The EMD procedure, according to the above specifications, is used in the next section for the decomposition of speech signal issued from Keele database as described in the next section.

3. EMD analysis of voiced speech signal

In the previous section we illustrate the efficiency of the empirical mode decomposition of a typical signal which is the sum of pure frequencies in detecting these frequencies. This approach is used to decompose the speech signal in order to analyze its formant frequencies.

We take as an example of speech signal, a vowel /o/ pronounced by a female speaker f1, extracted from the Keele University database and sampled at 20 kHz. Figure 5 shows the different modes obtained from the empirical mode decomposition of the vowel /o/ and the residue of the last algorithm step.

In our approach, we proceed to an LPC analysis of the IMFs represented in figure 5 and its comparison to results of the same analysis operated on speech signal. The results are depicted in figures 6 and 7.

The LPC analysis achieved for the first IMF shows a curve that fits approximately curve corresponding to speech signal but the peaks for IMF are sharper. The first analyzed IMF doesn't depict the low frequency composition of the signal. In fact it concerns the highest frequency.

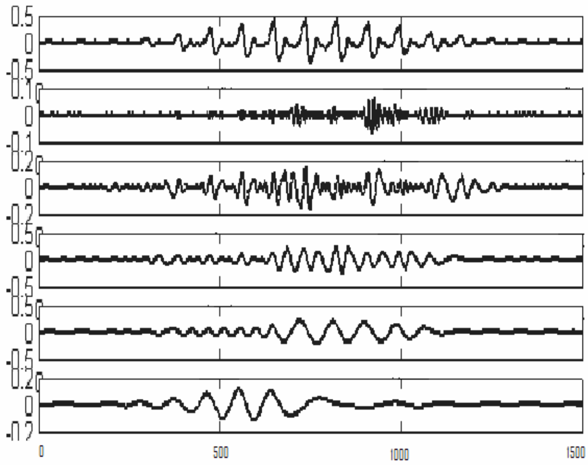

Figure 5: Illustration of the EMD: vowel /o/ speaker f1 and first five IMFs

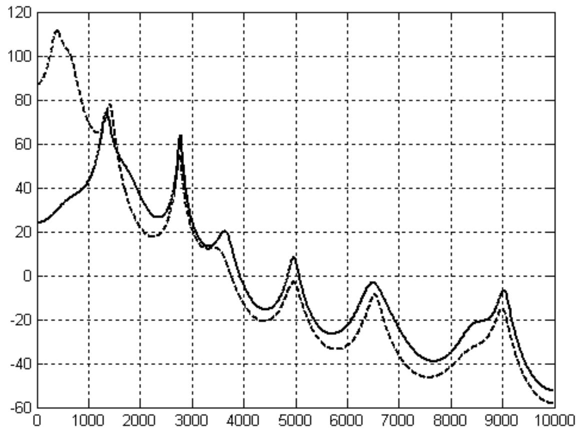

Figure 6: LPC analysis of vowel /o/ speaker f1 (dashed line) and of the first signal's IMF (solid line).

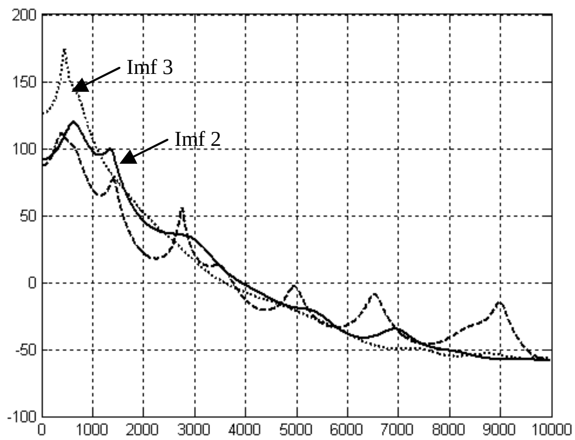

Figure 7: LPC analysis of vowel /o/ speaker f1 and IMFs of the signal (second and third).

This result, mode by mode, in a frequency profile can be interpreted as the frequency response of some equivalent filter. As evidenced in figures 6 and 7, the collection of all such filters tend to estimate the different resonant frequencies of the vocal tract.

An other example for speech signal is given to emphasize the efficiency of this method. It's about a vowel /a/ expressed by a male speaker m2. The achieved EMD is depicted in figure 8.

