

國立臺灣大學電機資訊學院電機電信電子產業研發碩士專班

碩士論文

Industrial Technology R&D Master Program in Electrical,
Communication and Electronics Engineering
College of Electrical Engineering and Computer Science
National Taiwan University

Master Thesis

加入人工訊號於經驗模態分解法中

在語音增強上的研究

Study of Empirical Mode Decomposition in Speech
Enhancement with Artificial Additive Signal



張鈞凱

Chun-Kai Chang

指導教授：陳永耀 博士

Advisor: Yung-Yaw Chen, Ph.D.

中華民國 98 年 7 月

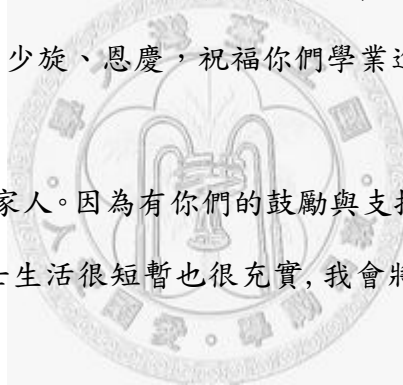
July 2009

致謝

時光匆匆，兩年的研究生生活也到一個段落，心中有無限的感動與感謝，非筆墨可以形容。感謝指導教授陳永耀老師的耐心教導與指引，讓我不會迷失方向。“主動積極、獨立思考”這八字更是我在研究上精神的支柱。感謝口試委員台大機械所顏家鈺教授與台大電機所連豐力教授，於口試期間的寶貴建議。

感謝學長凱翔、易道、士康、傑芳、黃璿，帶領我進入研究生的生活，並適時地給予建議與協助。感謝這兩年來一起努力奮鬥的同學，恕德、鈺堂、文謙、以及育民，因為有你們，我的碩士生活充滿歡笑與活力，祝福大家都有燦爛的前程。感謝學弟，政賢、群富、少旋、恩慶，祝福你們學業進步，畢業順利。

最後，感謝我最愛的家人。因為有你們的鼓勵與支持，讓我能全心全意在學業上奮鬥努力。兩年的碩士生活很短暫也很充實，我會將這經驗化為前進的動力，朝人生下一站邁進。



中文摘要

說話內容常會因背景聲音太大而聽不清楚，如何去將語音訊號中所含的噪音清除或抑制，就是所謂的語音增強技術。傳統在單通道語音增強技術中最常採用溫尼濾波器(Wiener filtering)或是頻譜相減法等方法，但大部分均是在頻域上做處理，經過時頻上的轉換，常有語音失真的情形。

黃鰲博士在 1998 年提出了一種新的訊號分析法希爾伯特-黃轉換(Hilbert Huang Transform, HHT),其方法是將訊號經由經驗模態分解法(Empirical Mode Decomposition, EMD),將資料變化的內部時間尺度作為特徵而分解成多個內建模態函數(Intrinsic Mode Functions, IMF)分量,這些分量經由希爾伯特轉換(Hilbert Huang Transform)可得到有物理意義的瞬時頻率。近年來經驗模態分解法被應用在語音增強上,針對白噪音分解後的特性,可對各個 IMF 分量中所含的噪音量做估測並消除。

在本論文我們針對基於經驗模態分解法的語音增強方法作研究。藉由在含噪訊號中加入人工訊號,噪音主要成分在分解過程中將集中在部份分量,移除這些分量以去除大部分噪音,在配合適應性中間值權重濾波器(Adaptive Center Weighted Average filter, ACWA filter)將語音中殘存的噪音消除。實驗顯示,此方法在低訊噪比下有很好的消噪效果,並且可以保留原先的語音特性。

關鍵字： 經驗模態分解法、消噪、希爾伯特轉換

Abstract

Degradation of the quality of speech caused by the background noise is common in most real situations. How to suppress and remove the noise content in a noisy speech is speech enhancement technique. In traditional signal-channel speech enhancement methods, Wiener filter and spectral subtraction are general methods. But these methods process in frequency domain, the distortion of signal often happen.

A new signal analyzing method, Hilbert-Huang Transform (HHT), was proposed by Norden E. Huang et al. in 1998. With EMD, signal can be decomposed into a finite number of intrinsic mode functions (IMFs) based on the local characteristic time scale of the signal. These IMFs with Hilbert transform obtain meaningful instantaneous frequencies. In recent years, EMD was used on speech enhancement. After EMD of white noise, noise component of each IMF can be estimated then remove it.

In this thesis, we research on speech enhancement with EMD. After adding an artificial signal to noisy signal, most noise component can concentrate on some IMFs. We can remove most noise by throwing away the IMFs. Adaptive center weighted average filter (ACWA filter) is used to whiten the residual noise in speech. These results of experiment show that the method has good performance of de-noising in low SNR situation and reserve the quality of original speech.

Keywords: Empirical Mode Decomposition, De-noising, HHT

Contents

中文摘要	i
Abstract	ii
Contents	iii
List of Figures.....	v
List of Tables	viii
Chapter 1 Introduction.....	1
1.1 Motivation.....	1
1.2 Background.....	1
1.3 Problem Foundation.....	2
1.4 Thesis Organization	2
Chapter 2 Preliminaries	4
2.1 Empirical Mode Decomposition.....	4
2.1.1 Instantaneous Frequency.....	6
2.1.2 Intrinsic Mode Function.....	11
2.1.3 Empirical Mode Decomposition Method.....	13
2.1.4 Hilbert Huang Spectrum	20
2.1.5 Application of EMD	21
2.2 Study of Speech Enhancement Methods.....	23
2.2.1 Spectral Subtraction	23
2.2.2 Wiener Filtering	24
2.2.3 Adaptive Noise Canceling	25
2.3 Speech Enhancement Methods by Empirical Mode Decomposition.....	28
2.3.1 Speech Enhancement Method with Filtering.....	29
2.3.2 Speech Enhancement Method with Thresholding	32
Chapter 3 Speech Enhancement with Additive Signal Base on Empirical Mode Decomposition	34
3.1 Problem Foundation.....	34
3.2 Analysis of Signal with Empirical Mode Decomposition	35
3.2.1 White Noise	35
3.2.2 Sinusoidal Signal	38

3.2.3	Summary	41
3.3	Proposed Speech Enhancement Method with Empirical Mode Decomposition	41
3.3.1	Architecture.....	41
3.3.2	Dynamic Filter	42
3.3.3	Adaptive Center Weighted Average Filter	48
Chapter 4	Experiment Results.....	51
4.1	Test Environment and Noisy Speech Database	51
4.2	Speech Quality Assessment	52
4.3	Experimental Results	53
Chapter 5	Conclusions.....	63
References		64



List of Figures

Figure 2.1	The flowchart of Hilbert-Huang transform [14].....	6
Figure 2.2	The phase plane for the analytic data of $x(t) = \alpha + \sin(t)$. (a) $\alpha = 0$; (b) $\alpha < 1$; (c) $\alpha > 1$ [5].....	10
Figure 2.3	(a) The unwrapped of phase function for the analytic data of $x(t) = \alpha + \sin(t)$. (b) The instantaneous frequency evaluated by equation 2.7.[5].....	10
Figure 2.4	A typical IMF which satisfies limitations: (1) the same numbers of zero crossings and extremes (2) local zero mean.[5]	13
Figure 2.5	The original data	14
Figure 2.6	The upper and lower envelopes (thin line) defined by the local maxima and minima and the mean value (think line) of the upper and lower envelopes. [11].....	15
Figure 2.7	The original data (think line) and h_1 (thin line).[11]	16
Figure 2.8	The first IMF after 12 sifting steps. [11]	18
Figure 2.9	The length-of-day data. [11].....	21
Figure 2.10	(a) The mean IMFs after EMD process (b) The standard deviation of the IMFs after EMD process. [11].....	22
Figure 2.11	The flowchart of spectral subtraction. [13]	23
Figure 2.12	The flowchart of Iterative Wiener filter. [13]	25
Figure 2.13	Adaptive noise cancellations [8].....	27
Figure 2.14	De-noising base on EMD-ACWA scheme [1].....	30
Figure 2.15	Theoretical noise-only model and actual IMF energies with respect to IMF number. [8].....	33

Figure 3.1	The waveform of the white noise.....	35
Figure 3.2	The FFT spectrum of the white noise.....	36
Figure 3.3	The waveform of $IMF_1 \sim IMF_6$ for the white noise.....	37
Figure 3.4	The FFT spectrums of $IMF_1 \sim IMF_6$ for the white noise.....	37
Figure 3.5	(a)The waveform of sinusoidal model , (b)The waveform of IMF_1 (c)The waveform of IMF_2	39
Figure 3.6	(a)The FFT spectrums of sinusoidal model (b) The FFT spectrums of IMF_1 (c) The FFT spectrums of IMF_2	39
Figure 3.7	The plot of ratio of magnitude as a function of f_2	40
Figure 3.8	Block diagram of proposal method	42
Figure 3.9	The flowchart of dynamic filter	42
Figure 3.10	The FFT of spectrum for noisy speech.....	43
Figure 3.11	The waveform of $IMF_1 \sim IMF_7$ for noisy speech.....	44
Figure 3.12	The waveform of $IMF_1 \sim IMF_7$ for noisy speech with an additive signal	45
Figure 3.13	(a)The waveform of the noisy speech (b)The waveform of noisy speech after dynamic filter	46
Figure 3.14	(a)The waveform of the noisy speech (b)The waveform of noisy speech after high-pass filter	47
Figure 3.15	The flowchart of ACWA filter	48
Figure 3.16	Frame blocking to each IMFs	49
Figure 3.17	(a)The waveform of noisy speech after dynamic filter (b) The waveform of the noisy speech after dynamic filter and ACWA filter	50
Figure 4.1	The waveforms of noisy speech producing result(a) Clean speech (b) Military vehicle noise (c) Noisy speech (-5dB).....	54

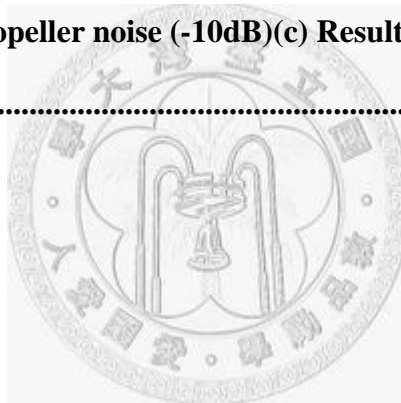
Figure 4.2 The waveforms of passing through two filter result(a) Noisy speech (-5dB) (b) Dynamic filter output(c) ACWA filter output.....55

Figure 4.3 The waveforms of noisy speech producing result(a) Clean speech (b)Military vehicle noise (c) Noisy speech(-10dB).....57

Figure 4.4 The waveforms of passing through two filter result(a) Noisy speech (-10dB)(b) Dynamic filter output(c) ACWA filter output.....58

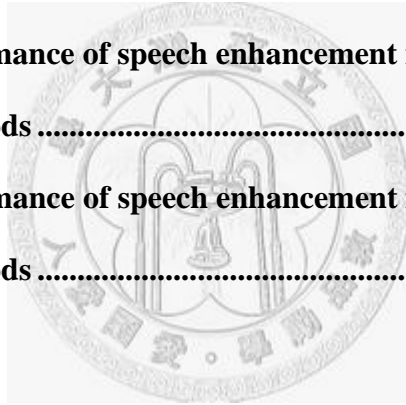
Figure 4.5 The waveforms using spectral spectrum method(a) Clear speech (b) Speech +propeller noise (-10dB)(c) Result base on spectral subtraction method.60

Figure 4.6 The waveforms using wiener filter method(a) Clear speech (b) Speech +propeller noise (-10dB)(c) Result base on wiener filter method.....60



List of Tables

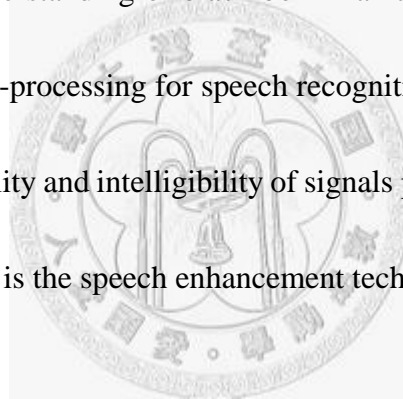
Table 4.1	Parameters used in the speech+ military vehicle noise situation....	54
Table 4.2	Parameters used in the speech+ propeller noise situation	55
Table 4.3	The performance of speech enhancement based on proposed method for experiment I.....	56
Table 4.4	Parameters used in the speech+ military vehicle noise situation....	57
Table 4.5	Parameters used in the speech+ propeller noise situation	58
Table 4.6	The performance of speech enhancement based on proposed method for experiment II	59
Table 4.7	The performance of speech enhancement for male speech base on three methods	61
Table 4.8	The performance of speech enhancement for female speech base on three methods	62



Chapter 1 Introduction

1.1 Motivation

In the real world, degradation of the quality of speech caused by the background noise is common. If speech signal need to record in noisy environment, the recorded voice will contain more background noise. For example, the reporter need to interview in a factory, understanding of black box in an aircraft accident, application on the hearing aid, and pre-processing for speech recognition or speaker identification. In these situations, the quality and intelligibility of signals play an important role. How to improve these problems is the speech enhancement technique.

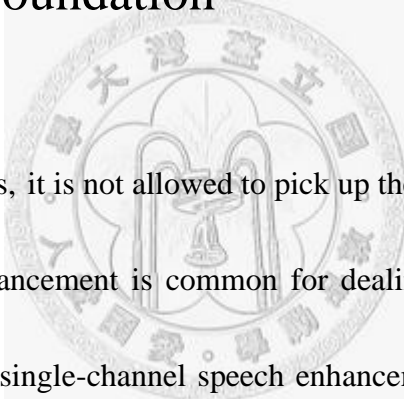


1.2 Background

Speech enhancement techniques can be classified as single-channel, dual-channel or multi-channel enhancement by the number of microphone. In traditional methods, the single-channel speech enhancement is used commonly such as Wiener filter and spectral subtraction. But these methods process in frequency domain, the distortion of signal often happen. And these methods can't have good

performance in low SNR situation. The dual-channel speech enhancement is used commonly such as adaptive noise cancelling. The method has two microphones as a primary and second channel. The primary channel contains speech with additive noise and a reference noise signal correlated to the noise is available in the secondary channel. In the multi-channel techniques, microphones arrays are used to reject the undesired noise signal.

1.3 Problem Foundation



In most real situations, it is not allowed to pick up the second sensor signal. The single-channel speech enhancement is common for dealing with kinds of situation. The traditional method of single-channel speech enhancement, such as Wiener filter and spectral subtraction, has two main problems (1) The aliasing of signal (2) In low situation; the performance of de-noising is bad. According to these problems, we proposed a de-noising method base on empirical mode decomposition to solve.

1.4 Thesis Organization

In Chapter 2, we will introduce Hilbert-Huang Transform (HHT), conventional

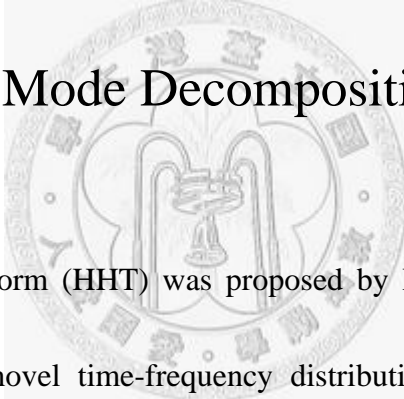
speech enhancement techniques and speech enhancement techniques base on empirical mode decomposition. Chapter 3 describes the proposed structure of speech enhancement system, including Dynamic Filter and Adaptive Center Weighted Average Filter. In Chapter 4, the experiment results of the proposed speech enhancement system are shown. Finally, conclusions are given in Chapter 5.



Chapter 2 Preliminaries

Speech enhancement is a popular problem. Many new methods are proposed in recent year. Section 2.1 introduces history and application of empirical mode decomposition. Section 2.2 introduces conventional speech enhancement methods. Section 2.3 introduces new speech enhancement method with empirical mode decomposition.

2.1 Empirical Mode Decomposition



Hilbert-Huang Transform (HHT) was proposed by Norden E. Huang et al. in 1998 [5]. It provides a novel time-frequency distribution for analyzing physical measurements. One of the main tools used in time-frequency signal processing is short time Fourier transform (STFT). STFT is the most basic method to represent the non-stationary property of the signal. It can be written as

$$X(m, w) = \sum_{n=-\infty}^{\infty} x(n)w(n-m)e^{-jwn} \quad (2.1)$$

By sliding the window along time axis, we can get a time-frequency distribution. The STFT starts with a prior-defined basis and convolves the signal with the basis to

get amplitude and frequency. Such an approach has a great advantage of having a solid mathematical foundation. Unfortunately, within the solid mathematical foundation, the method can not be adaptive. It is restricted by linear and stationary assumptions. Natural phenomena measurements are essentially nonlinear and non-stationary.

