UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Operation of the Hilbert-Huang Transform: basic overview with examples

Surnames, name	Antonino Daviu, Jose Alfonso (joanda@die.upv.es)
Department	Escuela Técnica Superior de Ingenieros Industriales Departamento de Ingeniería Eléctrica
Centre	Universitat Politècnica de València

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

1 Summary

This work is intended to provide a brief overview of the operation of the Hilbert-Huang Transform (HHT). This is a signal processing tool the use of which has gained an increasing importance recently in the area of non-stationary signals analysis.

The idea of the transform relies on decomposing the analyzed signal into a set of signals known as *Intrinsic Mode Functions* (IMFs). Each IMF covers a certain frequency range and it reflects the time evolution of the components included within that band. Unlike the DWT, the frequency band associated with each IMF is not known *a priori*, since the process carried out by the HHT is an adaptive filtering in which the components are progressively extracted as they appear in the analyzed signal. This process is known as *Empirical Mode Decomposition* (EMD).

The HHT was developed by NASA researchers and it has been applied with success in many applications, some of them critical, such as fault diagnosis in nuclear reactors. Due to its inherent advantages, its use has been extended to many other areas and applications such as biomedical diagnosis, electrical machines condition monitoring, seismic studies, financial applications, etc…

2 Introduction

The HHT was developed by Norden E. Huang and his collaborators [1, 2]. Their motivation relied on the fact that most traditional data-analysis methods are based on linear and stationary assumptions of the analyzed signals. Some approaches deal well with linear and non-stationary data, while others are designed for nonlinear but stationary and deterministic systems. Nonetheless, in many real-world applications, the data are both nonlinear and non-stationary. As Huang himself stated [1], "*…new approaches are urgently needed, for nonlinear processes need special treatment*". The HHT was conceived to produce physically meaningful representations of data from non-stationary and nonlinear processes [1]. The result is an empirical approach, rather than a theoretical tool, which has been successfully used for the analysis of such data.

The HHT has two different parts: 1) the Empirical Mode Decomposition (EMD) and 2)the Hilbert Spectral Analysis (HSA).

The EMD method is the main part of the HHT and it enables to decompose a signal into a set of components named '*Intrinsic Mode Functions' (IMFs)*. These IMFs are individual, nearly mono-component signals. Their nature guarantees a good behaviour of their Hilbert-transform. The EMD algorithm operates in the time domain and it is adaptive and highly efficient. Moreover, it is valid for nonlinear and non-stationary processes.

On the other hand, the HSA enables to obtain the time evolution of the instantaneous frequency of each IMF. This time-frequency representation of the IMFs is crucial to comprehend the inherent structure of the analysed data set. The main result of the HSA is an energy-frequency-time representation, known as the *Hilbert spectrum*.

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

3 Objetives

The present work has three main goals:

- To describe the operation of the Hilbert-Huang Transform. The intention is to make emphasis on the interpretation of the results of its application rather than focusing on its mathematical foundations.
- To provide several illustrative examples facilitating the comprehension of the HHT operation.
- To review the advantages and drawbacks of the transform, in comparison with classical FFT analysis and with other TFD tools.

This work can be useful for students involved in the signal processing area and, more specifically, on concrete applications of the HHT.

4 Development

4.1 Overview of the mathematical bases

As pointed out by C. Charlton-Pérez, R.B. Perez et al. in their excellent work reviewing the operation of the HHT [3], Huang introduced a new signal analysis technique based on the decomposition of a signal in terms of empirical modes and on their representation within the framework of the Complex Trace Method, introduced by Gabor several decades ago [4].

The formulation of the proposed methodology is the following: let $X(q)$ be a signal (with q representing either time or an spatial coordinate) that is the real part of a complex trace, $Z(q)$ (given by Equation 1), where the imaginary part, $Y(q)$, is the Hilbert transform of $X(q)$ (given by Equation 2, where *PV* is the Principal Value):

$$Z(q)=X(q)+iY(q)$$

Equation 1. Analytical signal Z(q).

$$Y(q) = \left(1/\pi\right) \cdot PV \cdot \int_{-\infty}^{\infty} X(q')dq'/(q-q')$$

Equation 2. Definition of function Y(q).

