

國立交通大學

電機學院 IC 設計產業研發碩士班

碩士論文

根據高階相關特性之非線性回音壓抑器

Nonlinear Acoustic Echo Suppression Based on Correlated Higher-orders
Nonlinear Echoes



研究生：王宏益

指導教授：謝世福 教授

中華民國九十七年三月

根據高階相關特性之非線性回音壓抑器

學生：王宏益

指導教授：謝世福

國立交通大學電機學院產業研發碩士班



爲了去補償在免持聽筒或是視訊會議系統中所發生喇叭的非線性回音，通常使用非線性音訊回音消除器去消除此回音。不過這些方式的收斂速度太慢而影響效能。使用非線性回音壓抑方式即改善此問題。這是因爲使用非線性回音壓抑時，不去找出實際的非線性路徑的部分。而是去找出非線性部份的能量頻譜密度。

這篇論文中，介紹兩種估計非線性回音能量頻譜密度的方式。第一種是以回音的相關性去估計出。第二種是使用非線性適應性濾波器去得到。然而使用第一種，會產生較大的誤差，是由於高階與一階回音不會是線性的關係。而第二種，需要大量的非線性濾波器去估計出高階的回音。在此篇論文將提出新的演算法，使用非線性適應濾波器去準確找出低階部分的回音，並且使用回音的相關性粗略估出高階部分的回音。此演算法能以少量的運算量較準確的估計出非線性回音的能量。另外對基底選擇提出分析方式，進一步的提升估計非線性回音的準確度。

Nonlinear Acoustic Echo Suppression Based on Correlated Higher-orders Nonlinear Echoes

Student : H. Y. Wang

Advisor : S. F. Hsieh

**Industrial Technology R & D Master Program of
Electrical and Computer Engineering College
National Chiao Tung University**

The logo of National Chiao Tung University is a circular emblem with a gear-like border. Inside the circle, there is a stylized building and the year '1896' at the bottom. The word 'Abstract' is written in a bold, serif font across the center of the logo.

Abstract

In order to compensate the nonlinear distortion in the hands -free telephones or teleconferencing system, the nonlinear acoustic echo cancellation can be used to cancel nonlinear acoustic echo. However, the convergence speed of these methods is too slow. Nonlinear Acoustic Echo Suppressions will resolve the question. This is because we are not interested in an exact identification of the nonlinear components of the echo path. We are rather aiming at estimates power spectral density of the nonlinear components.

In the thesis, two previous methods are introduced to estimate power spectral density of nonlinear residual error. First, use the linear property of echoes to estimate. Second, a nonlinear adaptive filter is used. We proposal the new method that can estimate low order echoes accurately and high order echoes roughly by less computational complexity than second method and estimate the nonlinear residual error more accurately than first method. And, we discuss how to choose echo basis.

Acknowledgments

時光飛逝，回首兩年多來的研究生活，真是讓我成長不少，首先要感謝指導教授謝世福老師，在老師的耐心與嚴謹的指導下，讓我在追求學識與做事態度，有了很大的震撼，讓我受益良多，老師謝謝您！除此之外，感謝實驗室的學長、同學和學弟，因為有你們，讓我在沮喪時有了動力，讓我的研究生活充滿色彩。最後要感謝我的家人，總是默默的陪伴我，給我支持與鼓勵，讓我在困境中完成。



Contents

中文摘要.....	I
English Abstract	II
Acknowledgments	III
Contents	IV
List of Figures	VI
List of Tables	VIII
1. Introduction	1
2. Overview of Nonlinear Acoustic Echo Cancellation	4
2.1 Memoryless Nonlinear AEC.....	5
2.1.1 Hammerstein model.....	5
2.1.2 Orthogonalized Power Filter model.....	6
2.2 Memory Nonlinear AEC.....	7
2.2.1 Wiener Model.....	7
2.2.2 Volterra model.....	8
2.3 Computation Reduction of Simplified Volterra Model.....	10
3. Estimate Nonlinear Residual Error for Nonlinear Acoustic Echo Suppression	14
3.1 Wiener Filter.....	15
3.2 Acoustic Echo Suppression Structures.....	16
3.2.1 AEC+NR structure.....	17
3.2.2 NR+AEC structure.....	18
3.3 Estimation nonlinear residual error for NAES.....	19
3.3.1 Based on Highly Correlated Nonlinear Residual Echo.....	21
3.3.2 Power Filter Model.....	22
3.4 Suppression of nonlinear residual error and background noise.....	24
3.4.1 AEC+NR suppression of noise and nonlinear residual error.....	24
3.4.2 NR+AEC suppression of noise and nonlinear residual error.....	25
3.5 Volterra structure for NAES.....	27
3.6 Using high order nonlinear echo to estimate nonlinear residual error.....	29
3.6.1 Basis Selection for proposed slope method.....	31

4. Computer Simulations	34
4.1 Parameters and speech signal of simulations.....	34
4.2 Compare two combined structures in three conditions.....	35
4.2.1 Background noise.....	35
4.2.2 Nonlinear residual error	38
4.2.3 Background noise and nonlinear residual error	39
4.3 Performance of Volterra Structure for NAES	39
4.4 Simulation of Highly Nonlinear Residual Errors.....	41
4.4.1 Single talk	41
4.4.2 Double Talk.....	45
4.5 Statistics Distribution of Higher-Order Nonlinear Residual Errors.....	46
5. Conclusions	56
Bibliography	58



List of Figures

1.1	The simplified diagram of hands-free telephone system	1
2.1	Hammerstein structure	5
2.2	Wiener model	7
2.3	Cascade model of the system and mirror adaptive system	8
2.4	Simplified second order Volterra filter	10
3.1	Linear AEC and nonlinear acoustic echo suppression structure	14
3.2	Block diagram representation of the statistical filtering problem	16
3.3	Block diagram of a combined echo canceling and noise reduction system	168
3.4	Block diagram of a combined noise reduction and echo canceling system	19
3.5	Block diagram of NR+AEC to suppress noise and nonlinear residual error	25
3.6	Block diagram of NR+AEC to suppress noise and nonlinear residual error	26
3.7	The Volterra structure for NAES	28
3.8	Block diagram of Sec. 3.3.1	289
3.9	Block diagram of the proposed method	30
3.10	Sigmoid function	32
4.1	Speech signal	35
4.2	Pseudo room impulse response	36
4.3	Combined structures for suppressing background noise (SNR=20dB)	37
4.4	Combined structures for suppressing background noise (SNR=-5dB)	37

4.5 Combined structures for suppressing nonlinear residual error (SNR=20dB)	38
4.6 Combined structures for suppressing nonlinear residual error and background noise	39
4.7 The Volterra structure for NAES using WGN	40
4.8 ERLE for speech signal & real system	43
4.9 ERLE for speech signal & polynomial system	44
4.10 ERLE for WGN & polynomial system.....	44
4.11 Statistic distribution of AEC1/slope1 in speech signal + real system	47
4.12 Statistic distribution of AEC13/AES3/slope1 in speech signal + real system	47
4.13 Statistic distribution of AEC13/AES3/slope3 in speech signal + real system	48
4.15 Statistic distribution of AEC13/AES3/slope1 in WGN signal + polynomial system.....	49
4.16 Statistic distribution of AEC13/AES3/slope3 in WGN signal + polynomial	50
4.17 Statistic distribution of AEC1/slope1 in speech signal + polynomial system	51
4.18 Statistic distribution of AEC13/AES3/slope1 in speech signal + polynomial system.....	51
4.19 Statistic distribution of AEC13/AES3/slope3 in speech signal + polynomial system.....	52
4.20 Statistic distribution of AEC1/slope1 in WGN signal + real system	53
4.21 Statistic distribution of AEC13/AES3/slope1 in WGN signal + real system	53
4.22 Statistic distribution of AEC13/AES3/slope1 in WGN signal + real system	54

List of Tables

2.1	The parameter of simplified Volterra Model	12
4.1	Notation of six algorithms.....	41
4.2	Simulation of double-talk	45
4.3	Correlation mismatch factor ε	55



Chapter 1

Introduction

In using hands-free telephone or teleconferencing, the speaker always hears his speech signal. This is because the system suffers from the annoying acoustic echo problem which is the far-end speech transmitted from the near-end microphone to near-end loudspeaker and back to far-end user from far-end loudspeaker. A simplified diagram of a hands-free telephone system is shown in Fig 1.1. If the far-end signal picked up from the near-end microphone can be cancelled, the acoustic echo problem would be overcome. If we estimate the echo path accurately, the echo would be cancelled. This method is acoustic echo cancellation (AEC) shown in Fig1.1. When using hands-free telephone or teleconferencing, the room impulse response can change very often. So, the AEC is time-variant to track the echo path to provide satisfactory speech communication quality in [1-5].

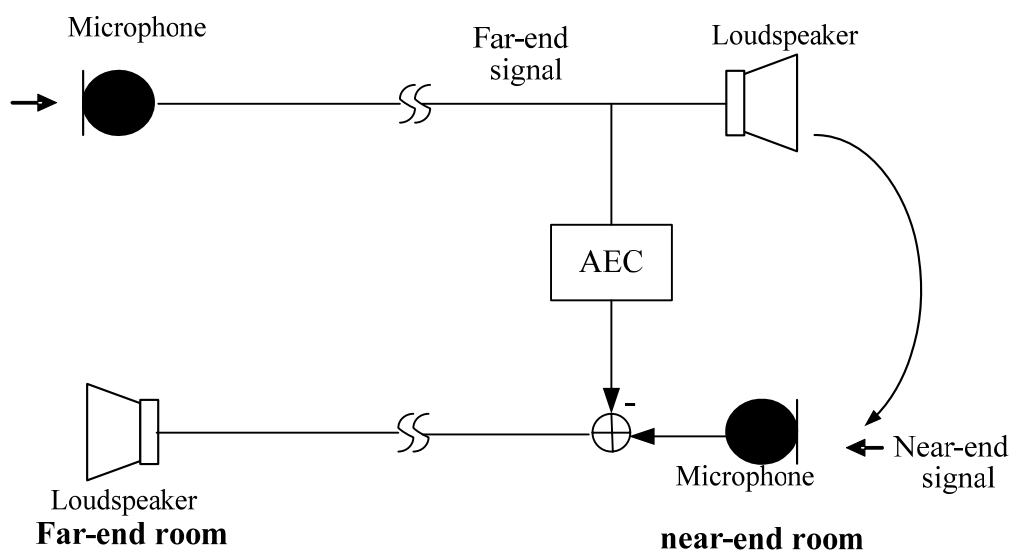


Fig. 1.1 The simplified diagram of hands-free telephone system

When the volume level is larger than the loudspeaker amplifier capability, the amplifier is overdriven. So, the loudspeaker power curve is not just linear. The linear AEC is not sufficient to estimate the acoustic echo channel. The performance is bad by the high level volume. To overcome the nonlinear problem, there are many methods that had been proposed. The nonlinearity of loudspeaker can be classified as nonlinearity with and without memory in [6-10]. In this thesis, we will introduce several nonlinear AEC methods.

In order to overcome the nonlinear effect of power amplifier, we always use high order structure of nonlinear AEC. However, the computation is more complex, and the convergence is too slow. The high order nonlinear AEC (NAEC) must adapt many coefficients simultaneously at the every iteration. The coefficients are interfering with each other. The object of nonlinear acoustic echo suppression (NAES) is to increase the attenuation of the nonlinearly distorted residual echo and the convergent speed. These time-variant estimates are used to approximately adjust the frequency-dependent gain value of the echo suppressor. In contrast to the application of adaptive filters to nonlinear echo cancellation, we are not interested in an exact identification of the nonlinear components of the echo path. We are rather aiming at estimates of power spectral density of the nonlinear components. This method uses linear relation between linear echo and nonlinear residual error to estimate power spectral density of nonlinear residual error that can save large calculation to suppress nonlinear residual error [11]. However, the linear echo is not linear with nonlinear residual error in fact. So, the high order nonlinear echoes can not be suppressed. Another method uses nonlinear adaptive filter to estimate high order nonlinear residual error that is more accurate than [12]. But, the method needs large calculation quantity.

Besides, we propose the new NAES method to combine the methods [11-12]. In fact, the nonlinear residual error is not linear with linear echo. The proposed method is used to overcome this drawback and save the calculation quantity than [12]. The new method is wanted to estimate nonlinear residual error to two parts. Low order nonlinear residual echoes estimate accurately by nonlinear adaptive filters and high order nonlinear residual echoes estimate roughly by correlate with low order echo.

This thesis is organized as follows. More details about nonlinear AEC with and without memory will be introduced in chapter 2. We will introduce nonlinear acoustic echo suppressions (NAES) and propose a new method that uses low order power filter to estimate low order nonlinear residual error and higher-order nonlinear echoes as basis to estimate high order nonlinear residual error. In chapter 4, we will show many computer simulations that have been discussed in chapter 2 and 3. Finally, we give a conclusion of our work.



Chapter 2

Overview of Nonlinear Acoustic Echo Cancellation

In Chapter 2 we will introduce several nonlinear AEC methods in the time domain including memoryless and memory. The traditional method to get over the acoustic echo is to eliminate all the signals from the far-end loudspeaker. However, this half-duplex communication is bad for user. The linear acoustic echo cancellation can overcome this difficulty and the hands-free telephone or teleconferencing can work at full-duplex. The loudspeakers for hands-free telephone or teleconferencing are usually small and cheap, so the loudspeaker will be saturated at high level speech. When the saturation effect happens, the loudspeaker is not linear any more. The residual error using only linear acoustic echo cancellation is very large. We will discuss the nonlinear acoustic echo cancellation to overcome this question.

To some loudspeakers, the nonlinear effects have memory. If using memoryless structures to model that, the cancellations don't eliminate echo perfect. The memory structures for canceling the memory echo are complex in general, i.e. Volterra model. As shown below, we will introduce the several memoryless and memory structures.

2.1 Memoryless Nonlinear AEC

2.1.1 Hammerstein model

The Hammerstein model, as shown in Fig. 2.1, is used to model the loudspeaker and acoustic channel in [6]. The structure has two layers. First layer is nonlinear order weights to model static loudspeaker part. Second layer is linear FIR to model dynamic echo path. The first layer can estimate the parameters of nonlinear part separately from linear part.

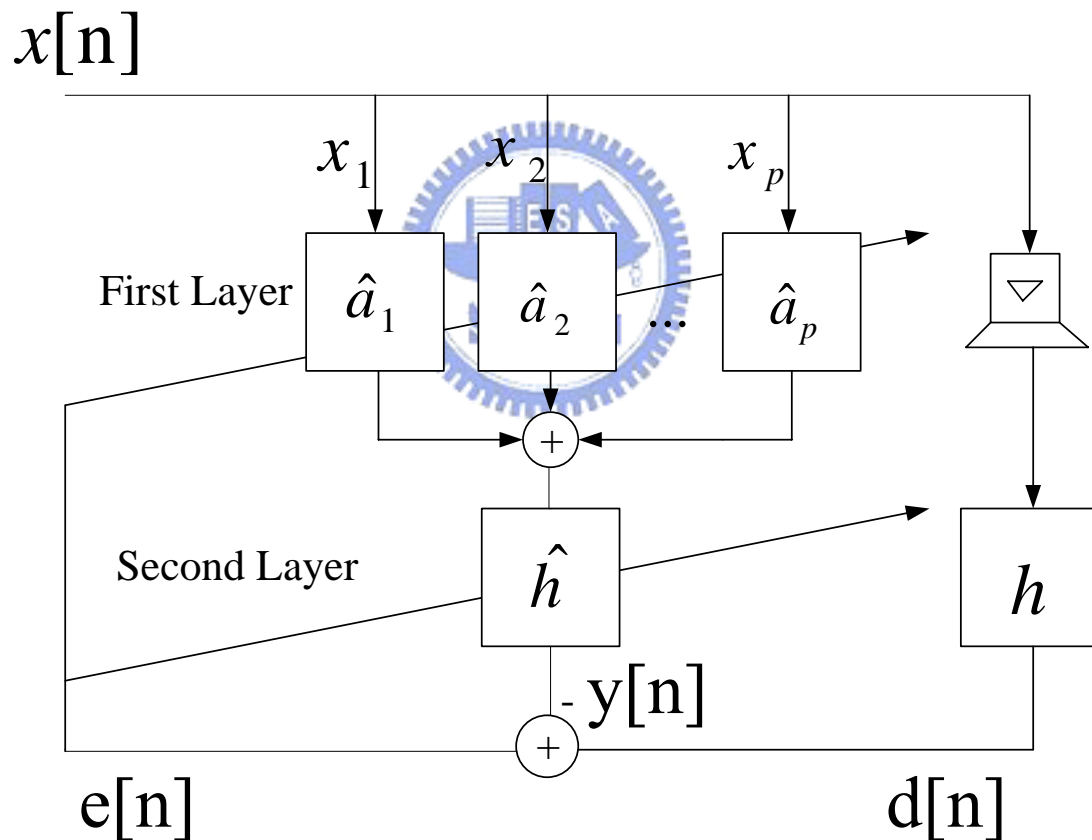


Fig. 2.1 Hammerstein structure

2.1.2 Orthogonalized Power Filter model

This method wants to use power filter to model nonlinear effect in [13-14]. A p -th order power filter defines by its input/output relation as follows

$$y[n] = \sum_{p=1}^P \sum_{l=0}^{N_p-1} \mathbf{h}_{p,l}[\mathbf{n}] \mathbf{x}^p[\mathbf{n}-l] \quad (2.1.5)$$

From (2.1.5) we notice that power filters can be considered as linear multiple input/single output systems, where the input of the p -th channel is given by the p -th power of $\mathbf{x}[n]$. The input of each channel is then filtered by as associated linear filter $h_{p,l}$ with channel length N_p . For compactness, we write (2.1.5) in matrix notation:

$$y[n] = \sum_{p=1}^P \mathbf{h}_p^T \mathbf{x}_p[\mathbf{n}] \quad (2.1.6)$$

With the vectors

$$\mathbf{x}_p[\mathbf{n}] = [x^p(n), x^p(n-1), \dots, x^p(n-N_p+1)]^T \quad (2.1.7)$$

$$\mathbf{h}_p = [h_{p,0}, h_{p,1}, \dots, h_{p,N_p-1}] \quad (2.1.8)$$

Note that the input signals of each channel are in general not mutually orthogonal, i.e., $E\{\mathbf{x}^i[\mathbf{n}]\mathbf{x}^j[\mathbf{n}]\} \neq 0$. Thus, a direct adaptive implementation of the no-orthogonalized power filter suffers from slow convergence. Therefore, a new set of mutually orthogonal input signals has been introduced [14]:

$$\mathbf{x}_{o,1}[\mathbf{n}] = \mathbf{x}[\mathbf{n}] \quad (2.1.9)$$

$$\mathbf{x}_{o,p}[\mathbf{n}] = \mathbf{x}^p[\mathbf{n}] + \sum_{i=1}^{p-1} q_{p,i} \mathbf{x}^i[\mathbf{n}] \quad (2.1.10)$$

for $1 < p \leq P$. The orthogonalization coefficients $q_{p,i}$ can be determined using the Gram-schmidt orthogonalization method. Applying standard gradient descent techniques, it directly follows from (2.1.6) that the normalized least mean square (NLMS) update of the coefficients of the orthogonalized structure by

$$h_{o,p}[n+1] = h_{o,p}[n] + \alpha_p \frac{\mathbf{x}_{o,p}[n]e[n]}{\mathbf{x}_{o,p}^T[n]\mathbf{x}_{o,p}[n]} \quad (2.1.11)$$

2.2 Memory Nonlinear AEC

In Sec.2.1, we assume the nonlinear effect of loudspeaker is memoryless. However, when the nonlinear memory effect of loudspeaker happens, the nonlinear memory AEC should be used.

