

Empirical Mode Decomposition for Speech Synthesis

Ram Kumar Panthi¹, Prof. Jyotsna Ogale²

M. Tech, (EC) SATI, Vidisha¹

Associate Prof, EC Department SATI, vidisha²

Abstract: A new method of Synthesis by Analysis for EMD multi-component signals of fast changing instantaneous attributes is introduced. It makes use of two recent developments for signal decomposition to obtain near mono-component signals whose instantaneous attributes can be used for synthesis. Furthermore, by extension and combination of both decomposition methods, the overall quality of the decomposition is shown to improve considerably summarize the use of Empirical mode decomposition (EMD) for denoising a speech signal. EMD, introduced by Huang et al in gives time-frequency representation of non-linear and non-stationary signals. It decomposes a signal into a sum of finite zero-mean, oscillating components called as intrinsic mode functions based on the local time characteristics of the signal. Main essence of the method is it's adaptive and data driven nature.

Keywords: Empirical mode decomposition, IMF, Hilbert transform, Denoising technique.

I. INTRODUCTION

The synthesis by analysis approach for sounds with fast changing attributes still poses problems. These can be traced back up to a large degree to the fundamental mathematical properties of the underlying Time-Frequency (TF) analysis methods. This paper introduces two recent methods for wide-band signal decomposition in the context of audio analysis and synthesis for such problematic signals. The requirement is that the components contained are sufficiently spaced apart in the spectrum. Similar to the spectral modeling synthesis introduced by Serra [1], the analysis decomposes a given signal into a sum of time-varying sinusoids plus residual. Here, stochastic components are split into frequency bands and not necessarily part of the residual. The precision of instantaneous phase information obtained by the analysis facilitates phase alignment for the synthesis, thus transients can be retained. Additionally, the involved decomposition methods eliminate the need to perform peak-continuation of spectral components. First, the analysis which is intended to be performed offline is shown, section 3 shows the method of resynthesis which can be performed online. Finally, the paper concludes with results on the Quality of the method and gives future directions for improvement.

II. ANALYSIS

Time-Frequency representations give insight into the complex structure of time series signals by revealing their comprising components within Temporal and spectral localization. The majority of algorithms performing such a representation on multi-component signals consist roughly of linear and quadratic ones. Representatives for the first group are the Short-Time Fourier and Wavelet Transformations and, respectively, the Wigner-Ville

Distribution for the latter one. The first group relies on the linear super-position Principle of base functions with which the signal to be analyzed is compared [2]. As such a basis is chosen a priori, presumptions are made in regards to the driving mechanisms of the data. In consequence, misfits in respect to the selected basis are assigned to various orders of harmonics thereof, thus coloring or possibly depriving the TF representation of physical meaning, especially if the data is the non-stationary result of non-linear driving mechanisms. Besides this, such integral transforms obey the Heisenberg-Gabor limit, forcing a trade-off for either time or frequency localization. Quadratic methods, on the other hand, avoid the use of basic functions as templates and generally provide a high-resolution TF representation for mono-component signals (defined below). However, for multi-component signals, the additional presence of interference terms between each pair of individual components can severely distort the representation. Removing them by means of filtering comes at the expense of TF resolution.

Alternatively, a signal can be regarded as the result of superimposed mono-components. A mono-component is a sinusoid whose attributes are instantaneous - amplitude and phase vary with time. It exhibits a well-behaved Hilbert-Transform (HT), so the derived analytic signal reflects these attributes uniquely and unambiguously. The question is, within the infinite possibilities to decompose a signal, how can multi-component signals be separated into such mono-components? In the last decade mainly two approaches towards this have emerged: the Empirical Mode Decomposition (EMD) [3] and the Hilbert Vibration Decomposition (HVD) [4]. Both are nonparametric and adaptive decompositions with base functions chosen a posteriori. The reason they will be shown in a little bit

more detail is that the proposed method makes use of both of them in a way to diminish their mutual downsides.

