Performance Investigation of Empirical Mode Decomposition in Biomedical Signals

Alexandros Karagiannis and Philip Constantinou

Mobile RadioCommunications Lab, National Technical University of Athens
Iroon Polytechneiou 9, Athens, Greece
{akarag,fkonst}@mobile.ntua.gr

Abstract. In this paper, the performance of Empirical Mode Decomposition (EMD) applied in biomedical signals is investigated and especially it is considered the case of electrocardiogram (ECG). Synthetic ECG signals corrupted with White Gaussian Noise (WGN) as well as real ECG records are employed and a variety of time series lengths is processed with EMD in order to extract the Intrinsic Mode Functions (IMF). Computation time is measured upon the completion of the process in simulation campaign stage and real records stage and the results are compared in both cases. Spectral characteristics of the time series as well as the tendency to exhibit extrema are the key factors with significant impact on both computation time as well as the total number of IMFs produced.

Keywords: Empirical Mode Decomposition, ECG, Intrinsic Mode Functions, extrema, time of computation.

1 Introduction

Extraction of meaningful information derived from signals corrupted by noise is accomplished with the utilization of data analysis methods. Many traditional methods for processing are employed under the assumption of stationarity of the sampled signal and the linearity of the physical process that produces these signals.

Fourier transform is a widely used technique, albeit ineffective in processing non-stationary data and signals originating from nonlinear systems, because of the properties of the basis functions that are incapable of following spectral changes in time. Wavelet analysis [1,2] and the Wigner-Ville distribution [3] are designed for non-stationary data extracted out of linear systems with intrinsic limitations in characterizing in a detailed fashion the time-frequency composition of signals from nonlinear processes.

A joint function in both time and frequency domains is considered to be an efficient approach towards overcoming the limitations of the traditional techniques [4]. However, Short-time Fourier transform [5], a characteristic tool for time-frequency representation, has fixed time and frequency resolution in order to follow signals with bursts and quasi-stationary components. Research on the field of time-frequency representations is ongoing and fruitful in terms of number of different methods that

J. Lin and K.S. Nikita (Eds.): MobiHealth 2010, LNICST 55, pp. 117–124, 2011.

tackle with certain limitations. Still most of the existing nonlinear time series analysis methods refer to stationary signals [6] and a necessary condition for the processing of nonlinear and nonstationary data, namely the adaptive basis, is often underestimated.

Most of the biosignals are related to the dynamic biological systems ruled by nonlinear equations and they are considered to be nonlinear and nonstationary. Dealing with nonlinearity and nonstationarity requires an adaptive nature of the processing method. A priori defined sophisticated basis functions face difficulties in meeting the requirement of adaptation which a data driven basis function implies.

A recently proposed method, the Hilbert-Huang Transform (HHT) [7], satisfies the condition of adaptation employed in processing of nonlinear and nonstationary data. The HHT consists of EMD and Hilbert Spectral Analysis (HSA) [8]. The lack of mathematical foundation and analytical expressions poses a problem for the theoretical study of the method. Nevertheless there has been an exhaustive validation in an empirical fashion especially in the time-frequency representations [9].

The core of the method is Empirical Mode Decomposition which resolves a signal into its components adaptively without using an a priori basis. The decomposition is based on the local time scale of data. The adaptive nature of the process successfully decomposes time series from nonlinear processes and nonstationary signals in the time domain. Each component extracted from the original signal through an iterative and sifting process is required to satisfy certain conditions in order to be characterized as IMF.

Application of EMD in time series data results in the production of a set of IMFs and a residual signal. The notion behind this procedure is that a subset of the IMFs is directly related to the underlying physical process. The decomposition is based on the assumption that data consists of different simple intrinsic modes of oscillations and after the production of the IMF set, well-behaved Hilbert transforms compute physically meaningful instantaneous frequencies, thus constructing the Hilbert spectrum.

Unlike wavelet processing, Hilbert-Huang transform decomposes a signal by direct extraction of the local energy associated with the time scales of the signal. This feature reveals the applicability of HHT in both nonstationary and nonlinear signals.

Literature references' variety reveals the extensive range of EMD applications in several areas of the biomedical engineering field. Particularly there are publications concerning the application of EMD in the study of Heart Rate Variability (HRV) [10], analysis of respiratory mechanomyographic signals [11], ECG enhancement artifact and baseline wander correction [12], R-peak detection [13], Crackle sound analysis in lung sounds [14] and enhancement of cardiotocograph signals [15]. The method is employed for filtering electromyographic (EMG) signals in order to perform attenuation of the incorporated background activity [16]. Numerous research papers have been published concerning applications of EMD in biomedical signals and especially towards the direction of optimizing traditional techniques of acquisition and processing of signals such as Doppler ultrasound for the removal of artifacts [17], the analysis of complex time series such as human heartbeat interval [18], the identification of noise components in ECG time series [19] and the denoising of respiratory signals [20].

