Distinction Between EMD & EEMD Algorithm for Pitch Detection in Speech Processing

Bhawna Sharma[#1] Sukhvinder Kaur[#2]
Scholar, M.Tech *Assistant Professor*
Department of Electronics and Communication Engineering
Swami Devi Dyal Institute of Engineering and Technology, Panchkula, Kurukshetra University
HARYANA, INDIA

***Abstract*— In this paper we describes the different algorithms for finding pitch markers in speech signal and it also explain how EEMD is better than EMD algorithm One of the major problem in EMD algorithm is mode mixing. EEMD algorithm helps in solving mode mixing problem. EEMD algorithm is a noise assisted data analysis (NADA) for extracting pitch information for the speech signal. In EEMD signal is decomposed into intermediate functions called IMF. Using these IMFs, information regarding pitch markers can be evaluated.**

***Keywords*— EMD, EEMD, IMF, NADA.**

I. INTRODUCTION

Speech coding is an important element in digital communication. Pitch detection and Pitch marking are very essential for speech processing. There are so many algorithms for pitch detection and pitch marking. These algorithms are very important for speech processing tasks. But there performance is not up to mark for many speech coding and synthesis applications.

The Empirical Mode Decomposition (EMD) has been proposed recently [3] as an adaptive time–frequency data analysis method. It has been proved quite versatile in a broad range of applications for extracting signals from data generated in noisy nonlinear and non stationary processes. As useful as EMD proved to be, it still leaves some annoying difficulties unresolved. One of the major drawbacks of the original EMD is the frequent appearance of mode mixing, which is defined as a single Intrinsic Mode Function (IMF) either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components. Mode mixing is often a consequence of signal intermittency. As discussed by Huang *et al.*,[1],[9] the intermittence could not only cause serious aliasing in the time–frequency distribution, but also make the physical meaning of individual IMF unclear. To alleviate this drawback, Huang *et al.*[1],[9] proposed the intermittence test, which can indeed ameliorate some of the difficulties. However, the approach has its own problems: first, the intermittence test is based on a subjectively selected scale. With this subjective intervention, the EMD ceases to be totally adaptive. Second, the subjective selection of scales works if there are clearly separable and definable timescales in the data. In case the scales are not clearly separable but mixed over a range continuously, as in the case of the majority of natural or man-made signals, the intermittence test algorithm with subjectively defined timescales often does not work very well.

To overcome the scale separation problem without introducing a subjective intermittence test, a new noise-assisted data analysis (NADA) method is proposed, the Ensemble EMD (EEMD), which defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude.

Section II discusses the different algorithm for extracting pitch information Section III defines Empirical mode decomposition. Section IV explain Ensemble Emperical mode decomposition. Section V describe mode mixing problem. Section VI explain steps for EEMD algorithm Finally section VII gives the conclusion.

II. DIFFERENT ALGORITHM FOR EXTRACTING PITCH INFORMATION

A. In a text-to-speech (TTS) conversion system based on the time-domain pitch-synchronous overlap-add

(TD-PSOLA) method, accurate estimation of pitch periods and pitch marks is necessary for pitch modification to assure optimal quality of synthetic speech. In general, there are two major tasks in pitch marking: pitch detection and location determination. In this paper, an adaptable filter, which serves as a band pass filter, is proposed for use in pitch detection to transform voiced speech into a sine-like wave. The pass band of the adaptable filter can be adapted based on the fundamental frequency. Based on the sine-like wave, a peak-valley decision method is proposed to determine the appropriate parts (positive part and negative part) of voiced speech for use in pitch mark estimation. In each pitch period, two possible peaks/valleys are searched, and dynamic programming is performed to obtain pitch marks. Experimental results indicate that our proposed method performs very well if correct pitch information is estimated.