2.2.1 Wiener Model

The Wiener model can model the memory loudspeaker in [15]. Wiener model has three layers. First layer and second layer model the nonlinear memory effect, and third layer models linear echo path. The Wiener model is shown Fig.2.2.

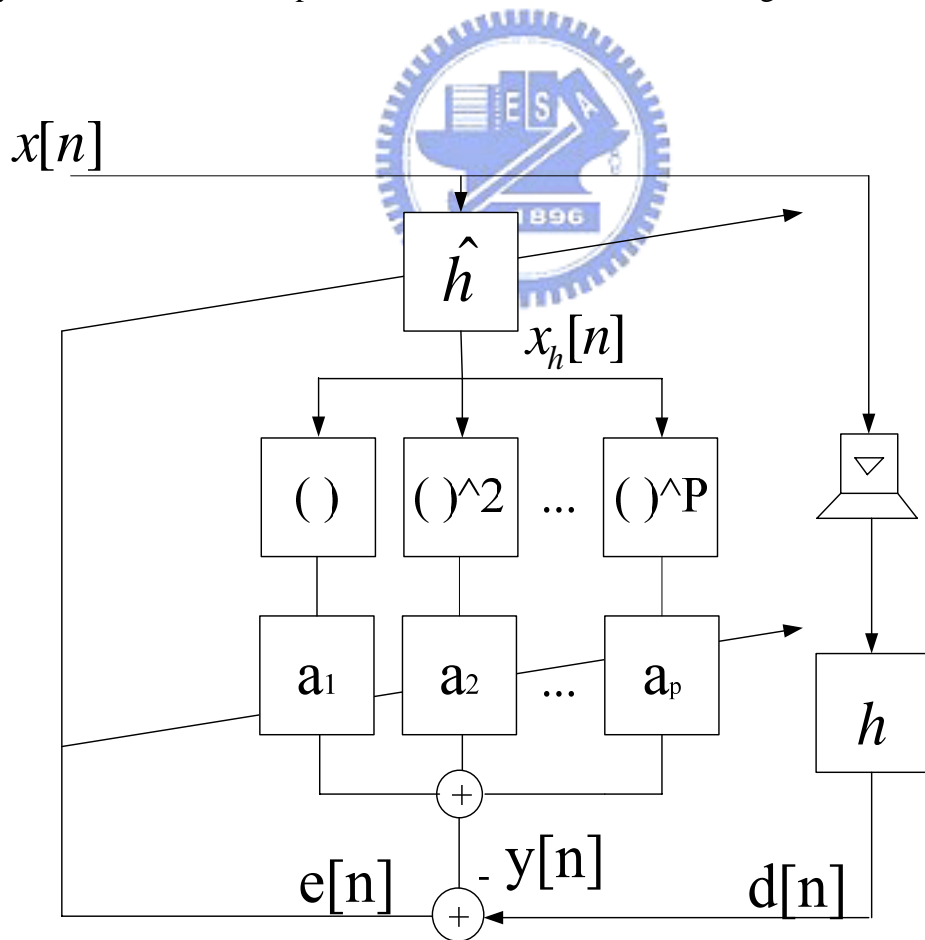


Fig. 2.2 Wiener model

2.2.2 Volterra Model

As show in Fig 2.3 the system is composed of two different modules organized in a cascaded structure: a first level is modeling the nonlinear loudspeaker effect based on polynomial Volterra structure, and a second level is modeling the room impulse response of the acoustic path with standard linear filter in.

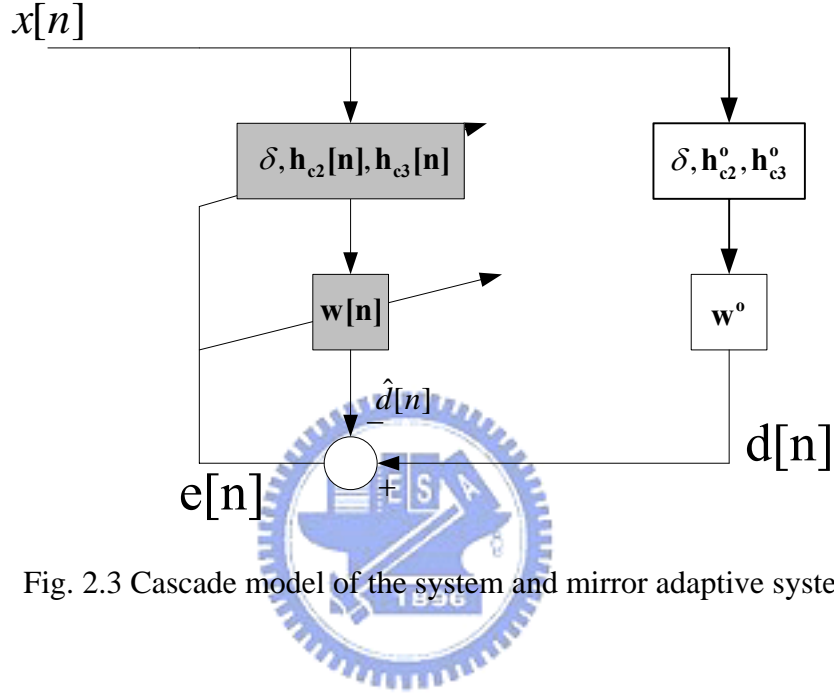


Fig. 2.3 Cascade model of the system and mirror adaptive system

The global include loudspeaker and acoustic path may be modeled by the parallel system (nonlinear/linear). The system, whose output is denoted by $d[n]$, is then modeled by a parallel structure with first-, second-, and third-order kernels

$$d[n] = \sum_{i=0}^L \mathbf{h}_1^o[n] \mathbf{x}[n-i] + \sum_{i=0}^L \sum_{j=i}^L \mathbf{h}_2^o[i, j] \mathbf{x}[n-i] \mathbf{x}[n-j] + \sum_{i=0}^L \sum_{j=i}^L \sum_{k=j}^L \mathbf{h}_3^o[i, j, k] \mathbf{x}[n-i] \mathbf{x}[n-j] \mathbf{x}[n-k] \quad (2.2.1)$$

$\hat{d}(n)$ is far-end signal pass through two layers adaptive filter to model nonlinear loudspeaker and linear channel.

$$\hat{d}(n) = \mathbf{X}_1^T[n] \mathbf{h}_1[n] + \mathbf{U}_2^T[n] \mathbf{h}_2[n] + \mathbf{U}_3^T[n] \mathbf{h}_3[n] \quad (2.2.2)$$

The input signal vectors for second- and third- order may be expressed as

products of matrices [16]. We define first the filter $\mathbf{h}_2[\mathbf{n}]$, corresponding to the second-order nonlinearity as the vector of dimension $L_2 \times 1$ (where $L_2 = L(L+1)/2$).

$$\mathbf{U}_2(\mathbf{n}) = \begin{bmatrix} x^2[n] & x^2[n-1] & \cdots & x^2[n-N+1] \\ x[n]x[n-1] & x[n-1]x[n-2] & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ x[n]x[n-L+1] & \vdots & & \vdots \\ x^2[n-1] & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x^2[n-L+1] & \vdots & \cdots & x^2[n-L-N+2] \end{bmatrix}_{L_2 \times N} \quad (2.2.3)$$

$$\mathbf{h}_2[\mathbf{n}] = \begin{bmatrix} h_2[0,0;n] \\ h_2[0,1;n] \\ \vdots \\ h_2[0,L-1;n] \\ h_2[1,1;n] \\ \vdots \\ h_2[L-1,L-1;n] \end{bmatrix}_{L_2 \times 1} \quad (2.2.4)$$



$$e(n) = d(n) - \hat{d}(n)$$

The update equation of a given vector $\mathbf{h}_1[\mathbf{n}], \mathbf{h}_2[\mathbf{n}], \mathbf{h}_3[\mathbf{n}]$, using the NLMS algorithm is given by [17-19].

$$\begin{aligned} \mathbf{h}_1[\mathbf{n}+1] &= \mathbf{h}_1[\mathbf{n}] + \frac{\mu_1 e[n] \mathbf{X}_1[\mathbf{n}]}{\mathbf{X}_1^T[\mathbf{n}] \mathbf{X}_1[\mathbf{n}] + \delta} \\ \mathbf{h}_2[\mathbf{n}+1] &= \mathbf{h}_2[\mathbf{n}] + \frac{\mu_2 e[n] \mathbf{U}_2[\mathbf{n}]}{\mathbf{U}_2^T[\mathbf{n}] \mathbf{U}_2[\mathbf{n}] + \delta} \\ \mathbf{h}_3[\mathbf{n}+1] &= \mathbf{h}_3[\mathbf{n}] + \frac{\mu_3 e[n] \mathbf{U}_3[\mathbf{n}]}{\mathbf{U}_3^T[\mathbf{n}] \mathbf{U}_3[\mathbf{n}] + \delta} \end{aligned} \quad (2.2.5)$$

The performance for Volterra structure is the best to the other structures, but the computer complexity is the most complexity.

2.3 Computation Reduction of Simplified Volterra Model

As the echo path inherently has memory, the nonlinear approaches have to be capable to model nonlinear system with memory. The Volterra filter is the common method to be deal with such system. Unfortunately, they suffer from high computational complexity. Therefore, we introduce an adaptive structure representing a simplified realization of a special second order Volterra filter in [20]. The simplified second Volterra filter is illustrated in Fig.2.4 and includes a linear branch and a second order nonlinear branch in parallel. The nonlinear branch consists of the cascade of an FIR filter, a multiplier, and a second stage filter.

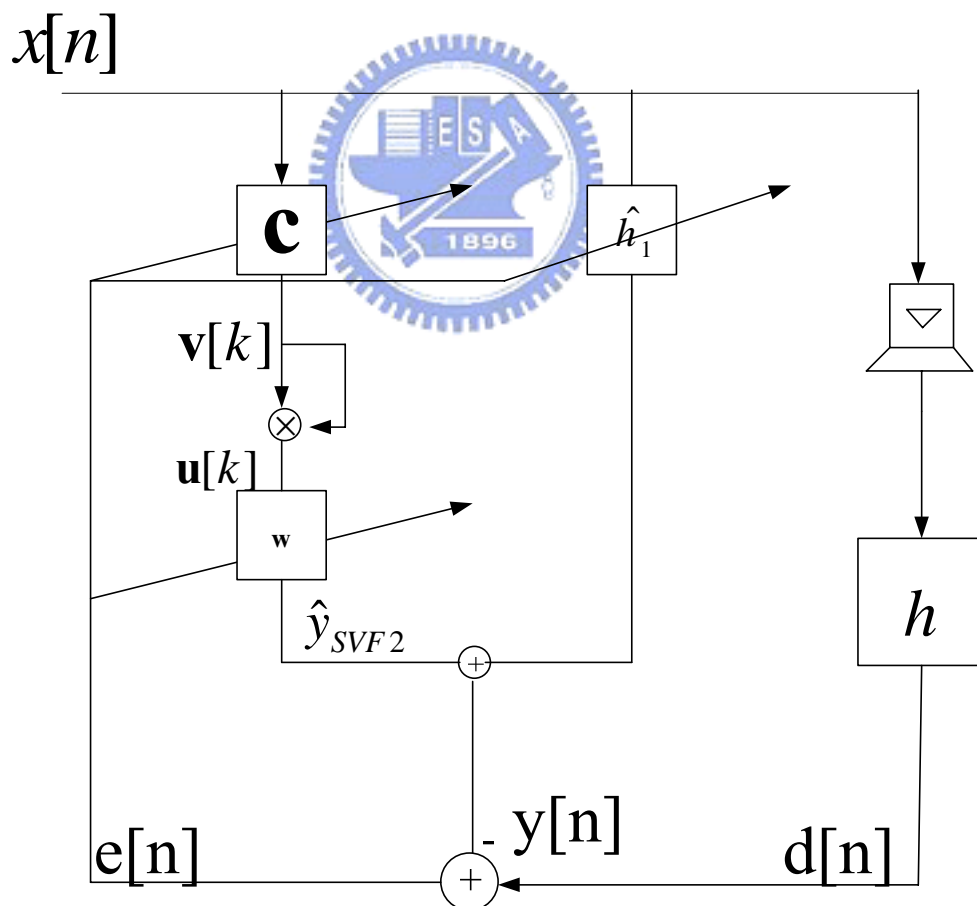


Fig. 2.4 simplified second order Volterra filter

The coefficients of Volterra second order is $\frac{N_v(N_v+1)}{2}$, the memory length is N_v .

However, the coefficients of simplified second order Volterra is N_v which is sum of $N_c + N_w$.

As the below, we will discuss the distribution for N_c and N_w at the total N_v coefficients. For example $N=5$, we expand three conditions $N_c = 4 N_w = 1$, $N_c = 2 N_w = 3$, $N_c = 1 N_w = 4$.

$$N_c = 4 N_w = 1:$$

$$v[n] = \sum_{i=0}^3 c_i x[n-i] \quad (2.2.6)$$

$$u[n] = \left(\sum_{i=0}^3 c_i x[n-i] \right)^2 = \sum_{i=0}^3 \sum_{j=0}^3 c_i c_j x[n-i] x[n-j] \quad (2.2.7)$$

$$\begin{aligned} y[n] &= w \cdot u[n] = \sum_{i=0}^3 \sum_{j=0}^3 w c_i c_j x[n-i] x[n-j] \\ &= w c_0 c_0 x[n] x[n] + (w c_1 c_0 + w c_0 c_1) x[n] x[n-1] + \\ &\quad (w c_2 c_0 + w c_0 c_2) x[n] x[n-2] + (w c_3 c_0 + w c_0 c_3) x[n] x[n-3] + \\ &\quad w c_1 c_1 x[n-1] x[n-1] + (w c_1 c_2 + w c_2 c_1) x[n-1] x[n-2] + \\ &\quad (w c_1 c_3 + w c_3 c_1) x[n-1] x[n-3] + w c_2 c_2 x[n-2] x[n-2] + \\ &\quad (w c_2 c_3 + w c_3 c_2) x[n-2] x[n-3] + w c_3 c_3 x[n-3] x[n-3] \end{aligned} \quad (2.2.8)$$

$$N_c = 2 N_w = 3$$

$$v[n] = \sum_{i=0}^1 c_i x[n-i] \quad (2.2.9)$$

$$u[n] = \left(\sum_{i=0}^1 c_i x[n-i] \right)^2 = \sum_{i=0}^1 \sum_{j=0}^1 c_i c_j x[n-i] x[n-j] \quad (2.2.10)$$

$$\begin{aligned} y[n] &= \sum_{k=0}^2 w_k \cdot u[n-k] = \sum_{k=0}^2 w_k \sum_{i=0}^1 \sum_{j=0}^1 c_i c_j x[n-k-i] x[n-k-j] \\ &= w_0 c_0 c_0 x[n] x[n] + (w_1 c_0 c_0 + w_0 c_1 c_1) x[n-1] x[n-1] + \\ &\quad (w_2 c_0 c_0 + w_1 c_1 c_1) x[n-2] x[n-2] + (w_0 c_1 c_0 + w_0 c_0 c_1) x[n] x[n-1] + \\ &\quad (w_1 c_1 c_0 + w_1 c_0 c_1) x[n-1] x[n-2] + (w c_1 c_2 + w c_2 c_1) x[n-1] x[n-2] + \\ &\quad (w_2 c_1 c_0 + w_2 c_0 c_1) x[n-2] x[n-3] + w c_2 c_2 x[n-2] x[n-2] + \\ &\quad w_2 c_1 c_1 x[n-3] x[n-3] \end{aligned} \quad (2.2.11)$$

$$N_c = 1 \quad N_w = 4:$$

$$v[n] = c \cdot x[n] \quad (2.2.12)$$

$$u[n] = (c \cdot x[n])^2 = c^2 \cdot x^2[n] \quad (2.2.13)$$

$$\begin{aligned} y[n] &= \sum_{k=0}^3 w_k c \cdot x[n] = \sum_{k=0}^2 w_k \sum_{i=0}^1 \sum_{j=0}^1 c_i c_j x[n-k-i] x[n-k-j] \\ &= w_0 c_0 c_0 x[n] x[n] + (w_1 c_0 c_0 + w_0 c_1 c_1) x[n-1] x[n-1] + \\ &\quad (w_2 c_0 c_0 + w_1 c_1 c_1) x[n-2] x[n-2] + (w_0 c_1 c_0 + w_0 c_0 c_1) x[n] x[n-1] + \\ &\quad (w_1 c_1 c_0 + w_1 c_0 c_1) x[n-1] x[n-2] + (w_0 c_1 c_2 + w_0 c_2 c_1) x[n-1] x[n-2] + \\ &\quad (w_2 c_1 c_0 + w_2 c_0 c_1) x[n-2] x[n-3] + w_0 c_2 c_2 x[n-2] x[n-2] + \\ &\quad w_2 c_1 c_1 x[n-3] x[n-3] \end{aligned} \quad (2.2.14)$$

Table 2.1 The parameter of simplified Volterra Model

	$N_c = 4 \quad N_w = 1$	$N_c = 2 \quad N_w = 3$	$N_c = 1 \quad N_w = 4$
h_{00}	$w_0 c_0 c_0$	$(w_0 + w_1 + w_2) c_0 c_0$	$(w_0 + w_1 + w_2 + w_3) c_0 c_0$
h_{01}	$w_0 c_0 c_1 + w_1 c_1 c_0$	$(2w_0 + 2w_1 + 2w_2) c_0 c_1$	0
h_{02}	$w_0 c_0 c_2 + w_1 c_1 c_0$	0	0
h_{03}	$w_0 c_0 c_3 + w_1 c_1 c_0$	0	0
h_{11}	$w_1 c_1 c_1$	$(w_0 + w_2) c_1 c_1$	0
h_{12}	$w_1 c_1 c_2 + w_2 c_2 c_1$	$w_2 c_1 c_2$	0
h_{13}	$w_1 c_1 c_3 + w_2 c_2 c_1$	0	0
h_{22}	$w_2 c_2 c_2$	0	0
h_{23}	$w_2 c_2 c_3$	0	0
h_{33}	$w_3 c_3 c_3$	0	0

When the N_c is more, the coefficients is more for nonlinear effect. From Table 2.1, modeling nonlinear memory effect of $N_c = 4 N_w = 1$ is the same with Volterra filter. However, the linear adaptive filter tap is one. $N_c = 1 N_w = 4$ is the same with linear adaptive filter. But, this condition can't estimate nonlinear effect at all. The total modeling nonlinear memory coefficients for simplified Volterra structure are shown in (2.2.15) and Volterra structure are shown in (2.2.16) .

$$\frac{(N_c + 1)N_c}{2} + N_c (N_w - 1) \quad (2.2.15)$$

$$\frac{(N_v + 1)N_v}{2} \quad (2.2.16)$$



Chapter 3

Estimate Nonlinear Residual Error for Nonlinear Acoustic Echo Suppression

Loudspeakers and amplifiers of mobile communication device may cause significant nonlinear distortion in the acoustic echo path, resulting in a limitation of the performance of only using linear acoustic echo cancellation [1-5]. There are several nonlinear acoustic echo cancellations [6-10] to overcome this distortion. But these methods converge too slowly for the room impulse response is changing on the going. So, we introduce nonlinear acoustic echo suppression (NAES). The convergence speed is faster than nonlinear acoustic echo cancellation (NAEC) [12]. Besides, NAES suppresses the residual echo that remains after a purely linear AEC is better than NAEC cancels the acoustic echo. There is linear acoustic echo cancellation and nonlinear acoustic echo suppression structure in Fig. 3.1.