HHT is a different kind of time-frequency distribution. It consists of two parts: (1) Empirical Mode Decomposition (EMD), and (2) Hilbert Transform. With EMD, any complicated data set can be decomposed into a finite number of intrinsic mode functions (IMFs). An IMF admits well-behaved Hilbert transforms. Since the decomposition is based on the local characteristic time scale of the data, it is suitable to analyze the nonlinear and non-stationary signals. With Hilbert transform, the IMFs yield instantaneous frequencies as functions of time that gives sharp identification of imbedded structures. The Hilbert-Huang transform process can be shown in Figure 2.1.

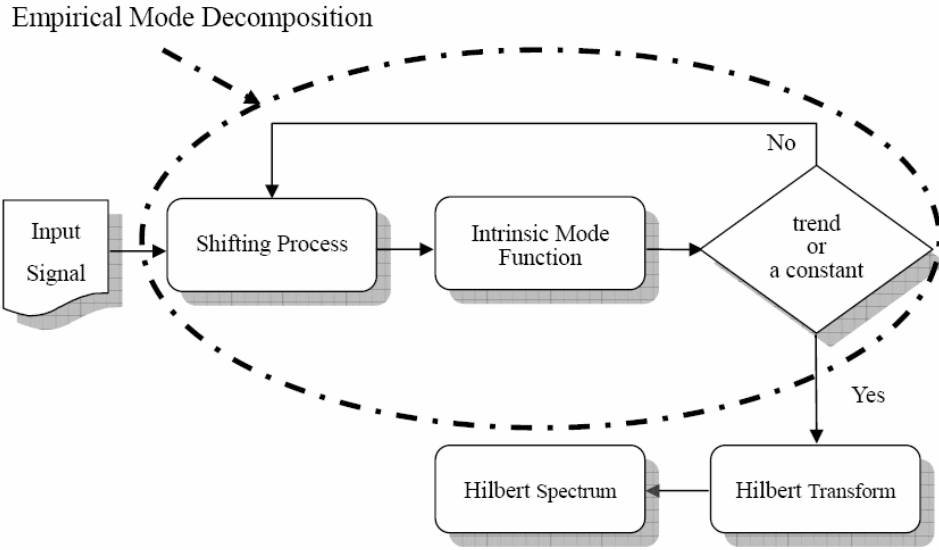


Figure 2.1 The flowchart of Hilbert-Huang transform [14]

2.1.1 Instantaneous Frequency

In 1930 Denis Gabor defined a generalization of the Euler's formula in the form of a complex analytic function. Given a real signal $u(t)$, the analytic function $\varphi(t)$ can be defined as:

$$\varphi(t) = u(t) + jv(t) \quad (2.2)$$

In which the imaginary part, $v(t)$, is the Hilbert transform of $u(t)$, given by

$$v(t) = H[u(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} d\tau \quad (2.3)$$

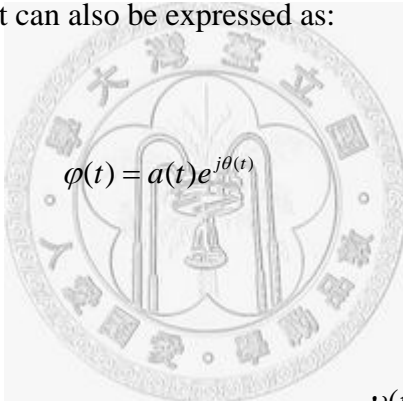
The above definition of the Hilbert transform can be written in convolution notation

as:

$$v(t) = u(t) * \frac{1}{\pi t} \quad (2.4)$$

Gabor used the analytic signal to introduce the Hilbert transform to signal processing for one-dimensional time series $u(t)$. When the Hilbert transform is applied to a general time series $u(t)$, we obtain another function of time $v(t)$.

The complex conjugate pair of functions $\{u(t), v(t)\}$ then comprised the above analytic function $\varphi(t)$ that can also be expressed as:



$$\varphi(t) = a(t)e^{j\theta(t)} \quad (2.5)$$

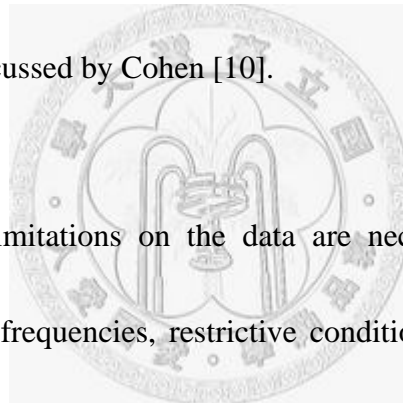
In which

$$a(t) = [u^2(t) + v^2(t)]^{1/2}, \quad \theta(t) = \arctan\left(\frac{v(t)}{u(t)}\right) \quad (2.6)$$

Here $a(t)$ is the instantaneous amplitude, and $\theta(t)$ is the phase function. Since Hilbert transform $v(t)$ is defined as the convolution of $u(t)$ and $1/\pi t$, it emphasizes the local properties of $u(t)$ even though the transform is global. In equation (2.5), the polar coordinate expression further clarifies the local nature of this representation. With equation (2.5), the instantaneous frequency of $u(t)$ is defined as

$$w(t) = \frac{d\theta(t)}{dt} \quad (2.7)$$

The definition of instantaneous frequency appears to be local for it is defined through differentiation rather than integration. The resulting instantaneous frequency may be able to describe nonlinear and non-stationary signals. There is still considerable controversy in defining the instantaneous frequency. Although Hilbert transform is valid under a very general condition, for the instantaneous frequency derived from an arbitrary time series to make physical sense, the function has to be ‘mono- component’ as discussed by Cohen [10].



In principle, some limitations on the data are necessary. In order to obtain meaningful instantaneous frequencies, restrictive conditions have to be imposed on the data: for any function to have a meaningful instantaneous frequency, the real part of its Fourier transform has to have only positive frequency. Here, some examples will be considered to illustrate the necessary limitations to obtain the meaningful instantaneous frequency. For a simple function as

$$x(t) = \sin t \quad (2.8)$$

Its Hilbert transform is $\cos(t)$. The polar coordinate representation of the analytic data,

the instantaneous frequency, and instantaneous amplitude of $x(t)$ can be obtained by equation (2.5)-(2.7). Then, the phase plot of $x - y$ is a circle of unit radius as in figure 2.2. The phase function is a straight line as shown in figure 2.3 (a) and the instantaneous frequency, shown in figure 2.3 (b), is a constant. If we add an arbitrary constant α to $x(t)$ as below

$$x(t) = \alpha + \sin t \quad (2.9)$$

The phase plot of $x - y$ is still a circle of unit radius, but the centre of the circle will be shifted by the amount of α as illustrated in figure 2.2. If $\alpha < 1$, the centre is still within the circle. Under the condition, the phase function and the instantaneous frequency will be very different as shown in figure 2.3 (a) and (b). If $\alpha > 1$, the centre is outside the circle. The phase function and the instantaneous frequency assume negative values as shown in figure 2.3 (a) and (b), which are meaningless.

These examples illustrate that, for a simple signal such a sine function, the instantaneous frequency can be defined only if we restrict the function to be symmetric locally with respect to the zero mean level.

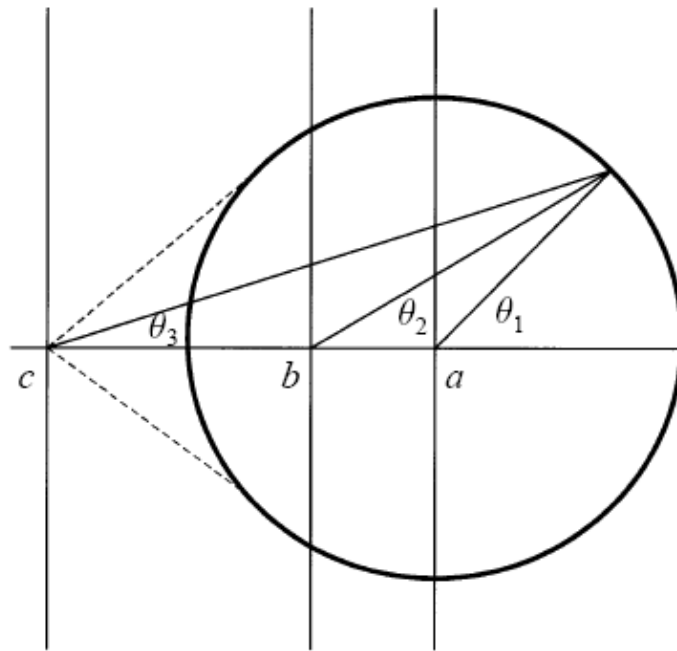


Figure 2.2 The phase plane for the analytic data of $x(t) = \alpha + \sin(t)$. (a) $\alpha = 0$; (b) $\alpha < 1$; (c) $\alpha > 1$ [5]

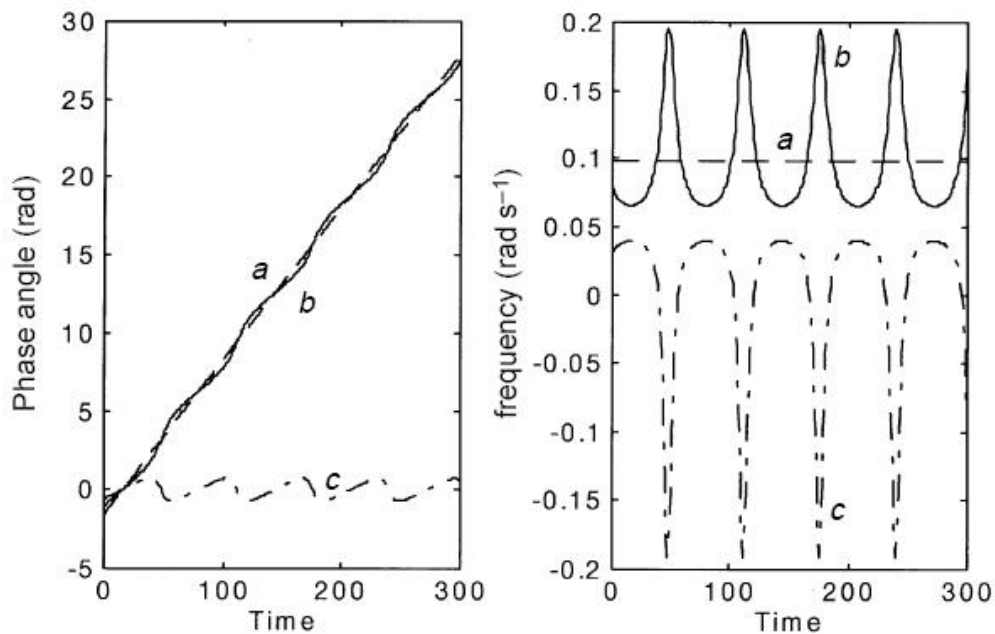
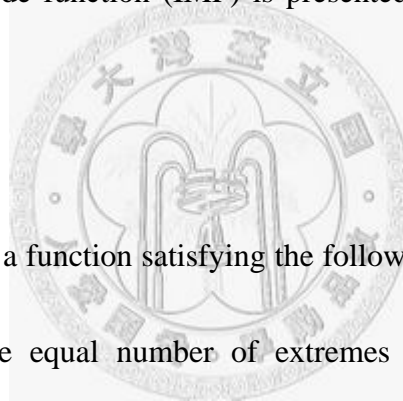


Figure 2.3 (a) The unwrapped of phase function for the analytic data of $x(t) = \alpha + \sin(t)$. (b) The instantaneous frequency evaluated by equation 2.7.[5]

2.1.2 Intrinsic Mode Function

The simple examples given above provide more physical interpretation of the restrictive conditions; they also suggest a practical way to decompose the data so that the components all satisfy the conditions. The necessary conditions to define a meaningful instantaneous frequency are that the functions are locally symmetric and have the same numbers of zero crossings and extremes. Therefore, a class of functions designated as intrinsic mode function (IMF) is presented with the following formal definition.

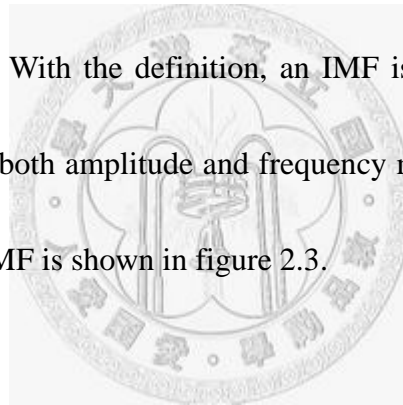


An IMF is defined as a function satisfying the following limitations: (1) Original data series must have the equal number of extremes and zero crossings or the difference of them must not be larger than one; and (2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The first condition is similar to the traditional narrow band requirements for a stationary Gaussian process. The second condition is a new idea. It modifies the global requirement to a local one. As a result, the instantaneous frequency will not have the undesired fluctuations induced by asymmetric waveforms. For a

non-stationary data, the ‘local mean’ involves a ‘local time scale’ to compute the mean, which is impossible to define. Thus, Huang uses the local mean of envelopes defined by the local maxima and the local minima to force the local symmetry instead. This is an approximation to avoid the definition of a local average time scale.

The name ‘intrinsic mode function’ is adopted because it represents the oscillation mode imbedded in the data. With the definition, the IMF in each cycle, defined by the zero crossings, involved only one mode of oscillation, no complex riding waves are allowed. With the definition, an IMF is not restricted to a narrow band signal, and it can be both amplitude and frequency modulated. In fact, it can be non-stationary. A typical IMF is shown in figure 2.3.



For a complicated data, we can obtain more than one instantaneous frequency at a time locally. We will introduce the empirical mode decomposition method to reduce the data into IMFs.

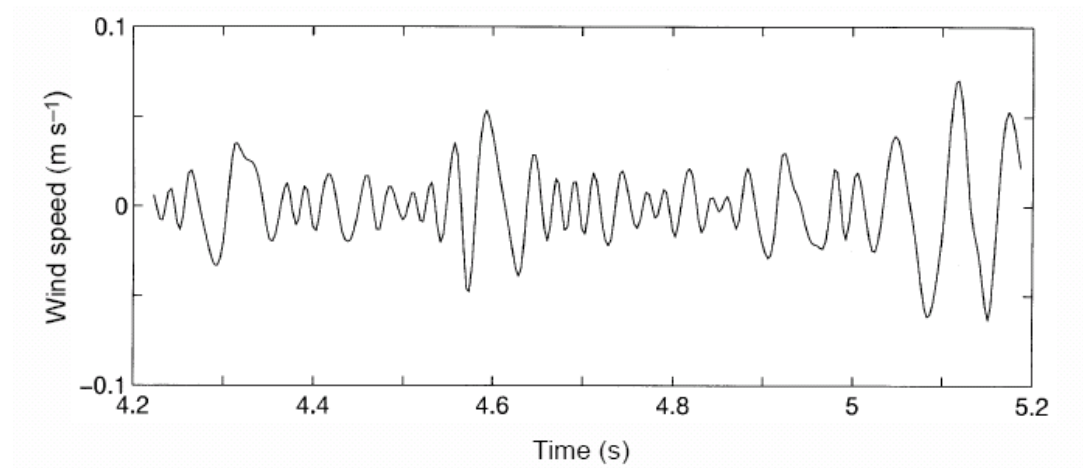


Figure 2.4 A typical IMF which satisfies limitations: (1) the same numbers of zero crossings and extremes (2) local zero mean.[5]

2.1.3 Empirical Mode Decomposition Method

To apply the concept of the instantaneous frequency to real physical signals, it is necessary preprocessing to decompose the signal into mono component contributions.

These components called intrinsic mode function (IMF) admit well-behaved Hilbert transform and allow the calculation of a meaningful instantaneous frequency. In contrast to all of the previous methods, this new method is intuitive, direct, and adaptive, with a posteriori-defined basis derived from the data.

The decomposition is based on assumptions: (1) the signal has at least two extremes – one maximum and one minimum; (2) the characteristic time scale is defined by the time lapse between the extremes; and (3) if the data were totally devoid

of extremes but contained only inflection points, then it can be differentiated once or more times to reveal the extremes.