The complex conjugate pair $(X(q); Y(q))$ defines the amplitude, $a(q)$, and phase $\theta(q)$, as an analytical function of the q-variable [3, 5] (see Equation 3), with the instantaneous frequency defined by Equation 4.

$$Z(q) = a(q) \cdot e^{i\theta(q)}$$

Equation 3. Analytical function Z(q).

$$\omega(q) = d\theta(q)/dq$$

Equation 4. Instantaneous frequency.

The Complex Trace Method enables to define the concepts of instantaneous amplitude, phase and frequency, in such a way that the original signal can be expressed in terms of a Fourier-like expansion based on these concepts [1,3,5]. This process, along with the instantaneous frequency definition, work well for mono-component signals. Nevertheless, in many real applications the signals are multi-component (and often noise-polluted). In these situations, the Complex Trace method fails due to the fact that the Hilbert transform processing of those noisy waves generates spurious amplitudes at negative frequencies [3, 5]. The approach developed by Huang [1,2] enabled the signal analysis to avoid generating unphysical results. To this end, the Hilbert transform is not directly applied to the signal itself but to each of the members of an empirical decomposition of the signal into Intrinsic Mode Functions (IMFs) [3, 5]. As R.B. Perez stated [3, 5], these IMFs are "*individual, nearly mono-component signals with 'Hilbert-friendly' waveforms, to which the instantaneous frequency defined by Equation 4 can be applied*".

4.2 Operation of the HHT

Empirical Mode Decomposition (EMD) Algorithm

The Empirical Mode Decomposition (EMD) method is the basis of the HHT. By means of the EMD, the analysed data set is decomposed into a set of almost orthogonal components named *Intrinsic Mode Functions* (IMFs). The algorithm to create the IMFs is rather simple as well as smart, as explained in [3, 5]: firstly, the local extrema of the data are identified; they are used to create upper and lower envelopes which enclose the signal completely. A running mean is created from this envelope. When subtracting this mean from the data, a new function is obtained; this function must have the same number of zero crossings and extrema (i.e. it must exhibit symmetry across the q-axis). Otherwise, if the function so constructed does not satisfy this criterion, the process continues until an acceptable tolerance is reached [1,3]. The process results in the first IMF, $i_1(q)$, which contains the highest frequency oscillations found in the data (the shortest time scales). The IMF1, is then subtracted from the original data, and this difference $R1$ is taken as if it were the original signal and then the sifting process is applied to the new signal. The process of finding modes, $i_j(q)$, is carried out until the last mode ,the residue Rn, is found. It contains the trend (i.e. , the "time –varying" mean). Hence, the signal $X(q)$ can be characterised as indicated by Equation 5. The scheme of the whole EMD algorithm is plotted in Figure 1.

$$X(q) = \sum_{j=1}^{n} i_j(q) + Rn$$

Equation 5. Characterization of the signal X(q) in terms of the IMFs and the residual.

Hilbert Spectral Analysis (HSA)

When the IMFs have been obtained, the Hilbert transform can be applied to each IMF, computing the instantaneous frequency and amplitude. After applying the Hilbert transform to each IMF, the signal can be expressed according to Equation 6 ($a_j(q)$ and $w_j(q)$ are, respectively, the instantaneous amplitude and frequency corresponding to each IMF $i_j(q)$).

$$X(q) = \mathrm{Re}\left(\sum_{j=1}^{n} a_j(q) \cdot e^{i \int \omega_j(q) \cdot dq} \right)$$

Equation 6. Characterization of X(q) in terms its instantaneous amplitude and frequency

Equation 6 allows the representation of the instantaneous amplitude and frequency as functions of q in a 3-D plot or map. The time-frequency representation of the amplitude is named *Hilbert spectrum*, $H(w,q)$ [5, 6].