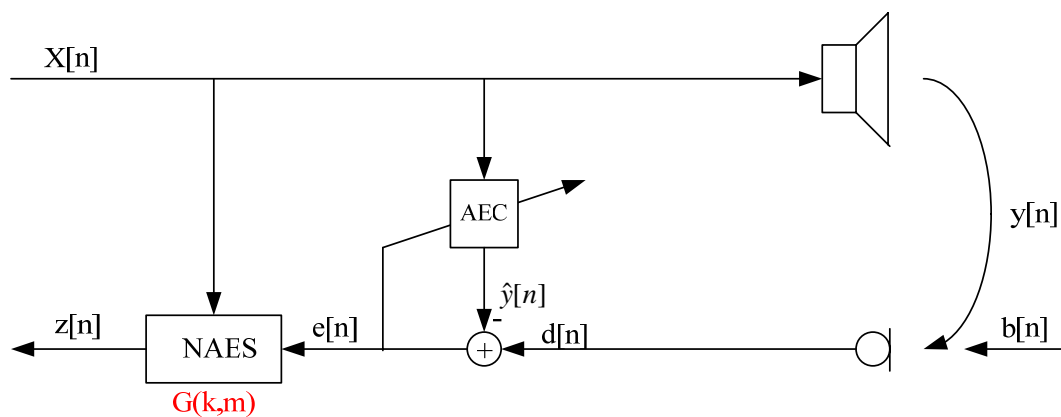


Fig. 3.1 Linear AEC and nonlinear acoustic echo suppression structure

The near-end microphone picks up the signal $d[n]$ which includes the acoustic echo speech $y[n]$ and near-end speech $b[n]$ in silent room. The acoustic echo speech $y[n]$ is far-end speech passing through nonlinear loudspeaker and room impulse response. We use a linear adaptive filter to find room impulse response, and want to get replica acoustic echo $\hat{y}[n]$ to cancel the desire signal $d[n]$. However, this adaptive filter can't find the loudspeaker's nonlinear channel. The residual error is very large by only using linear acoustic echo cancellation (AEC) when a high level's power speech is injected into a small loudspeaker. That is because the loudspeaker has been operated at saturation region. So, we want to find the filter $G(k, m)$ after AEC to suppress nonlinear residual error. If the optimum gain can be found, the output signal of suppressed signal will be close to the near-end signal $b[n]$.

3.1 Wiener Filter



The foundation of nonlinear acoustic echo suppression is Wiener filter [1]. Consider the block diagram of Fig. 3.2 built around a linear discrete-time filter. The filter input consists of a time series $u(0), u(1), u(2), \dots$, and the filter is itself characterized by the impulse response w_0, w_1, w_2, \dots . At some discrete time n , the filter produces an output denoted by $y(n)$. The output is used to provide an estimate of a desired response designated by $d(n)$. With the filter input and the desired response representing single realizations of respective stochastic processes, the estimation is ordinarily accompanied by an error with statistical characteristics of its own. In particular, the estimation error, denoted by $e(n)$, is defined as the difference between the desired response $d(n)$ and the filter output $y(n)$.

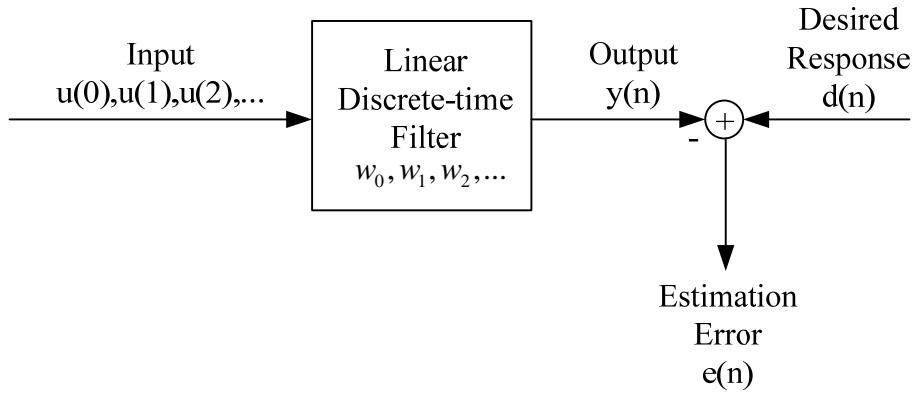


Fig. 3.2 Block diagram representation of the statistical filtering problem

The optimum linear filter in Fig.3.2 is shown in Eq. (3.1.13). Assuming that $S_{UU}(\Omega) \neq 0$ for all Ω , we find the following transfer function of the noncausal Wiener filter:

$$H_{opt}(\Omega) = \frac{S_{UD}^*(\Omega)}{S_{UU}(\Omega)} \quad (3.1.13)$$

3.2 Acoustic Echo Suppression Structures

The integration [21] of noise reduction and echo cancellation has advantage of utilizing the synergy among its components. An important issue is the placement of these two algorithms. These structures can only be implemented in frequency domain because noise reduction requires frequency-domain implementation using FFT. Since the performance of the NLMS algorithm degrades significantly in the presence of high-level background noise, an immediate suggestion would be place the noise reduction prior to echo cancellation. However, the drawback is that the noise reduction introduces nonlinearity into the echo path.

In this section, we only take care of the linear echo and background noise. Only using the AEC to cancel the echo is not enough to assure the quality of auditory. This

is because another side user hears the background noise and the background noise interferes with the AEC algorithm. We will introduce two combined structures as below.

3.2.1 AEC+NR structure

The combined system is shown in Fig. 3.3 [21]. We use a conventional echo canceller, consisting of a time variant FIR-filter adapted by the NLMS algorithm, and of a combined residual echo and noise reduction filter implemented in the frequency domain. $x[n]$ denotes the far-end speech, $b[n]$ the near-end speech and $v[n]$ the noise. The microphone signal $d[n]$ is made up of the echo $y[n]$ as well as of the near-end speech and noise,

$$d[n] = y[n] + b[n] + v[n] \quad (3.2.1)$$

The estimated echo $\hat{y}[n]$ is subtracted from $d[n]$ forming the echo compensated signal $e[n]$,

$$\begin{aligned} e[n] &= y[n] - \hat{y}[n] + b[n] + v[n] \\ &= \Delta y[n] + b[n] + v[n] \end{aligned} \quad (3.2.2)$$

Depending on the effectiveness of the echo canceller, the residual echo $\Delta y[n] = y[n] - \hat{y}[n]$ must be more or less attenuated by the filter G . The output signal of the system is denoted by $z[n]$. If the AEC is perfect, the residual error $e[n]$ is shown in (3.2.3).

$$e[n] \approx b[n] + v[n] \quad (3.2.3)$$

The suppression gain G suppresses the noise $v[n]$, and the output $z[n]$ is approximated to $b[n]$.

$$Z(w) = G(w) \cdot E(w) \quad (3.2.4)$$

The suppress gain comes from the Wiener filter concept in Eq. (3.13) of Sec.3.1.

$$G(w) = \frac{S_{EB}}{S_{EE}} \quad \text{if } b \text{ and } n \text{ are uncorrelated} \quad (3.2.5)$$

$$= \frac{S_{BB}}{S_{EE}}$$

$S_{BB}(w)$ is power spectral density of near-end speech and the $S_{EE}(w)$ is power spectral density of residual error.

The error signal $e[n]$ is then processed by noise reduction to get the noise free signal $z[n]$. With an optimal echo, the echo is completely canceled by the first filter, leaving the useful signal and noise unchanged. The output from the AEC is ideally $b[n] + v[n]$. The second stage aims at reducing noise through the Wiener gain filter $\frac{S_{BB}}{S_{BB} + S_{VV}}$. A disadvantage of this integrated structure is that the AEC has to process noisy signals. In practice, the AEC system is adaptive. The coefficients of the AEC are disturbed by the ambient noise which is omnipresent.

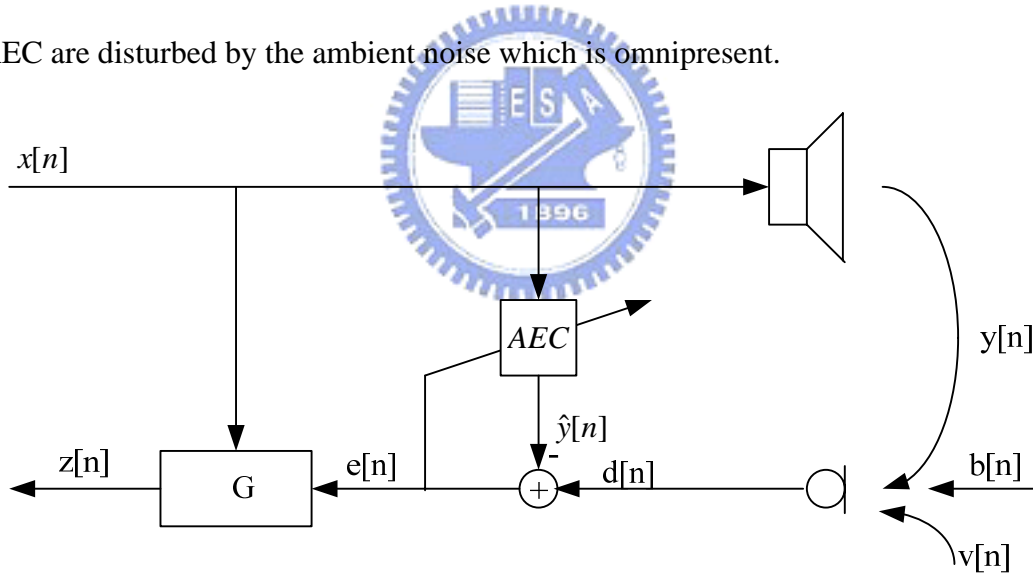


Fig. 3.3 Block diagram of a combined echo canceling and noise reduction system

3.2.2 NR+AEC structure

Fig. 3.4 [21] shows the implementation of the integrated noise reduction and echo cancellation where echo cancellation precedes noise reduction. $x[n]$ is processed by the AEC filter H to generate the echo $y[n]$, which is subtracted from

$d'[n]$ that is the microphone signal $d[n]$ after suppressing the noise to generate $e[n]$.

The effects of noise on the AEC can be minimized by placing noise reduction upstream from this system. The noise reduction operation enhances the signal-to-noise (SNR), which can improve the AEC behavior. However, the noise reduction causes nonlinear distortion, and disturbs the AEC.

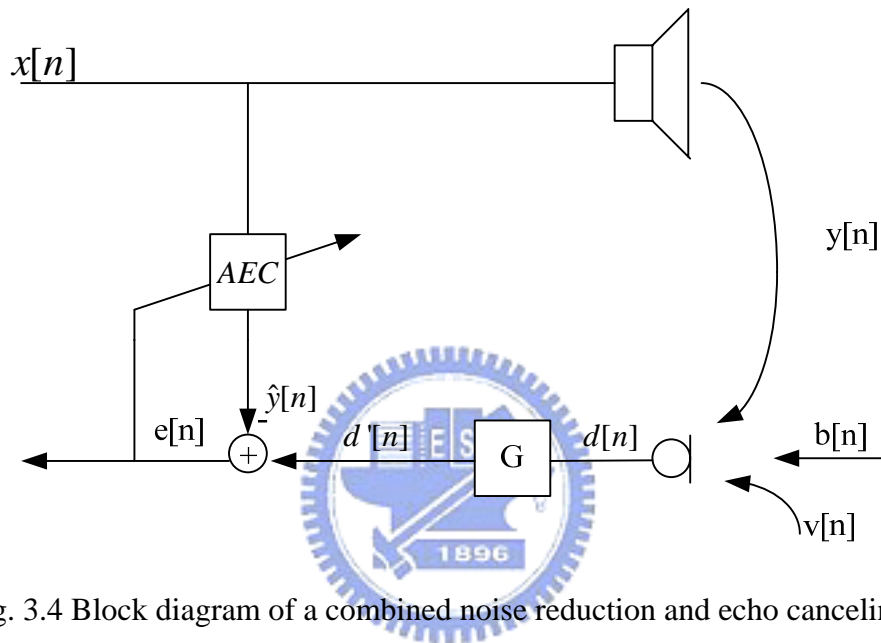


Fig. 3.4 Block diagram of a combined noise reduction and echo canceling system

3.3 Estimation nonlinear residual error for NAES

One of most basic filter to echo suppression is Wiener filter in Fig3.3.

The Wiener filter gain is shown in (3.3.1).

$$G = \frac{S_{EZ}}{S_{EE}} \quad (3.3.1)$$

If the $Z = B$, the suppression performs optimally. E concludes three parts which are nonlinear residual echo Y_{nl} , near-end speech b , and background noise N . Assuming the residual echo, near-end speech, and near-end noise are uncorrelated and ignoring the noise, we have

$$\begin{aligned}
G &= \frac{S_{EB}}{S_{EE}} \\
&= \frac{S_{BB}}{S_{BB} + S_{Y_{nl}Y_{nl}} + S_{NN}}
\end{aligned} \tag{3.3.2}$$

In fact, we don't know S_{BB} . So, we want to use S_{EE} , $S_{Y_{nl}Y_{nl}}$, and S_{NN} to estimate G .

$$G = \frac{S_{EE} - S_{Y_{nl}Y_{nl}} - S_{NN}}{S_{EE}} \tag{3.3.3}$$

S_{EE} can be easily estimated from E . $S_{Y_{nl}Y_{nl}}$ is more difficult to estimate.

We assume the background noise is very small in the quiet room [21].

$$S_{NN} \approx 0 \tag{3.3.4}$$

This would be the case in quiet offices or in cars that are not moving and when the engine is switched off. The nonlinear residual echo suppression filter is used to reduce the nonlinear echo further. For the transfer function of this filter, a Wiener filter is often applied:

$$\begin{aligned}
H_{RE} &= \frac{\hat{S}_{EB}}{\hat{S}_{EE}} \\
&= \frac{\hat{S}_{EE} - \hat{S}_{Y_{nl}Y_{nl}}}{\hat{S}_{EE}} \\
&= 1 - \frac{\hat{S}_{Y_{nl}Y_{nl}}}{\hat{S}_{EE}}
\end{aligned} \tag{3.3.5}$$

It should be noted that any impact of the residual echo suppression filter on residual echoes also affects the local speech signal. When applying Eq.(3.3.5), the estimated power spectral densities \hat{S}_{EE} and $\hat{S}_{Y_{nl}Y_{nl}}$ contain estimation errors. Therefore, the quotient may become larger than one. To prevent that, the filter transfer function can be used in Eq.(3.3.6) where H_{\min} determines the maximum attenuation of the filter. The overestimation parameter [21] β can be used to control the ‘‘aggressiveness’’ of the filter.

$$H_{RE} = \max \left[1 - \beta \frac{\hat{S}_{Y_{nl}Y_{nl}}}{\hat{S}_{EE}}, H_{\min} \right] \quad (3.3.6)$$

In order to estimate the short-term power spectral density of the error signal \hat{S}_{EE} first-order IIR smoothing of the squared magnitudes of the frequency domain error signals E is applied in order to estimate the short-term power spectral density. m represents the frame index.

$$\hat{S}_{EE}(m) = (1 - \gamma)|E|^2 + \gamma\hat{S}_{EE}(m-1) \quad (3.3.7)$$

Because the disturbed error Y_{nl} is not accessible, the estimation of the short-term power spectral density $\hat{S}_{Y_{nl}Y_{nl}}$ can't be approximated in the same manner as \hat{S}_{EE} .

As the above, we will discuss how to estimate $\hat{S}_{Y_{nl}Y_{nl}}$. If the $\hat{S}_{Y_{nl}Y_{nl}}$ can be estimated accurately, the low level of echo signal and minimum distortion of near-end speech can be achieved.



3.3.1 Based on Highly Correlated Nonlinear Residual Echo

This method [11] proposes a new residual-echo model based on the spectral correlation between the residual echo and the echo replica. For this method there should have ambient length for near-end at beginning. At first, find the ratio of AEC residual error $|E(k, m)|$ to AEC replica $|Y(k, m)|$ when there is only single talk. k, m represent the frame index, and the frequency bin, respectively. The $|E(k, m)|$ is approximated to $|E_{nl}(k, m)|$ in quiet room.

$$|E(k, m)| \approx |E_{nl}(k, m)| \approx a(k, m) \cdot |Y(k, m)| \quad (3.3.8)$$

Let us consider approximating $a(k, m)$ by $\hat{a}(k)$ using averaged absolute

values of the residual echo and the echo replica. Then,

$$\hat{a}(k) = \frac{\overline{|E(k, m)|}_{\text{single talk}}}{\overline{|Y(k, m)|}} \quad (3.3.9)$$

where overline $\overline{\cdot}$ means an average operation. When there is no near-end speech, i.e. $B(k, m) = 0$. By approximating $a(k, m)$ in (3.3.8) with a regression coefficient $\hat{a}(k)$, the residual echo $|E_{nl}(k, m)|$ is modeled as the product of $\hat{a}(k)$ and $|Y(k, m)|$.

$$|E_{nl}(k, m)| \approx |\hat{E}_{nl}(k, m)| \triangleq \hat{a}(k) \cdot |\hat{Y}(k, m)| \quad (3.3.10)$$

So, we can find suppression magnitude gain $G_{o,p,m}(k, m)$.

$$\begin{aligned} \hat{G}_{opt}(k, m) &= \frac{S_{BB}(k, m)}{S_{EE}(k, m)} \\ &= \frac{S_{EE}(k, m) - S_{E_{nl}E_{nl}}(k, m)}{S_{EE}(k, m)} \end{aligned} \quad (3.3.11)$$

However, this method has drawback. The property between the nonlinear residual error and linear echo is not linear in real system.

3.3.2 Power Filter Model

We will find a frequency-depend gain $G(k, m)$ which is cascade the residual error from AEC [12]. If the optimum gain is found, the suppressed error will be close to the near-end signal. The output $Z(k, m)$ of the AES with input $e[n]$ reads in the frequency domain.

$$Z(k, m) = G(k, m)E(k, m) \quad (3.3.17)$$

Here, $E(k, m)$ denotes the STFT of $e[n]$, where k represents the block time index and m represents frequency bin. For the power spectral density,

$$S_{ZZ}(w) = G(w)S_{EE}(w) \quad (3.3.18)$$

If we want to find optimum value $G(w)$, $S_{ZZ}(w)$ is equal to near-end background noise $S_{BB}(w)$, and $S_{EE}(w)$ is equal to AEC residual error which includes $S_{BB}(w)$ and the high order nonlinear echo signal $S_{nl}(w)$.

$$\begin{aligned} G_{opt}(k, m) &= \frac{S_{ZZ}(k, m)}{S_{EE}(k, m)} \\ &= \frac{S_{BB}(k, m)}{S_{BB}(k, m) + S_{nl}(k, m)} \\ &= \frac{S_{EE}(k, m) - S_{nl}(k, m)}{S_{EE}(k, m)} \end{aligned} \quad (3.3.19)$$

In the above equation, $S_{EE}(w)$ can be obtained from AEC, but $S_{nl}(w)$ can not. We will try to find the approximate the high order nonlinear echo power spectral density. This method uses power filter model to replace the loudspeaker nonlinear effect.

The input/output relation of a p th-order power filter is given by

$$y[n] = \sum_{p=1}^P \sum_{n=0}^{N-1} h_{p,n} x^p(k, m) \quad (3.3.20)$$

where $h_{p,n}$ denotes the filter coefficients of the p th channel having input $x^p[n]$.

Aiming at a frequency-domain implementation of power filters, we give the short-time Fourier transfer (STFT) representation of (3.3.20):

$$Y(k, m) = \sum_{p=1}^P H_{p,m} X_p(k, m) \quad (3.3.21)$$

Here, $X_p(k, m)$ denotes the STFT of $x^p(k)$ of length M . k, m represent k th frame and m th frequency bin. $H_{p,m}$ is the coefficient of the p th power filter corresponding to the m th frequency bin. In order to un-correlating the channel inputs, we will use equivalent orthogonal structure (EOS) [14].