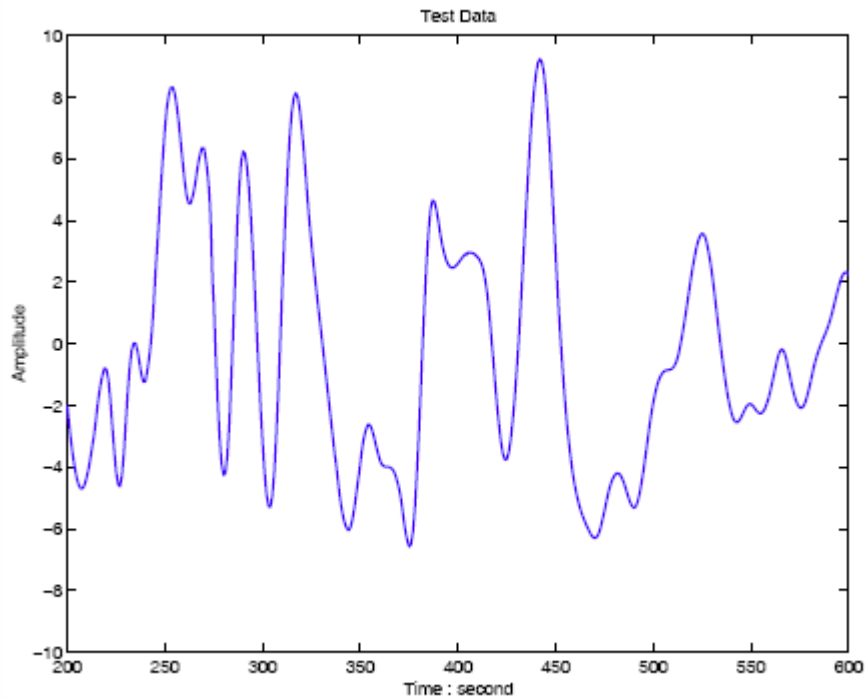


Figure 2.5 The original data. [11]

An IMF represents a simple oscillatory mode as a counterpart to the simple harmonic function, but it is much more general: instead of constant amplitude and frequency, the IMF can have a variable amplitude and frequency as functions of time. We can decompose any function as follows: take the test data as given in Figure 2.4; locate all the local extremes, then connect all the local maxima by a cubic spline function as shown in the upper envelope. Repeat the procedure for the local minima to form a lower envelope. The mean value of upper and lower envelopes is designated as

m_1 , shown in Figure 2.5, and the difference between the data and m_1 is the first component h_1 shown in Figure 2.6; i.e.,

$$h_1 = x(t) - m_1 \quad (2.10)$$

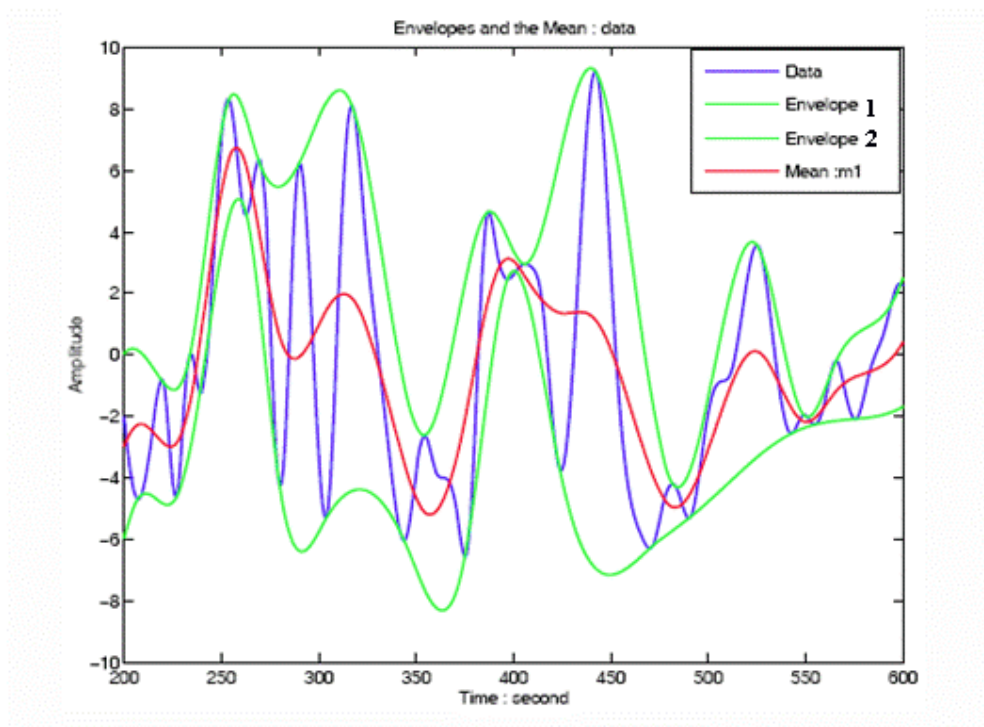


Figure 2.6 The upper and lower envelopes (thin line) defined by the local maxima and minima and the mean value (thick line) of the upper and lower envelopes. [11]

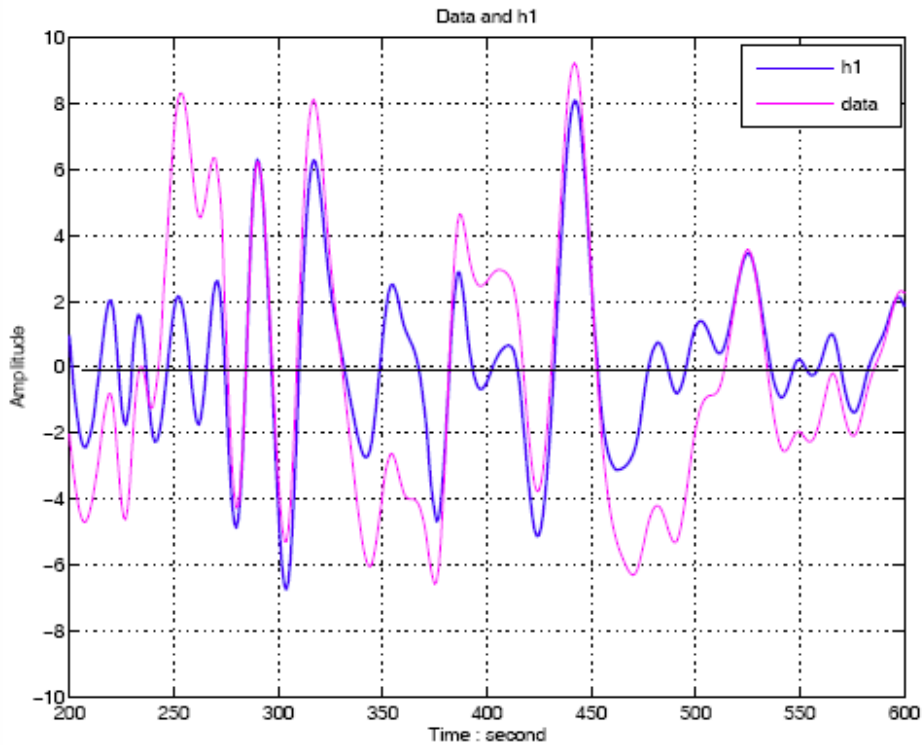


Figure 2.7 The original data (thick line) and h_1 (thin line).[11]

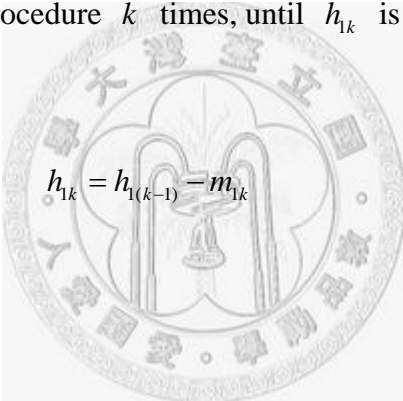
Ideally, h_1 should satisfy the definition of an IMF by construction of h_1 described above. In changing the local zero from a rectangular to a curvilinear coordinate system some inflection points could become additional extremes. New extremes generated this way actually reveal the hidden modes missed in the initial treatment. The sifting process can recover signals representing low amplitude riding waves with repeated siftings.

The sifting process serves two purposes: to eliminate riding waves and to make the wave profiles more symmetric. While the first condition is absolute necessary for

Hilbert transform to give a meaningful instantaneous frequency, the second condition is also necessary in case the neighboring wave amplitudes having too large a disparity. As a result, the sifting process has to be repeated many times to reduce the extracted signal an IMF. In the subsequent sifting process, h_1 is treated as the data for the next round of sifting; therefore,

$$h_{11} = h_1 - m_{11} \quad (2.11)$$

We can repeat the sifting procedure k times, until h_{1k} is an IMF; that is



$$h_{1k} = h_{1(k-1)} - m_{1k} \quad (2.12)$$

Then, it is designated as

$$c_1 = h_{1k} \quad (2.13)$$

The first IMF component from the data is shown in Figure 2.7.

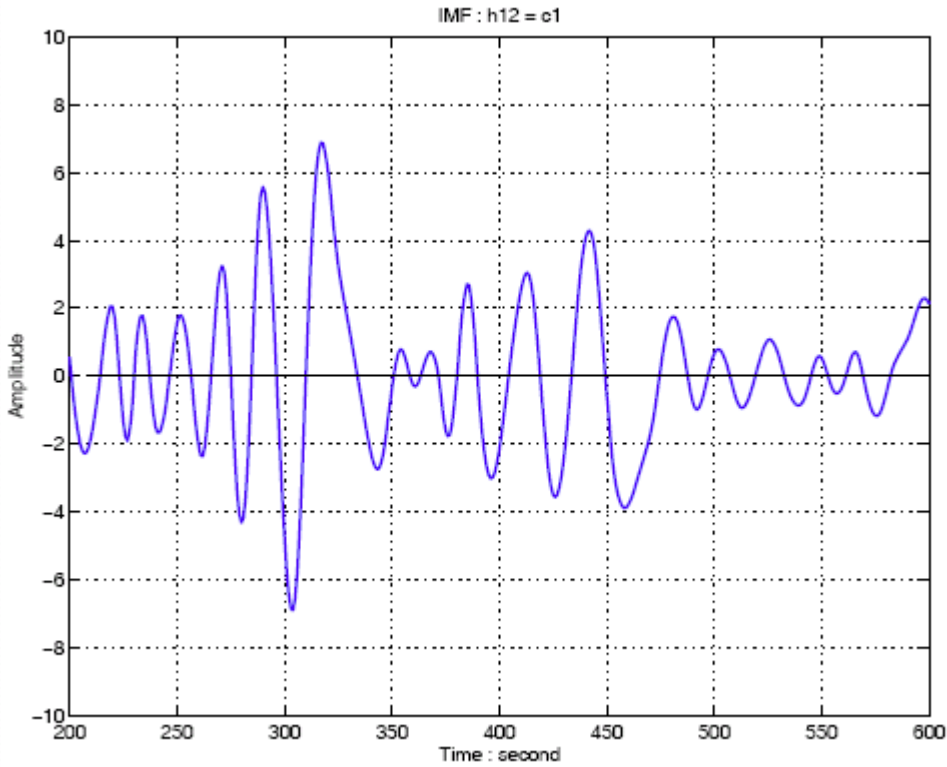


Figure 2.8 The first IMF after 12 sifting steps. [11]

To guarantee that the IMF retains physical sense of both amplitude and frequency modulations, a critical decision has to be made: when to stop. The threshold can be assigned for the variance between two consecutive results:

$$SD = \sum_{t=0}^T \left[\frac{|(h_{1(k-1)}(t)) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \right] \quad (2.14)$$

A typical value of SD can be set between 0.2 and 0.3.

Overall, c_1 should contain the finest scale or the shortest period component of the signal. It follows that c_1 can be separated from the rest of the data by

$$r_1 = x(t) - c_1 \quad (2.15)$$

Since the residue, r_1 , contains all longer period variations in the data, it is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated to all the subsequent r_j 's, and the result is

$$\begin{aligned}
 r_2 &= r_1 - c_2 \\
 &\dots \\
 r_n &= r_{n-1} - c_n
 \end{aligned}
 \tag{2.16}$$

The sifting process should stop when the residue, r_n , becomes a constant, a monotonic function, or a function contains only a single extreme, from which no more IMF can be extracted. By summing up Equations (2.15) and (2.16), we finally obtain

$$x(t) = \sum_{j=1}^n c_j + r_n
 \tag{2.17}$$

Thus, sifting process produces a decomposition of the data into n -intrinsic modes, and a residue, r_n . When apply the EMD method, a mean or zero reference is not required; EMD needs only the locations of the local extreme. The sifting process generates the zero reference for each component. Without the need of the zero reference, EMD avoids the troublesome step of removing the mean values for the large non-zero mean.

2.1.4 Hilbert Huang Spectrum

Once the intrinsic mode function components have been determined, one will have no difficulty in applying the Hilbert transform to each IMF component. After performing the Hilbert transform on each IMF component, the original signal can be reconstructed as

$$x(t) = \sum_{k=1}^n a_k(t) e^{j \int w_k(t) dt} \quad (2.18)$$

where n is the number of decomposed IMFs, $a_k(t)$, $w_k(t)$ are the instantaneous amplitude and frequency of each component, respectively, and the residue, $r_n(t)$, is omitted. Equation 2.18 allows the instantaneous amplitude and frequency to be represented as functions of time in a three-dimensional plot. This frequency-time distribution of the amplitude is known as the Hilbert spectrum, $H(w, t)$.

It should be emphasized that the Hilbert spectrum allows for time-varying amplitudes and frequencies, and the expansion is adaptive. The basic functions are directly extracted from data, and it is suitable for analyzing nonlinear and non-stationary signals. Moreover, we will extract features from the results of the Hilbert spectrum.

2.1.5 Application of EMD

The components of the EMD are usually physically meaningful, for the characteristic scales are defined by the physical data. Huang collected length-of-day data and analyzed the data. The length-of-day data can be shown in Figure 2.8, which measure the deviation from the fixed cycle of 24 hours. The mean and the standard deviation of the IMFs, given in Figure 2.9, were obtained after EMD method. Thus, these IMF results are physically meaningful. The first component c_1 represents the very short period of perturbation caused by large-scale storms to the earth's rotational speed. The second component c_2 represents the half-monthly tides; the eighth component c_8 , the annual tidal variations. We believe the components of the EMD which are also physically meaningful in speech processing.

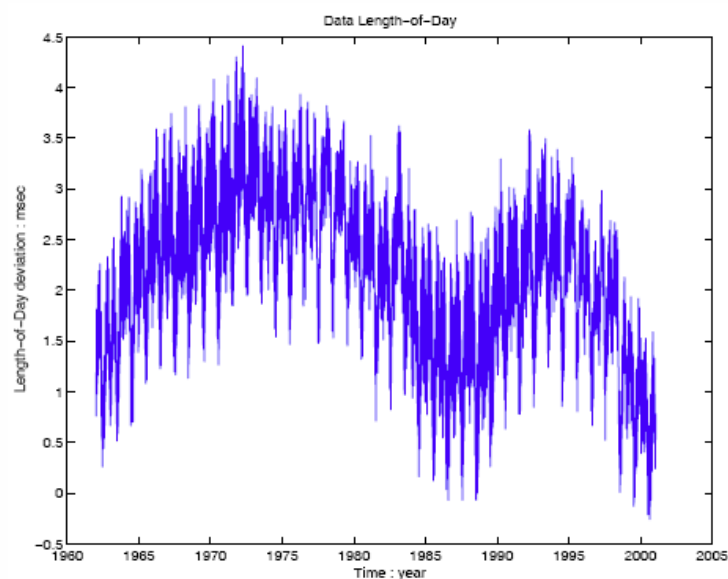


Figure 2.9 The length-of-day data. [11]

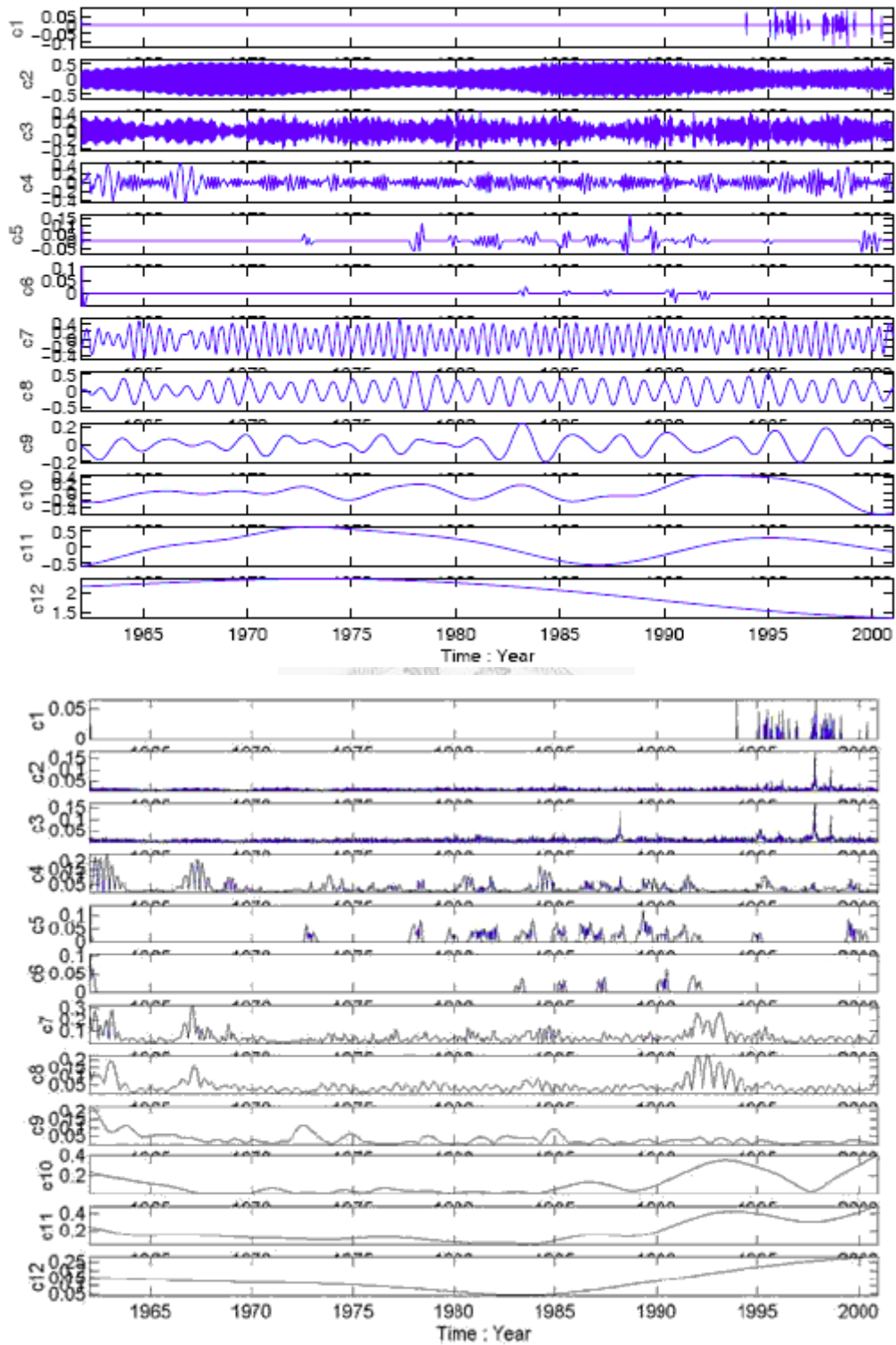


Figure 2.10 (a) The mean IMFs after EMD process (b) The standard deviation of the IMFs after EMD process. [11]

2.2 Study of Speech Enhancement Methods

2.2.1 Spectral Subtraction

Spectral Subtraction has two basic assumptions: (1) The noise is uncorrelated with the speech signal. (2) The noise is short-term stationary, so we can estimate the noise during silent frame. The noise spectrum is subtracted from the transformed noisy speech.

$$\hat{S}_s(\omega; m) = \begin{cases} \left[|S_y(\omega; m)|^a - k |\hat{S}_d(\omega; m)|^a \right]^{\frac{1}{a}} e^{j\varphi_y(\omega; m)} & \text{if } |S_y(\omega; m)|^a > |\hat{S}_d(\omega; m)|^a \\ 0 & \text{otherwise} \end{cases} \quad (2.19)$$

where $\hat{S}_s(\omega; m)$ is the short-term spectrum of enhanced speech, $S_y(\omega; m)$ is the short-term spectrum of noisy speech, $\hat{S}_d(\omega; m)$ is the short-term spectrum of estimated noise, $\varphi_y(\omega; m)$ is the phase of noisy speech. k, a is the vary parameters.