III. INSTANTANEOUS ATTRIBUTES OF MONO-COMPONENTS

One way to obtain the instantaneous attributes of a mono-component signal $x(t)$, the amplitude $A(t)$ and the phase $\Phi(t)$, is by constructing its complex valued analytic signal $X(t)$. This can be achieved by composing the original time-domain signal $x(t)$ with its imaginary Hilbert-Transformed version $\hat{x}(t)$ (the quadrature projection). As a result, the

Instantaneous amplitude and phase can be determined

$$\begin{aligned} \text{As } \sqrt{x^2(t) + \hat{x}^2(t)} &= A(t), \\ \Phi(t) &= \arctan \frac{\hat{x}(t)}{x(t)} \end{aligned} \quad (1)$$

Throughout the rest of the paper $\Phi(t)$ denotes the unwrapped instantaneous phase function.

IV Empirical Mode Decomposition

Empirical mode decomposition (EMD) decomposes a signal $x(t)$ into a finite number of Intrinsic Mode Functions (IMFs),

$$x(t) = \sum_{i=1}^L h_i(t) + r(t) \quad (1)$$

Where $r(t)$ is a remainder which is a non zero-mean slowly varying function with only few extreme. Decomposition is based on the characteristics of the signal itself. IMFs are zero-mean oscillating signals satisfying the following conditions:

- A. The number of extreme and the number of zero crossings must either be equal or differ at most by one,
- B. At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero. Steps for finding the IMFs of a signal are as follows-

- Identify local maxima and minima of $x(t)$.
- 2) Form the upper and lower envelope $u(t)$ and $l(t)$ by cubic spline interpolation of the extrema points.
- 3) Calculate the mean of the upper and lower envelop, $m_1(t)$ using $m_1(t) = u(t) + l(t) / 2$.
- 4) Subtract mean from the signal $x(t)$ to obtain $d_1(t)$. If $d_1(t)$ is a zero-mean function, then the iteration stops and $d_1(t)$ is accepted as first IMF, ie $h_1(t) = d_1(t)$.
- 5) If not, use $d_1(t)$ as the new data and repeat steps 1-4 until an IMF is obtained.
- 6) Once the first IMF $h_1(t)$ is obtained, residual signal is defined as

$$r_1(t) = x(t) - h_1(t) \quad (2)$$

Residual signal contains information about the lower frequency components and is taken as the input signal to obtain next IMFs. At the end, a monotonic function with

only few extreme is obtained from which no further decomposition can be done.

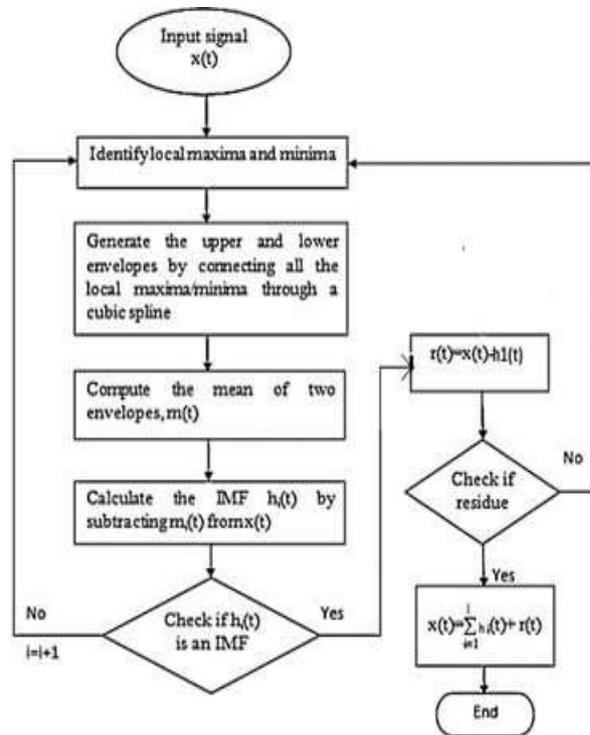


Fig 3.1 Flow chart

IV. CASE STUDY

At the core of the EMD is a sifting process that creates almost mono-components [4]. The sifting is performed by identifying the innate undulations belonging to different relative frequency scales and recursively discerning waves riding on top of each other using repeated approximation. By means of this scale separation, intrinsic modes of oscillations are extracted from signal $s(t)$. These are called intrinsic mode functions (IMFs), $h_k(t)$, if they fulfil: (a) for $h_k(t)$ the number of extreme points (min/max) and zero crossings are equal or differ at most by one (b) the mean of the lower envelope defined by the local minima and the respective upper envelope of $h_k(t)$ is at any point zero. With the global residue (or trend) $r(t)$, $s(t)$ can be expressed as:

$$\sum_{k=1}^n h_k + r(t) = A_k(t) \cos _k(t) + r(t) \quad (2)$$

Where n is the number of IMFs extracted. As equation 2 suggests, an IMF has variable amplitude and frequency as functions of time and therefore constitutes the opposite of a mono-harmonic signal. Figure 1 exemplifies such decomposition. The objective to find IMFs is performed by a sifting process, starting with

$$r(t) = r_0(t) = s(t) \text{ and } i = k = 0:$$

- A1 find all local minima and maxima of $r(t)$
- A2 create interpolant $e_{\min}(t)$ through the local minima and respectively $e_{\max}(t)$ through the local maxima
- A3 set $m(t)$ as local average with

$$m(t) = \frac{E_{min} + E_{max}}{2} \quad A4$$

Define a “proto-mode” function

$$p_i(t) = r_p(t) - m(t),$$

Set $r_p(t) = p_i(t)$ and $i = i + 1$

A5 repeat steps A1-4 until $p_i(t)$ meets stopping criterion S; then an IMF is found,

$$h_k(t) = p_i(t)$$

A6 set $r(t) = r(t) - h_k(t)$; if stopping criterion T is fulfilled then terminate, else $i = 0$, $k = k + 1$ and $r_p(t) = r(t)$; restart from step A1. Here, steps A1-4 create a k-level IMF and step A5 controls the global sifting process. In this way, the EMD repeatedly removes a wave riding on top of the local residue $r(t)$ as it identifies the wave through local extreme points and treats the residue as global trend. At the whole, the behaviour of the EMD is similar to a filter-bank: performing as high-pass filter for the first IMF and as band-pass for successive IMFs. Yet the characteristic is that the cut-off/centre frequencies are non-stationary. Albeit the EMD is still of algorithmic nature, some theoretical work has been put in to describe its behaviour. When analyzing white noise-like wideband signals, the EMD behaves like a dyadic filter-bank [5], while for bi component signals of harmonics there exists a theoretical limit for separation of

$$A1/A2 = (F1/F2)^2 = 1 \quad [6].$$

Hence, the EMD does not perform well when the components' frequencies are close or differ little in amplitude. The existence of a plethora of implementations for the EMD make further theoretical assessment difficult as some tackle core issues of the algorithm like the choice of the interpolation technique or the construction of the envelope differently. As suggested by Huang [3], the cubic spline interpolation is used here. The condition criteria for the envelope are currently not completely understood [7], leading to various contributions how extreme points ought to be chosen. Instead of finding the local extreme of $s(t)$ itself, it is proposed to find them in the inverse of the second derivative of $s(t)$. Hence, first the “frequency resolution” is increased for riding waves that are partially immersed in the local trend and thus do not produce local extrema (e.g saddle points), and second, for pure sinusoids the positions stay the same. This approach, however, comes at the danger of producing artificial vibrations, especially in lower IMFs. Consequently it is applied for the first IMFs only ($k \leq 5$) where most of the high frequency contents of $s(t)$ are to be expected. In figure 2 an example is given where this method helps to uncover positions of extreme. Regarding the stopping criteria: for the number of IMFs generated can be either set to a fixed amount of iterations, commonly $k \approx \log_2 N$ with N being the number of data points of $s(t)$, or an indicator that the residue still contains oscillations. Here, the former criterion was applied as for the used test signal ($N \approx$