Another important concept is the *marginal spectrum*. Its mathematical definition is given by Equation 7, where Q is the total data length [7]. While the Hilbert spectrum provides a measure of the amplitude contribution for each frequency and time, the marginal spectrum offers a measure of the total amplitude (or energy) contribution from each frequency [5, 7]. The frequency in the marginal spectrum indicates only the likelihood that an oscillation with such a frequency exists; the exact occurrence time of that oscillation is given in the full Hilbert spectrum [2].

$$h(w) = \int_0^Q H(w,q) \cdot dq$$

Equation 7. Mathematical definition of marginal spectrum

Figure 2 summarizes the results of the application of the HHT to a certain signal $s(t)$, considering two IMFs. Each IMF is a waveform which extracts the components of the original signal within a certain frequency range. In order to know the frequencies covered by each IMF, its Hilbert spectrum is computed; it provides a time-frequency-amplitude representation of the corresponding IMF.

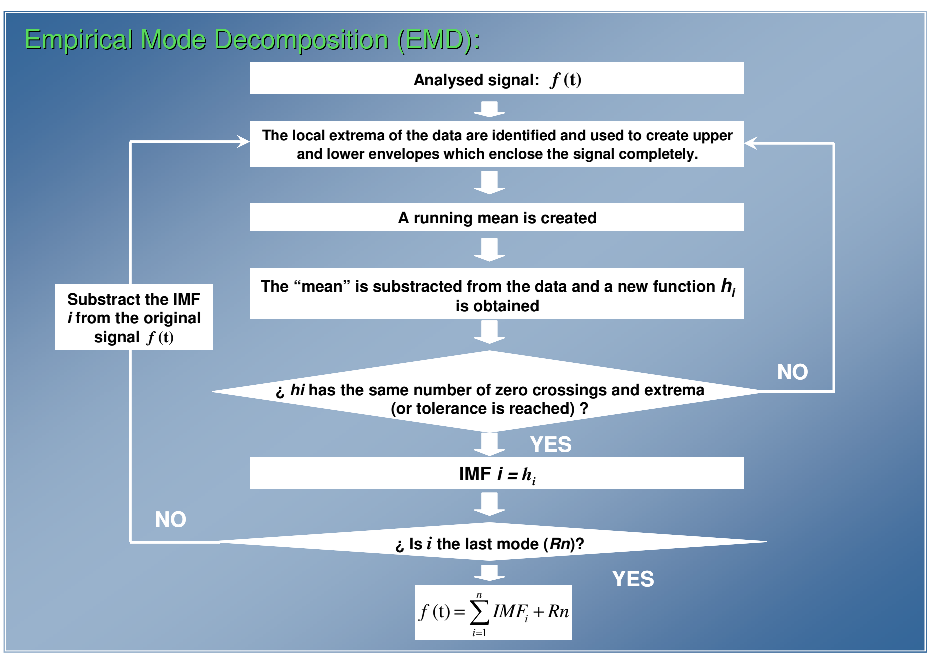


Figure 1. Filtering process carried out by the DWT

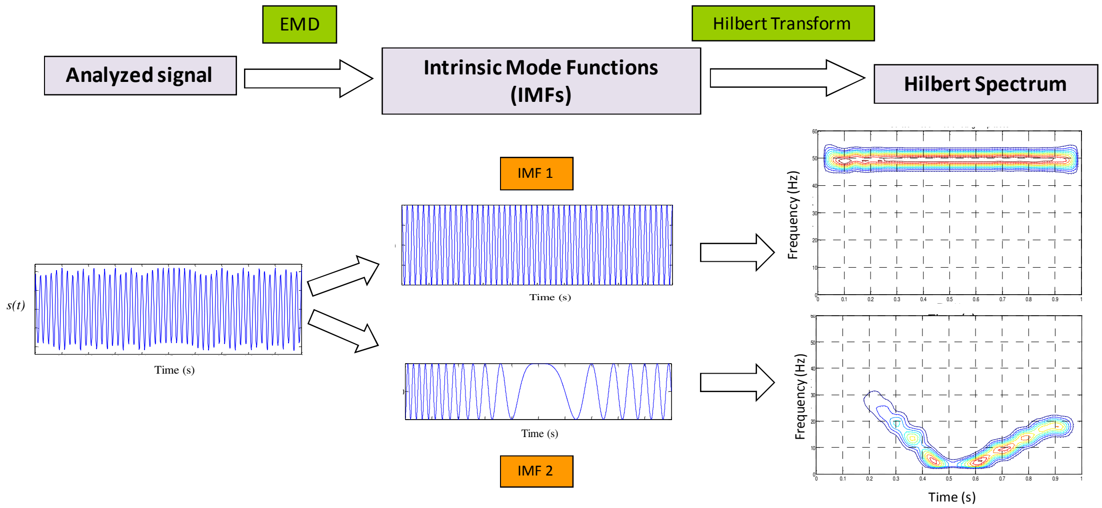


Figure 2. Schematic representation of the HHT results

4.3 Examples

In this section, several didactic examples of the operation of the HHT are presented. They are useful to understand how the transform works in a very simple way. In all the examples, the HHT is carried out and the next results shown for each IMF: IMF waveform, IMF Hilbert spectrum and IMF marginal spectrum.

4.3.1 Example 1: HHT of the addition of two sinusoidal signals