B. A speech data base, consisting of eight utterances spoken by three males, three females, and one child was constructed. Telephone, close talking microphone, and wideband recordings were made of each of the utterances. For each of the utterances in the data base; a "standard" pitch contour was semi automatically measured using a highly sophisticated interactive pitch detection program.[8] The "standard" pitch contour was then compared with the pitch contour that was obtained from each of the seven programmed pitch detectors. The algorithms used in this study were 1) a centre clipping, infinite-peak clipping, modified autocorrelation method (AUTOC), 2) the cepstral method (CEP), 3) the simplified inverse filtering technique (SIFT) method, 4) the parallel processing time-domain method (PPROC), 5) the data reduction method (DARD), 6) a spectral flattening linear predictive coding (LPC) method, and 7) the average magnitude difference function (AMDF) method.

C. This paper describes a new approach to pitch marking. Unlike other approaches that use the same combination of features for the whole signal, we take into account the signal properties and apply different features according to some heuristic. Basically we use a special type of energy contour for pitch marking. Where the energy information turns out to be not suitable as an indicator we resort to the fundamental wave computed from a contiguous *F*0 contour in combination with detailed voicing information. Our experiments demonstrate that the proposed pitch marking algorithm clearly improves the quality of synthesised speech generated by a concatenative text-to-speech system that uses TD-PSOLA for prosodic modifications.

D. Robust fundamental frequency estimation in adverse conditions is important in various speech processing applications. In this paper a new pitch detection algorithm (PDA) based on the autocorrelation of the Hilbert envelope of the LP residual is compared to another well established algorithm from Goncharoff. A set of evaluation criteria is collected on which the two PDA algorithms are compared. In order to evaluate the algorithms in adverse conditions a suited reference database was constructed. This reference database consists of parts of the SPEECON speech database where recordings of 60 speakers were selected and manually pitch marked. The recordings cover several adverse conditions as noise in the car cabin and reverberations of office rooms. The evaluation highlights the good performance of the new algorithm in comparison but shows, that low SNR conditions and strong reverberation are still a demanding challenge for future pitch detection algorithms.

E. The work by Veldhuis also proposes the use of DP to determine the pitch marks. It suggests the use of three different consistency requirements for selecting candidate pitch markers and also as part of the DP cost function. The three proposed consistency requirements are the characteristic-property requirement, the waveform consistency requirement and the pitch-consistency requirement. The characteristic property requirement demands that candidate pitch markers be positioned at either higher maxima of the absolute value of the signal or at the first zero crossing before the maximum positive peak or based on some other signal property. On the other hand, the waveform-consistency requirement tests the similarity between signal portions around adjacent pitch markers, whereas the pitch-consistency requirement selects pitch markers whose distance is close to the estimated pitch period.

F. Goncharoff and Gries present both an algorithm for pitch period estimation and another algorithm for pitch phase 1 hypotheses using dynamic programming (DP). The DP cost function is computed with automatically trained artificial neural networks (ANNs) which combine the outputs of heuristic functions measuring the similarity of adjacent period hypotheses. The speech signal is normalized to zero mean value, and positive zero crossings are determined in the sections of speech marked as voiced using external frame-based voiced-unvoiced decisions. The zero crossings are identified based on heuristic criteria.

III THE EMPIRICAL MODE DECOMPOSITON

This section starts with a brief review of the original EMD method. The detailed method can be found in the works of Huang *et al*[1] and Huang *et al.*[2] Different to almost all previous methods of data analysis, the EMD method is adaptive, with the basis of the decomposition based on and derived from the data. In the EMD approach, the data x(t) is decomposed in terms of IMFs, c_j , i.e.,

$$x(t) = \sum_{j=1}^{n} c_j + r_n \qquad (3.1)$$

Where r_n is the residue of data *x*(*t*), after *n* number of IMFs are extracted. IMFs are simple oscillatory functions with varying amplitude and frequency, and hence have the following properties:

a. Throughout the whole length of a single IMF, the number of extrema and the number of zero-crossings must either be equal or differ at most by one (although these numbers could differ significantly for the original data set).

b. At any data location, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. In practice, the EMD is implemented through a sifting process that uses only local extrema. From any data $r_j - 1$, say, the procedure is as follows:

(i) Identify all the local extrema (the combination of both maxima and minima) and connect all these local maxima (minima) with a cubic spline as the upper (lower) envelope;

(ii) Obtain the first h component by taking the difference between the data and the local mean of the two envelopes; and