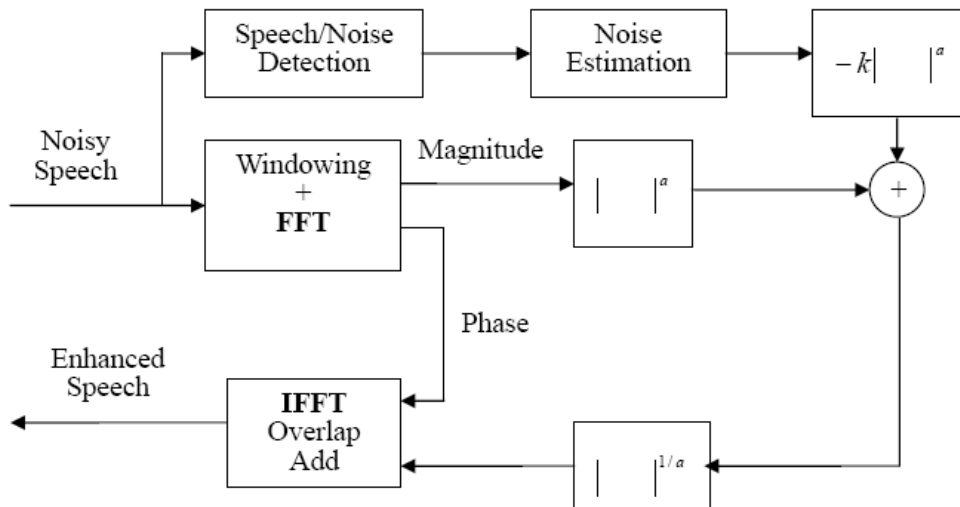


Figure 2.11 The flowchart of spectral subtraction. [13]

Above-mentioned two methods are useful in high SNR situation. But SNR is low; the error of voice activity detection often leads to incorrect result. The performance of de-noising is bad. Moreover, signal aliasing is a problem for transforming from frequency to time domain.

2.2.2 Wiener Filtering

Wiener filter has the same two assumptions with spectral subtraction. In Figure 2.12, $H^+(\omega; m, k)$ is a gain function, the function is a number between 0 and 1. The main idea is: When noisy speech pass through the filter, the noise component of noisy speech is filtered.

$$H^+(\omega; m, r) = \frac{\hat{\Gamma}_s(\omega; m, r)}{\hat{\Gamma}_s(\omega; m, r) + \hat{\Gamma}_d(\omega; m, r)} \quad (2.20)$$

where $\hat{\Gamma}_d(\omega; m, r)$ is the power density spectrum of estimated noise, $\hat{\Gamma}_s(\omega; m, r)$ is the power density spectrum of estimated speech, k is the number of iteration. For getting $\hat{\Gamma}_s(\omega; m, r)$, the method of all-pole model is used.

$$\hat{\Gamma}_s(\omega; m, r) = \frac{g_r^2}{\left| 1 - \sum_{i=1}^M \hat{a}(i; m) e^{-j\omega i} \right|^2} \quad (2.21)$$

where $\hat{a}(i; m)$ is the linear prediction coefficient of speech, i is the i^{th} order coefficient, m is the m^{th} frame, g_r can be get from Parseval's method. If k is too big, signal

aliasing will happen. In general, r is set to four.

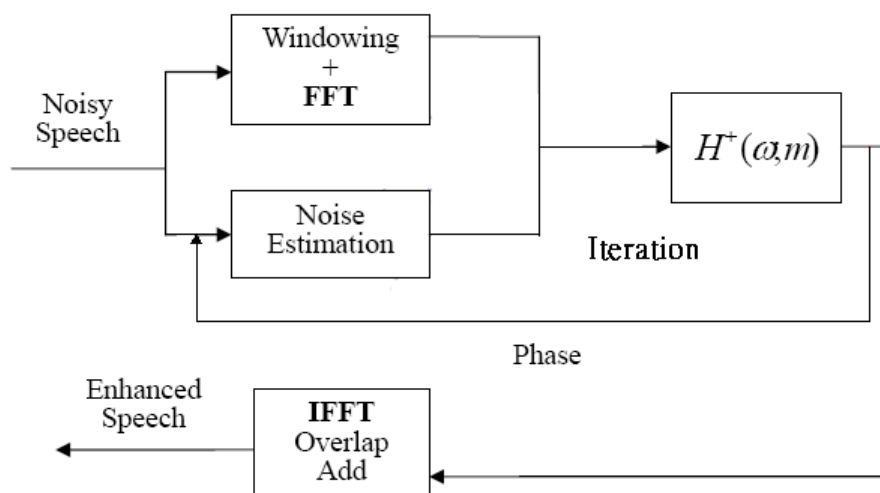


Figure 2.12 The flowchart of Iterative Wiener filter. [13]

2.2.3 Adaptive Noise Canceling

As the name implies, adaptive noise canceling relies on the use of noise canceling by subtracting noise from a received signal, an operation controlled in an adaptive manner for the purpose of improved signal-to-noise ratio. Basically, an adaptive noise canceller is dual-input, closed-loop adaptive control system as illustrated in Figure 2.12. The two inputs of the system are derived from a pair of sensors: a primary sensor and a reference sensor. Specifically, we have the following:

The primary sensor receives an information-bearing signal $x(n)$ corrupted by

additive noise $n_0(n)$, as shown by

$$d(n) = x(n) + n_0(n) \quad (2.22)$$

The signal $x(n)$ and the noise $n_0(n)$ are uncorrelated with each other; that is,

$$E[x(n)n_0(n-k)] = 0, \quad \text{for all } k \quad (2.23)$$

Where $x(n)$ and $n_0(n)$ are assumed to be real valued.

The reference sensor receives a noise $n_1(n)$ that is uncorrelated with the signal $x(n)$ but correlated with the noise $n_0(n)$ in the primary sensor output in an unknown way; that is,

$$E[x(n)n_1(n-k)] = 0, \quad \text{for all } k \quad (2.24)$$

and

$$E[n_0(n)n_1(n-k)] = p(n) \quad (2.25)$$

Where, as before, the signals are real valued and $p(n)$ is an unknown cross-correlation for lag k .

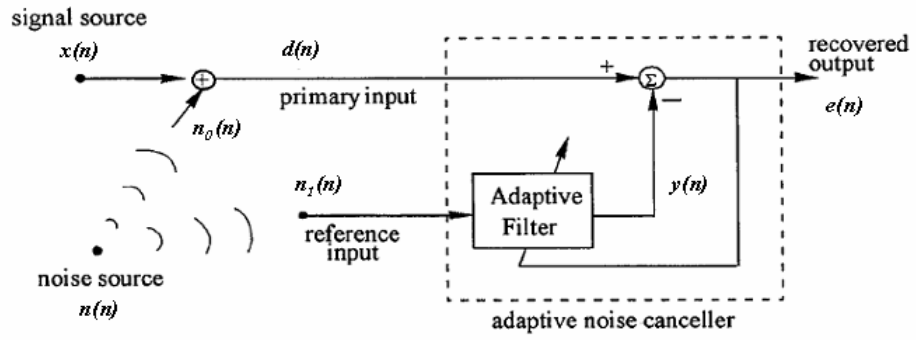


Figure 2.13 Adaptive noise cancellations [8].

The reference signal $x(n)$ is processed by an adaptive filter to produce the output signal:

$$y(n) = \sum_{k=0}^{M-1} \hat{w}_k(n) n_1(n-k) \quad (2.26)$$

where the $\hat{w}_k(n)$ are the adjustable tap weights of adaptive filter. The filter output $y(n)$ is subtracted from the primary signal $d(n)$, serving as the “desired” response for the adaptive filter. The error signal is defined by

$$e(n) = d(n) - y(n) \quad (2.27)$$

Thus, substituting Equation (2.22) in (2.27), we get

$$e(n) = s(n) + n_0(n) - y(n) \quad (2.28)$$

The error signal is in turn used to adjust the tap weights of the adaptive filter, and

the control loop around the operations of filtering and subtraction is thereby closed.

Note that the information-bearing signal $x(n)$ is indeed part of the error signal $e(n)$, as indicated in Equation (2.28).

The error signal $e(n)$ constitutes the overall system output. From Equation (2.28) we see that the noise component in the system output is $n_0(n) - y(n)$. Now, the adaptive filter attempts to minimize the mean-square value of the error signal $e(n)$.

The information-bearing signal $x(n)$ is essentially unaffected by the adaptive noise canceller. Hence, minimizing the mean-square value of the error signal $e(n)$ is equivalent to minimizing the mean-square value of the output noise $n_0(n) - y(n)$.

With the signal $x(n)$ remaining essentially constant, it follows that the minimization of the mean-square value of the error signal is indeed the same as the maximization of the output signal-to-noise ratio of the system.

2.3 Speech Enhancement Methods by Empirical Mode Decomposition

Traditional speech enhancement schemes are based on linear methods. Furthermore, real signals are often nonlinear and non-stationary. Performances of

de-noising are limited. Recently, the EMD has been introduced for analyzing data from nonlinear and non-stationary processes like speech. The major advantage is that the basis functions are derived from the signal itself. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal, call intrinsic mode function, starting from high-frequency IMFs to low-frequency IMFs. Base on the characteristic of EMD, most speech signal are often concentrated on lower-frequency components, and decrease toward high-frequency modes. We can reconstruct the signal with only few IMFs that speech dominated using an energy criterion. Further, in recent research of EMD, two main approaches for speech enhancement are proposed: filtering and thresholding.

2.3.1 Speech Enhancement Method with Filtering

Khaldi K. et al proposed an adaptive speech de-noising scheme combine EMD and adaptive center weighted average (ACWA) filter [2]. The ACWA filter is basically used in the image enhancement domain. Main idea is the IMFs are less noisy than the noisy speech for speech enhancement. By combination of two methods, we can have better performance than only use ACWA filter.

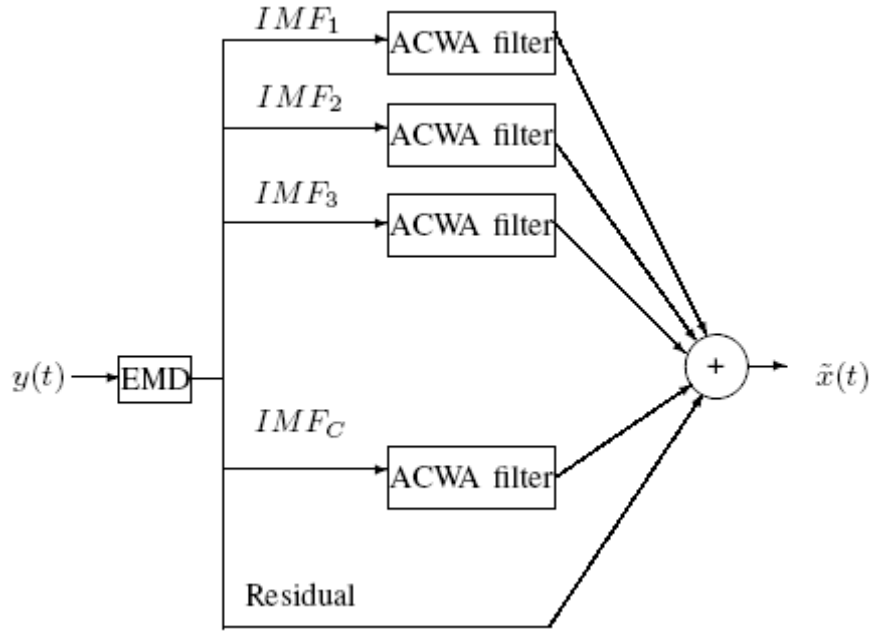


Figure 2.14 De-noising base on EMD-ACWA scheme [1].

In figure 2.14, the noisy signal $y(t)$ described by an additive model speech signal $x(t)$ and noise speech $b(t)$ is given by :

$$y(t) = x(t) + b(t) \quad (2.29)$$

The noisy signal is decomposed into a sum of IMFs as follows:

$$y(t) = \sum_{j=1}^c IMF_j(t) + r_c(t) \quad (2.30)$$

The extracted IMFs include the noise since each IMF, indexed by j , can be approximated as follows:

$$IMF_j(t) = f_j(t) + b_j(t) \quad (2.31)$$

Finally, the estimated signal, $\tilde{x}(t)$, is given by :

$$\tilde{x}(t) = \sum_{j=1}^c \tilde{f}_j(t) + r_c(t) \quad (2.32)$$

where $\tilde{f}_j(t)$ is a temporal processing using ACWA filter . The de-noising of the IMF

by the ACWA filter is given as follows [1]:

$$\tilde{f}_j(t) = \begin{cases} F_{mean} + W_j(IMF_j(t) - F_{mean}), & \text{if } F_{var} \geq \sigma_j^2 \\ F_{mean}, & \text{otherwise} \end{cases} \quad (2.33)$$

where F_{mean} and F_{var} denote respectively the average and the variance of the IMF

computed over a sliding window of length L, and σ_j^2 designates the variance of noise

contained in the IMF indexed by j.

The noise level σ_j is estimated as in [3],[4] as following:

$$\sigma_i = 1.4826 * Median\{|IMF_i(n) - Median\{IMF_i(n)\}|\} \quad (2.34)$$

In this thesis, we use the method to decrease the residual white noise.

2.3.2 Speech Enhancement Method with Thresholding

Yannis K. et al proposed an alternative de-noising procedure inspired by wavelet thresholding [6]. Main idea is the white noise contained in each IMF having linear distribution of energy in each mode.

The noisy signal is decomposed into a sum of IMFs as follows:

$$y(t) = \sum_{j=1}^C IMF_j(t) + r_c(t) \quad (2.35)$$

The noise-only IMF energies can be approximated according to

$$\hat{E}_k = \frac{E_1^2}{\beta} \rho^{-k}, \quad k = 2, 3, 4, \dots \quad (2.36)$$

where E_1^2 is the energy of the first IMF and, β , ρ are parameters .In figure 2.15 noise-only model and actual IMF energy are display.Multiples of the IMF-dependent universal thresholds is setting base on the energy of IMFs.

$$T_k = C \sqrt{E_k 2 \ln N} \quad (2.37)$$

where C is constant, N is length of data

$$\tilde{f}_j(t) = \begin{cases} \text{sgn}(\text{IMF}_j(t))(|\text{IMF}_j(t)| - T_j), & |\text{IMF}_j(t)| > T_j \\ 0, & |\text{IMF}_j(t)| \leq T_j \end{cases} \quad (2.38)$$

where $\tilde{f}_j(t)$ is a temporal processing using soft threshold.

The estimated signal, $\tilde{x}(t)$, is given by :

$$\tilde{x}(t) = \sum_{j=M_1}^{M_2} \tilde{f}_j(t) + \sum_{j=M_2+1}^C f_j(t) \quad (2.39)$$

where the introduction of M_1, M_2 parameters and gives us flexibility on the exclusion of the noisy low-order IMFs and on the optional threshold of the high-order ones, which in white Gaussian noise conditions contain little noise energy.