The adaptation of the channels $\hat{H}_{o,p,m}(k)$ is performed independently for each nonlinear channel ($p > 2$) with respect to the channel dependent error signal

$$E_{o,p}(k,m) = E(k,m) - \hat{Y}_{o,p}(k,m) \quad (3.3.27)$$

In comparison to the application of adaptive power filters to nonlinear echo cancellation, they are not interested in an exact identification of the nonlinear components of the echo path $H_{o,p,m}(k,m)$ by their adaptive counterparts $\hat{H}_{o,p,m}(k,m)$ [15]. They are rather aiming at estimates of $S_{Y_{o,p}}(k,m)$.

3.4 Suppression of nonlinear residual error and background noise

The combined structure from Sec.3.2 can suppress the background noise. If the loudspeaker is ideal, the combined structures of AEC and NR could cancel the linear echo and suppress the background noise. However, the loudspeaker for mobile communication is cheaper and small, so the nonlinear effect is obvious. The methods in Sec. 3.3 ignore the background noise, and only cancel the linear echo and suppress the nonlinear residual error. In this section, we will suppress the noise first, and then use this concept in the NR+AEC structure of Section 3.3.1.

3.4.1 AEC+NR suppression of noise and nonlinear residual error

The NR+AEC structure to suppress the background noise and nonlinear residual error is shown in Fig.3.7. The microphone receives the linear echo, nonlinear echo, near-end speech signal, and noise signal. In optimal condition, the AEC can cancel the linear echo and suppression gain G_{ANnl} suppress the nonlinear residual error and the background noise in (3.4.1).

$$G_{ANnl} = \frac{S_{BB_ANnl_avg}}{S_{EE_avg}} \quad (3.4.1)$$

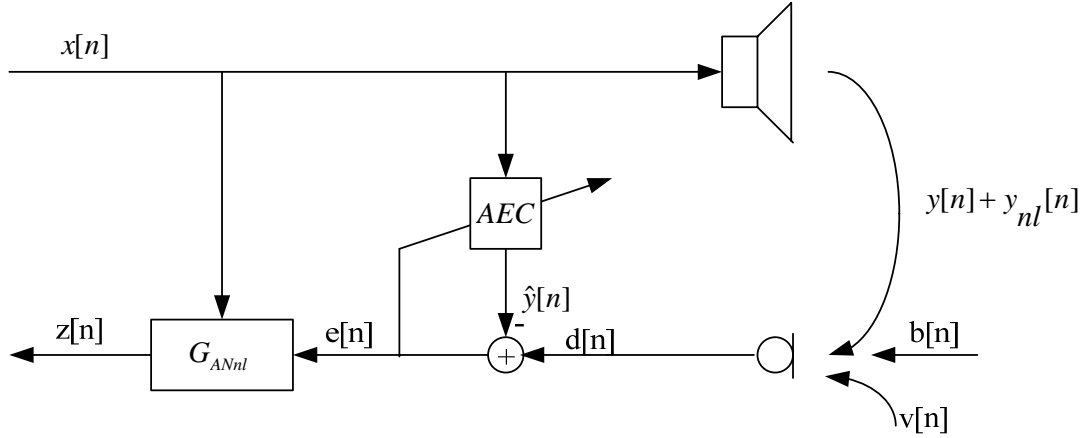


Fig. 3.5 Block diagram of NR+AEC to suppress noise and nonlinear residual error

S_{E_avg} is the average power spectral density of the residual error from AEC in (3.4.2).

$$S_{E_avg} = \beta \times S_{E_avg} + (1 - \beta) \times S_E \quad (3.4.2)$$

In ideal condition, the $S_{B_ANnl_avg}$ is approximated to S_{B_avg} . However, we can't know S_{B_avg} . We can estimate $S_{B_ANnl_avg}$ to replace S_{B_avg} . $S_{B_ANnl_avg}$ is used S_{E_avg} to eliminate S_{N_avg} and $S_{Nonlinear_avg}$. The background noise N can be estimated first when there is no far-end speech signal.

$$S_{B_ANnl_avg} = S_{E_avg} - S_{N_avg} - S_{Nonlinear_avg} \quad (3.4.3)$$

$$S_{N_avg} = \beta \times S_{N_avg} + (1 - \beta) \times S_N \quad (3.4.4)$$

$$S_{Nonlinear_avg} = \beta \times S_{Nonlinear_avg} + (1 - \beta) \times S_{Nonlinear} \quad (3.4.5)$$

3.4.2 NR+AEC suppression of noise and nonlinear residual error

In Section 3.4.1, the combined structure cancels echo first, and suppresses

background noise continuous. However, background noise disturbs adaptive filter for echo cancellation. So, NR+AEC structure overcomes this disturbance in Fig. 3.6.

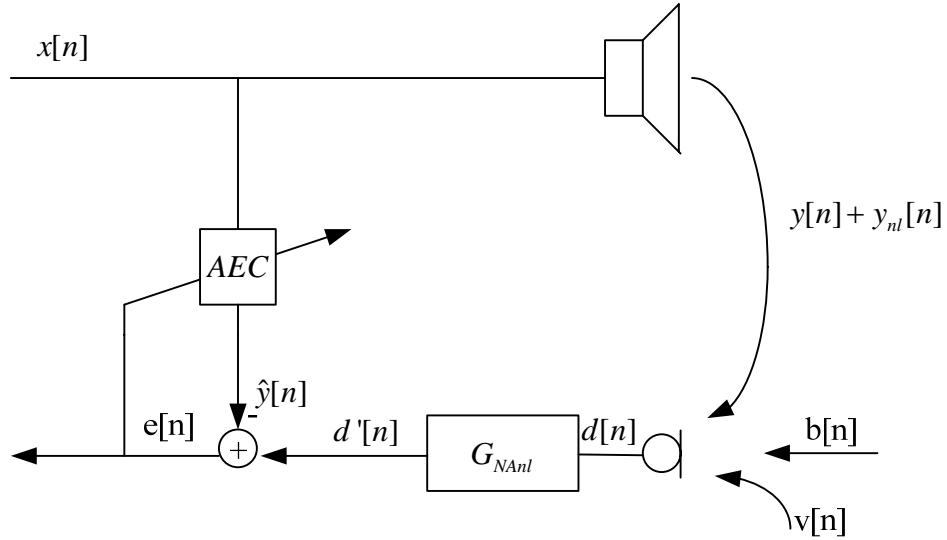
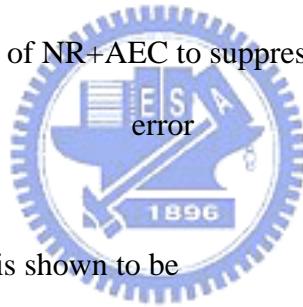


Fig. 3.6 Block diagram of NR+AEC to suppress noise and nonlinear residual



The suppression gain G_{NAml} is shown to be

$$G_{NAml} = \frac{S_{D'-avg}}{S_{D-avg}} \quad (3.4.6)$$

S_{D-avg} is the average power spectral density of the signal received from microphone in (3.4.7).

$$S_{D-avg} = \beta \times S_{D-avg} + (1 - \beta) \times S_D \quad (3.4.7)$$

In ideal condition, the $S_{D'-avg}$ is approximated to S_{B-avg} and the average power spectral density of linear echo S_{y-avg} . However, we can't know S_{B-avg} and S_{y-avg} .

We can estimate $S_{D'-avg}$ which is used S_{D-avg} to eliminate S_{N-avg} and $S_{Nonlinear-avg}$ to replace optimal $S_{D'-avg}$ in Eq. (3.4.8).

$S_{D'_{avg}}$ includes linear echo. The linear echo has been removed from $S_{B_{ANnl_{avg}}}$. So, $S_{D'_{avg}}$ is larger than $S_{B_{ANnl_{avg}}}$, and G_{NAnnl} will cause larger distortion than G_{ANnl} for near-end speech signal.

$$S_{D'_{avg}} = S_{D_{avg}} - S_{N_{avg}} - S_{Nonlinear_{avg}} \quad (3.4.8)$$

NR+AEC structure can suppress background noise before adaptive filter. However, the suppression gain will cause distortion for microphone signal. The adaptive filter can't estimate real room impulse response. The AEC will not cancel echo perfect. In

3.5 Volterra structure for NAES

The high order echo basis of the nonlinear acoustic echo suppression can estimate the power spectral density of nonlinear residual error. However, the nonlinear memory effect maybe arises in some microphone. The nonlinear acoustic echo suppression can't suppress this error and we use the Volterra structure in Section 2.2.3 to overcome nonlinear memory effect. In this section, we use second order Volterra structure to suppress the nonlinear memory effect in frequency domain [22].

First, get the Fourier transfer of the second order Volterra for far-end speech signal. The Fourier transfer for linear far-end speech signal is shown in(3.5.1).

$$X(w) = \sum_{k=0}^{N-1} x[k] e^{-j \frac{2\pi kw}{N}} \quad (3.5.1)$$

The Fourier transfer of the second order Volterra for far-end speech signal is shown in(3.6.2). An second order discrete Volterra filter with input $x[k]$, frequency $Xv(w)$ and memory length L can be described as

$$Xv(w) = \sum_{k=0}^{N-1} \sum_{p=0}^{L-1} \sum_{q=i}^{L-1} x[k-p] x[k-q] e^{-j \frac{2\pi kw}{N}} \quad (3.5.2)$$

The nonlinear echo basis and Volterra structure is shown in Fig.3.13

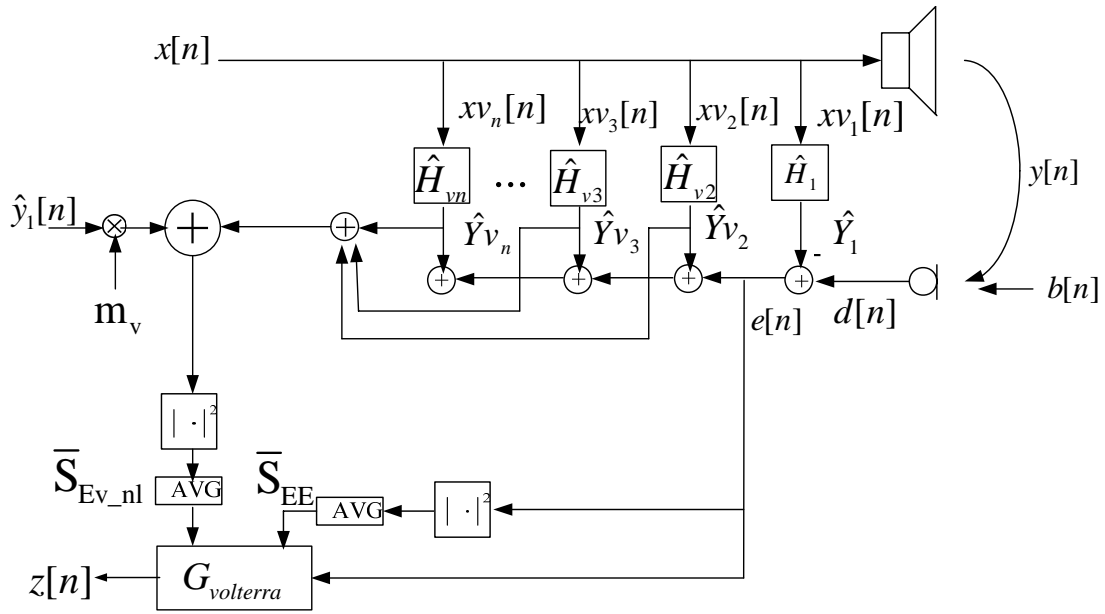


Fig. 3.7 The Volterra structure for NAES

The Volterra structure can estimate second order nonlinear residual error $\hat{Y}_{v_2}(k)$, and use the value to estimate slope $m_v(k)$. $m_v(k)$ is more accurate than $\hat{a}(k)$ in Eq. (3.3.9) to estimate nonlinear residual error, because the second order residual error has been removed.

$$m_v(k) = \frac{E_{b=0}(k) - \hat{Y}_{v_2}(k)}{\hat{Y}_1(k)} \quad (3.5.3)$$

Then, we will find nonlinear residual error from $m_v(k)$. The nonlinear residual error is equivalent to the linear echo $\hat{Y}_1(k, m)$ multiplied the fixed value for every frame $m_v(k)$ plus second order Volterra echo.

$$Ev_{nl}(k, m) = \hat{Y}_1(k, m) \cdot m_v(k) + \hat{Y}_{v_2}(k, m) \quad (3.5.4)$$

For finding suppression gain, we use $S_{Ev_{avg}}$ and $S_{Enlv_{avg}}$ are the power spectrum density and average from linear residual error $Ev(k, m)$ and nonlinear residual error $Ev_{nl}(k, m)$.

$$G_volterra(k,m)=\frac{S_{Ev_avg}-S_{Enlv_avg}}{S_{Ev_avg}} \quad (3.5.5)$$

The suppression error is the power spectrum from instant linear residual error

$$Z(k,m)=G_volterra(k,m) \cdot S_{Ev_instant} \quad (3.5.6)$$

Using Volterra structure for NAES can suppress nonlinear residual error more accurate than Hammerstein structure.

3.6 Using high order nonlinear echo to estimate nonlinear residual error

In Section 3.3.1, it is assumed that the nonlinear residual errors of orders more than two are linearly related with linear echo \hat{Y} . In fact, they are not. The assumption can be wrong. Fig. 3.8 is a block diagram of Section 3.3.1. However, if we could estimate higher order of residual error, the slope $m(k)$ is more accurate to estimate nonlinear residual error

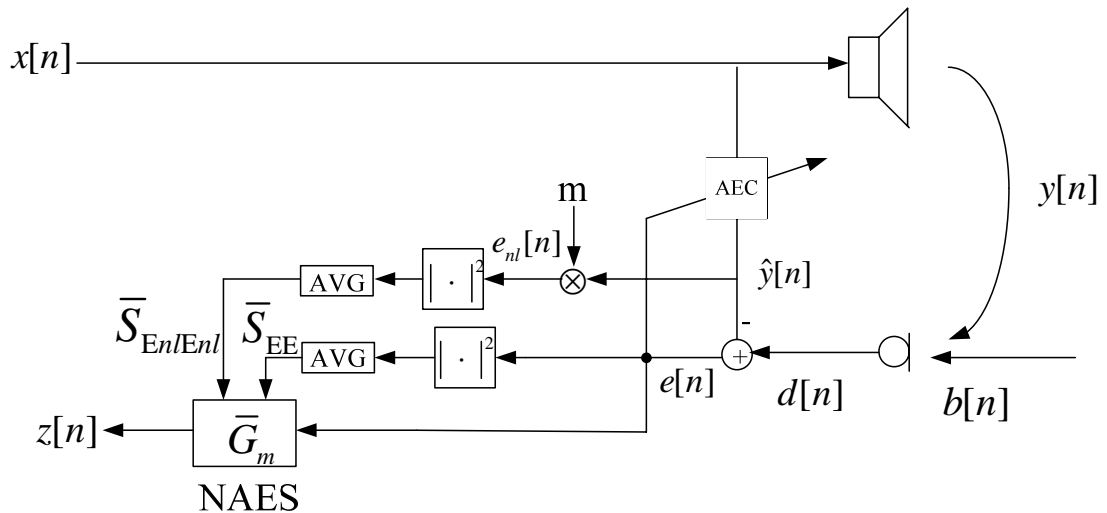


Fig. 3.8 Block diagram of Sec3.3.1

The nonlinear power filter can estimate nonlinear residual error to NAES in Section 3.3.2. So, we exploit this ideal to find medium order nonlinear echoes accurately and use slope method in Section 3.3.1 to estimate high order nonlinear echoes roughly.

We propose the new method as shown in Fig.3.9. The block diagram includes first, and third-order adaptive filters. For finding $m_{pro}(k)$ first, the difference to Section 3.3.1 is removing the third order echo.

$$m_{pro}(k) = \frac{E_{b=0}(k) - \hat{Y}_3(k)}{\hat{Y}_1(k)} \quad (3.6.1)$$

The nonlinear residual error includes two part that are third-order echo, and above fifth-order echoes.

$$E_{nl}(k, m) = \hat{Y}_3(k, m) + \hat{Y}(k, m) \cdot m_{pro}(k) \quad (3.6.2)$$

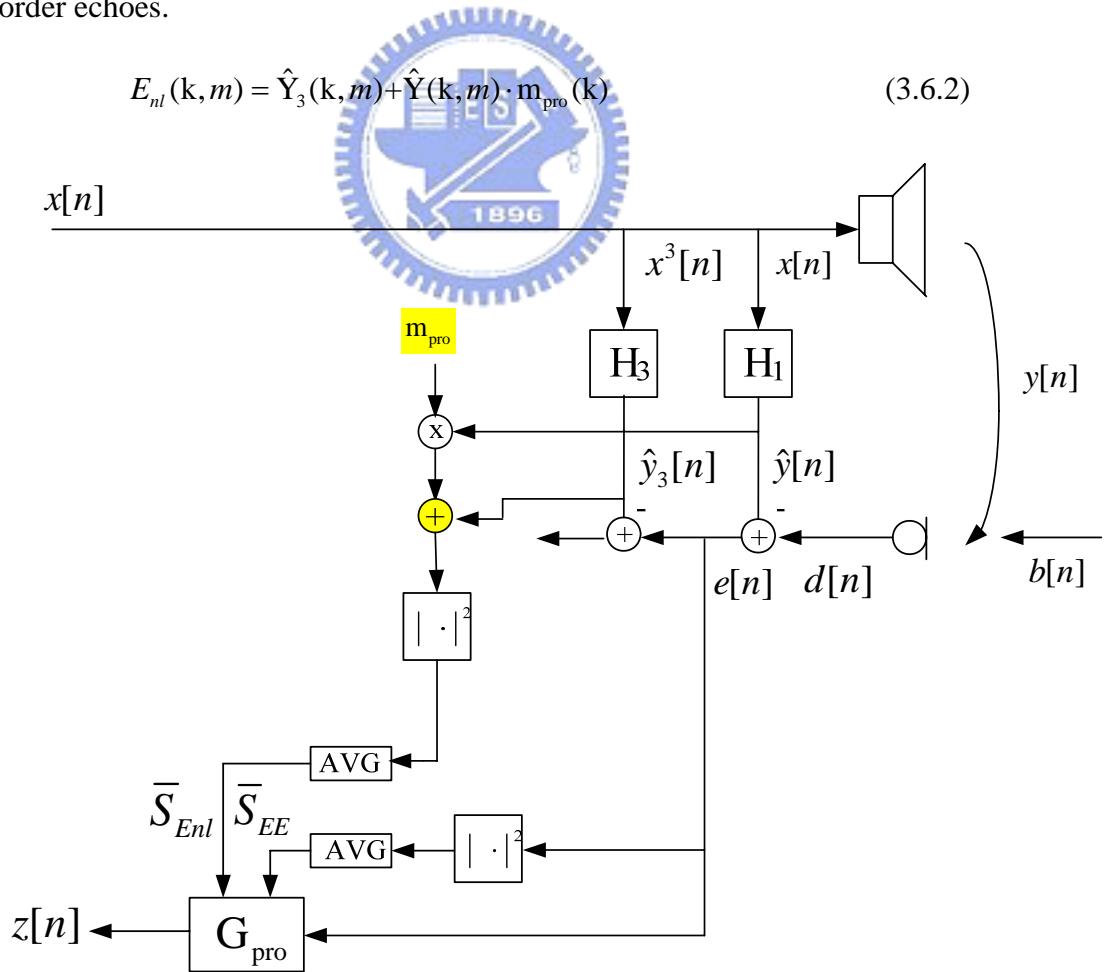


Fig. 3.9 Block diagram of the proposed method

For finding suppression gain, we use $S_{E_{avg}}$ and $S_{E_{nl_{avg}}}$ are the power spectrum density and average from linear residual error $E(k,m)$ and nonlinear residual error $E_{nl}(k,m)$.

$$G_{proposed}(k,m) = \frac{S_{E_{avg}} - S_{E_{nl_{avg}}}}{S_{E_{avg}}} \quad (3.5.7)$$

The suppression error is the power spectrum from instant linear residual error.

$$Z(k,m) = G_{proposed}(k,m) \cdot S_{E_{instant}} \quad (3.5.8)$$

The good combined system wants to cancel or suppress echo large and keep the near-end speech signal.