22050), the final residue always showed a non-oscillatory trend. For S a number of stopping criteria have been suggested, the original recommendation being to set the number of iterations to the order of tens. Accordingly, the number of iterations was set to $i = 30$. However, they are terminated before the cubic splines interpolation leads to degenerated results. This condition is met as soon as the area under the cubic splines increases in succeeding iterations i . This is an indicator for large overshoots of the interpolation caused by ill-conditioned extrema points. Thus, the possibility of degenerated envelopes creating artificial vibrations for succeeding IMFs is reduced. A major problem that exists for the EMD is the phenomenon of mode-mixing that results in a) an IMF containing signals of widely disparate scales or b) signals of similar scale residing in different IMF components [7]. This happens when the intermittency in the extrema detected belongs to different signals as caused when parts of the riding wave are completely immersed in the local trend. Several methods have been proposed to alleviate this problem; commonly the aim is to emphasize “lost” extrema points of the riding wave. In general, there are two approaches to this: either calculates the mean of an ensemble of decompositions that have different instances of noise added to the signal (EEMD) [8], or add masking signals in the decomposition that approximate the riding waves in the problematic areas [7,9]. Initial attempts to use the EEMD resulted in less mode mixing of type b) but more of type a), when components reside closely in a frequency band with similar amplitudes. Therefore, the use of masking signals has been chosen. Suppose that a masking signal $\hat{r}(t)$ that contains for the sake of brevity, the HVD is only superficially presented here. As opposed to the EMD, the HVD is entirely based on the HT. Therefore, the HVD does not depend on a dissimilar harmonics amplitude ratio as does the EMD. The method is based on the observation that, in a multi-component signal, the instantaneous attributes of the component with the highest energy change more slowly in comparison to the sum of those of the underlying components.

In order to rid a signal of these fast oscillating instantaneous attributes and thereby performing the decomposition, the instantaneous attributes derived by means of the HT are low-pass filtered. The filtered result is seen to constitute a mono-component. The residue can again be used in the decomposition process leading to a set of basic functions that similarly express $s(t)$ as in equation 2. By applying the HVD for only one iteration (to obtain the singular highest energy component) on $r(t)$ from the EMD, the masking signal \hat{r} is generated. The reason the HVD is not used principally for the decomposition is that the HT is very sensitive to false spikes or random noise that leads to the distortion of transients in the instantaneous attributes or smearing [4]. The EMD, on the other hand, is capable of decomposing noisy signals [5]. Also, due to practical limitations of precise low-pass filtering in the HVD, the number of extracted components is limited [4]. However, in general the HVD is able to

better separate components in a narrow band than the EMD. By combining both methods this way the HVD helps increasing the frequency resolution of the EMD and reducing mode-mixing errors. To compare the performance of this approach to the original EMD one, the quality of the decomposition of a bi-component signal was measured in the same way as discussed in [6]. Due to the Recursive nature of the EMD such comparison gives insight into the overall decomposition performance for complex multi-component signals. Figure 3 allows the comparison of the ability of both methods to identify a high frequency signal $x_h(t)$ within a composition $x(t)$ of $x_h(t)$ and a low frequency signal $x_l(t)$. As can be seen in plot 3 a), the proposed

V. EXPERIMENTAL RESULT

The quality of the resynthesis depends very much on the effectiveness of the post-processing and the presence of mode-mixing in the obtained components. For example, the resynthesis of a monophonic (synthetic) bass-drum (cp. table 1) without post-processing led to a change of the originally sinusoidal signal to a more square wave-like one due to the errors introduced by the Hilbert FIR. With post-processing, the resynthesized audio had no perceivable differences.

For the piano sample, the decomposition introduced mode mixing errors in the decay phase of the sound as extreme of the previously correctly tracked high-frequency components were immersed in lower frequency harmonics. This resulted in perceivable phase distortions (bursts) when pitch-shifting or time-stretching; on reducing the value of the post-processing coefficients the

resynthesis expectedly introduced perceivable glissandi around such sections of mode-mixing. When disregarding these sections the results were satisfactory as the timbre of the sound was preserved (1octave, 2 x time-stretches) once the post-processing removed the unwanted modulations. Since the EMD is able to decompose noisy signals, the sample of a (real) snare could be decomposed into separate IMFs containing noise (dyadic frequency bands) and a tonal component.

Similarly, the sample of a (real) cowbell was successfully decomposed into fundamental and harmonics. For all of these percussive samples, the fundamental could be well separated without the phenomenon of mode-mixing. Depending on the used operator the results of the pitch-shifting can sound convincing, especially since no artifacts of blurred transients were introduced. if the harmonics were treated as formants.