Figure 3 represents the waveform of a signal based on the addition of two sinusoidal signals with respective frequencies 50 Hz and 15 Hz and respective amplitudes 5 and 1. Figure 4 shows the results of the application of the HHT to the previous signal, considering two IMFs in the decomposition. Figure 4(a) corresponds to IMF1, depicting: the IMF1 waveform (Figure 4(a), top), the IMF1 Hilbert spectrum (Figure 4(a), middle) and the IMF1 marginal spectrum ((Figure 4(a), bottom). Figure 4(b) is analogous, but for the IMF2.

That figure illustrates the way in which the HHT works: IMF1 extracts the largest component present in the signal (namely, the 50 Hz sinusoidal component with amplitude 5). This can be perfectly noticed in the IMF1 waveform (Figure 4(a), top). The Hilbert spectrum of this IMF1 (Figure 4(a), middle), as expected, reveals a single line at 50 Hz for every time instant (recall that the Hilbert spectrum is just a time-frequency-amplitude representation of the IMF). Finally, the IMF1 marginal spectrum shows a single peak at 50 Hz. On the other hand, IMF2 extracts the evolution of the rest of components in the analyzed signal (in this case, the 15 Hz sinusoidal component with amplitude 1). This can be observed in Figure 4(b), top, that depicts the IMF2 waveform revealing a sinusoidal component with lower frequency (and amplitude) than that in Figure 4(a), top. This is confirmed by the Hilbert spectrum of IMF2 (Figure 4(b), middle) which shows a single line at 15 Hz for every time instant. Accordingly, the marginal spectrum of IMF2 reveals a single frequency peak at 15 Hz.

This example shows the adaptive filtering nature of the HHT, extracting the components present in the signal in the different IMFs.