The linear AEC slightly cancels the far-end speech and keeps the near-end speech signal. The nonlinear AEC hears the noise even louder than the near-end speech. This is because the near-end speech would interfere with the adaptive filter to estimate the nonlinear and linear channels in the double-talk situation. The linear AEC and nonlinear AEC are not using suppression structure, so we don't discuss G_{min} in Eq. (3.3.6). The nonlinear AES can greatly suppress the nonlinear residual error. However, the near-end speech signal is seriously lossy. That is because the suppression gain causes the large disturbance. So, we should set G_{min} for suppression gain. With $G_{min}=0.25$, the near-end speech is little lossy, in spite of the far-end speech is less canceling than $G_{min}=0$.

3.6.1 Basis Selection for proposed slope method

From Section 3.2, we only estimate the linear echo from linear AEC. So, there is only one choice for finding slope basis. The choice is the linear echo. For proposed method, there are several bases to be selected, because we use nonlinear AEC to estimate linear echo and lower order nonlinear echo. Next, we will discuss selecting

basis for proposed method.

A commonly used function for modeling saturation is a sigmoid function $\varphi(u)$ in (3.5.9) [15]. It's popular to model nonlinear effect of loudspeaker. Fig.3.10 is used parameter $\alpha = 3.5, \beta = 1$. We try to fix the sigmoid function by fifth-order polynomial function in Eq. (3.5.10).

$$\varphi(u) = \left(\frac{2}{1 + e^{-\alpha u}} - 1 \right) \beta \quad (3.5.9)$$

$$s(x) = 3.4761x - 3.1740x^3 + 2.3999x^5 - 0.8233x^7 \quad (3.5.10)$$

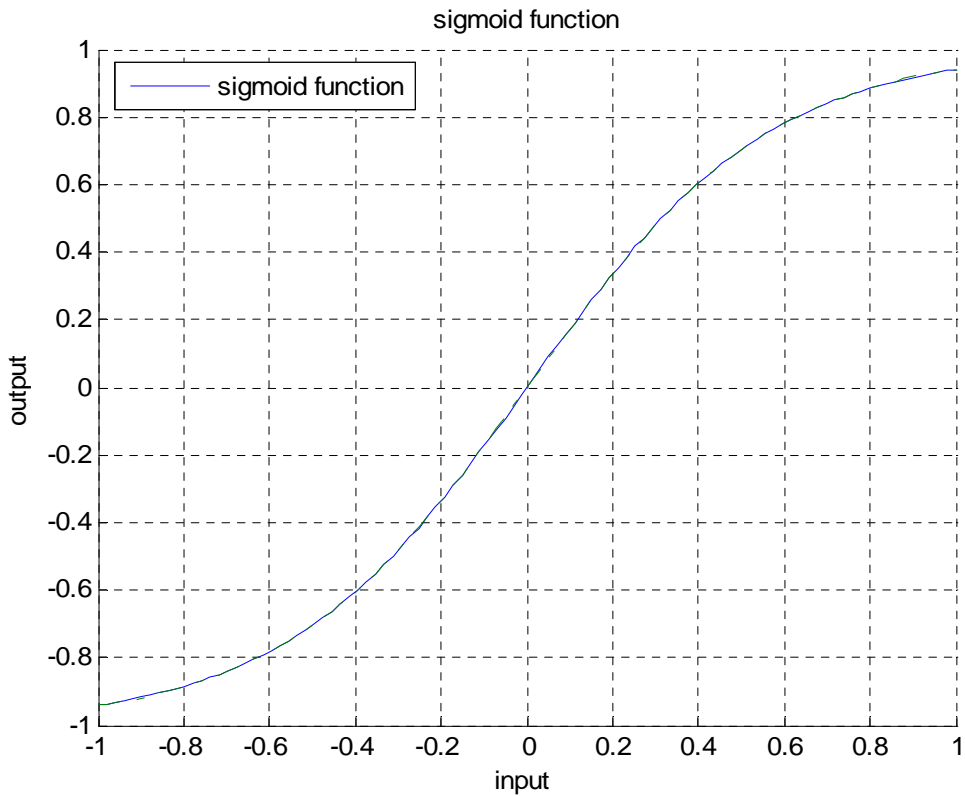


Fig. 3.10 Sigmoid function

The nonlinear AEC includes linear and third order adaptive filter. If the nonlinear AEC is ideal, the residual error has only fifth and seventh order echoes. We will select

first or third echo as basis to estimate summation of fifth and seventh order echoes. However, we can't which order echo is more linear than summation echoes in every frequency bin. For selecting basis, we only observe the statistic distribution and ERLE in chapter 4. From simulation, we can say the third echo is linear than first echo.

In order to accurately estimate the linear property, we define correlation mismatch factor ε which is variance of estimate nonlinear residual error and real nonlinear residual error. If the correlation mismatch factor ε is less, the order nonlinear echo is more linear.

$$\varepsilon = \left| \overline{E}_{nl} \right| - \left| \overline{E}_{nl_est} \right| \quad (3.5.16)$$

$$E_{nl_est} = m \times \hat{Y} \quad (3.5.17)$$

The ε of third basis is smaller than first basis. So, we will select the highest order echo that we can get as basis.



Chapter 4

Computer Simulations

We will show the simulation of chapter 3 to verify the algorithms in this chapter. First, we define some parameters and speech signal in Section 4.1. Second, compare two combined structures that are AEC+NR and NR+AEC in three conditions. The conditions include background noise, nonlinear residual error, and both of two in Section 4.2. Third, we use Volterra structure to estimate second order memory echo for nonlinear AES. Fourth, comparing six methods in ERLE are linear AEC, linear & third order AEC, linear AEC & third order AES, linear AEC & slope1 AES, AEC1 & third order AES & slope1, and AEC1 & third order AES& slope 3 in Section 4.3. Finally, in Section 4.4, we analyze the statistical distributions and correlation mismatch factor ε of three slope methods that are discussing in Section 4.3, and select nonlinear order echo to as basis.

4.1 Parameters and speech signal of simulations

For nonlinear systems, the sigmoid function is commonly used [15], is shown in Fig. 3.10. The sigmoid function is $\varphi(u) = \left(\frac{2}{1 + e^{-\alpha u}} - 1 \right) \beta$. In the following nonlinear model for sigmoid function, the parameters are $\alpha = 3.5$ and $\beta = 1$. We give a real speech signal as far-end signal, and then the output of near-end microphone which we get is defined as desired signal in Fig. 4.1.

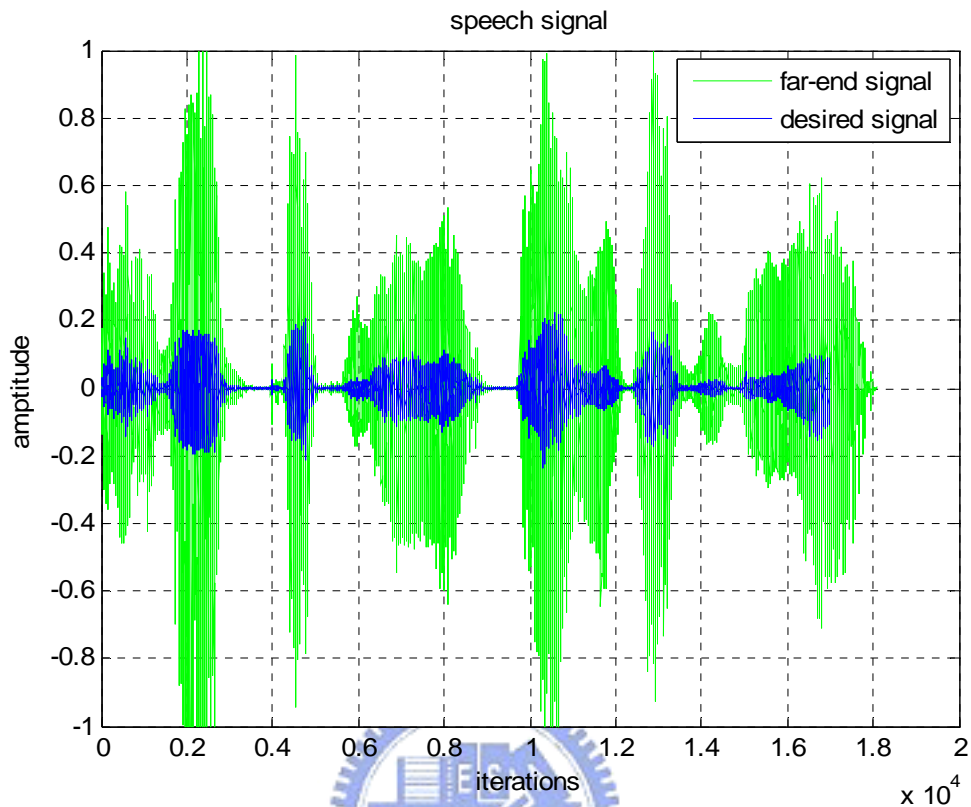


Fig. 4.1 Speech signal

4.2 Compare two combined structures in three conditions

The two combined structures are AEC+NR, and NR+AEC. NR+AEC structure can suppress background noise or nonlinear residual error first that avoid the noise disturbing operation of AEC. However, if we use NR first, there is a distortion for desired signal. Then, the AEC can't estimate the real impulse response any more. We will discuss the two structures in three different noise conditions.

4.2.1 Background noise

The input signal is white gauss signal. The loudspeaker is perfect and room impulse response is exponential decay with 128 taps in Fig. 4.2. The SNR is 20dB and -5dB. The adaptive filter is linear and the length is 128.

As evaluation criterion we use the echo return loss enhancement (ERLE) defined as

$$\text{ERLE} = 10 \log \frac{E\{d^2(k)\}}{E\{z^2(k)\}} [dB] \quad (4.2.1)$$

If the output signal after the cancellation or suppression structures is less correlated than input signal, the performance of the structures is better. From Eq. (4.2.1), ERLE is large for good cancellation or suppression structures.

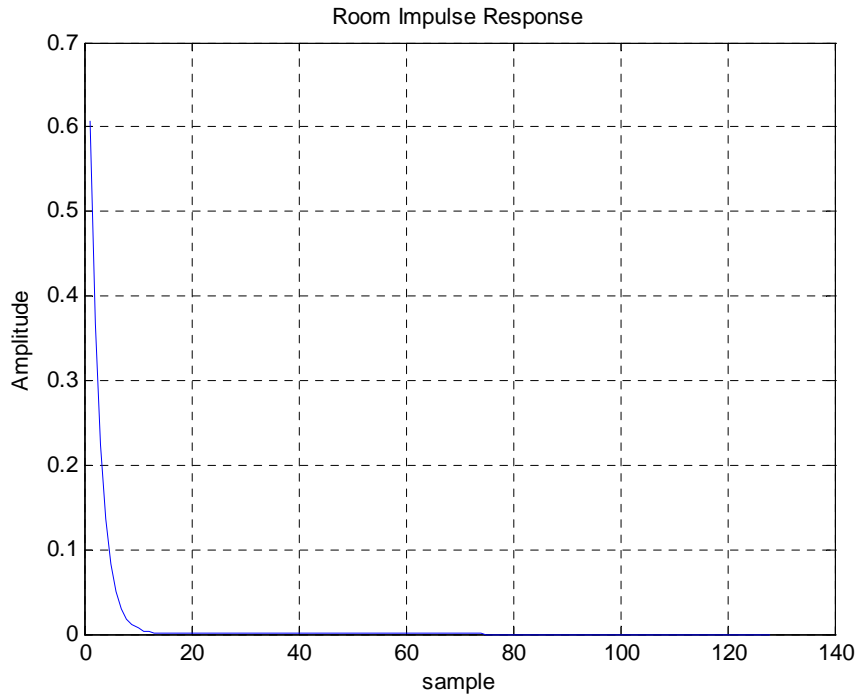


Fig. 4.2 Pseudo room impulse response

The Fig. 4.3 and Fig. 4.4 are shown the simulation in SNR = 20dB, and SNR = -5dB. In Fig. 4.3, AEC+NR structure is better than NR+AEC structure by 3.5dB in SNR = 20dB. However, in Fig. 4.4 AEC+NR structure is worse than NR+AEC structure by 1.5dB in SNR = -5dB. This is because using NR first will cause nonlinear distortion obviously to disturb AEC in high SNR and using AEC first that the background noise will disturb AEC in low SNR. The background noise is large (SNR = -5dB) in Fig. 4.4, so the performance is not good in both structures.

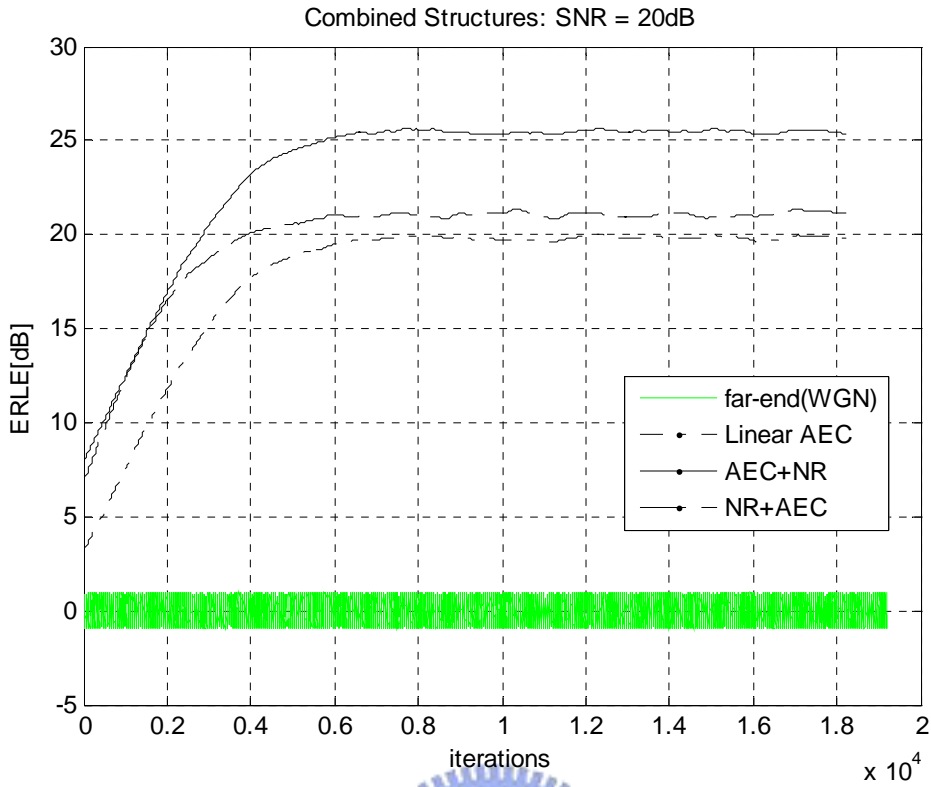


Fig. 4.3 Combined structures for suppressing background noise (SNR=20dB)

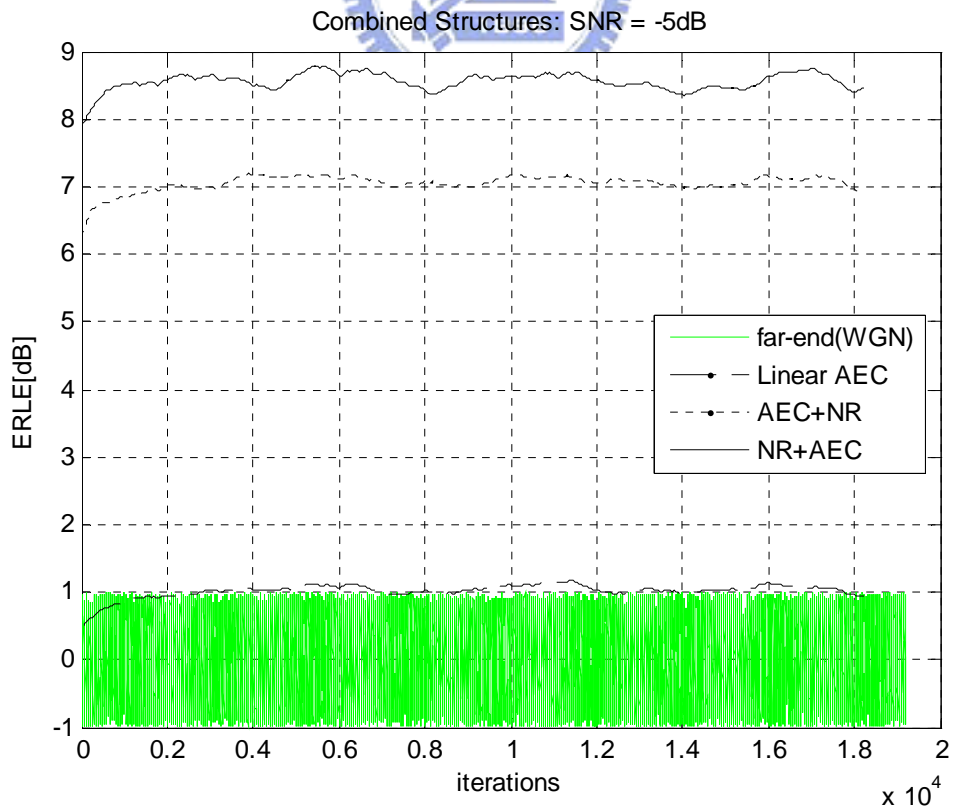


Fig. 4.4 Combined structures for suppressing background noise (SNR=-5dB)

4.2.2 Nonlinear residual error

We use two combined structures to suppress nonlinear residual error. The environment is same as Section 4.2.1 except for loudspeaker that has nonlinear effect and SNR = 50dB. The nonlinear effect uses the sigmoid function in Fig. 3.10 and Eq. (3.5.9). In order to estimate nonlinear residual error for NR structure, we use “Based on Highly Nonlinear Residual Echo” in Section 3.3.1. In Fig.4.5, AEC+NR structure is better than NR+AEC structure by 3dB.