Expectedly for extreme settings, the resynthesis of the snare drum produced audible artifacts for the noise components if they were altered by time-stretching or heavy post-processing, since they were interpreted as sinusoidal. Hence, their modeling as noise partials would be preferable. The additions to the original EMD method presented here have shown that the quality of the decomposition can be improved considerably. With it, a post-processing method has been introduced that helps to remove some of the errors introduced by the Hilbert-Transform and to condition the IMFs for synthesis by removing low-energy modulations of phase and amplitude. Finally, a rough summary of the quality of the synthesized sounds has been given.

Table1 parameters of original signals before processing and after processing synthesized signals

signals	Standerd deviation of original signal	Standerd deviation of reconstructed signal	Mean of original signal	Mean of reconstructed signal	MSE between original and reconstructed signal	MSE between power spectrum of original and reconstructed signal
Bass drum	0.5666	0.6170	0.1225	0.0445	6.5766e-007	5.0380e-007
bell	0.5666	0.6170	0.1225	0.0445	6.5766e-007	5.0380e-007
paino	0.5666	0.6170	0.1225	0.0445	6.5766e-007	5.0380e-007
snare	0.5666	0.6170	0.1225	0.0445	6.5766e-007	5.0380e-007

Table2 intrinsic mode function (imf) of signals & noise reduction coefficients

Signals	nr	nc	Imf1	Imf2	Imf3	Imf4	Imf5	Imf6	Imf7	Imf8	nrc	ncc	nrcp	nccp
Bass drum	8	65528	65528	65528	65528	65528	65528	65528	65528	65528	8	8883	8	8883
bell	8	65528	65528	65528	65528	65528	65528	65528	65528	65528	8	8883	8	8883
paino	8	65528	65528	65528	65528	65528	65528	65528	65528	65528	8	8883	8	8883
snare	8	65528	65528	65528	65528	65528	65528	65528	65528	65528	8	8883	8	8883