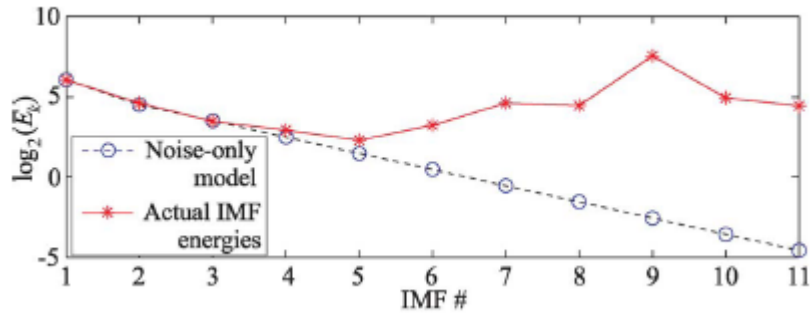


Figure 2.15 Theoretical noise-only model and actual IMF energies with respect to IMF number. [8].

Chapter 3 Speech Enhancement with Additive Signal Base on Empirical Mode Decomposition

In the chapter, a novel speech enhancement method is proposed. By adding an artificial signal to empirical mode decomposition of noisy speech, the background noise can be separated from the noisy speech. Then adaptive center weighted average filter (ACWA filter) is used to whiten the residual noise. These procedures will be introduced in follow sections.

3.1 Problem Foundation

In recent study of speech enhancement base on EMD, most background noise is white noise. The performance of result is well just for the case of white noise. In fact, the white noise doesn't exist in real situation. A real noise like a rotation of motor has often dominant frequency signal and white noise arising from oscillation of air. Therefore we use signals from kinds of real situation as background noises to test performance of de-noising in this thesis. The code of EMD comes from matlab central [18].

3.2 Analysis of Signal with Empirical Mode Decomposition

Decomposition

3.2.1 White Noise

For understanding EMD of characteristics, we use the algorithm on some artificial signals to understand the action of each IMFs. White noise is a random signal with a flat power spectral density. Figure 3.1 and 3.2 shows the white noise and the FFT spectrum.

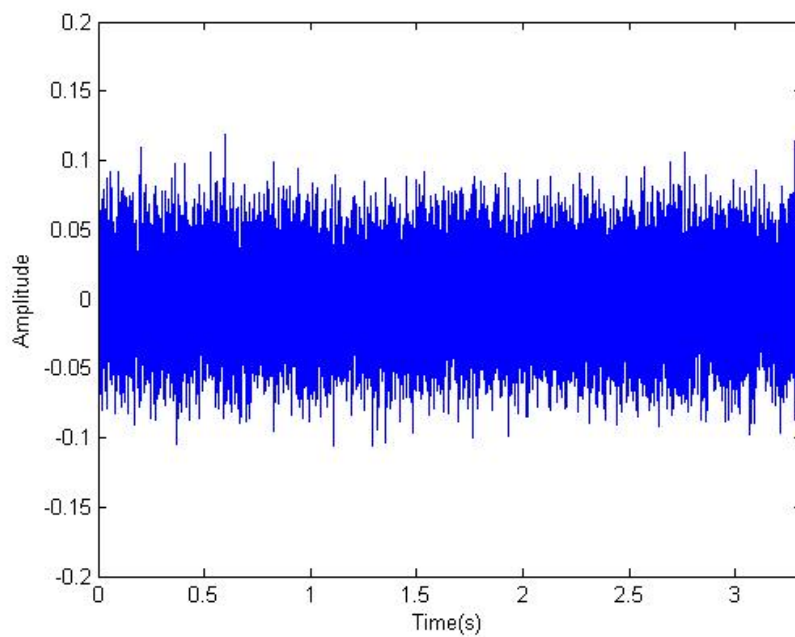
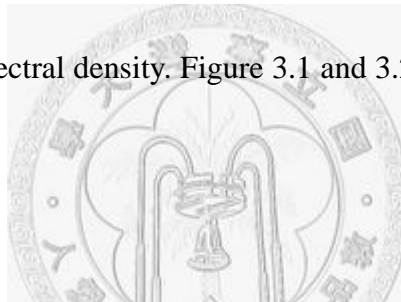


Figure 3.1 The waveform of the white noise.

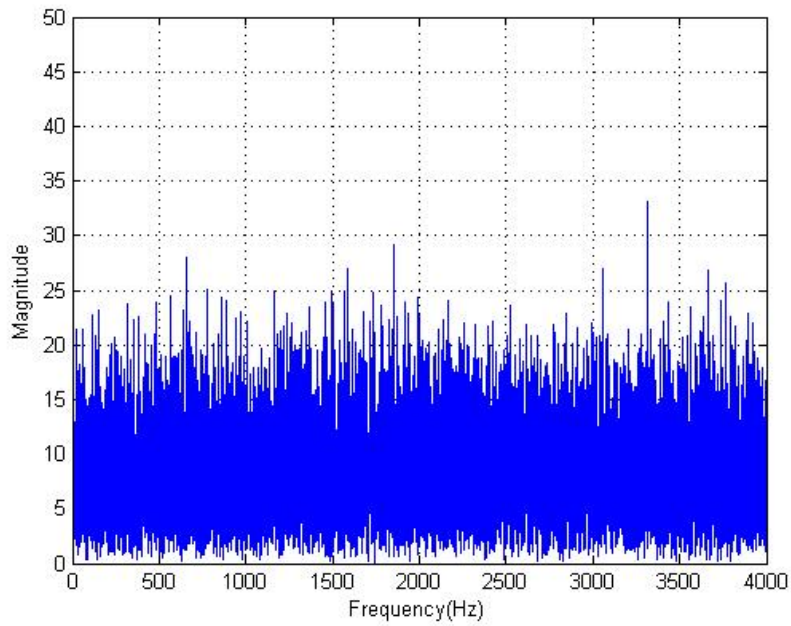


Figure 3.2 The FFT spectrum of the white noise

In other words, the signal contains equal power within a fixed bandwidth at any center frequency. After EMD of white noise, Figure3.3 and Figure3.4 shows IMF1~IMF6 and the FFT spectrums.

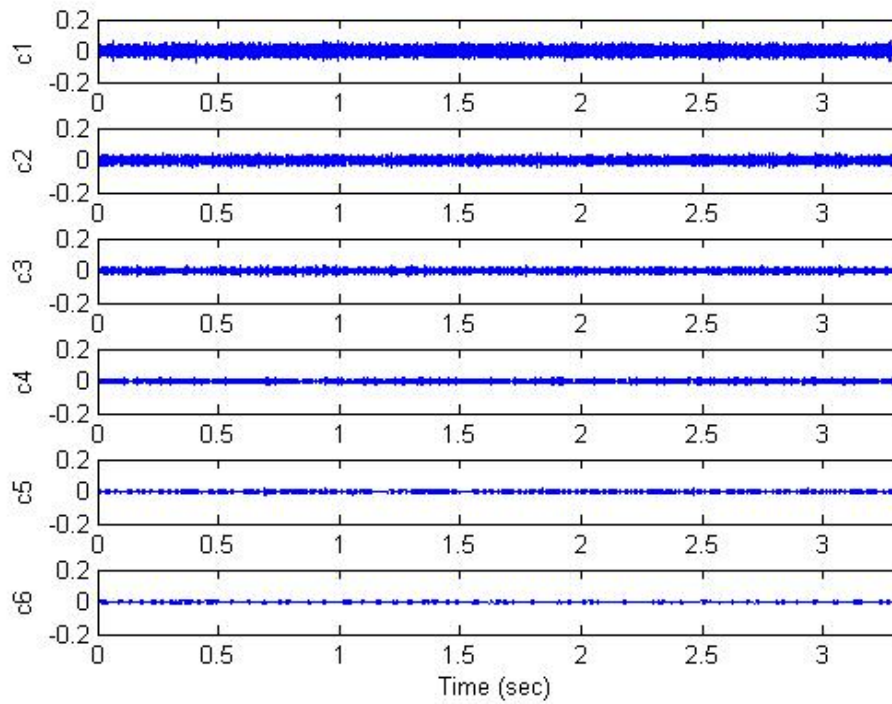


Figure 3.3 The waveform of $IMF_1 \sim IMF_6$ for the white noise.

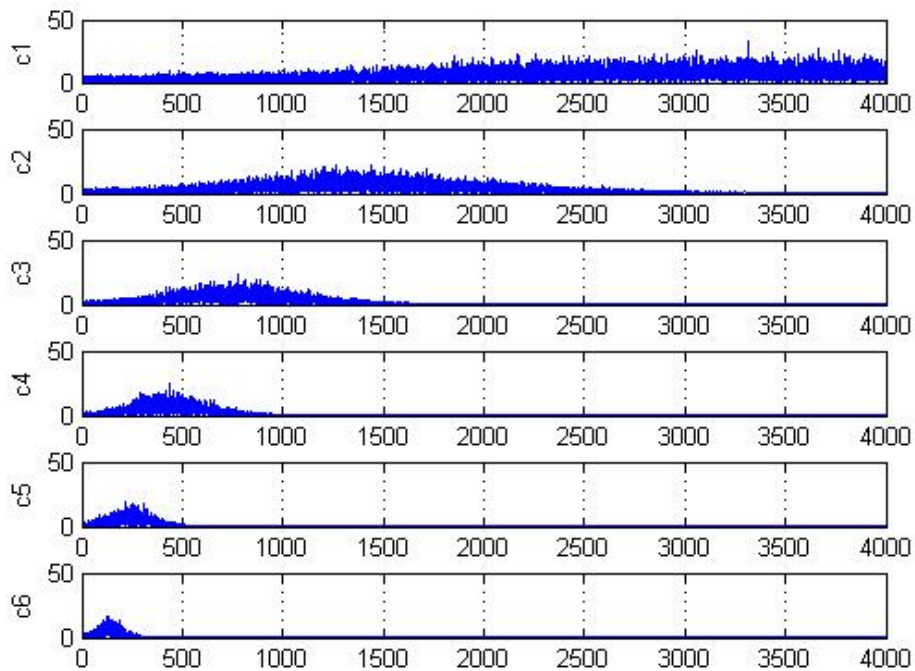
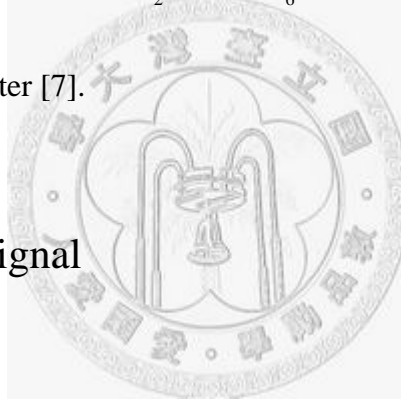


Figure 3.4 The FFT spectrums of $IMF_1 \sim IMF_6$ for the white noise

From observation of Figure3.3 and Figure3.4, the characteristics of EMD have two points[11],[12]: (1) Amplitude of each IMFs decrease with the decomposition, the amplitude of IMF_6 approximates zeros and hears nothing.(2) More specifically, for the IMF_1 , shows a temporal behavior in which the frequency of the mode main oscillates around 2000-4000 Hz with varying amplitude and frequency. The analysis of IMF_2 and IMF_3 shows the frequency of the mode oscillates around 1000-2000 Hz and 500-1000 Hz, respectively. EMD can be interpreted as a filter bank of overlapping band-pass filters for IMF_2 to IMF_6 . The IMF_1 correspond essentially to a half-band high-pass filter [7].



3.2.2 Sinusoidal Signal

In 3.2.1, we find out the EMD like a filter bank. Then, some questions of interest need to confirm about performance of EMD. The sinusoidal model is given below:

$$x(t) = a * \sin 2\pi f_1 t + b * \sin 2\pi f_2 t \quad (3.1)$$

In Figure3.5, the combination pure tone signals are decomposed into two individual tones [9]. When does the EMD retrieve the two individual tones?

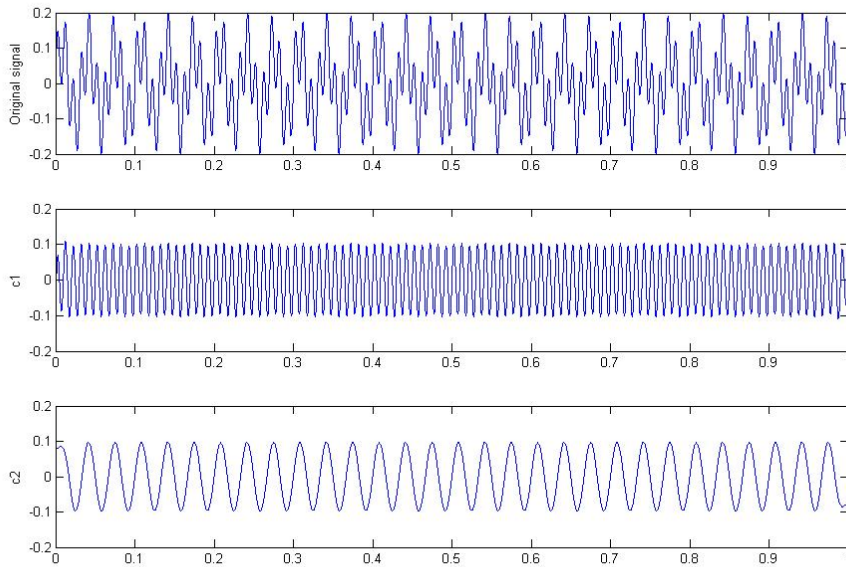


Figure 3.5 (a)The waveform of sinusoidal model , (b)The waveform of IMF_1 (c)The waveform of IMF_2

We can observe FFT spectrum in Figure3.6.

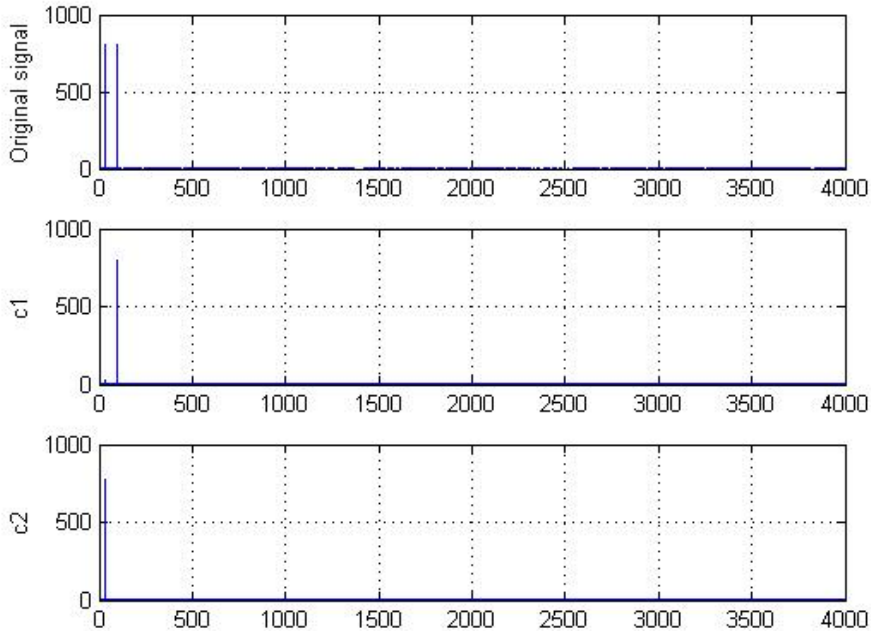


Figure 3.6(a)The FFT spectrums of sinusoidal model (b) The FFT spectrums of IMF_1 (c) The FFT spectrums of IMF_2

After setting $a=b=0.1$, $f_1=100\text{Hz}$, $f_2=10\sim 100\text{Hz}$, we compute the ratio of magnitude on f_2 between original signal and IMF_1 .

$$\text{Ratio of magnitude} : \frac{M_{IMF_1}(f_2)}{M_{origin}(f_2)} \quad (3.2)$$

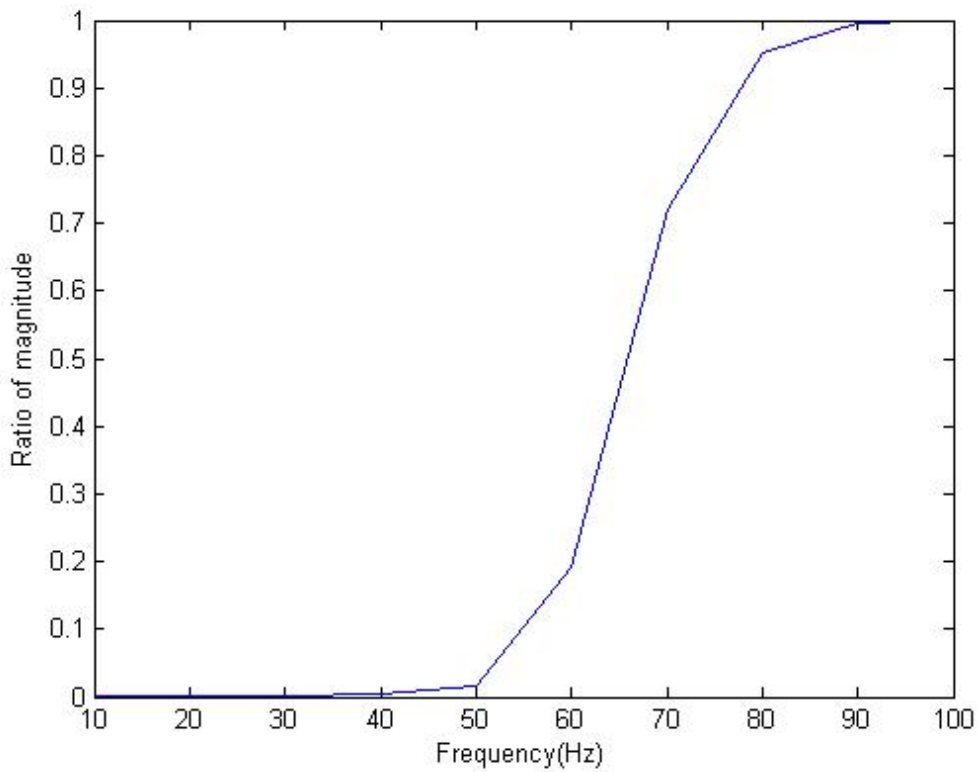


Figure 3.7The plot of ratio of magnitude as a function of f_2

In Figure3.7, when f_2 is equal to 50 Hz, the component of $\sin 2\pi f_2 t$ is almost disappear. After changing the proportion between f_1 and f_2 , the situation still exist.

so if $\frac{f_1}{f_2} > 2$, the sinusoidal model can be separated by EMD.