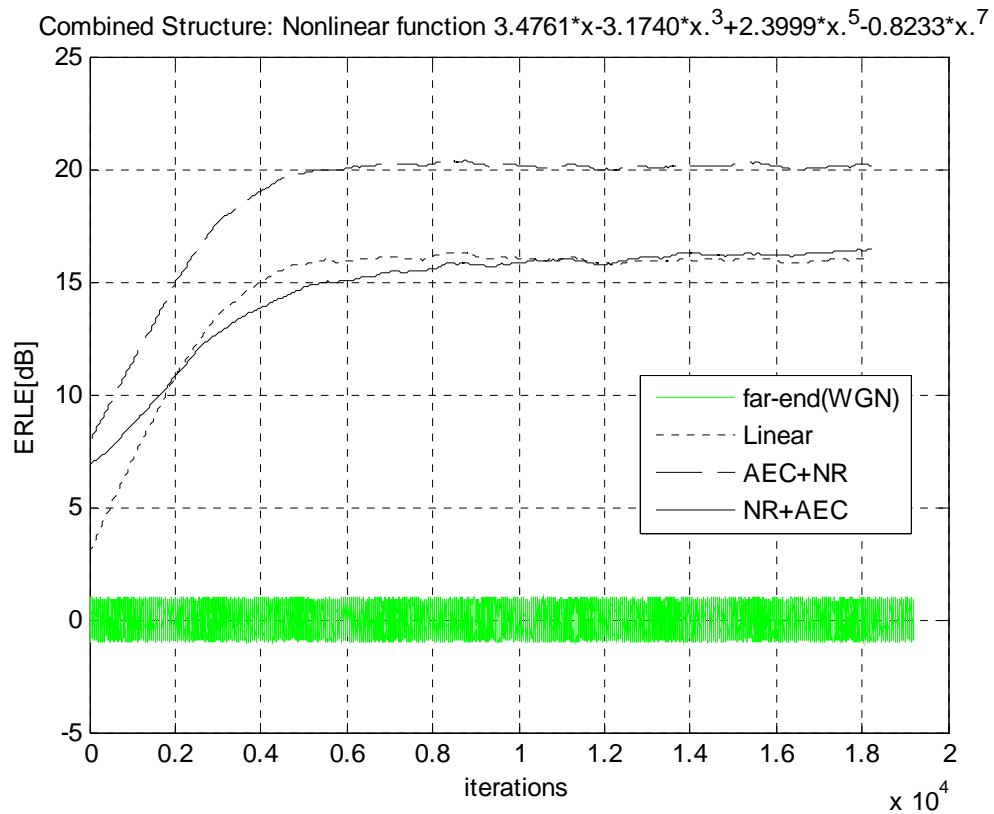


Fig. 4.5 Combined structures for suppressing nonlinear residual error
(SNR=20dB)

4.2.3 Background noise and nonlinear residual error

In Section 3.4, we introduce two combined structures AEC+NR and NR+AEC to suppress background noise and nonlinear residual error in simultaneous. The simulation result is shown in Fig.4.4. The AEC+NR structure is better than NR+AEC structure by 3dB.

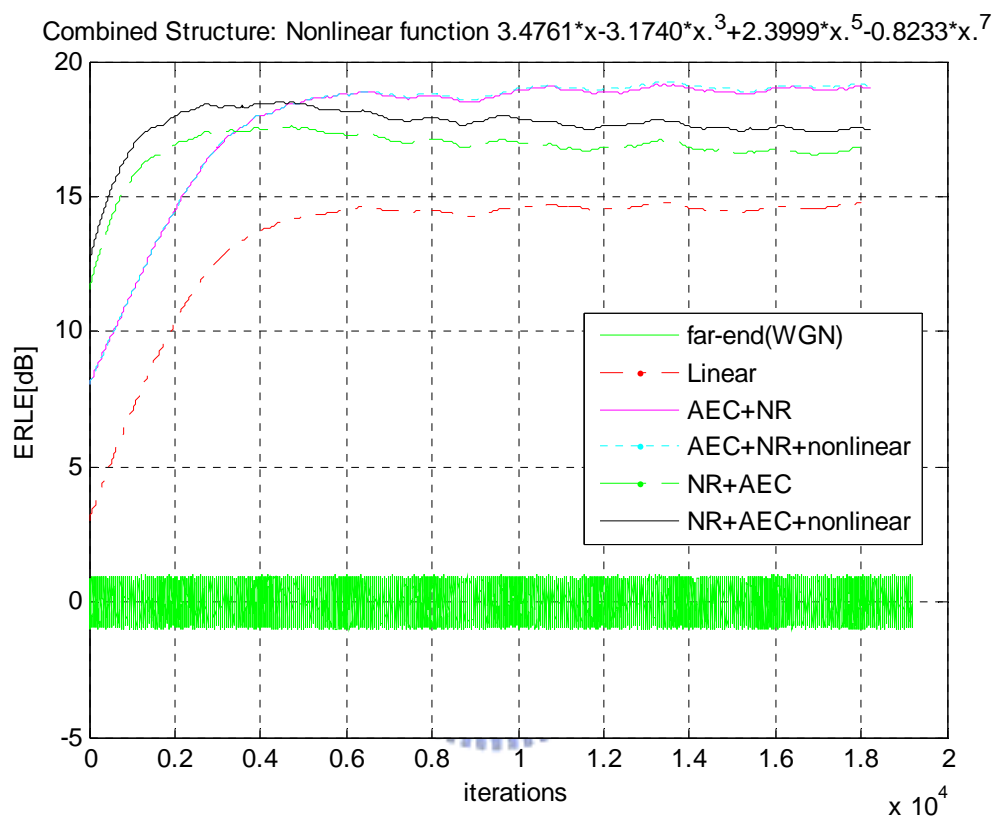


Fig. 4.6 Combined structures for suppressing nonlinear residual error and background noise

4.3 Performance of Volterra Structure for NAES

In Section 3.6, we introduce use the Volterra structure in Section 2.2.3 can cancel the nonlinear memory effect. The simulation result is shown in Fig.4.5 and Fig.4.6.

The linear filter \hat{H}_1 length is $N=64$. The second order Volterra filter \hat{V}_2 of length $V = 4$. For input signal of WGN, the Volterra structure is better than NAEC. However, for input signal of speech, the Volterra structure is worse than NAEC. This is because

the convergence speed for Volterra structure is slow. The linear filter H length is $N=128$, and the nonlinear filter is memory polynomial shown in (4.5.1). Volterra structure for NAES is not better than power filter, because the Volterra structure needs to adaptive too many coefficients.

$$S = x[n] + 0.05x^3[n] + 0.1\{x[n-1]x[n-2] + x[n-1]x[n-3] + x[n]x[n-4] + x[n-2]x[n-3] + x[n-2]x[n-4] + x[n-3]x[n-4]\} \quad (4.5.1)$$

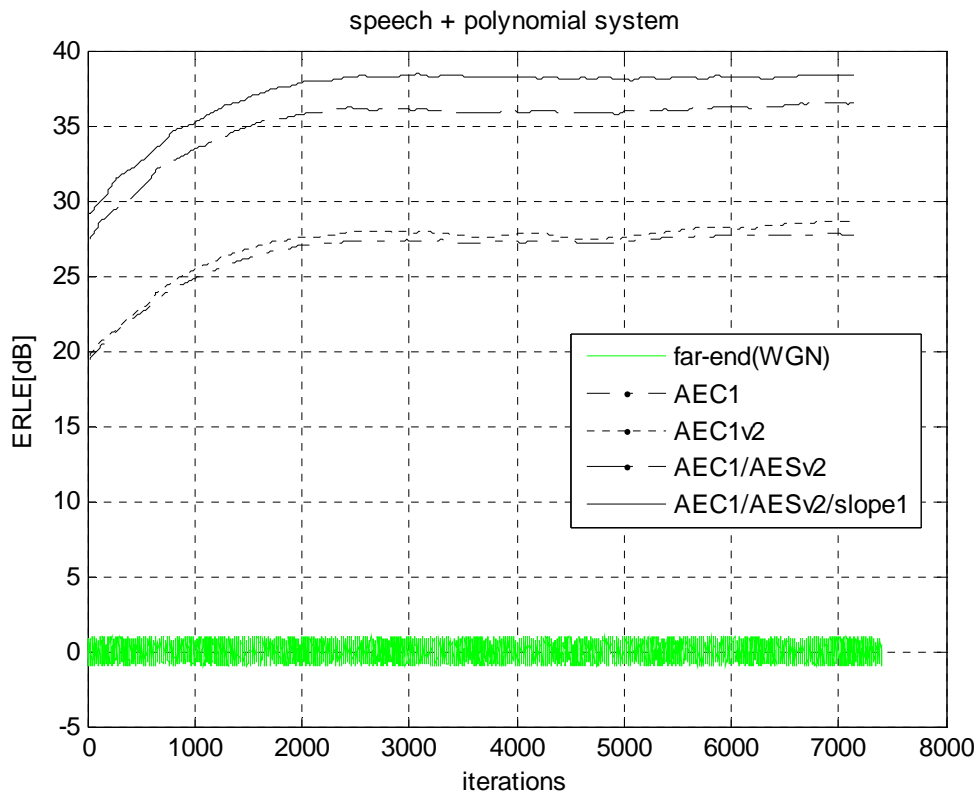


Fig. 4.7 The Volterra structure for NAES using WGN

4.4 Simulation of Highly Nonlinear Residual Errors

4.4.1 Single talk

The linear AEC is only used to cancel the linear echo only. The nonlinear residual echo of more than two orders is not cancelled at all. The Nonlinear AEC method can cancel the high orders' residual echo. The performance of NAEC is dependent on the adaptive filter number you used. We implement the third order NAEC. We use two concepts of "slope" and "power filter" in Section 3.3.1 and Section 3.3.2 and use the proposed method in Section 3.5 to run the simulation. We arrange the six algorithms' notation in Table 4.1.

Table 4.1 Notation of six algorithms

	notation	comment	Reference
1	AEC1	AEC	[1]
2	AEC13	NAEC	[3]
3	AEC1/AES3	AEC/NAES	[9]
4	AEC1/slope1 \rightarrow 3+	AEC/ slope(linear echo)	[10]
5	AEC1/AES3/slope1 \rightarrow 5+	AEC/ NAES/ slope(linear echo)	Proposed
6	AEC1/AES3/slope3 \rightarrow 5+	AEC/ NAES/ slope(third echo)	Proposed

AEC1 uses linear AEC to cancel the linear echo that can't cancel the nonlinear residual error anymore. AEC13 uses linear AEC and third-order nonlinear AEC to cancel linear echo and third echo that still exists the nonlinear residual error more than fifth-order nonlinear residual error. The performance of AEC13 is better than AEC1. AEC1/AES3 uses linear AEC to cancel the linear echo and suppress third-order

nonlinear residual error by “power filter” method in Section 3.2.2. The performance of AEC1/AES3 is better than AEC1/AEC3. That is because the convergence speed of AEC1/AES3 is quicker than AEC1/AEC3. AEC1/slope1→3+ uses linear AEC to cancel linear echo and suppress all order nonlinear error by “slope” method of linear echo basis in Section 3.2.1. If the total high order nonlinear residual error is larger than the third-order nonlinear echo, the performance of AEC1/slope1→3+ is better than AEC1/AEC3. AEC1/AES3/slope1→5+ and AEC1/AES3/slope3→5+ use linear AEC to cancel the linear echo, third-order suppression to suppress third echo, and suppress more than fifth-order nonlinear residual error by “slope” method of linear echo or third-order echo basis. AEC1/AES3/slope1→5+ and AEC1/AES3/slope3→5+ are better than AEC1/AES3 and AEC1/slope1→3+, because AEC1/AES3/slope1→5+ and AEC1/AES3/slope3→5+ can suppress third-order echo accurately and suppress more than fifth-order nonlinear residual error. For real speech signal to loudspeaker, the high order nonlinear residual error is more linearly with the third-order echo.

In Fig. 4.8, we use the real speech signal and real system. The nonlinear effect is of the total more than fifth-order nonlinear residual error is larger than third-order echo. So, AEC1/slope1 is better than AEC1/AES3. AEC13/AES3/slope1 can suppress third-order echo accurately than AEC1/slope1, so the performance is better. The nonlinear residual error that is more than fifth-order echo is more linearly with the third-order echo than first-order echo. So, AEC13/AES3/slope3 is better than AEC13/AES3/Slope1.

In Fig. 4.9, we use speech signal as far-end signal and nonlinear system is polynomial function $s(x) = 3.4761x - 3.1740x^3 + 2.3999x^5 - 0.8233x^7$. The

proposed method is the best than the others.

In Fig. 4.10, we use white gauss noise as far-end signal and nonlinear system is polynomial function $s(x) = 3.4761x - 3.1740x^3 + 2.3999x^5 - 0.8233x^7$. The third-order echo is large. If we can't estimate third-order echo accurately, the ERLE will bad. So, AEC1/AES3 is better than AEC1/slope1. The proposed method is the best than the others.

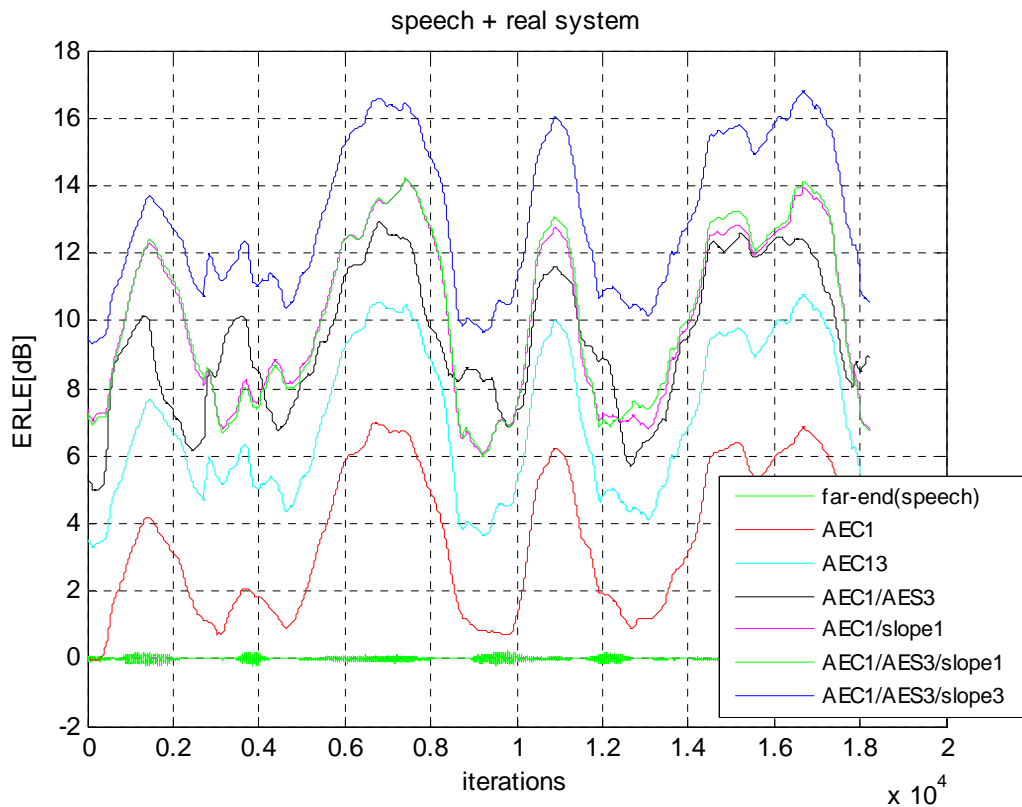


Fig. 4.8 ERLE for speech signal & real system

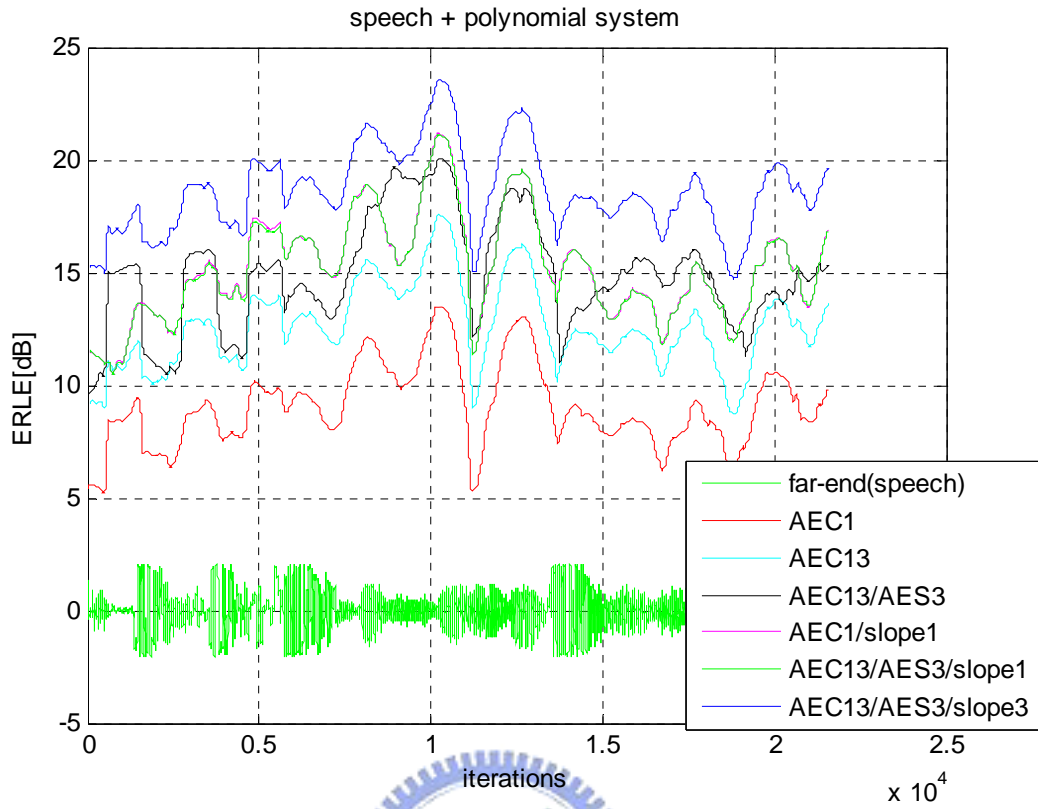


Fig. 4.9 ERLE for speech signal & polynomial system

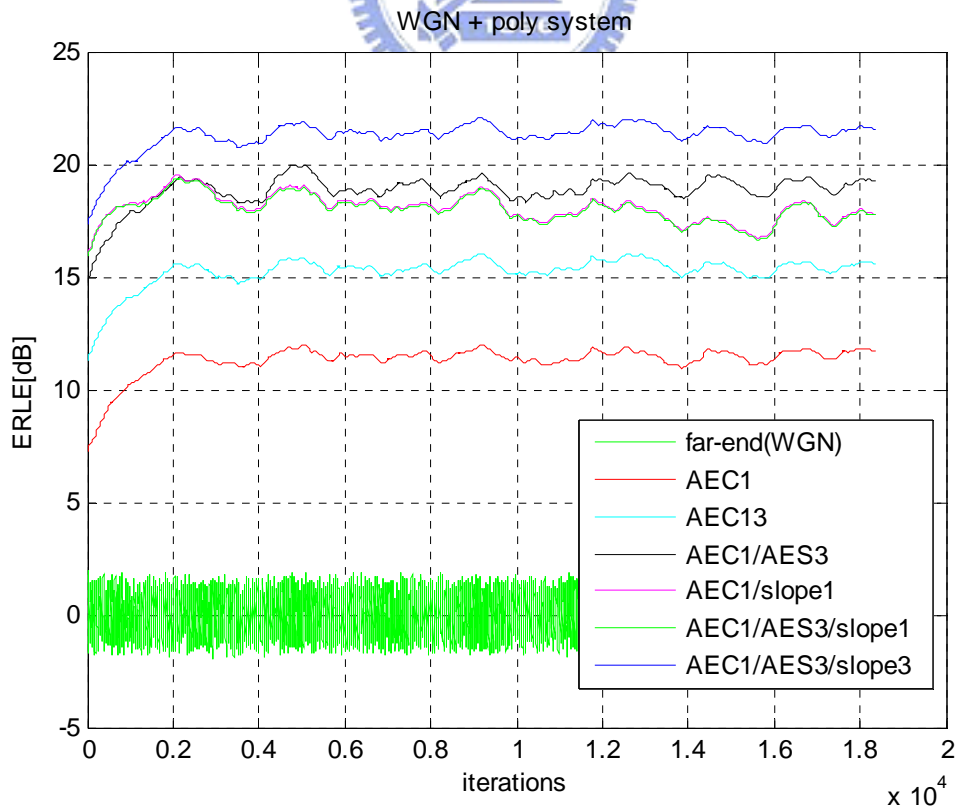


Fig. 4.10 ERLE for WGN & polynomial system

4.4.2 Double Talk

In order to avoid severe degradation of near-end speech signals during double-talk situations, the actual values of $G(k,m)$ are usually limited to a desired level of attenuation G_{\min} . We set $G_{\min} = 0.25dB$, implying a maximum attenuation of about 12dB. The simulation result is shown in Table 4.3. AEC1 can cancel the linear echo in spite of the near-end speech signal disturbs AEC. We hear a slightly low level of echo from the near-end speech signal. Using AEC13, we hear the loud noise that is almost covered the near-end speech signal. This is because the near-end speech is critically disturbing the nonlinear AEC. For suppression structures, we only discuss $G_{\min} = 0.25$, because $G_{\min} = 0$ is critically distorting the near-end speech. AEC1/slope1 can suppress the echo better than AEC1, but the near-end speech signal is poorly disturbed than AEC1. AEC13/AES3/slope1 and AEC13/AES3/slope3 can suppress the nonlinear residual error better than AEC1/slope1.