The contribution of this work is twofold. First, it proposes a mixed scheme based on EMD with an additional pre-processing stage, prior to the EMD application, to be defined according to signal's special characteristics. In this paper, both simulated and experimental ECG time series are employed and some types of filters are used as a filtering stage. Number of IMFs is measured for both filtered and non filtered cases in order to comparatively evaluate the introduction of the pre-processing stage in terms of number of extracted IMFs. Secondly, this work investigates the role of basic parameters such as time series length and signal to noise ratio (SNR) involved at the determination of the total computation time.

2 Synthetic and Experimental ECG Time Series

Synthetic electrocardiogram time series (Fig. 1) are produced artificially by a (software) generating function. The amplitude is expressed in normalized voltage values, taking into account the magnitude scales of the various complexes from experimental time series. For comparison purposes MIT-BIH ECG signals [21] are used as reference signals.

Simulated ECG time series are corrupted by White Gaussian Noise in order to produce noisy ones in multiple SNR levels from 0dB to 35dB. Data length varies from 500 to 8000 samples in numerical experiments where simulated noisy ECG time series are incorporated.

Two series of tests are conducted targeting to measure the total number of IMFs as a function of time series length and SNR in simulated ECG time series and the total computation time after the application of EMD until the completion of the whole process. A small sample incremental step is chosen for simulations targeted to evaluate the impact of time series length on the number of extracted IMFs. On the other hand, the maximum number of samples for the process of estimating the computation time in respect to time series length and SNR is 8000 and the step is 1000.

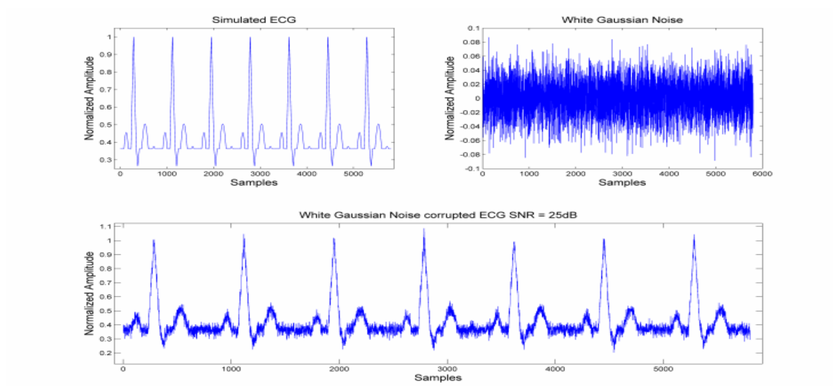


Fig. 1. Simulated ECG and addition of white Gaussian noise in 25dB SNR

In the case of real MIT-BIH ECG time series, the SNR as independent variable is not considered since these recordings already contain noisy components which are generally uncontrollable and unknown. However, time series lengths are selected without any restrictions sharing a common range with the simulated ones.

3 Results

3.1 Number of IMFs as a Function of Time Series Length and SNR

The study for the effect of time series length on the total number of IMFs is carried out comparatively by employing two widely used ECG filters. The variable of SNR as the second independent variable is determined by the process of noise addition to the simulated ECG time series and the production of the White Gaussian noisy simulated ECG time series. The composite impact study of time series length and SNR into the total number of IMFs reveal the indirect relation of noise level with the outcome of EMD method through the determination of the extrema and the distribution of them at the time series. An intrinsic mode function set is considered to be related at least for a subset with the underlying physical process that produces the biomedical signal. Application of Hilbert Transform on the IMFs set delineates the instantaneous frequencies and forms a time-frequency distribution of the signal for each IMF. The number of the IMF set is critical in the sense that any procedure applied on the time series may distort the signal characteristics resulting in a different IMF set. Currently, there is no robust methodology (optimum algorithm thresholds or filter's characteristics) in order to produce the optimum IMF set in terms of richest physical meaning or undistorted instantaneuous frequencies.

In the first case, a Savitzky Golay filter is incorporated due to the well known behavior in the preservation of the time series extrema. Various synthetic ECG time series of multiple lengths are processed with EMD and the total number of IMFs is monitored. Secondly, a low pass filter with frequency cutoff equal to 40Hz [22] is employed which tends to affect high frequency content of the time series and the extracted IMFs exhibit attenuated power in higher frequencies, resulting in a degraded peaky nature compared to the unfiltered one. Distortion in the tendency of the time series to exhibit peaks results in a significant impact on the total number of IMFs produced at a wide range of SNR levels. Low pass filters are used in electro cardiology to limit artifact for routine cardiac rhythm monitoring and reduce 50 or 60 Hz power line noise.