3.2.3 Summary

From two kinds of signal with EMD, we find out characteristics of EMD:

(1) The IMFs show an action of band-pass filter from high to low frequency.(2)The frequency range of each IMF is according half frequency of prior IMF. By two characteristics, we can separate into the frequency range we wanted with EMD.

3.3 Proposed Speech Enhancement Method with Empirical Mode Decomposition

3.3.1 Architecture



From 3.2, we properly know the characteristics of EMD. By these characteristics, a method of speech enhancement with an additive signal base on empirical mode decomposition is proposed. Basic architecture is given in Figure 3.8. The flowchart of speech enhancement is separated into two parts: (1) The dynamic filter is used to remove dominant noise component. (2) The ACWA filter is used to smooth the residual noise component. In the following sections, the functions of each unit will be introduced.

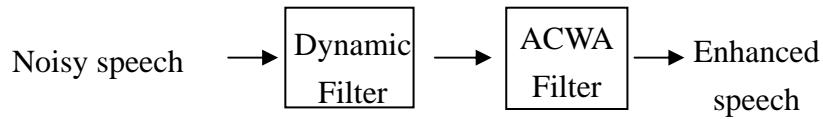


Figure 3.8 Block diagram of proposal method

3.3.2 Dynamic Filter

In the section of dynamic filter, the dominant frequency of noise will be removed.

The flowchart of dynamic filter is given in Figure 3.9.

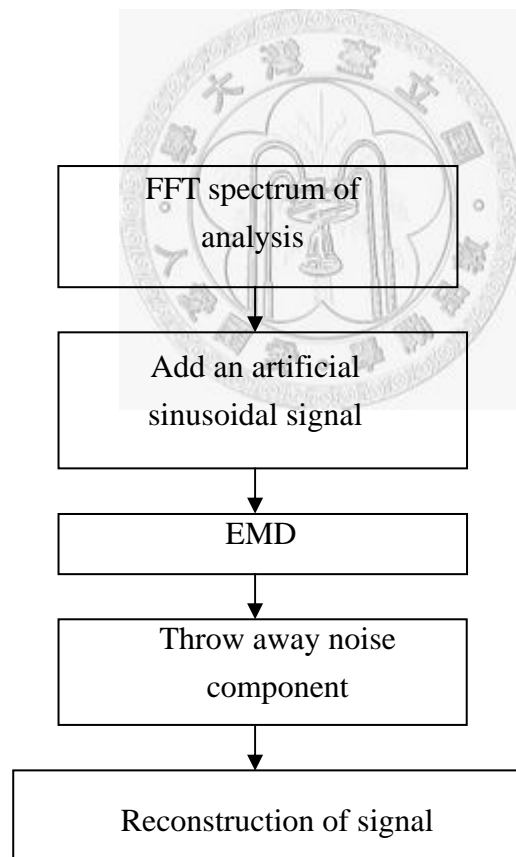


Figure 3.9 The flowchart of dynamic filter

The noisy speech $y(t)$ is combination between speech $s(t)$ and noise $n(t)$

$$y(t) = s(t) + a * n(t) \tag{3.3}$$

where a is constant for change amplitude of noise .

The FFT of $y(t)$ is given in Figure 3.10. After these magnitudes arranging from big to small, we set a threshold maximum of magnitude dividing ten. Finding out maximum of these frequencies was corresponded to these magnitudes. The double of maximum frequency is set as 964 Hz.

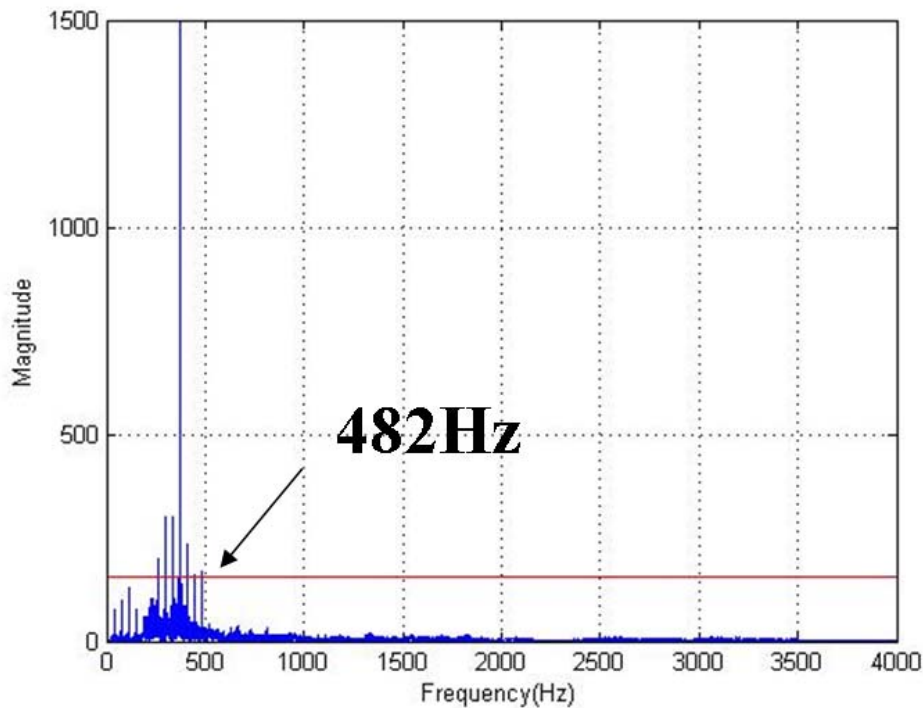


Figure 3.10 The FFT of spectrum for noisy speech

After EMD of $y(t)$ originally, the IMF_1 contains most noise on low frequency in Figure 3.10 because the energy on low frequency is more than the energy on high frequency.

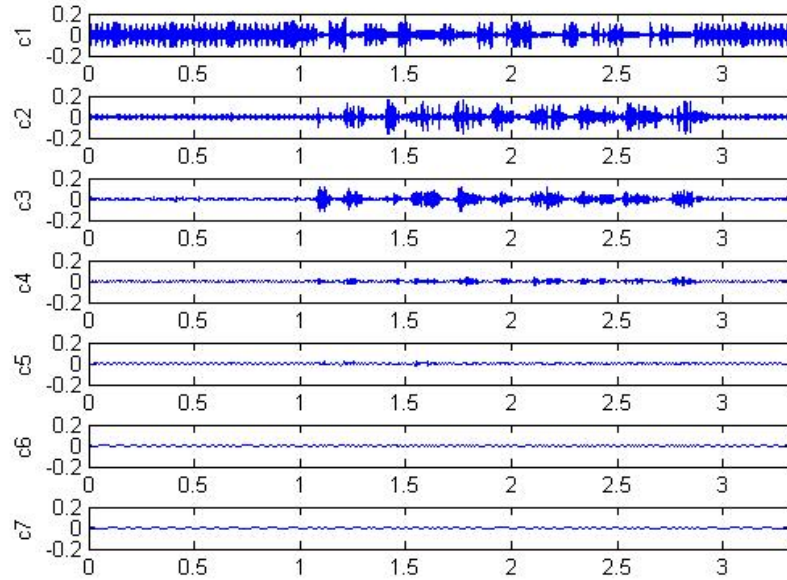


Figure 3.11 **The waveform of $IMF_1 \sim IMF_7$ for noisy speech**

Set an artificial signal $x(t)$ in 3.4

$$x(t) = g * (\sin 2\pi f_1 t + \sin 2\pi f_2 t) \tag{3.4}$$

$y_1(t)$ is the signal adding an artificial signal $x(t)$ to $y(t)$

$$y_1(t) = y(t) + x(t) \tag{3.5}$$

where g is the maximum amplitude of speech signal $y(t)$.

In this case, the f_1 is set on 2000 Hz (half of maximum frequency) f_2 is set on 964 Hz by prior estimation, we can set the lowest frequency of IMF_1, IMF_2 on $\frac{f_1}{2}$ and $\frac{f_2}{2}$. As expected the signal $x(t)$ will be decomposed into IMF_1, IMF_2 . We add IMF_1 and IMF_2 up then subtract the additive signal $x(t)$ then get a new signal $K(t)$ in (3.6).

$$K(t) = IMF_1 + IMF_1 - g * (\sin 2\pi f_1 t + \sin 2\pi f_2 t) \quad (3.6)$$

The EMD of noisy speech with an additive signal is given in Figure 3.12.

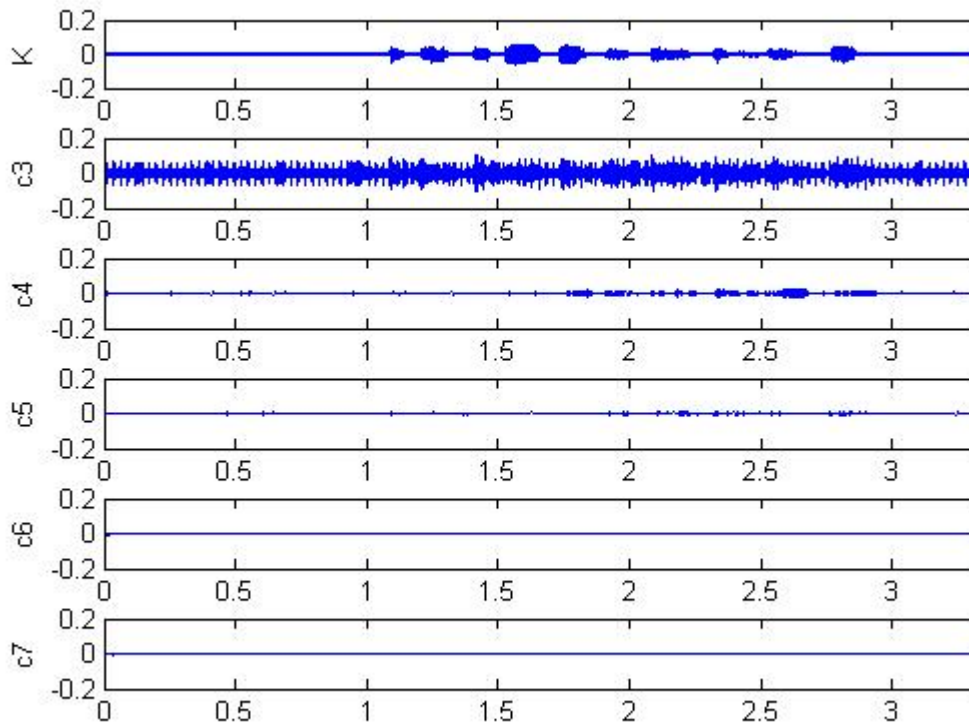


Figure 3.12 The waveform of $IMF_1 \sim IMF_7$ for noisy speech with an additive signal

Then find out the energy of each IMFs, and throw away the IMF that energy bigger than energy of $K(t)$.

$$Energy_i = \sum_{i=3}^C IMF_i^2(t) \quad (3.7)$$

In this case, we throw away the IMF_3 . Finally, the sum of these residual IMFs is the reconstruction of signal $y_2(t)$ in Figure 3.13.

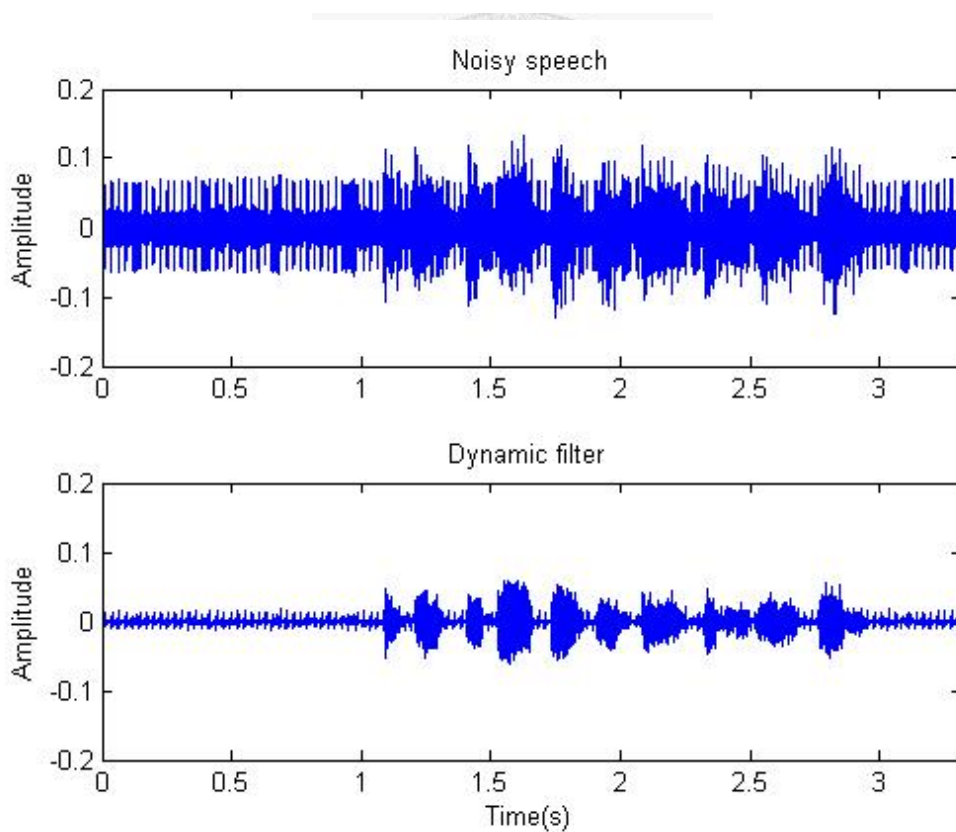


Figure 3.13 (a)The waveform of the noisy speech (b)The waveform of noisy speech after dynamic filter

In fact, the dynamic filter is like a high-pass filter. We design a high-pass filter with a cutoff frequency of 482 Hz and use the filter to the noisy speech. The result is shown in Figure 3.14. After comparing to two results, the dynamic filter can reserve for more the quality of speech on each frequency bank than high-pass filter because of characteristics of EMD. In other words, the distortion of signal is less with dynamic filter than high-pass filter.

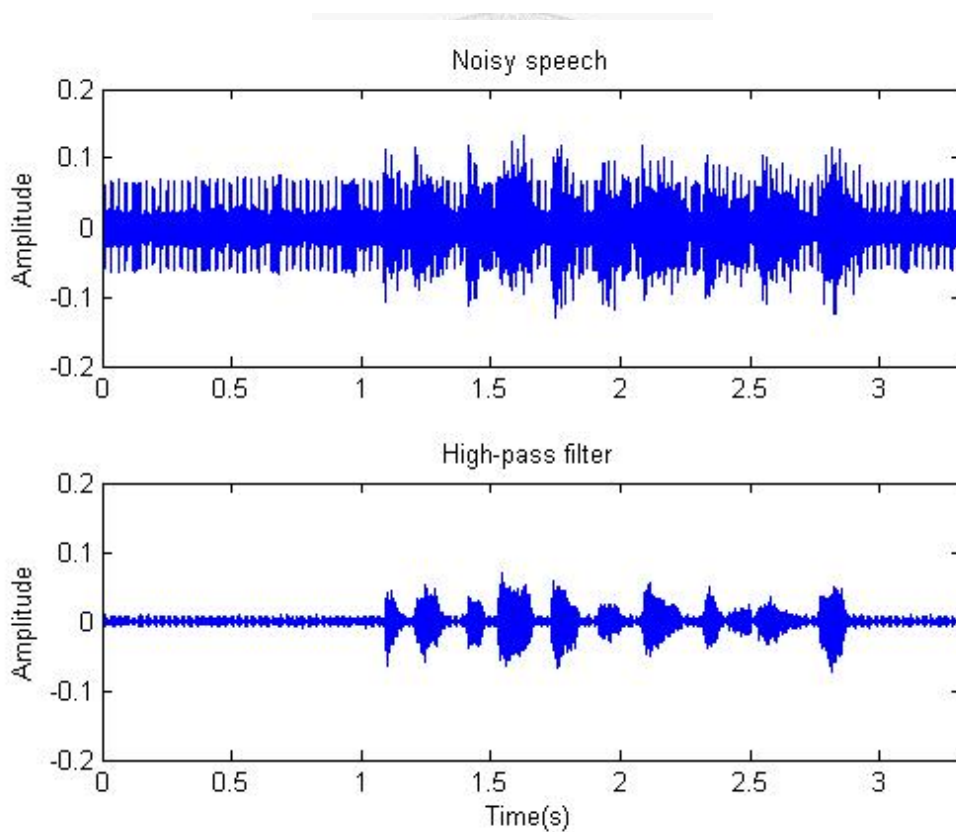


Figure 3.14 (a)The waveform of the noisy speech (b)The waveform of noisy speech after high-pass filter

3.3.3 Adaptive Center Weighted Average Filter

After the noisy speech pass through dynamic filter, little white noise still exist.