Table 4.2 Simulation of double-talk

	Gmin = 0	Gmin = 0.25 (-12dB)
AEC1	Echo: slightly cancelled	
	DT: not distortion	
AEC13	Echo: mismatch noise	
	DT: not distortion	
AEC1/slope1	Echo: almost cancelled	Echo: little cancelled
	DT: lossy	DT: little lossy
AEC1/AES3 /slope1	Echo: almost cancelled	Echo: moderately cancelled
	DT: almost silent	DT: slightly lossy
AEC1/AES3 /slope3	Echo: almost cancelled	Echo: moderately cancelled
	DT: almost silent	DT: slightly lossy

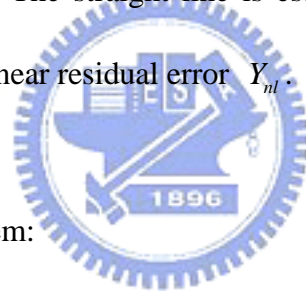
In double-talk situation, we want to cancel or suppress the echo large and keep the near-end speech signal. From Table 4.2, the near-end speech of AEC1/AES3/slope1 and

AEC1/AES3/slope3 for $G_{min} = 0.25$ is better than the others except AEC1 and AEC13. However, AEC1 and AEC13 can slightly cancel the echo only or cause the adaptive filter mismatch. The AEC13/AES3/slope1 or AEC13/AES3/slope3 is the better choice than the others.

4.5 Statistics Distribution of Higher-Order Nonlinear Residual Errors

In Section 4.4.1, we discuss the ERLE for six methods and the “slope” method in Section 3.3.1 has been used in AEC1/slope1, AEC1/AES3/slope1, and AEC1/AES3/slope2. In order to analyze the three methods, we will observe statistical distributions. If the residual error is linearly than which order echo, the residual error is closer than estimate value. The straight line is estimated nonlinear residual error

\hat{Y}_{nl} and the point is real nonlinear residual error Y_{nl} .



(1) speech signal + real system:

First, we input speech signal and real system that is real loudspeaker and real room impulse response.

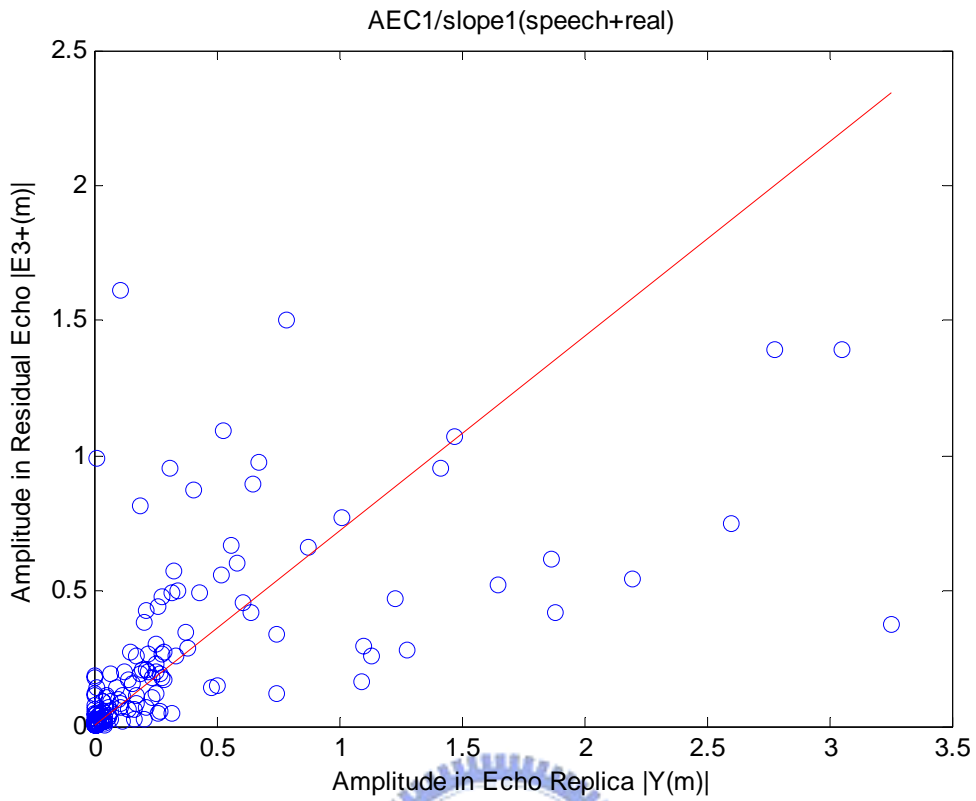


Fig. 4.11 Statistic distribution of AEC1/slope1 in speech signal + real system

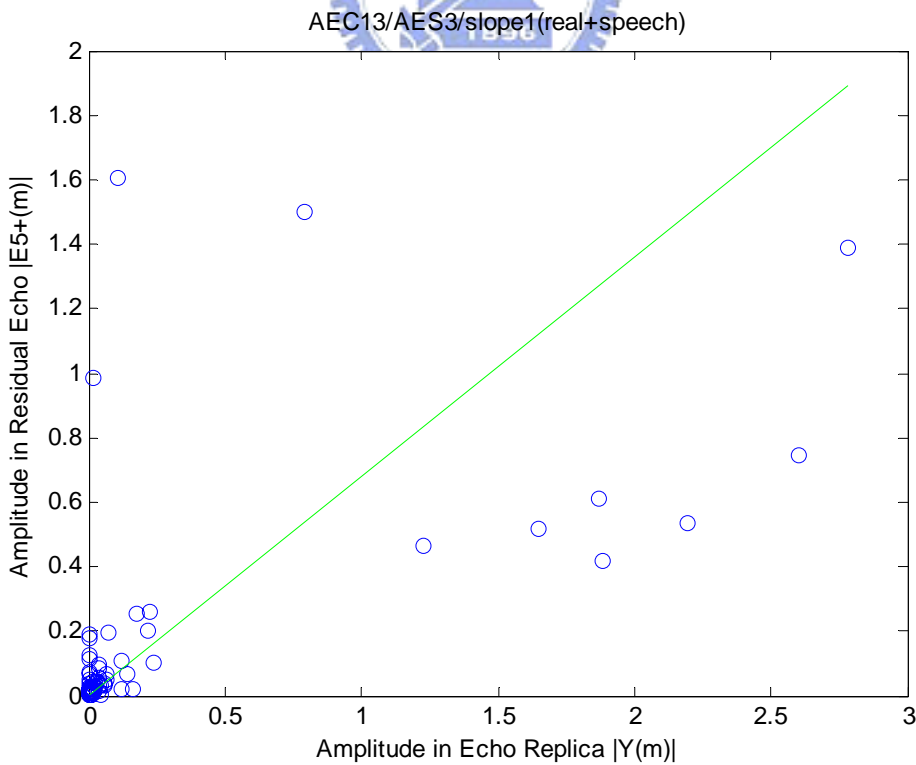


Fig. 4.12 Statistic distribution of AEC13/AES3/slope1 in speech signal + real system

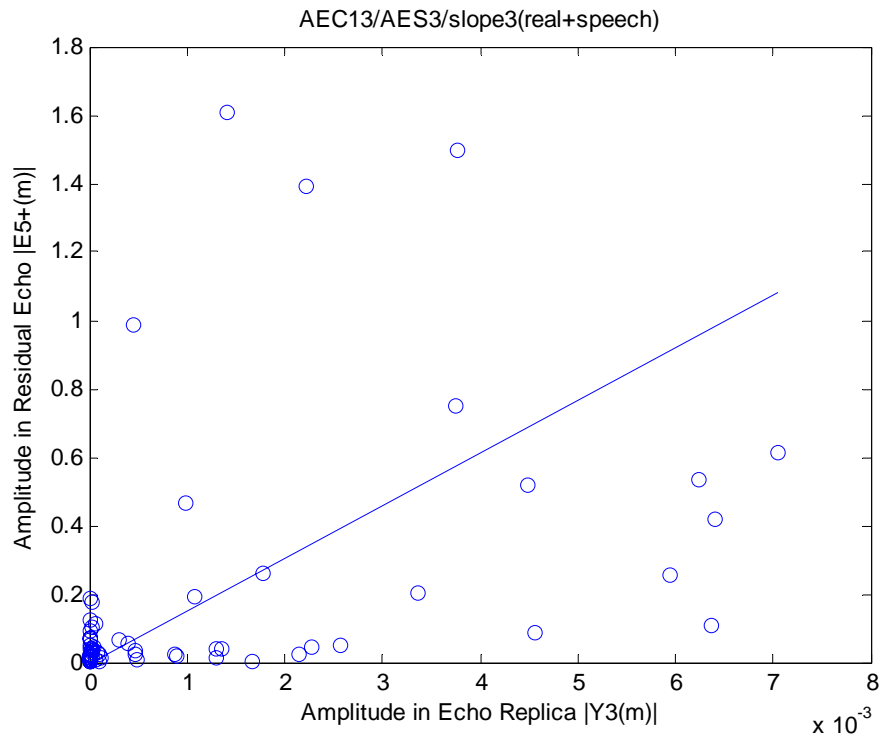
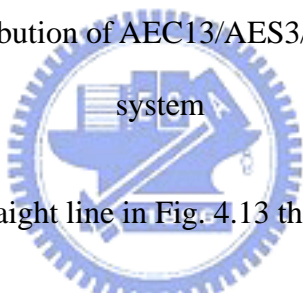


Fig. 4.13 Statistic distribution of AEC13/AES3/slope3 in speech signal + real



The point is closer than straight line in Fig. 4.13 than Fig. 4.11 and Fig. 4.12.

(2) WGN signal + polynomial system:

Second, we input white gauss noise signal and the pseudo system includes two parts that are polynomial function in Fig. 3.10 and exponential function in Fig. 4.2.

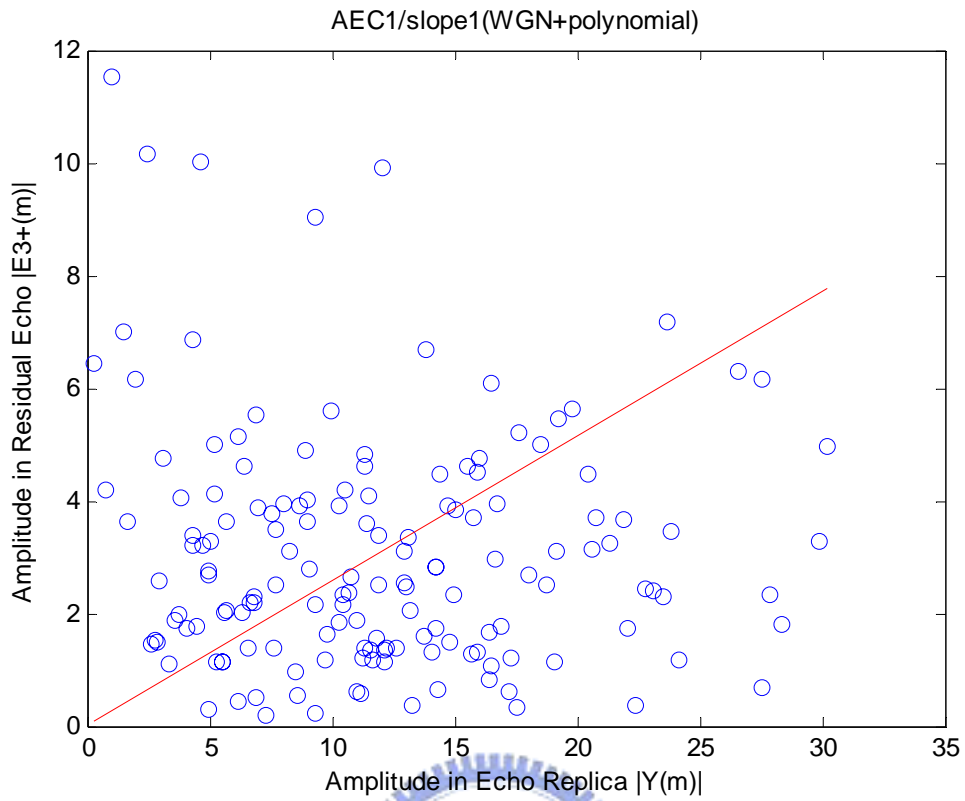


Fig. 4.14 Statistic distribution of AEC1/slope1 in WGN signal + polynomial system

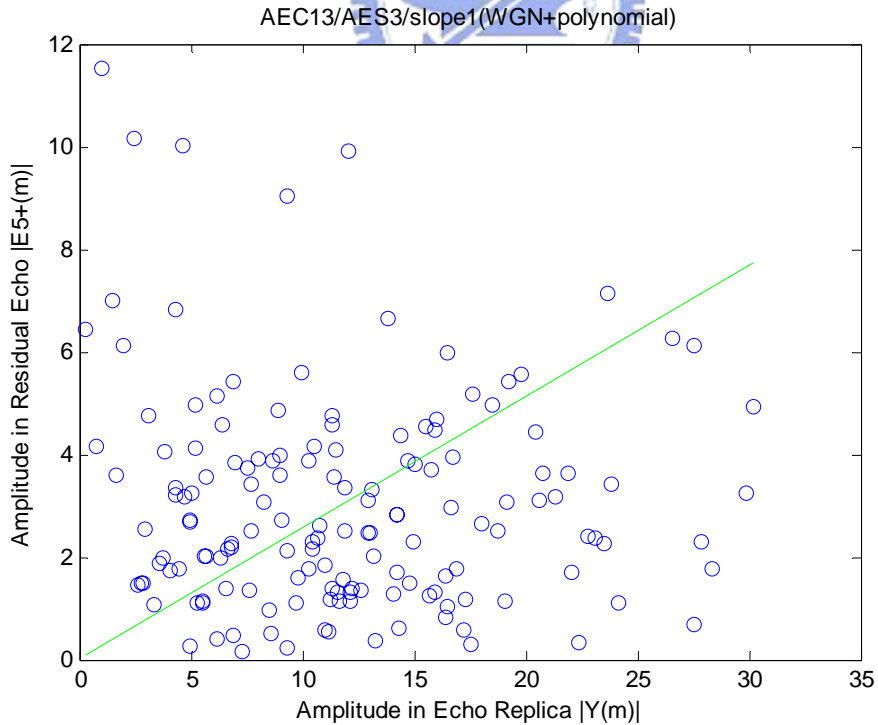


Fig. 4.15 Statistic distribution of AEC13/AES3/slope1 in WGN signal + polynomial system

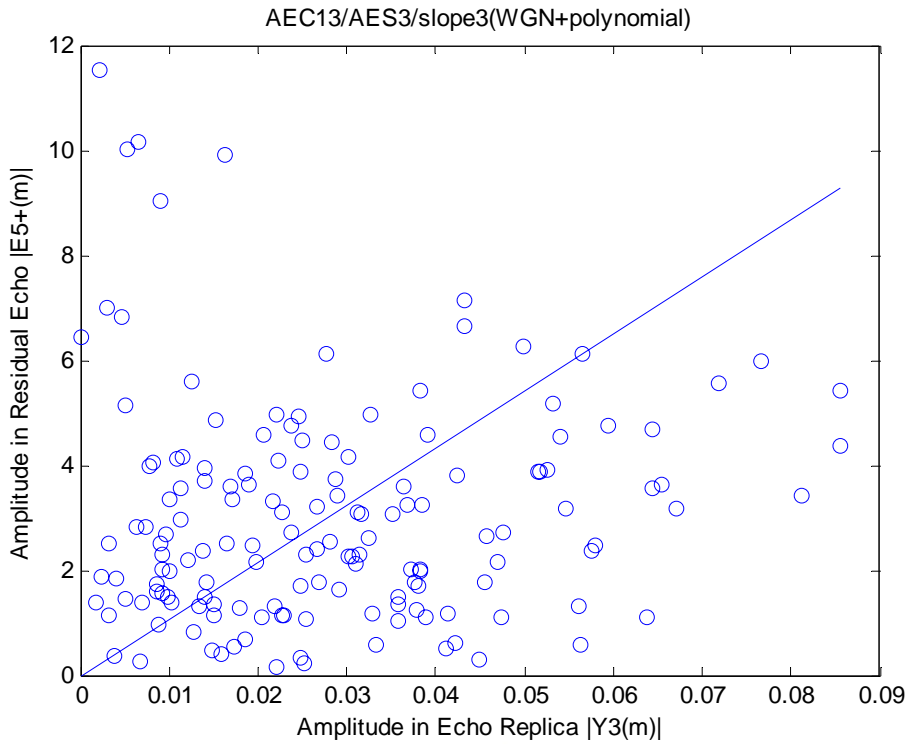


Fig. 4.16 Statistic distribution of AEC13/AES3/slope3 in WGN signal + polynomial

The point is closer than straight line in Fig. 4.16 than Fig. 4.14 and Fig. 4.15.

(3) speech signal + polynomial system:

Third, we input speech signal and the pseudo system includes two parts that are polynomial function in Fig. 3.10 and exponential function in Fig. 4.2.