(iii) Treat h as the data and repeat steps (i) and (ii) as many times as is required until the envelopes are symmetric with respect to zero mean under certain criteria. The final h is designated as c_j

A complete sifting process stops when the residue, r_n , becomes a monotonic function from which no more IMFs can be extracted. Based on this simple description of EMD Wu and Huang [1] have shown that, if the data consisted of white noise which has scales populated uniformly through the whole timescale or time–frequency space, the EMD behaves as a dyadic filter bank: the Fourier spectra of various IMFs collapse to a single shape along the axis of logarithm of period or frequency. Then the total number of IMFs of a data set is close to LOG_2 N with N the number of total data points. When the data is not pure noise, some scales could be missing; therefore, the total number of the IMFs might be fewer than $LOG_2\ N.$ Additionally, the intermittency of signals in certain scale would also cause mode mixing.

The Empirical Mode Decomposition (EMD) has been proposed recently [3] as an adaptive time–frequency data analysis method. It has been proved quite versatile in a broad range of applications for extracting signals from data generated in noisy nonlinear and non stationary processes. As useful as EMD proved to be, it still leaves some annoying difficulties unresolved. One of the major drawbacks of the original EMD is the frequent appearance of mode mixing, which is defined as a single Intrinsic Mode Function (IMF) either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components.

IV. EEMD

To overcome the scale separation problem without introducing a subjective intermittence test, a new noise-assisted data analysis (NADA) method is proposed, the Ensemble EMD (EEMD), which defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. It should be noted here that we use word ‘single’ instead of word ‘data’ in this paper because the purpose of this paper is to decompose the whole targeted data but not to identify the particular part that is known *a priori* as containing interesting information. Since there is added noise in the decomposition method, we refer the original data as ‘signal’ in most occasions.

With this ensemble approach, we can clearly separate the scale naturally without any *a priori* subjective criterion selection. This new approach is based on the insight gleaned from recent studies of the statistical properties of white noise [1] which showed that the EMD is effectively an adaptive dyadic filter bank a when applied to white noise. More critically, the new approach is inspired by the noise-added analyses initiated by Flandrin *et al.*7 and Gledhill.8 Their results demonstrated that noise could help data analysis in the EMD. The principle of the EEMD is simple: the added white noise would populate the whole time–frequency space uniformly with the constituting components of different scales.[8] When signal is added to this uniformly distributed white background, the bits of signal of different scales are automatically projected onto proper scales of reference established by the white noise in the background. Of course, each individual trial may produce very noisy results, for each of the noise-added decompositions consists of the signal and the added white noise. Since the noise in each trial is different in separate trials, it is canceled out in the ensemble mean of enough trials. The ensemble mean is treated as the true answer, for, in the end, the only persistent part is the signal as more and more trials are added in the ensemble. The critical concept advanced here is based on the following observations:

(i) A collection of white noise cancels each other out in a time–space ensemble mean; therefore, only the signal can survive and persist in the final noise-added signal ensemble mean.

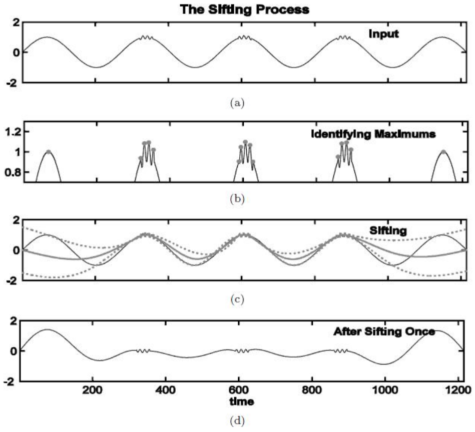
(ii) Finite, not infinitesimal, amplitude white noise is necessary to force the ensemble to exhaust all possible solutions; the finite magnitude noise makes the different scale signals reside in the corresponding IMF, dictated by the dyadic filter banks, and render the resulting ensemble mean more meaningful.

(iii) The true and physically meaningful answer to the EMD is not the one without noise; it is designated to be the ensemble mean of a large number of trials consisting of the noise-added signal.