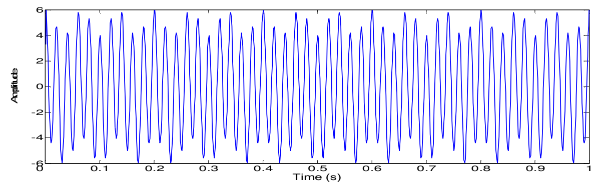


Figure 3. Signal based on the addition of two sinusoidal signals

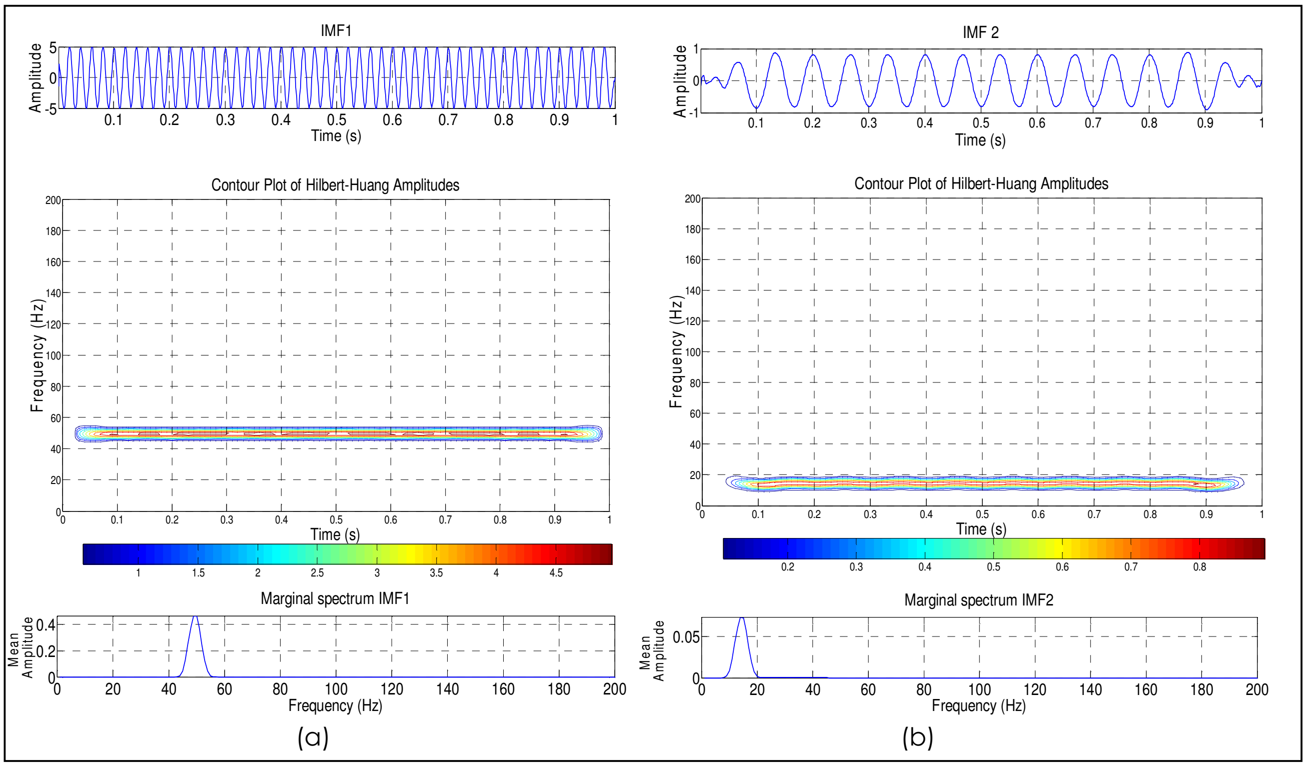


Figure 4. HHT of the previous signal: (a) IMF1: waveform (top), Hilbert spectrum (middle) and marginal spectrum (bottom); (b) IMF2: waveform (top), Hilbert spectrum (middle) and marginal spectrum (bottom).

4.3.2 Example 2: HHT of the concatenation of sinusoidal signals

The second example consists of the HHT analysis of the signal plotted in Figure 5, which is based on the concatenation of two sinusoidal signals with respective frequencies 50 Hz and 15 Hz and respective amplitudes 5 and 1. As a consequence of the operation of the EMD algorithm, the HHT analysis leads to a single IMF (IMF1) which reflects the evolution of both components. Figure 6 represents the HHT results: IMF1 waveform (top), Hilbert spectrum of the IMF1 (middle) and marginal spectrum of IMF1 (bottom).

The waveform of the IMF1 (Figure 6, top) corresponds to that of the original signal, based on the concatenation of the two aforementioned components (initially, that of 50 Hz, which is present during the initial 0,5 s and, then, the 15 Hz component, present between 0,5 and 1 s. The Hilbert spectrum of IMF1 (Figure 6, middle) is especially interesting; a single line at 50 Hz reveals the presence of the first frequency component during the initial 0,5s. A second trace at 15 Hz shows the occurrence of the second component during the last 0,5 s. The lower color intensity of this second trace is due to the lower amplitude of the 15Hz frequency component. This Hilbert spectrum illustrates rather well the time-frequency nature of the tool, since it informs not only on which frequency components are present in the analyzed signal, but also when they occur. Finally, the marginal spectrum of IMF1 (Figure 6, bottom) shows two peaks at the corresponding frequencies present in the signal (15 Hz and 50 Hz), the amplitudes of which reflect the amplitude of the associated sinusoidal signals.