Using of adaptive center weighted average filter (ACWA filter) can get better performance of de-noising.

The flowchart of ACWA filter is given in Figure 3.15.

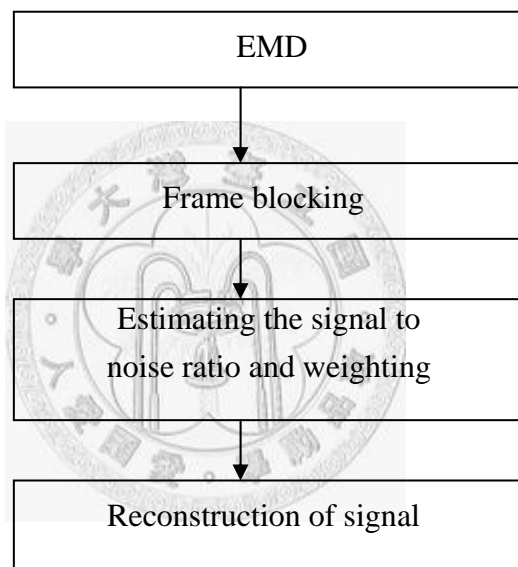


Figure 3.15 The flowchart of ACWA filter

After EMD of $y_2(t)$, $y_2(t)$ can be described as:

$$y_2(t) = \sum_{i=1}^c IMF_i(t) + r_c \quad (3.8)$$

Frame blocking to each IMFs in Figure 3.16

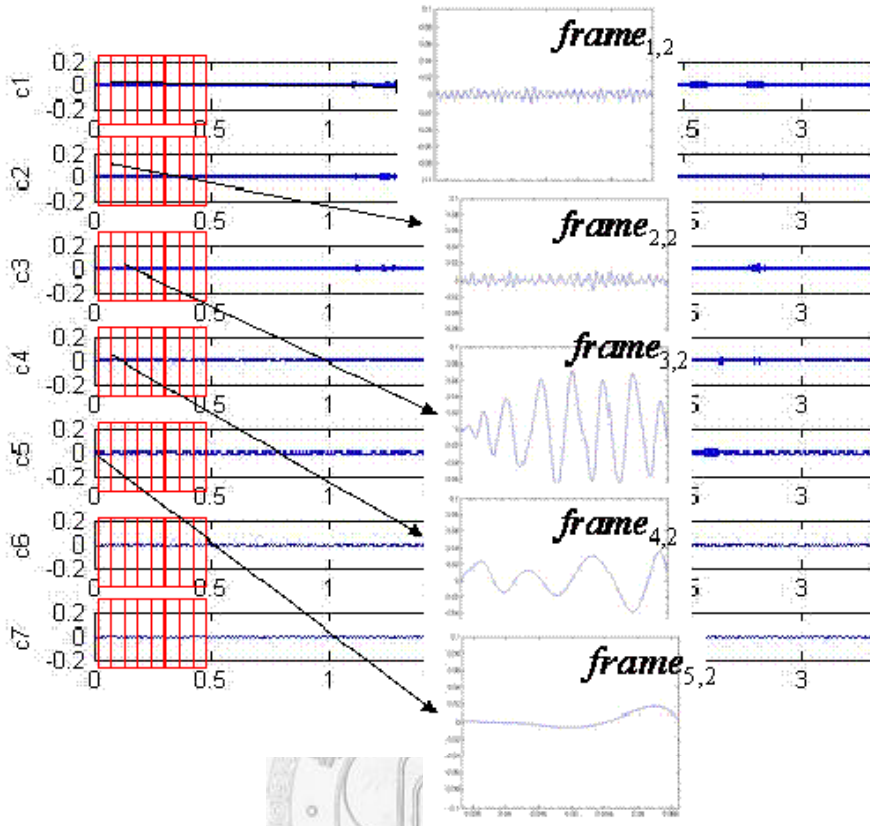


Figure 3.16Frame blocking to each

where $f_{i,j}$ is the j^{th} frame of the i^{th} IMF .

The noise level σ_j is estimated as following:

$$\sigma_i = 1.4826 * \text{Median}\{|IMF_i(n) - \text{Median}\{IMF_i(n)\}|\} \quad (3.9)$$

The de-noising of the IMF by the ACWA filter is given as follows:

$$W_{i,j} = \left(1 - \frac{\sigma_i^2}{\text{var}(f_{i,j})}\right) \quad (3.10)$$

$$f'_{i,j} = \begin{cases} \text{mean}(f_{i,j}) + W_{i,j}(f_{i,j} - \text{mean}(f_{i,j})), & \text{if } \text{var}(f_{i,j}) \geq \sigma_i^2 \\ \text{mean}(f_{i,j}) & , \text{otherwise} \end{cases} \quad (3.11)$$

where $\text{mean}(f_{i,j})$ and $\text{var}(f_{i,j})$ denote respectively the average and the variance of frames σ_j^2 designates the variance of noise contained in the IMF indexed by j.

New IMF is as following equation 3.12

$$IMF'_i = [f'_{i,1}, f'_{i,2} \dots f'_{i,end}] \quad (3.12)$$

Finally, the enhanced speech is in Figure 3.17

$$\hat{y}(t) = \sum_{i=1}^C IMF'_i(t) + r_c \quad (3.13)$$

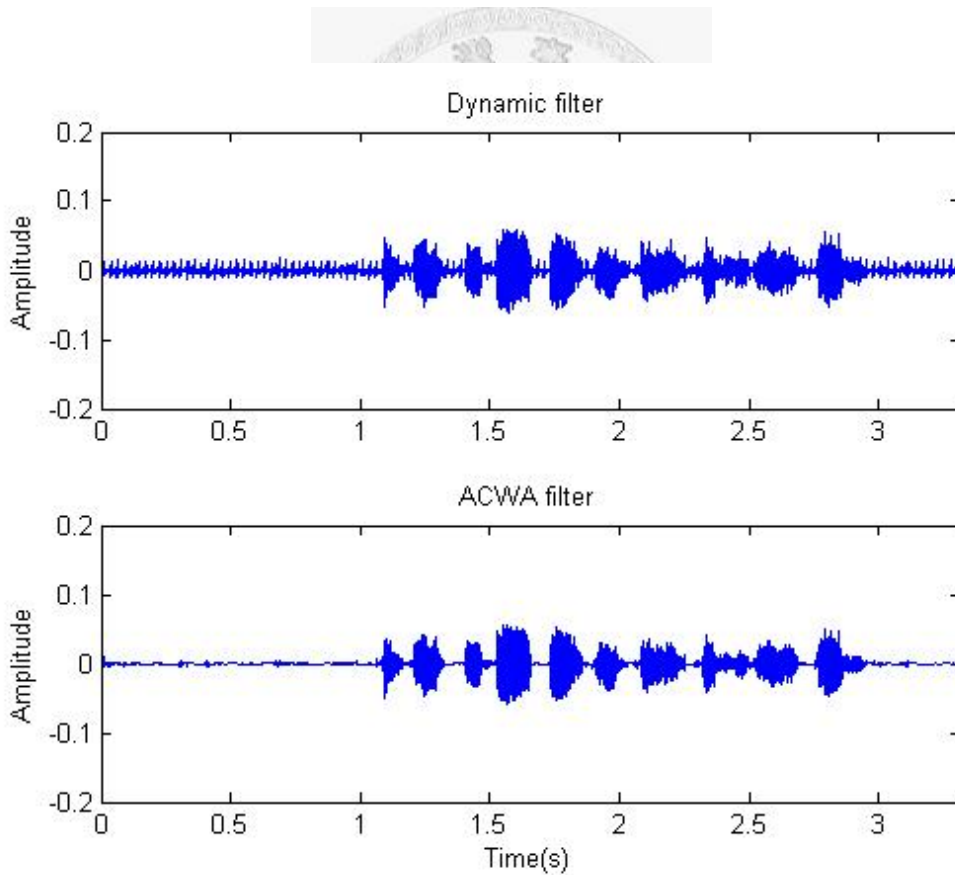
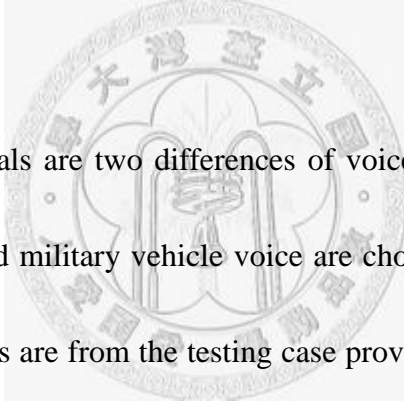


Figure 3.17 (a)The waveform of noisy speech after dynamic filter (b) The waveform of the noisy speech after dynamic filter and ACWA filter

Chapter 4 Experiment Results

In order to confirm the performance of this method, the simulation of proposed method was carried out in this chapter. We will introduce the test environment, the noisy speech database, and speech quality assessment. Finally, we evaluate the performance of proposed method in two experiments with various values of the SNR.

4.1 Test Environment and Noisy Speech Database



The clean speech signals are two differences of voices spoken by a male and a female. Propeller voice and military vehicle voice are choice as additive background noises. The propeller noises are from the testing case provided by the Aviation Safety Council. The military vehicle noises are from Noisex-92 database [17]. The content of speech is a Mandarin speech sentence with length of 3.3 seconds. In the experiment, clean speech signals and noise signals were sampled at 8 kHz with 16-bits quantization. The matlab code of EMD, spectral subtraction and wiener filter come from matlab central [18].

4.2 Speech Quality Assessment

The performance of speech enhancement in variable noise level condition is measured by objective evaluation:

$$\text{SNR}_{\text{in}} = 10 \log \left[\frac{\sum_{t=1}^N s^2(t)}{\sum_{t=1}^N n^2(t)} \right], \quad (4.1)$$

where the “ SNR_{in} ” is the SNR value of the input noisy speech signal standing for the amount of the additive noise, N is the length, $x(t)$ is clean speech signal, $n(t)$ is the additive noise signal.

$$\text{SNR}_{\text{out}} = 10 \log \left[\frac{\sum_{n=1}^N s^2(t)}{\sum_{n=1}^N [s(t) - \hat{y}(t)]^2} \right], \quad (4.2)$$

where the “ SNR_{out} ” is the SNR value of the output enhanced speech signal standing for the efficiency of the proposed method, N is the length, $x(t)$ is clean speech signal, $\hat{y}(t)$ is the enhanced speech signal.

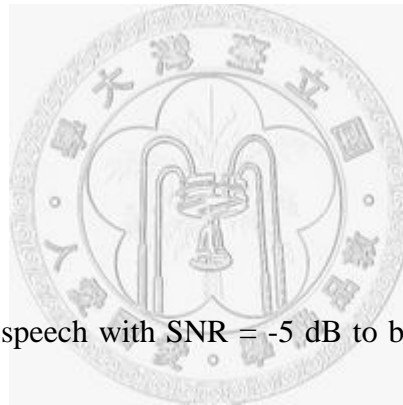
4.3 Experimental Results

Before we evaluate the performance of proposed method, we produce a noisy speech to be a test case.

$$y(t) = s(t) + a * n(t) \quad (4.1)$$

where $s(t)$ is the clean speech signal, $n(t)$ is the additive noise and $y(t)$ is the noisy speech signal. We can adjust the constant a to get desired SNR value.

Experimental results I:



We choose the noisy speech with SNR = -5 dB to be the first experimental test case. Figure 4.1 shows the time waveforms of a female speech, military vehicle noise, and noisy speech, respectively.

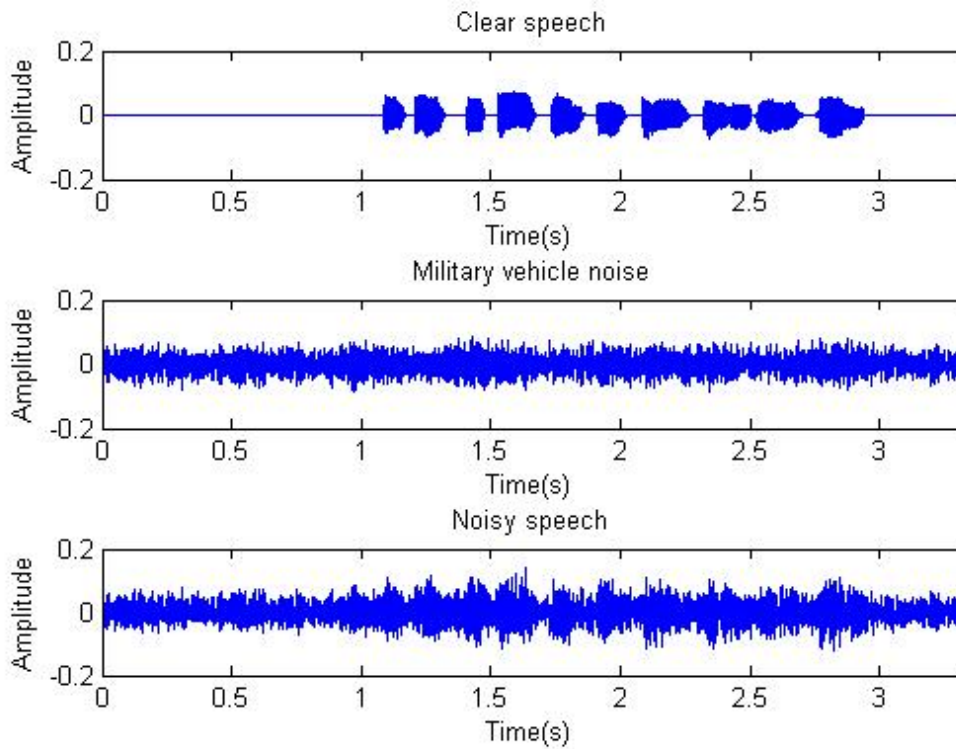


Figure 4.1 The waveforms of noisy speech producing result (a) Clean speech (b) Military vehicle noise (c) Noisy speech (-5dB).

The parameters used in the experiment I are shown in Table 4.1 and Table 4.2.

Table 4.1 Parameters used in the speech+ military vehicle noise situation

Dynamic filter	f_1, f_2	2000Hz,590Hz
	Throw away component	IMF_3, IMF_4
ACWA filter	Frame	265 points

Table 4.2 Parameters used in the speech+ propeller noise situation

Dynamic filter	f_1, f_2	2000Hz,964Hz
	Throw away component	IMF_3
ACWA filter	Frame	265 points

The results of passing through two filter based on the proposed method are shown in Figure 4.2.

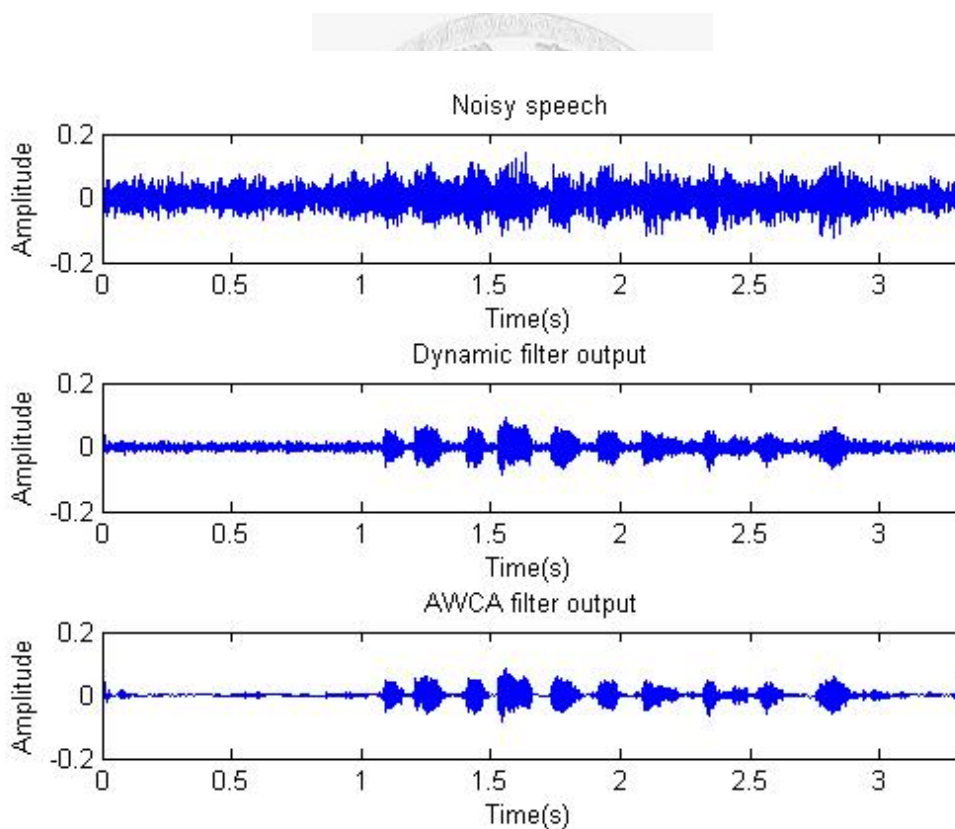


Figure 4.2 The waveforms of passing through two filter result(a) Noisy speech (-5dB) (b) Dynamic filter output(c) ACWA filter output.