This EEMD proposed here has utilized many important statistical characteristics of noise. We will show that the EEMD utilizes the scale separation capability of the EMD, and enables the EMD method to be a truly dyadic filter bank for any data. By adding finite noise, the EEMD eliminated largely the mode mixing problem and preserve physical uniqueness of decomposition. Therefore, the EEMD represents a major improvement of the EMD method. In the following sections, a systematic exploration of the relation between noise and signal in data will be presented. Wu and Huang[1] have revealed that the EMD serves as a dyadic filter for various types of noise. This implies that a signal of a similar scale in a noisy data set could possibly be contained in one IMF component. It will be shown that adding noise with finite rather than infinitesimal amplitude to data indeed creates such a noisy data set; therefore, the added noise, having filled all the scale space uniformly, can help to eliminate the annoying mode mixing problem first noticed by Huang et al.2 Based on these results, we will propose formally the concepts of NADA and noise-assisted signal extraction (NASE), and will develop a method called the EEMD, which is based on the original EMD method, to make NADA and NASE possible.

V. MODE MIXING PROBLEM

“Mode mixing” is defined as any IMF consisting of oscillations of dramatically disparate scales, often caused by intermittency of the driving mechanisms. When mode mixing occurs, an IMF can cease to have physical meaning by itself, suggesting falsely that there may be different physical processes represented in a mode. Even though the final time–frequency projection could rectify the mixed mode to some degree, the alias at each transition from one scale to another would irrecoverably damage the clean separation of scales. Such a drawback was first illustrated by Huang *et al.*2 in which the modelled data was a mixture of intermittent high frequency oscillations riding on a continuous low-frequency sinusoidal signal. An almost identical example used by Huang *et al.*2 is presented here in detail as an illustration. The data and its sifting process are illustrated in Fig.1. The data has its fundamental part as a low-frequency sinusoidal wave with unit amplitude. At the three middle crests of the low-frequency wave, high-frequency intermittent oscillations with an amplitude of 0.1 are riding on the fundamental, as panel (a) of Fig.1 shows. The sifting process starts with identifying the maxima (minima) in the data. In this case, 15 local maxima are identified, with the first and the last coming from the fundamental, and the other 13 caused mainly by intermittent oscillations (panel (b)).



Fig(1): The very first step of the sifting process. Panel (a) is the input; panel (b) identifies local maxima (gray dots); panel (c) plots the upper envelope (upper gray dashed line) and low envelope (lower gray dashed line) and their mean (bold gray line); and panel (d) is the difference between the input and the mean of the envelopes

As a result, the upper envelope resembles neither the upper envelope of the fundamental (which is a flat line at one) nor the upper one of the intermittent oscillations (which is supposed to be the fundamental outside intermittent areas).[5] Rather, the envelope is a mixture of the envelopes of the fundamental and of the intermittent signals that lead to a severely distorted envelope mean (the thick gray line in panel (c)). Consequently, the initial guess of the first IMF (panel (d)) is the mixture of both the low-frequency fundamental and the high-frequency intermittent waves, as shown in Fig.2. An annoying implication of such scale mixing is related to unstableness and lack of the uniqueness of decomposition using the EMD. With stoppage criterion given and end-point approach prescribed in the EMD, the application of the EMD to any real data results in a unique set of IMFs, just as when the data is processed by other data decomposition methods. This uniqueness is here referred to as "the mathematical uniqueness," and satisfaction to the mathematical uniqueness is the minimal requirement for any decomposition method. The issue that is emphasized here is what we refer to as "the physical uniqueness." Since real data almost always contains a certain amount of random noise or intermittences that are not known to us, an important issue, therefore, is whether the decomposition is sensitive to noise. If the decomposition is insensitive to added noise of small but finite amplitude and bears little quantitative and no qualitative change, the decomposition is generally considered stable and satisfies the physical uniqueness; and otherwise, the decomposition is unstable and does not satisfy the physical uniqueness. The result from decomposition that does not satisfy the physical uniqueness may not be reliable and may not be suitable for physical interpretation. For many traditional data decomposition methods with prescribed base functions, the uniqueness of the second kind is automatically satisfied. Unfortunately, the EMD in general does not satisfy this requirement due to the fact that decomposition is solely based on the distribution of extrema. To alleviate this drawback, Huang *et al.*2 proposed an intermittence test that subjectively extracts the oscillations with periods significantly smaller than a preselected value during the sifting process. The method works quite well for this example. However, for complicated data with scales variable and continuously distributed, no single criterion of intermittence test can be selected. Furthermore, the most troublesome aspect of this subjectively pre-selected criterion is that it lacks physical justifications and renders the EMD non adaptive. Additionally, mode mixing is also the main reason that renders the EMD algorithm unstable: any small perturbation may result in a new set of IMFs Obviously, the intermittence prevents EMD from extracting any signal with similar scales. To solve these problems, the EEMD is proposed, which will be described in the following sections.