VLRESULT

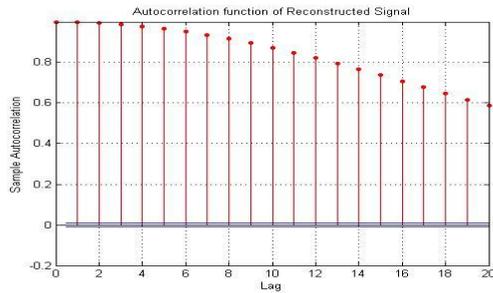


Fig.1 autocorrelation function of original signal

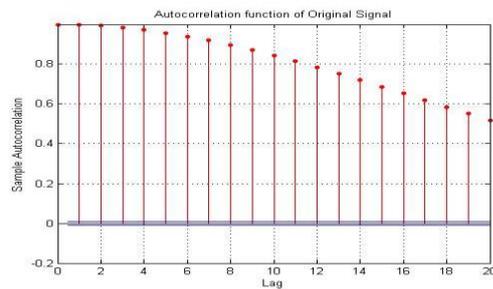


Fig.2 autocorrelation function of reconstructed signal

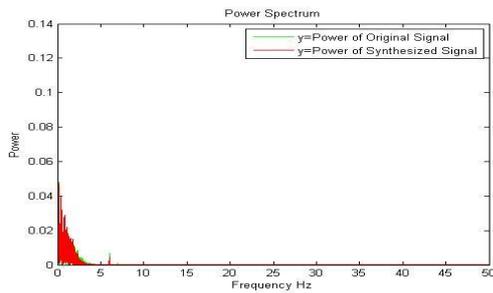


fig 3 power specterm

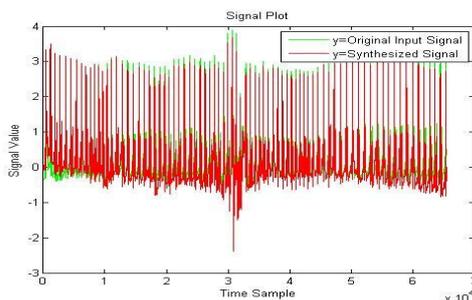


Fig.4 signal plot

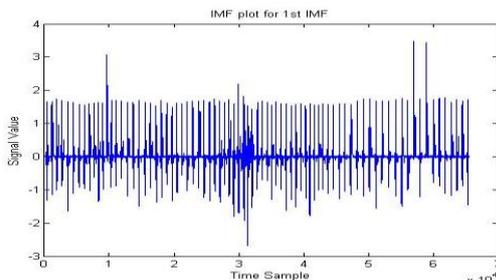


Fig.5 IMF plot for 1st IMF

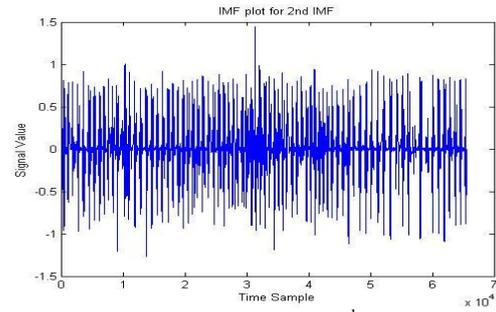


Fig.6 IMF plot for 2nd IMF

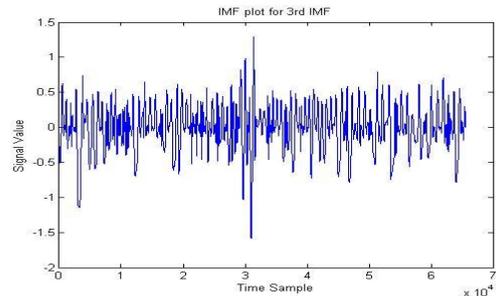


Fig.7 IMF plot for 3rd IMF

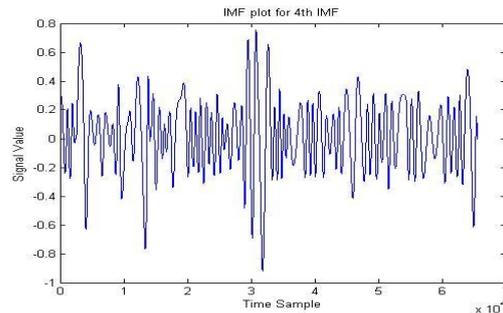


Fig.8 IMF plot for 4th IMF

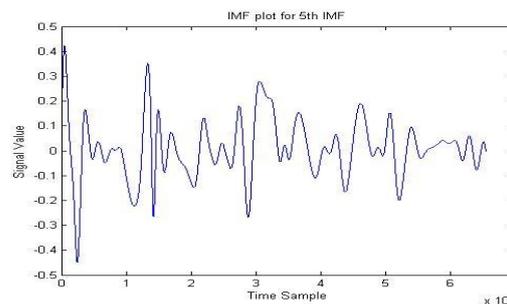


Fig.9 IMF plot for 5th IMF

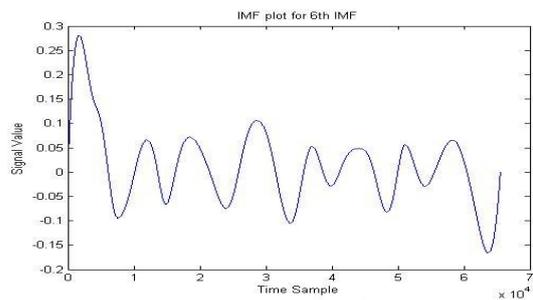


Fig.10 IMF plot for 6th IMF

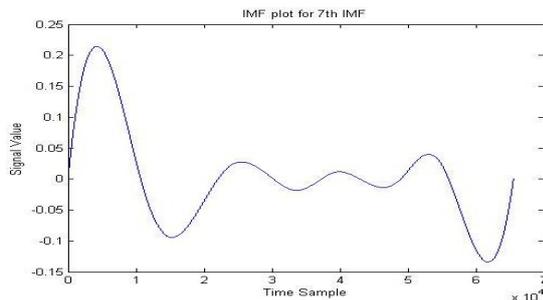


Fig.11 IMF plot for 7th IMF

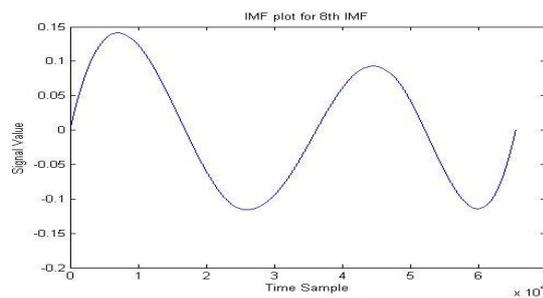


Fig.12 IMF plot for 8th IMF

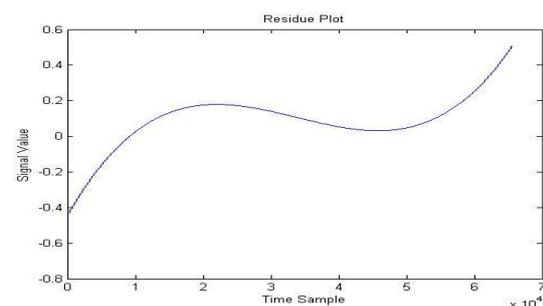


Fig.13 residue plot

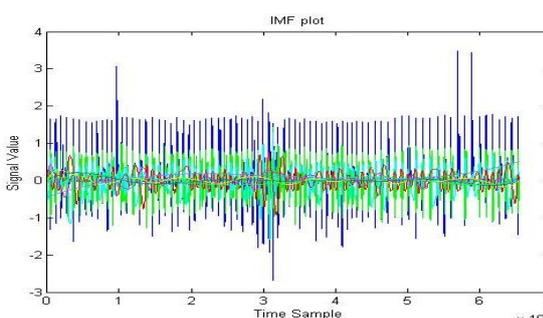


Fig.14 IMF plot

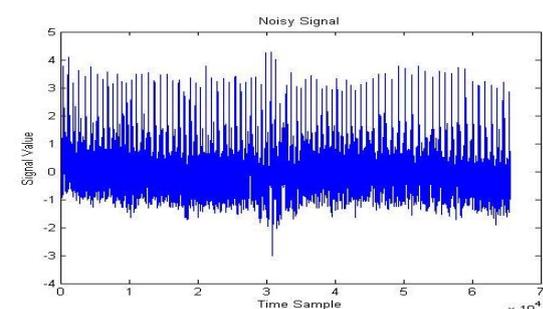


Fig.15 noisy signal

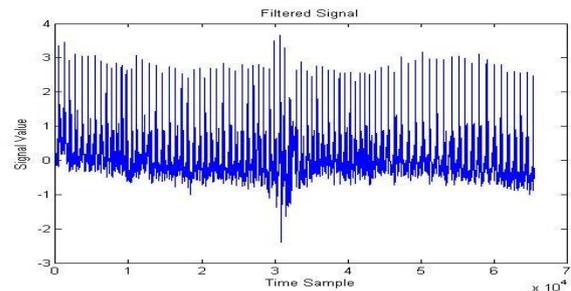


Fig .16 filtered signal

- Standard Deviation of Original Signal = 0.5666
- Standard Deviation of Reconstructed Signal= 0.6170
- Mean of Original Signal= 0.1225
- Mean of Reconstructed Signal= 0.0445
- Mean Square Error between Original and Reconstructed Signal= 6.5766e-007
- Mean Square Error between Power Spectrum of Original and Reconstructed Signal= 5.0380e-007

VI. CONCLUSION

Hence, to conclude, the shown approach can be very well used to EMD wide-band signals with partials that have fast changing instantaneous attributes and are sufficiently spaced apart in the spectrum. An improvement to this approach would be to add a dedicated noise model to the sinusoidal one in order to be able to alter the behavior of noisy partials properly. As in the case of EMD a piano sound, there are remaining problems regarding the quality of the decomposition, most importantly the frequency and amplitude resolution. However, this may change with future developments of the EMD and To confirm the effectiveness of the method, a male speech is recorded and taken as the original signal. White Gaussian noise is used to model the background noise. Noisy speech signal is applied to the EMD algorithm taken from The obtained IMFs are filter using the principle of soft filter to recover an estimate of the original signal. al filter rule is used. All the simulations are done in MATLAB environment