Figure 2 and 3 depict the results of the application of EMD on White Gaussian noisy simulated ECG time series of various lengths and SNR levels compared with the non application case.

Concerning the Savitzky-Golay case, the filter seems to have minimum effect on the filtered and non filtered case in various SNR levels in terms of number of extracted IMFs. Actually, these curves indicate the good performance of this filtering method in preserving the extremas in ECG signals and stress the high sensitivity of interpolation techniques implemented in the EMD algorithm in terms of accurate detection of peaky values.

3.2 Computation Time

The computation time for the EMD processing of simulated and real records ECG time series is related to the nature of the processed signal and the tendency to exhibit extremas. Apart from the number of extremas, another significant parameter is considered to be the distribution of the peaks in the whole range of time series length and inside the various ECG complexes.

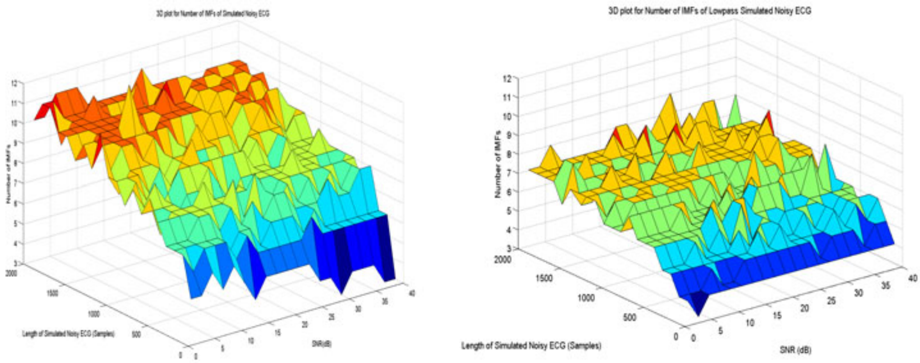


Fig. 2. 3D plots of the number of IMFs as a function of the SNR and the time series length for a simulated White Gaussian Noise corrupted ECG without the filtering with lowpass filter (left figure) and at the right figure with the application of the lowpass filter

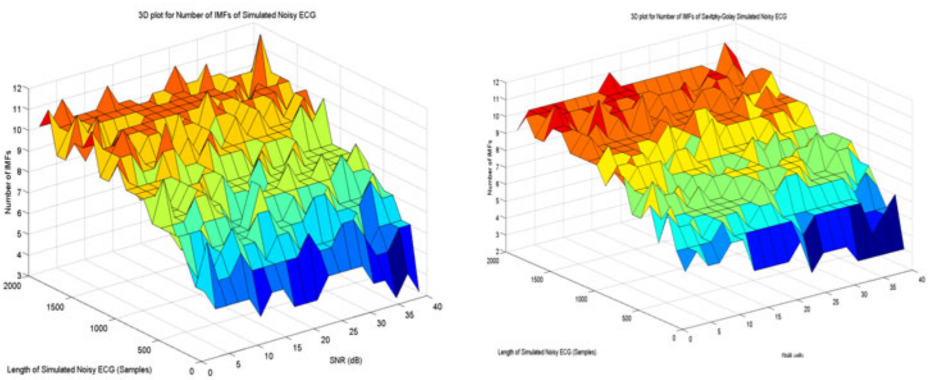


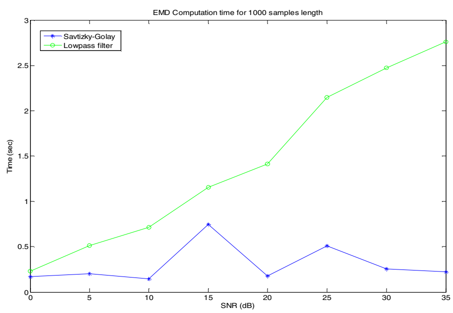
Fig. 3. 3D plots of the number of IMFs as a function of the SNR and the time series length for a simulated White Gaussian Noise corrupted ECG without filtering with Savitzky-Golay (left figure) and at the right figure with the filtering of Savitzky-Golay

An indirect way to estimate computation time is through the number of extracted IMFs. Specifically, computation time is monotonically increasing as the IMF set is growing and the fashion of this relation is quasi proportional. It is also affected by the total number of iterations required for the extraction of the IMF set. This goes down to implementation issues concerning the EMD algorithm and the thresholds used in termination criterion and even in the maximum number of iterations allowed.