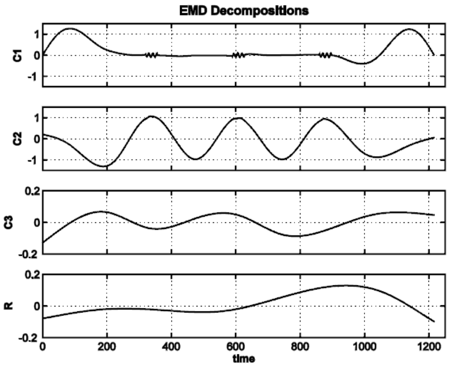


Figure 2: The intrinsic mode functions of the input displayed in Fig.1(a).

VI STEPS FOR ENSEMBLE EMPIRICAL MODE DECOMPOSITION

To improve the accuracy of measurements, the ensemble mean is a powerful approach, where data are collected by separate observations, each of which contains different noise. To generalize this ensemble idea, noise is introduced to the single data set, *x*(*t*), as if separate observations were indeed being made as an analog to a physical experiment that could be repeated many times[2]. The added white noise is treated as the possible random noise that would be encountered in the measurement process. Under such conditions, the *i*th "artificial" observation will be

$$x_i(t) = x(t) + w_i(t) \qquad (3.2)$$

In the case of only one observation, each multiple-observation ensembles is mimicked by adding not arbitrary but different realizations of white noise, $w_i(t)$,that single observation as given in Eq. (3.2). Although adding noise may result in smaller signal-to-noise ratio, the added white noise will provide a relatively uniform reference scale distribution to facilitate EMD; therefore, the low signal–noise ratio does not affect the decomposition method but actually enhances it to avoid the mode mixing. Based on this argument, an additional step is taken by arguing that adding white noise may help to extract the true signals in the data, a method that is termed EEMD, a truly NADA method. Before looking at the details of the new EEMD, a review of a few properties of the original EMD is presented:

(i) The EMD is an adaptive data analysis method that is based on local characteristics of the data, and hence, it catches nonlinear, non stationary oscillations more effectively.

(ii) The EMD is a dyadic filter bank for any white (or fractional Gaussian) noise-only Series.

(iii) When the data is intermittent, the dyadic property is often compromised in the original EMD as the example in Fig.2 shows.

(iv) Adding noise to the data could provide a uniformly distributed reference scale, which enables EMD to repair the compromised dyadic property; and

(v) The corresponding IMFs of different series of noise have no correlation with each other. Therefore, the means of the corresponding IMFs of different white noise series are likely to cancel each other. With these properties of the EMD in mind, the proposed EEMD is developed as follows:

(a) Add a white noise series to the targeted data;

(b) Decompose the data with added white noise into IMFs;

(c) Repeat step 1 and step 2 again and again, but with different white noise series each time; and

(d) Obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result.