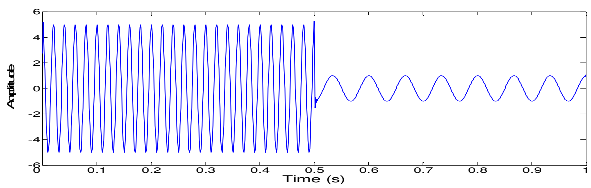


Figure 5. Signal based on the concatenation of two sinusoidal signals

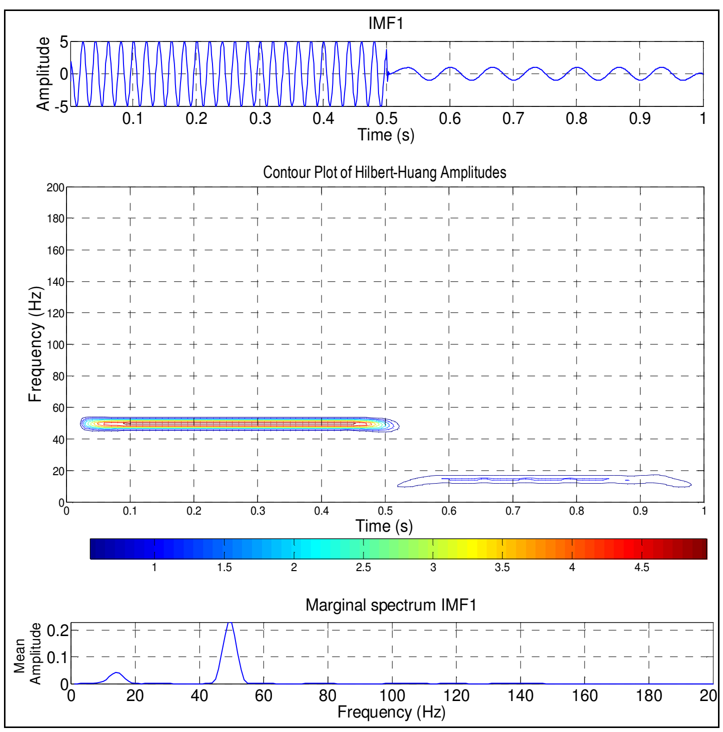
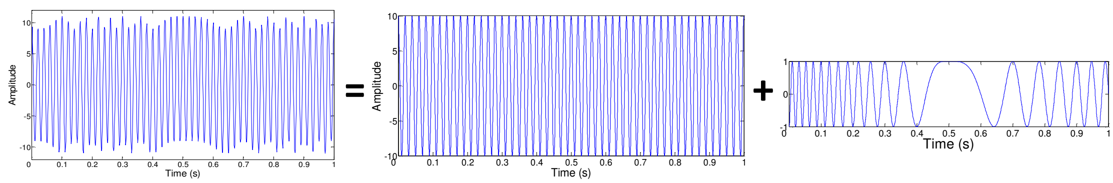


Figure 6. HHT of the previous signal: waveform of the IMF1 (top), Hilbert spectrum of IMF1 (middle) and marginal spectrum of IMF1 (bottom).

4.3.3 Example 3: Addition of a sinusoidal signal and two chirp signals

The third example is based on the HHT analysis of the signal plotted in Figure 7 (a). This signal is based on two waveforms: a 50 Hz sinusoidal signal with amplitude equal to 10 (Figure 7 (b)) and a signal based on the concatenation of two chirp functions (Figure 7(c), the first chirp with linearly decreasing frequency (from 50 Hz to 0 Hz) and the second chirp with frequency increasing linearly from 0 to 50 Hz).

Figure 8 represents results of the HHT analysis considering two IMFs. Figure 8 (a) plots the IMF1 waveform (top), the Hilbert spectrum of IMF1 (middle) and the IMF1 marginal spectrum (bottom). Figure 8 (b) is analogous but for IMF2.