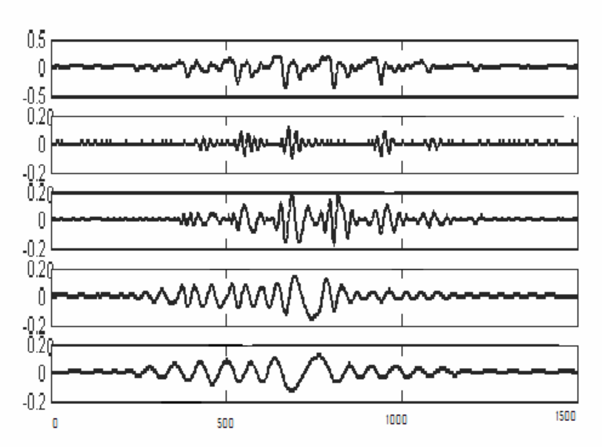

Figure 8: Illustration of the EMD: vowel /a/ speaker m2 and the different IMFs.

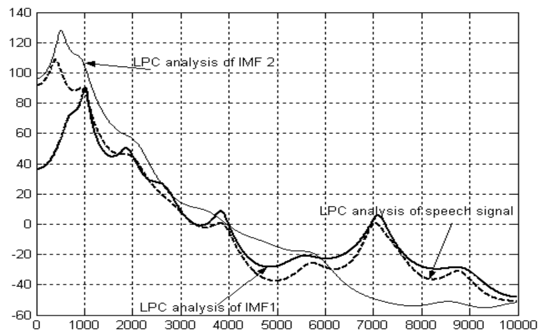

Figure 9: LPC analysis vowel /a/ speaker m2 and the two first IMFs. Solid line, concerns LPC analysis of the first IMF, the dashed line represents vowel /a/ of speaker m2, and the third curve concerns the second IMF.

Figure 9 depicts the LPC analysis of the corresponding speech signal and its 3 first IMFs. We note that peaks given by IMFs are more distinguishable than those releated to speech.

4. Conclusion

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. LPC analysis of IMFs shows the frequency components.

If we represent all the LPC analysis, we may obtain a complete description of the speech production model. A, study of the residue can be considered and compared to the frequency representation of the glottal pulse.

Besides, we can look for a new time-frequency attributes obtained from the EMD analysis and based on an instantaneous frequency calculation of each component of the decomposition.

5. References

[1] P. Flandrin, Temps-fréquence, Hermes, 1993.

[2] L. Cohen, Time-frequency analysis, Prentice Hall, 1995.

[3] I. Daubechies, Ten lectures on wavelet, SIAM, 1992.

[4] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. Roy. Soc. London A, Vol. 454, pp. 903-995,1998.

[5] I. M. Chagnollean and R. G. Baranniuk, Empirical mode decomposition based time-frequency attributes, 69th SEG Meeting, Houston, United States, November 1999.

[6] R. Fournier, Analyse stochastique modale du signal stabilométrique. Application à l'étude de l'équilibre chez l'Homme, Thèse de Doctorat, Univ. Paris XII Val de Marne, 2002.

[7] E. P. Sousa et al., Assessment of cardiovascular autonomic control by the empirical mode decomposition , 4th Int. Workshop on Biosignal Interpretation, Como (I), pp. 123-126, 2002.

[8] R. Balocchi, D. Menicucci, E. Santarcangelo, L. Sebastiani, A. Gemignani, B. Ghelarducci, and M. Varanini, Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode Decomposition, Qauntitative Biology, October 2003.

[9] P. Flandrin et P. Gonçalvès, sur la decomposition modale empirique, GRETSI 2003, Paris, Semptembre 2003.

[10] G. Rilling, P. Flandrin and P. Gonçalvès, On empirical mode decomposition and its algorithms, IEEE-EURASIP Workshop on nonlinear signal and image processing NSIP-03, Grado(I), 2003.

[11] J. C. Nunes, O. Niang, Y. Bouaoune, E. Deléchelle and Ph. Bunel, Décomposition empirique multimodale bidimensionnelle modifiée pour l'analyse d'images, GRETSI 2003, Paris, Septembre 2003.

[12] P. Flandrin, G. Rilling and P. Gonçalvès, Empirical mode decomposition as a filter bank, IEEE signal processing letters, Vol.11, No.2: pp. 112-114.