REFERENCES

- [1] X Serra and J Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition," Computer Music Journal, vol. 14, no. 4, pp. 12–24, 1990.
- [2] Franz Hlawatsch and G. Faye Boudreaux-Bartels, "Linear and Quadratic Time Frequency Representations," IEEE Signal Processing Magazine, vol. 9, no. 2, pp. 21–67, 1992.
- [3] Norden E. Huang, Zhaohua Wu, Steven R. Long, Kenneth C. Arnold, Xianyao Chen, and Karin Blank, "ON INSTANTANEOUS FREQUENCY," Advances in Adaptive Data Analysis, vol. 1, no. 2, pp. 177–229, Dec. 2009.
- [4] Michael Feldmann, Hilbert Transform Applications in Mechanical Vibration, Hoboken, N.J.:Wiley, 1 edition, 2011.
- [5] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical Mode Decomposition as a Filter Bank," IEEE Signal Processing Letters, vol. 11, no. 2, pp. 112–114, Feb. 2004.
- [6] Gabriel Rilling and Patrick Flandrin, "One or two frequencies? The empirical mode decomposition answers," Signal Processing, IEEE Transactions, vol. 56, no. 1, pp. 85–95, 2008.

- [7] Xiyuan Hu, Silong Peng, and Wen-liang Hwang, "EMD Revisited : A New Understanding of the Envelope and Resolving the Mode Mixing Problem in AM-FM Signals," *Agenda*, no. June, 2011.
- [8] Zhaohua Wu and Norden E Huang, "Ensemble empirical mode decomposition for high frequency ECG noise reduction.," *Advances in Adaptive Data Analysis*, vol. 55, no. 4, pp. 193–201, 2009.
- [9] Ryan Deering and James F Kaiser, "The use of a masking signal to improve empirical mode decomposition," *Time*, vol. 4, no. January, pp. 485–488, 2005.DAFX-4
- [10] Y. Zhang, Y. Gao, L. Wang, J. Chen, and X. Shi, "The removal of wall components in doppler ultrasound signals by using the empirical mode decomposition algorithm," *IEEE Trans. Biomed. Eng.*, vol. 9, pp. 1631–1642, Sept. 2007.
- [11] L. Hadjileontiadis, "Empirical mode decomposition and fractal dimension filter," *IEEE Eng. Med. Biol. Mag.*, pp. 30 – 39, Jan. 2007. [12] B. Ning, S. Qiyu, Y. Zhihua H. Daren, and H. Jiwu, "Robust image watermarking based on multiband wavelets and empirical mode decomposition," *IEEE Trans. Image Processing*, pp. 1956 – 1966, Aug. 2007.
- [13] Md. K. I. Molla and K. Hirose, "Single-mixture audio source separation by subspace decomposition of hilbert spectrum," *IEEE Trans. on Audio, Speech and Language Processing*, pp. 893 – 900, Aug. 2007.
- [14] S. Mallat, *A wavelet tour of signal processing*, Academic press, second edition, 1999.
- [15] A. Antoniadis and J Bigot, "Wavelet estimators in nonparametric regression: A comparative simulation study," *Journal of statistical software*, vol. 6, pp. 1–83, 2001. [16] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [17] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic press, third edition, 2006.
- [18] H. C. Huang and N. Cressie, "Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data," *Technometrics*, vol. 42, pp. 262–276, 2000.
- [19] P. Flandrin, G. Rilling, and P. Gonc,alv`es, EMD equivalent filter banks, from interpetation to applications (in N. E. Huang and S. Shen, *Hilbert-Huang Transform and Its Applications*), World Scientific Publishing Company, first edition, 2005.
- [20] 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*, vol. 460, pp. 1597–1611, June 2004.
- [21] N. E. Huang Z. Wu, Statistical significance test of intrinsic mode functions, (in N. E. Huang and S. Shen, *Hilbert-Huang Transform and Its Applications*), World Scientific Publishing Company, first edition, 2005.
- [22] A. O. Boudraa and J. C. Cexus, "Denoising via empirical mode decomposition," in *ISCCSP2006*, 2006.
- [23] Y. Mao and P. Que, "Noise suppression and flaw detection of ultrasonic signals via empirical mode decomposition," *Russian Journal of Nondestructive Testing*, vol. 43, pp. 196–203, 2007.
- [24] T. Jing-tian, Z. Qing, T. Yan, L. Bin, and Z. Xiao-kai, "Hilbert-huang transform for ECG de-noising," in *1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2007)*, 2007.