(a) (b) (c)

Figure 7. Signal (a) based on the addition of a sinusoidal signal (b) and two chirp signals (c)

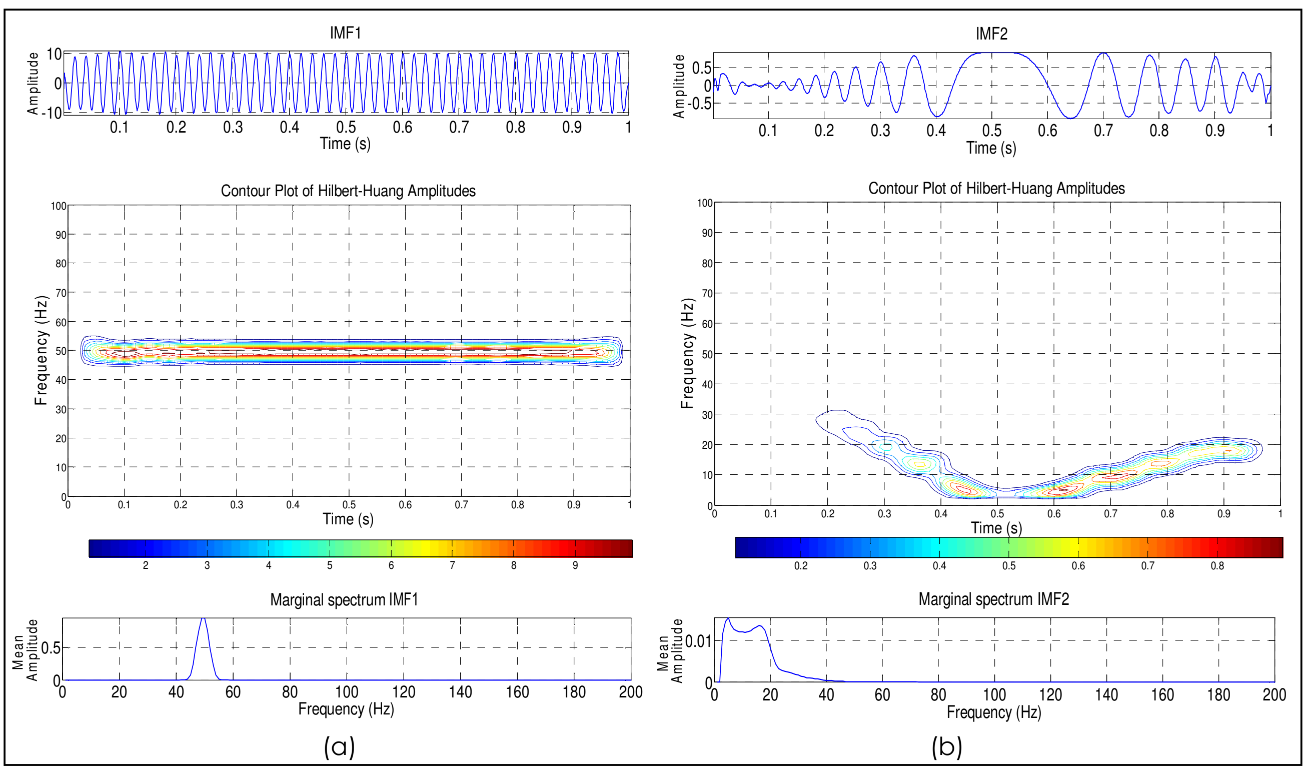


Figure 8. HHT of the previous signal: (a) IMF1: waveform (top), Hilbert spectrum (middle) and marginal spectrum (bottom); (b) IMF2: waveform (top), Hilbert spectrum (middle) and marginal spectrum (bottom)

The results plotted in Figure 8 are rather logical. IMF1 extracts the largest component present in the analyzed signal, i.e., the 50 Hz sinusoidal component with amplitude equal to 10. Figure 8(a), top confirms this fact, since IMF1 waveform corresponds to sinusoidal function with the aforementioned amplitude and frequency. Accordingly, the IMF1 Hilbert spectrum shows a single line at 50 Hz for every time instant. Of course, the IMF1 marginal spectrum is also coherent with these facts, revealing a single peak at 50 Hz.