For the simulated ECG time series, the length and SNR are controlled independent variables whilst for the experimental ECG the level of noise superimposed in the signal is generally unknown. The independent variable in real ECG records is the time series length.

Multiple time series lengths of simulated ECG are studied providing an overview of the computation time variation as a function of the length for various filtered time series. For demonstration reasons the minimum and maximum number of samples (1000, 8000) are depicted in figure 4. Due to the improvement of signal to noise ratio, noise levels are decreased and the number of noise samples superimposed to ECG time series is reduced. Spline interpolation scheme produces less uniform waveforms for the EMD envelopes resulting in significant increase of iterations required for the IMFs decomposition. The increase of computation time in 8000 samples time series with the increase of SNR is significant in lowpass fitltered time series compared to the corresponding increase in computation times monitored in Savitzky-Golay filtered time series. The combination of two components, the SNR variation and the impact of filtering in the peaky nature of the time series results in the composite behavior of the time series depicted in figure 4.

In figure 5, the variation of computation time in real ECG records is presented as a function of length along with the standard deviation. The pattern of the variation in computation time is dependent on the specific characteristic of the record sample (possible pathologies or ECG distortions) and its spectral characteristics. There is a tendency of computation time to increase as the time series length increases but the pattern of this tendency is not common in every ECG record.



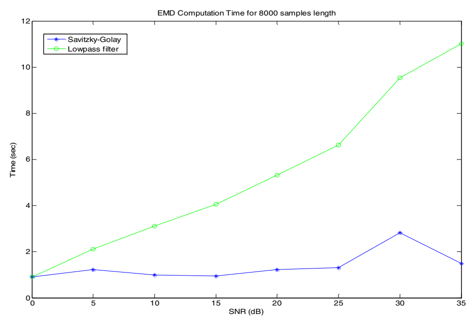


Fig. 4. Comparison results of EMD Computation Time for 1000 and 8000 samples of Simulated ECG time series

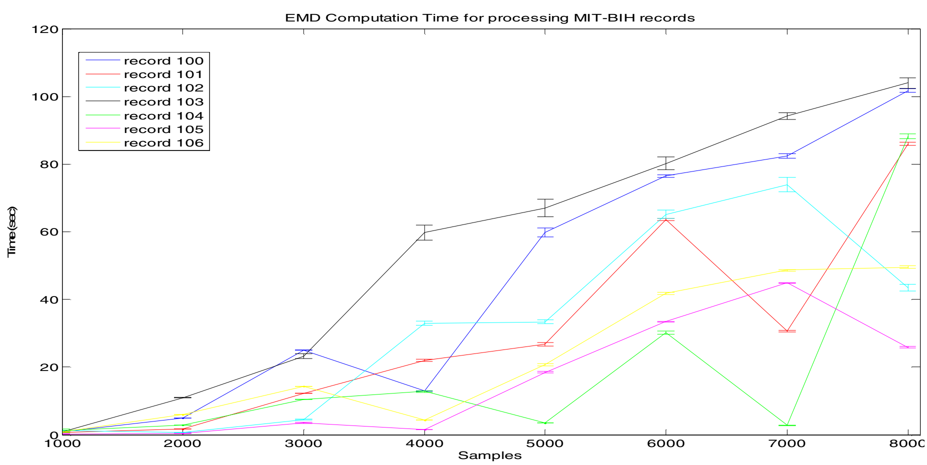


Fig. 5. EMD Computation Time in experimental MIT-BIH ECG time series for various time series lengths

4 Discussion – Conclusions

The introduction of a filtering stage before EMD application is studied and the effects of this scheme are investigated. Simulation results reveal that according to the type of filtering performed on data sets, variations are observed in the total number of IMFs reflecting to changes in total processing time as well. Filtering stage affects the spectral characteristics of the input signal and distortion of the time series' statistical and spectral content have an effect in the performance of EMD algorithm. Based on the inherent properties of the time series to be processed, one may select an appropriate pre-processing stage in order to achieve smaller number of IMFs and optimized processing time without changing in a significant degree the physical content of IMFs.

Total computation time is an essential aspect in transferring EMD algorithm from units with adequate processing power to embedded level in an efficient programming way. In multiple time series lengths, it is observed that computation time is monotonically increasing with the increase of SNR values. It is expected that the time series length is another significant parameter with a quasi proportional relation to the total computation time in simulation approach. However, this is partially verified in experimental ECG time series due to the involvement of various factors such as pathological situations in the signal and artifacts that seriously affect the application of EMD algorithm. Nevertheless, the increase tendency of computation time with the increase of length still exists.