Table 4.3 The performance of speech enhancement based on proposed method for experiment I

	Male speech+ military vehicle noise	Male speech+ propeller noise	Female speech+ military vehicle noise	Female speech+ propeller noise
SNR _{in} [dB]	-5	-5	-5	-5
SNR _{out} [dB]	2.98	4.1	3.33	2.21

From Table 4.3, the performance of speech enhancement based on proposed method is good. The SNR has more improvement for four situations. In the other word, the background noise was suppressed and removed well. After hearing these enhanced speeches actually, these speeches almost the same with their original speech. The quality of clear speech can be heard in these enhanced speeches.

Experimental results II:

We choose the noisy speech with SNR = -10 dB to be the second experimental test case. Figure 4.3 shows the time waveforms of clear speech, military vehicle noise, and noisy speech, respectively.

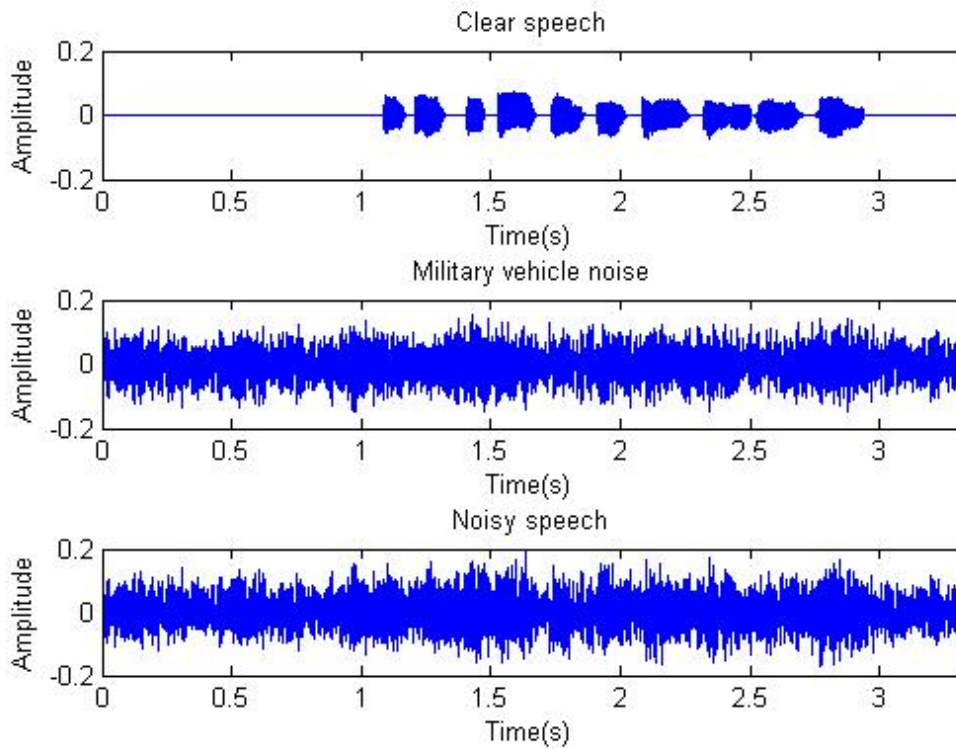


Figure 4.3 The waveforms of noisy speech producing result
(a) Clean speech (b) Military vehicle noise (c) Noisy speech(-10dB).

The parameters used in the experiment II are shown in Table 4.3 and Table 4.4.

Table 4.4 Parameters used in the speech+ military vehicle noise situation

Dynamic filter	f_1, f_2	2000Hz,590Hz
	Throw away component	IMF_3, IMF_4
ACWA filter	Frame	265 points

Table 4.5 Parameters used in the speech+ propeller noise situation

Dynamic filter	f_1, f_2	2000Hz,964Hz
	Throw away component	IMF_3
ACWA filter	Frame	265 points

The results of passing through two filter based on the proposed method are shown in Figure 4.4.

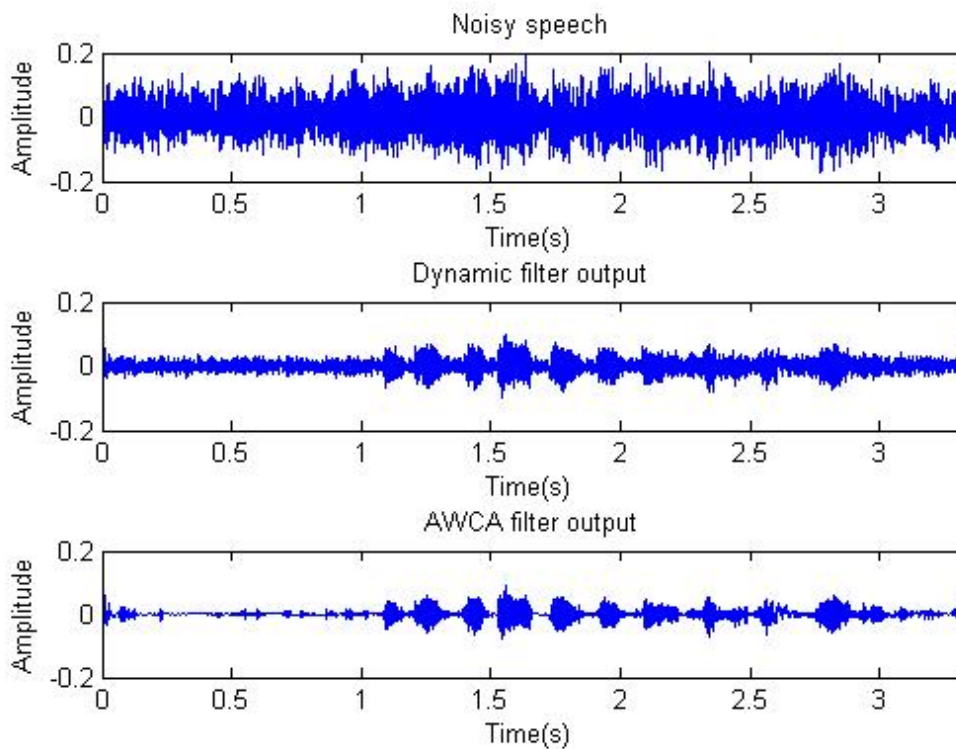


Figure 4.4 The waveforms of passing through two filter result(a) Noisy speech (-10dB)(b) Dynamic filter output(c) ACWA filter output.

**Table 4.6 The performance of speech enhancement based on proposed method
for experiment II**

	Male speech+ military vehicle noise	Male speech+ propeller noise	Female speech+ military vehicle noise	Female speech+ propeller noise
SNR _{in} [dB]	-10	-10	-10	-10
SNR _{out} [dB]	1.06	3.35	1.78	2.64

Although background noise has larger volume than experiment I, the performance of speech enhancement based on proposed method is still good from Table 4.6. The background noise was suppressed and removed well. After hearing these enhanced speeches actually, these speeches almost the same with their original speech. The quality of clear speech still retains these enhanced speeches. For comparing performance of speech enhancement, we use spectral subtraction and wiener filter methods to the noisy speech. These results are shown in Figure.4.5 and Figure.4.6.

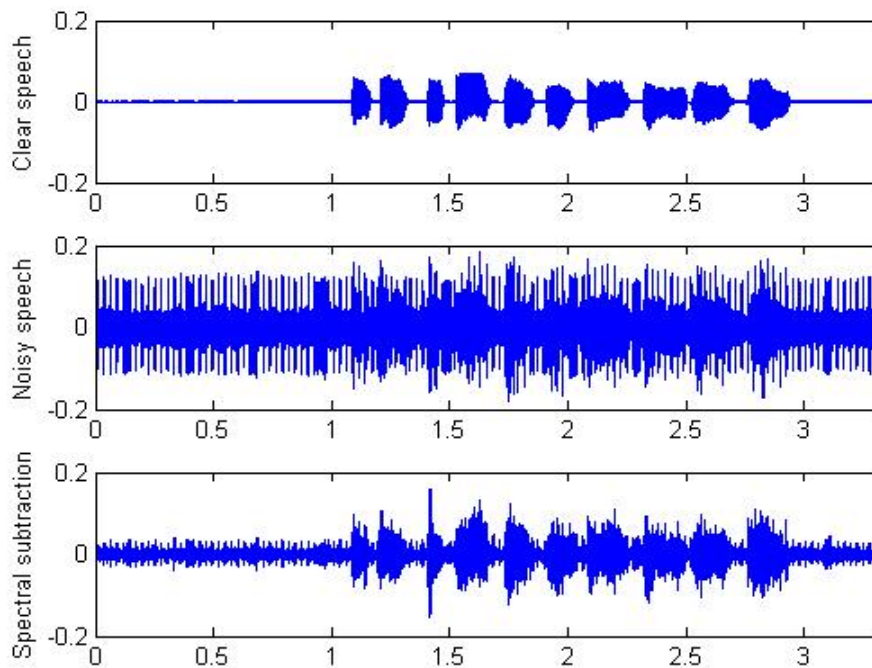


Figure 4.5 The waveforms using spectral spectrum method (a) Clear speech (b) Speech +propeller noise (-10dB) (c) Result base on spectral subtraction method.

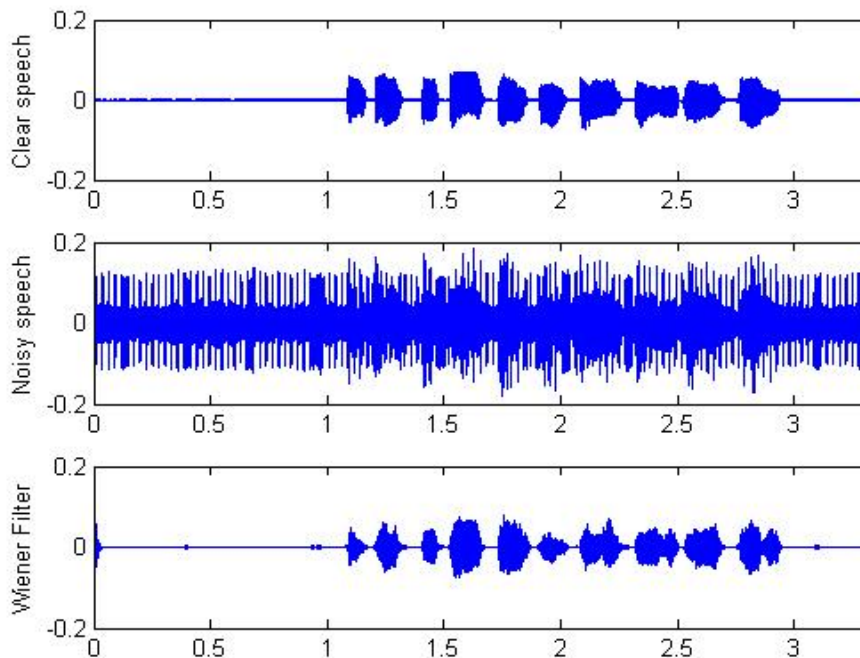


Figure 4.6 The waveforms using wiener filter method (a) Clear speech (b) Speech +propeller noise (-10dB) (c) Result base on wiener filter method.

The improvement of SNR basing on three methods is shown in Table 4.7. The performance of proposed method is better than the spectral spectrum method but worse than the wiener filter method in low SNR situation. After hearing these enhanced speeches actually, the quality of enhanced speech using spectral spectrum method is worst. The quality of enhanced speech is the same with proposed method and wiener filter method.

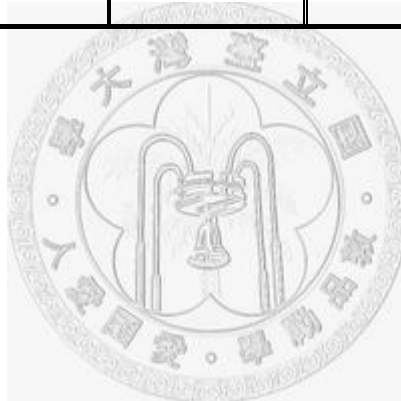
Table 4.7 The performance of speech enhancement for male speech

base on three methods

	Male speech+ propeller noise(-5dB)	Male speech+ military vehicle noise(-5dB)	Male speech+ propeller noise (-10dB)	Male speech+ military vehicle noise(-10dB)
Spectral Subtraction	2.17dB	1.82dB	-0.98dB	-1.27dB
Wiener Filter	4.66dB	3.52dB	3.1dB	2.06dB
Proposed method	2.98dB	4.1dB	3.35dB	1.06dB

**Table 4.8 The performance of speech enhancement for female speech
base on three methods**

	Female speech+ propeller noise(-5dB)	Female speech+ military vehicle noise(-5dB)	Female speech+ propeller noise (-10dB)	Female speech+ military vehicle noise(-10dB)
Spectral Subtraction	2.12dB	1.93dB	-0.82dB	-1.83dB
Wiener Filter	5.02dB	3.56dB	3.34dB	2.61dB
Proposed method	2.21dB	3.33dB	2.64dB	1.78dB



Chapter 5 Conclusions

In this thesis, a novel idea for signal channel speech enhancement method based on dynamic filter and ACWA filter is proposed. In present method two problems are shown: (1) The aliasing of signal (2) In low SNR situation, the performance of de-noising is bad. Therefore, we proposed the novel de-noising method basing on empirical mode decomposition with an additive artificial signal. Because of the characteristic of EMD, the additive artificial signal can be separated into first two IMFs. The setting of f_2 make noise component into later IMF. In dynamic filter, the dominant noise component is removed then suppresses residual noise in ACWA filter. From the two experimental results, the improvement of SNR in the enhanced speech signal is observed. Hence, the effectiveness of the method is confirmed. Comparing to the present methods, the performance of proposed method is better than spectral spectrum method. In common with wiener filter, the speech quality was most reserved without distortion of signal. By the proposed method, we can understand the speech content clearly.

References

- [1] K. Khaldi and A. O. Boudraa, "Speech denoising by Adaptive Weighted Average filtering in the EMD framework," *IEEE Int. Conf. Signals, Circuits and Systems*, Nov, 2008.
- [2] J.S. Lee. , "Digital image enhancement and noise filtering by using local statistics," *IEEE Trans. Pattern Anal. Mach. Int.*, vol.2, issue 4, pp. 165-168, Mar.1980.
- [3] A.O. Boudraa and J.C. Cexus., "Denoising via empirical mode decomposition," *In Proc. IEEE ISCCSP, Marrakech, Morocco*, 2006.
- [4] A.O. Boudraa, J.C. Cexus, and Z. Saidi. , "EMD-based signal noise reduction," *Int. J. Sig. Process.*, vol.1, issue 1, pp. 33-37, 2004.
- [5] N. E. Huang, Z. Shen and S. R. Long., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A(454)*, pp. 903-995, 1998.
- [6] Y. Kopsinis, S. McLaughlin, "Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding," *IEEE Trans. Signal Processing*, vol.57, NO.4, April.1994.
- [7] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, pp. 112–114, Feb. 2004.
- [8] B. Widrow et al., "Adaptive Noise Canceling: Principles and Applications," *Proceedings of the IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
- [9] G. Rilling and P. Flandrin, "One or two frequencies? The empirical mode decomposition answers," *IEEE Trans. Signal Process.*, pp. 85–95, Jan. 2008.
- [10] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*,

vol.77, issue 7, pp.941-981, July 1989.

- [11] N. E. Huang and S.P. Shen, "Hilbert-Huang transform and its applications," *World Scientific*, 2005.
- [12] 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, Jun. 2004.
- [13] 葉向林, “聽障者之語音增強與轉換,” 國立清華大學電機工程研究所碩士論文, 2004.
- [14] 陳厚君, “經驗模態分解法之語音辨識,” 國立中央大學電機工程研究所碩士論文, 2004.
- [15] 劉致廷, “座艙語音記錄器之單通道適應性噪音濾除,” 國立台灣大學電機工程研究所碩士論文, 2007.
- [16] 賴亦桓, “運用經驗模態分解法於語者辨識,” 國立台灣大學電機工程研究所碩士論文, 2006.
- [17] NOISEX-92: <http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html>
- [18] Matlab central: <http://www.mathworks.com/matlabcentral/>.