On the other hand, IMF2 extracts the rest of components present in the signal. This is, the component based on the concatenation of the two chirp functions. IMF2 waveform confirms this idea (see similarity between Figure 8(b), top and Figure 7 (c); the slight deviations are due to inherent border effects of the analysis). Figure 8 (b), middle shows how the Hilbert spectrum of IMF2 is coherent with the time evolution of the frequency of such component based on the two chirps: first, its frequency decreases from near 50 Hz to 0 Hz and later increases from 0 Hz to near 50 Hz, leading to a V-shaped pattern that characterizes the time-frequency evolution of such component. Finally, Figure (b), bottom shows a blurred marginal spectrum due to the varying nature of the frequency of the IMF2 (between 0 and 50 Hz).

This example illustrates rather well the adaptive filtering nature of the HHT (since it separates the components present in the signal in different IMFs corresponding to different frequency bands). Moreover, it illustrates the time-frequency decomposition process carried out by the transform (since it indicates the frequencies present and when they occur).

5 Closing

This work presents a basic overview of the operation of the Hilbert-Huang Transform (HHT). This is a cutting-edge time-frequency decomposition tool that has been used with success for the analysis of non-stationary signals. The HHT is based on two parts; the EMD algorithm which enables to decompose the analyzed signal into IMFs and the HSA process, which computes the time-frequency map associated with each IMF. Both parts are explained in this work, briefly reviewing their mathematical foundations.

The main purpose of the work is to introduce the operation of the HHT under an engineering perspective, rather than deepening in mathematical details. Moreover, some didactic examples illustrating the operation of the transform are included: the HHT analysis of a signal based on the addition of two sinusoidal signals, the HHT analysis of a signal based on the concatenation of two sinusoidal signals and the HHT analysis of a signal based on the addition of a sinusoidal signal and a signal based on the concatenation of two chirp functions. All three didactic examples are useful to understand the adaptive filtering nature of the transform, as well as its ability to obtain a time-frequency map of the analyzed signal.

As a conclusion of the ideas exposed in this work we can summarize some of the the *advantages* of the HHT in the following points:

- Adaptive nature; it avoids fixed limits of the bans covered by each IMF.
- Possible higher flexibility in comparison with other TFD tools.
- Easy interpretation of the results

With regards to its *drawbacks* we can remark, among other, the following ones:

- Number of necessary IMFs is not known a priori.
- Possible constraints when distinguishing different components with near frequencies.
- Possible masking problems when high energy components are present in the signal.

6 References

6.1 Textbooks:

[1] Huang, N.E; Shen, S.S.P: "Hilbert-Huang Transform and its Applications". World Scientific Publishing, 2005.

6.2 Journal papers:

[2] Huang , N.E. ; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung C.C. and Liu, H.H.: "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis", Proc. Royal Society of London, A, Vol. 454, 1998, pp. 903-995.

[3] Charlton-Perez, C.; Perez, R.B.; Protopopescu, V.; Worley B.A.: "Detection of unusual events and trends in complex non-stationary data streams" Annals of Nuclear Energy, Vol. 38, No. 2–3, February–March 2011, pp. 489-510

[4] Gabor, D: "Theory of Communication," Journal of Institute of Electrical Engineers, Vol. 93, No. 26, November 1946, pp. 429-457.

[5] Antonino-Daviu, J.; Jover Rodriguez, P.; Riera-Guasp, M.; Arkkio, A.; Roger-Folch, J.; Pérez, R.B.: "Feature Extraction for the Prognosis of Electromechanical Faults in Asynchronous Motors using the Hilbert-Huang Transform" Energy Conversion and Management, Vol. 50, No. 7, July 2009, pp. 1810-1820.

[6] Peng, Z.K.; Tse, P.W.; Chu, F.L.: "A Comparison Study of Improved Hilbert-Huang Transform and Wavelet Transform: Application to Fault Diagnosis for Rolling Bearing", Mechanical Systems and Signal Processing, Vol. 19, 2005, pp. 974-988.

[7] Yu, D.; Cheng, J.; Yang, Y.: "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings", Mechanical Systems and Signal Processing, Elsevier, Vol. 19, 2005, pp. 258-270.