The effects of the decomposition using the EEMD are that the added white noise series cancel each other in the final mean of the corresponding IMFs; the mean IMFs stay within the natural dyadic filter windows and thus significantly reduce the chance of mode mixing and preserve the dyadic property. To illustrate the procedure, the data in Fig. 1 is used as an example. If the EEMD is implemented with the added noise having an amplitude of 0.1 standard deviation of the original data for just one trial, the result is given in Fig. 3. Here, the low-frequency component is already extracted almost perfectly.[6] The high-frequency components, however, are buried in noise. Note that high-frequency intermittent signal emerges when the number of ensemble members increases, as Fig. 4 displays. Clearly, the fundamental signal C5 is represented nearly perfect, as well as the intermittent signals, if C2 and C3 are added together. The fact that the intermittent signal actually resides in two EEMD components is due to the average spectra of neighbouring IMFs of white noise overlapping, as revealed by Wu and Huang.6 Thus sometimes, the combination of two adjutant components to form one IMF is necessary. The need for this type of adjustment is easily determined through an orthogonality check. Whenever two IMF components become grossly unorthogonal, one should consider combining the two to form a single IMF component. This provides the first example to demonstrate that the NADA, using the EEMD significantly, improves the capability of extracting signals in the data, and represents a major improvement of the EMD method.

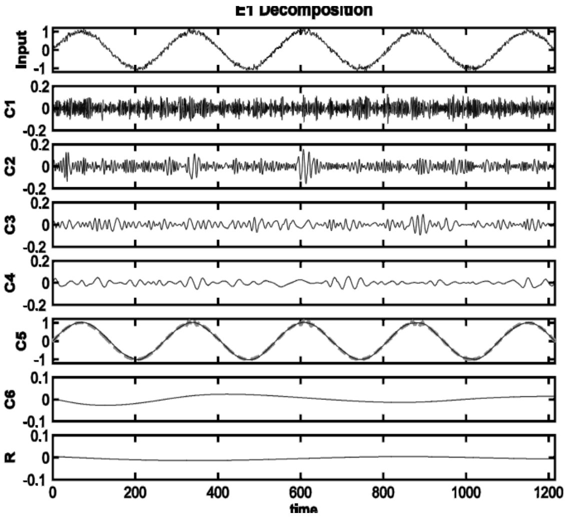


Figure 3: The modified input (the top panel), its intrinsic mode functions (C1–6), and the trend of (*R*). In panel C5, the original input is plotted as the bold dashed gray line for comparison.

VII CONCLUSION

In this study, we have proposed a better and effective method for determining pitch markers in speech signal which operates using the Ensemble Emperical mode Decomposition. Results prove that mode mixing problem can be solved by EEMD algorithm and preserve physical uniqueness of decomposition. Therefore, the EEMD represents a major improvement of the EMD method. Further we can use this algorithm for comparing pitches of different musical instruments. [4],[7]

VIII REFERENCES

[1] Zhaohua Wu nd Norden E Huang , Empirical mode Decomposition : A noise Assisted data analysis method , Advances in adaptive data analysis Vol .1, No.1(2009) 1-41, c , World scientific publishing company.

[2]G.Schlotthauer,M.E Torres and H.L.Rufiner''A new algorithm for instantaneous FO Speech extraction based on ensemble empirical mode decomposition". Proc. Of the 17th European signal processing conference (EUSIPCO 2009), pp . 2347 – 2351, Aug. 2009

[3] D.G. Childers and C.K. Lee." Voice quality factor analysis senthesis and perception". J. Acoust.Soc. Amer, volume 90, pp. 2394-2410, 1991.

[4] Peyman Heydarian , Joshua D. Reiss , Extraction of long trerm structures in musical signals using the empirical mode decom,osition , Proc. Of the 8th Int. conference on digital audio effects (DAFX-05) , Madrid , Spain , September , 20-22, 2005

[5] Jau-Hung Chen and Yung – An Kao , pitch marking based on an adaptable filter and a peak valley estimation method , computational linguistics and Chinese language processing Vol.6, No.2, February 2001 , pp. 31-42 c computational linguistics society ofR.O.C

[6] M.Dziubinski and B.Kostek , High accuracy and Octave error immune pitch detection algoritm rchives of acoustics, 29,1-21(2004).

[7]Mei-Lin Su and Keh-Shih Chuang , An ECG signal enhancement based on improved EMD , PIERS proceedings, Taipei , March 25-28 , 2013.

[8]Thomas Ewender and Beat Pfister , accurate pitch marking for Prosodic modifications of speech segments , INTERSPEECH , 2010.

[9]http://www.worldscientific.com/doi/abs/10.1142/S1793536909000047?journalCode=aada