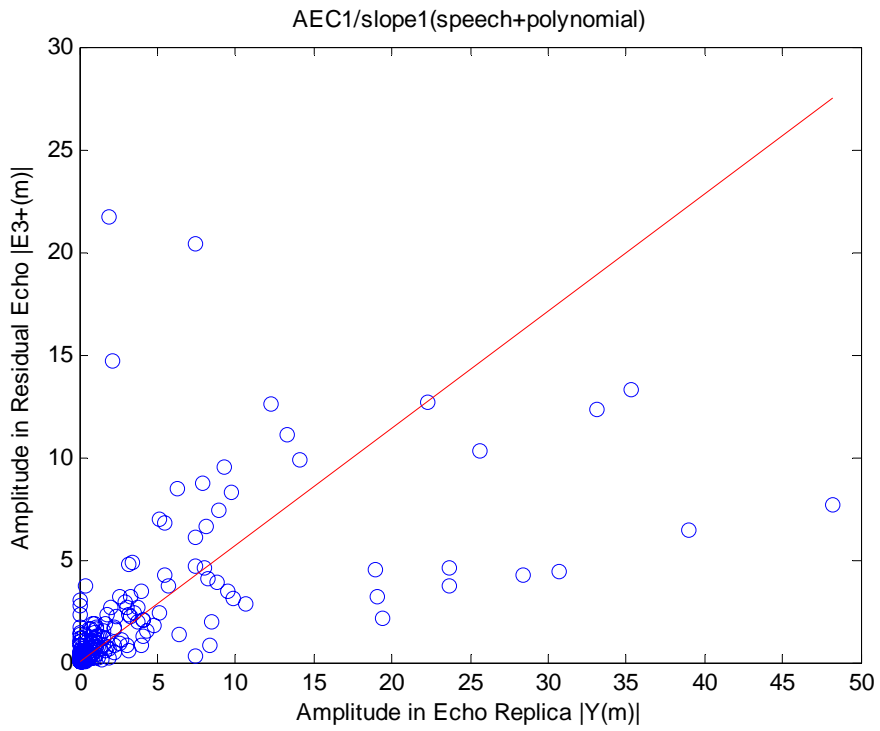


Fig. 4.17 Statistic distribution of AEC1/slope1 in speech signal + polynomial

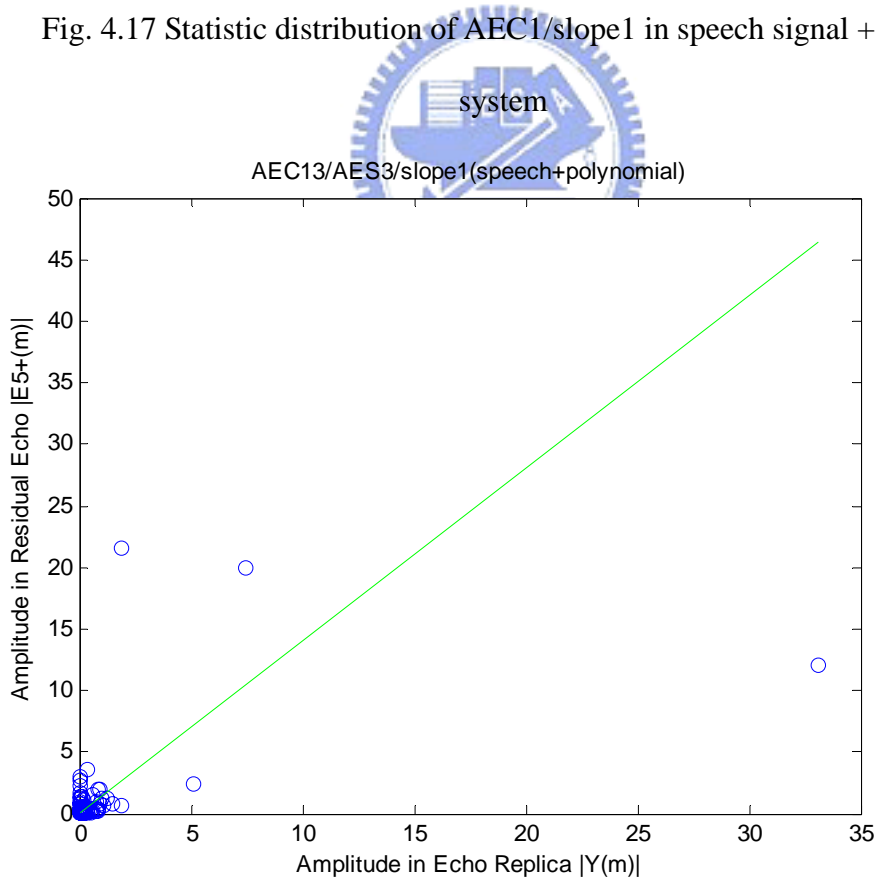


Fig. 4.18 Statistic distribution of AEC13/AES3/slope1 in speech signal + polynomial system

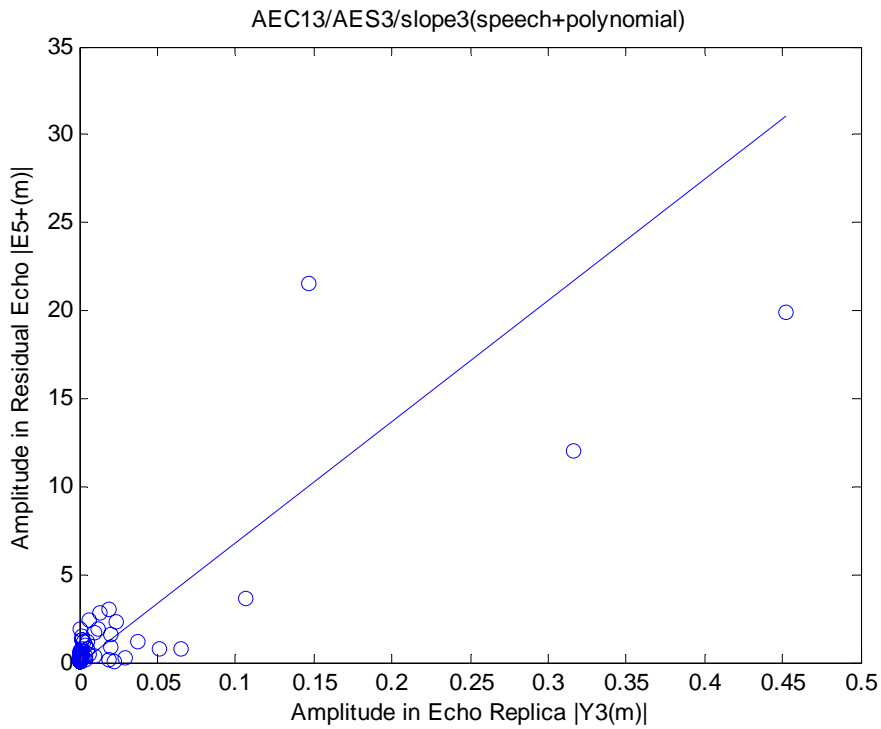


Fig. 4.19 Statistic distribution of AEC13/AES3/slope3 in speech signal + polynomial system

The point is closer than straight line in Fig. 4.19 than Fig. 4.17 and Fig. 4.18.

(4) WGN signal + real system:

Fourth, we input white gauss noise signal and real system that is real loudspeaker and real room impulse response.

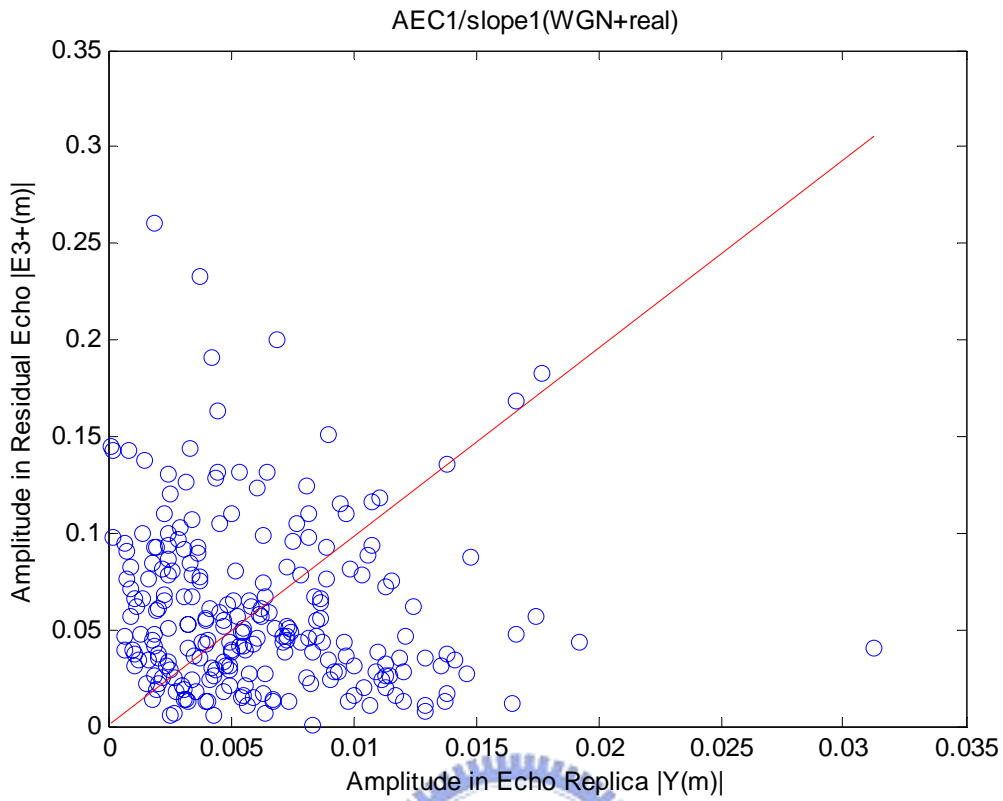


Fig. 4.20 Statistic distribution of AEC1/slope1 in WGN signal + real system

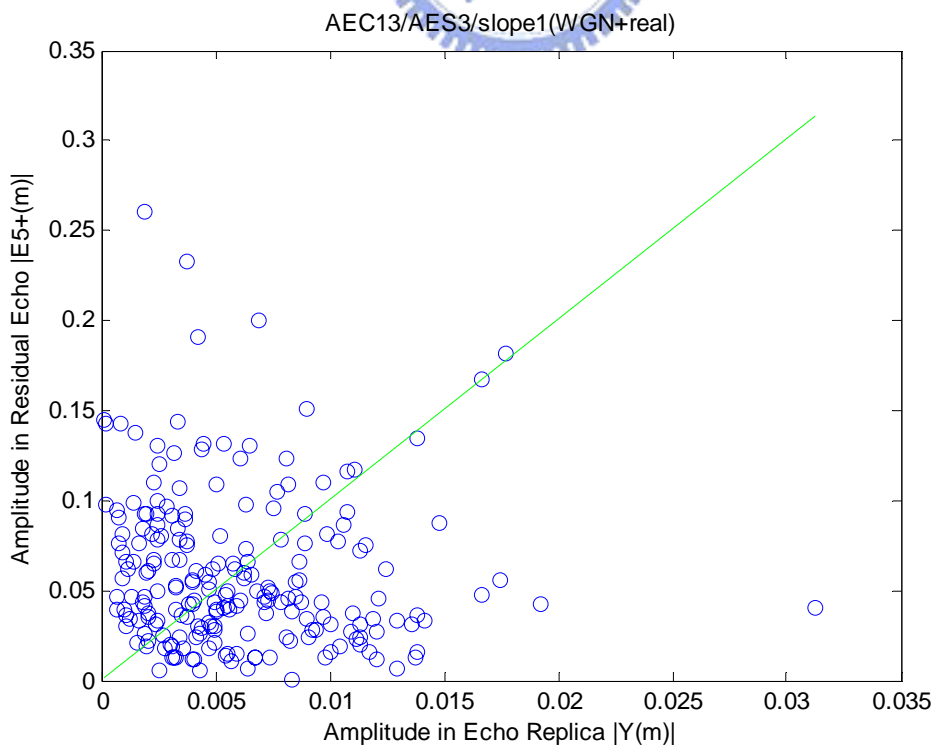


Fig. 4.21 Statistic distribution of AEC13/AES3/slope1 in WGN signal + real

system

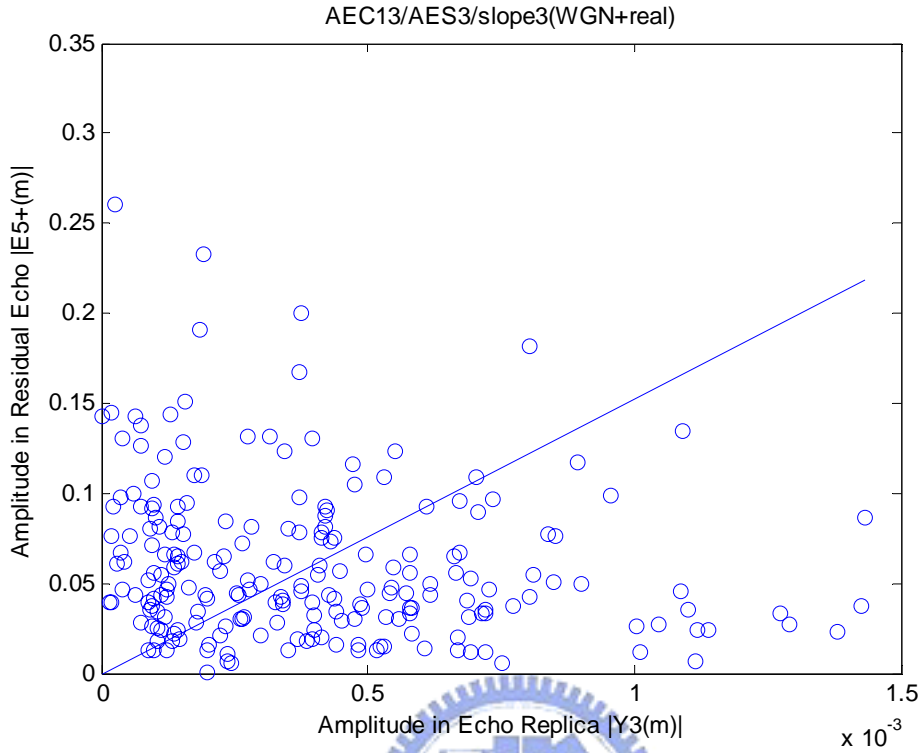


Fig. 4.22 Statistic distribution of AEC13/AES3/slope1 in WGN signal + real

system

The point is closer than straight line in Fig. 4.22 than Fig. 4.20 and Fig. 4.21.

From the above statistical distribution, we can find roughly the AEC13/AES3/slope3 is more linear than the AEC13/AES3/slope1 and AEC1/slope1. In order to accurately find which one can suppress nonlinear residual error more, we define correlation mismatch factor ε which is variance of estimate nonlinear residual error and real nonlinear residual error. If the correlation mismatch factor ε is less, the method is more linear.

$$\varepsilon = \bar{Enl}_{real} - \bar{Enl}_{estimate} \quad (4.5.1)$$

$$\bar{Enl}_{estimate} = \left(m \cdot \hat{Y}_n \right)_{avg} \quad (4.5.2)$$

Table 4.3 Correlation mismatch factor ε

	AEC1/slope1	AEC1/AES3/slope1	AEC1/AES3/slope3
Speech & Real	5.4690e-004	5.0143e-004	5.0079e-004
WGN & Polynomial	0.0121	0.0123	0.0058
Speech & polynomial	3.8030e-004	3.1465e-004	3.1465e-004
WGN & Real	3.4303e-005	3.3021e-005	2.3639e-005

The correlation mismatch factor ε of AEC1/AES/slope3 is less than AEC1/slope1 and AEC/AES/slope1 and AEC1/AES/slope3 can suppress more nonlinear residual error than the others. Table 4.3 can prove the ERLE of AEC1/AES/slope3 is the best than the others in Section 4.4.1.



Chapter 5

Conclusion

In this thesis, we introduce different structures for estimating the nonlinear residual error and near-end speech signal. However, the method using the linear echo to estimate the nonlinear residual error in Section 3.3.1 is not accurate. The method using the NAEC to estimate the nonlinear residual error needs large computational operation in Section 3.3.2. We integrate the two methods to propose the new method that can estimate the nonlinear residual error more accurate and save the calculating operation. From the simulation results in Fig. 4.8-11, we know the new method is better than the others methods. Besides, the statistics distribution and correlation mismatch factor can prove the proposed method better than the others. We provide the input signal of WGN or speech signal and the system of polynomial or real loudspeaker for the simulation, and the new method are better than the others methods.

In addition, we introduce the other suppression structure which is NR+AEC structure. In low SNR condition, the NR+AEC structure is better than AEC+NR structure. The first NR suppresses the large noise, so the noise has less affect on AEC. The AEC can estimate the echo path accurately. In high SNR condition, the NR+AEC structure is worse than AEC+NR structure. This is because the NR+AEC structure causes the nonlinear disturbance by the NR. So, AEC can't estimate the accurate echo path. We also extend the NAES to other nonlinear adaptive filter which is Volterra model. Volterra model can estimate nonlinear memory residual error that the loudspeaker is time-variant. So, the performance is better than Hammerstein model that can estimate memoryless residual error only.

The simulation result shows the Volterra model is better than Hammerstein model for steady input signal, but the input signal is real speech model for communication that the performance is worse than Hammerstein model. This is because the convergence speed of Volterra structure is too slow. When the length of input signal period isn't enough to adaptive the nonlinear channel, the performance will come down.

Using higher-order nonlinear echoes as bases to estimate nonlinear residua error is not only can estimate lower-order nonlinear echoes accurately, but also estimate high-order nonlinear echoes roughly.



Bibliography

- [1] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 1996.
- [2] E. Hansler, "The hands-free telephone problem; an annotated bibliography," *Signal Processing*, vol. 27 pp. 259-271, 1992.
- [3] S. M. Kuo, Y. C. Huang, and Z. Pan, "Acoustic noise and echo cancellation microphone system for videoconferencing," *IEEE Trans. Consumer Electronics* vol. 41, no. 4, pp. 1150-1158, Nov. 1995.
- [4] W. Hsu, F. Chui, and D. A. Hodges, "An acoustic echo canceler," *IEEE Journal Solid State Circuits* vol. 24, no. 6, pp. 1639-1646, Dec. 1989.
- [5] S. Gudvangen, and S.J. Flockton, "Modelling of acoustic transfer functions for echo cancellers," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 42 no. 1 pp.47-51, Feb 1995.
- [6] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Automatic Control*, vol. 11, issue 13, pp. 546-550. Jul. 1966.
- [7] N. J. Bershad, P. Celka, and S. McLaughlin, "Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions," *IEEE Trans. Signal Process*, vol. 49, pp. 1060-1071, May 2001.
- [8] V. J. Mathews, "Adaptive Volterra filters using orthogonal structures," *IEEE Signal Processing Letter*, vol. 3, issue 12, pp. 307-309, Dec. 1996.

- [9] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," ICASSP proceeding, vol. 2, pp. 877–880, Nov. 1999.
- [10] A. Guerin, G. Faucon, and Le Bouquin-Jeannes, R., "Nonlinear acoustic echo cancellation based on Volterra filters," IEEE Trans. Speech and Audio Processing vol. 11, no. 6, pp. 672-683, Nov. 2003.
- [11] O. Hoshuyama and A. Sugiyama, "An Acoustic Echo Suppressor Based On A Frequency-Domain Model Of Highly Nonlinear Residual Echo," IEEE Proc. ICASSP'06, pp. V-269-272, 2006.
- [12] F. Kuech and W. Kellermann, "Nonlinear Residual Echo Suppression Using a Power Filter Model of the Acoustic Echo Path," IEEE Proc. ICASSP'07, pp.I-73-76, 2007.
- [13] R. Raich, Hua Qian, and G.T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," IEEE Trans. Vehicular Technology vol. 53, issue5, pp. 1468-1479, 2004
- [14] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear Acoustic Echo Cancellation Using Adaptive Orthogonalized Power Filters," IEEE Proc. ICASSP'05, pp. III-105-108, 2005

- [15] J. P. Costa, A. Lagrange, and A. Arliaud, “Acoustic echo cancellation using nonlinear cascade filters,” ICASSP, pp.v389-v392, 2003.
- [16] S. Boyd, L. Chua, and C. Desoer, “Analytical foundations of Volterra series,” IMA J. Inform. Contr., vol. 1, pp. 243-282, 1984.
- [17] V. J. Mathews, “Adaptive Volterra filters using orthogonal structures,” IEEE Signal Processing Letter, vol. 3, issue 12, 99. 307-309, Dec. 1996.
- [18] A. Stenger, L. Trautmann, and R. Rabenstein, “Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters,” ICASSP proceeding, vol. 2, pp. 877–880, Nov. 1999.
- [19] A. Guerin, G. Faucon, and Le Bouquin-Jeannes, R., “Nonlinear acoustic echo cancellation based on Volterra filters,” IEEE Trans. Speech and Audio Processing vol. 11, no. 6, pp. 672-683, Nov. 2003.
- [20] F. Kuech, and W. Kellermann, “Nonlinear Line Echo Cancellation Using A Simplified Second Order Vollterra Filter” IEEE Proc. ICASSP’98, pp. II-1117-1120. 1998.
- [21] E. Hansler, and G. Schmidt, *Acoustic Echo and Noise Control A Practical Approach*, New Jersey: Willy, 2004.

[22] F. Kuech, W. Kellermann, "Partitioned block frequency-domain adaptive second-order Volterra filter" IEEE Tran., vol.53, issue 2, part 1, pp. 564-575, Feb. 2004