EMD implementation takes into account the termination criterion, a significant parameter to be optimized in order to avoid numerous iterations for the extraction of IMFs. Research effort is still to be undertaken to investigate in what degree tight restrictions in number of iterations drain the physical content of IMFs. An optimization procedure for both termination criterion and number of iterations is an open issue in this field.

In low SNR levels, noise is prevalent in the time series, reflecting in a smoother spline and generally faster extraction due to smaller number of iterations. In high SNR values, it is observed a clear tendency towards the increase of computation time raising the issue of the magnitude of noise to be added in the signal in noise assisted data analysis methods.

References

1. Daubechies, I.: Ten Lectures on Wavelets. CBMS-NSF Series in Applied Mathematics 61, 357 (1992)
2. Mallat, S.: A Wavelet Tour of Signal Processing, p. 637. Academic Press, London (1998)
3. Wigner, E.P.: On the quantum correction for thermodynamic equilibrium. Phys. Rev. 40, 749–759 (1932)
4. Cohen, L.: Time-Frequency Analysis, p. 299. Prentice Hall, Englewood Cliffs (1995)
5. Gabor, D.: Theory of Communications. J. IEE 93, 429–457 (1946)
6. Diks, C.: Nonlinear Time Series Analysis, p. 220. World Scientific Press, Singapore (1999)

7. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc. R. Soc. London 454, 903–995 (1998)
8. Hahn, S.: Hilbert Transforms in Signal Processing, p. 442. Artech House, Boston (1995)
9. Huang, N.E., Wu, M.C., Long, S.R., Shen, S.S.P., Qu, W., Gloersen, P., Fan, K.L.: A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proc. R. Soc. A 459, 2317–2345 (2003), doi:10.1098/rspa.2003.1123
10. Echeverría, J.C., Crowe, J.A., Woolfson, M.S., Hayes-Gill, B.R.: Application of empirical mode decomposition to heart rate variability analysis. Med. Biol. Eng. Comput. 39(4), 471–479
11. Torres, A., Fiz, J.A., Jané, R., Galdiz, J.B., Gea, J., Morera, J.: Application of the Empirical Mode Decomposition method to the Analysis of Respiratory Mechanomyographic Signals. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France
12. Blanco-Velasco, M., Weng, B., Barner, K.E.: ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Comput. Biol. Med. 38(1), 1–13 (2008)
13. Nimunkar, A.J., Tompkins, W.J.: R-peak detection and signal averaging for simulated stress ECG using EMD. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 1261–1264 (2007)
14. Charleston-Villalobos, S., Gonzalez-Camarena, R., Chi-Lem, G., Aljama-Corrales, T.: Crackle Sounds Analysis by Empirical Mode Decomposition. Engineering in Medicine and Biology Magazine 26(1), 40–47 (2007)
15. Krupa, B.N., Mohd Ali, M.A., Zahedi, E.: The application of empirical mode decomposition for the enhancement of cardiotocograph signals. Physiol. Meas. 30, 729–743 (2009)
16. Andrade, A.O., Nasuto, V., Kyberd, P., Sweeney-Reed, C.M., Kanijn, F.R.V.: EMG signal filtering based on Empirical Mode Decomposition. Biomedical Signal Processing and Control 1(1), 44–55 (2006), doi:10.1016/j.bspc.2006.03.003
17. Zhang, Y., Gao, Y., Wang, L., Chen, J., Shi, X.: The removal of wall components in Doppler ultrasound signals by using the empirical mode decomposition algorithm. IEEE Trans. Biomed. Eng. 54(9), 1631–1642 (2007)
18. Yeh, J.R., Sun, W.Z., Shieh, J.S., Huang, N.E.: Intrinsic mode analysis of human heartbeat time series. Ann. Biomed. Eng. 38(4), 1337–1344 (2010)
19. Karagiannis, A., Constantinou, P.: Noise components identification in biomedical signals based on Empirical Mode Decomposition. In: 9th International Conference on Information Technology and Applications in Biomedicine, ITAB 2009 (2009), 10.1109/ITAB.2009.5394300
20. Karagiannis, A., Loizou, L., Constantinou, P.: Experimental respiratory signal analysis based on Empirical Mode Decomposition. In: First International Symposium on Applied Sciences on Biomedical and Communication Technologies, ISABEL 2008 (2008), 10.1109/ISABEL.2008.4712581
21. `http://www.physionet.org/physiobank/database/mitdb`
22. Mark, J.B.: Atlas of Cardiovascular Monitoring, p. 130. Churchill Livingstone, New York (1998)