Two-Staged Nonlinear Acoustic Echo Cancellation and Its Convergence Analysis

Ko-Chiang Li[1], Shih-Fu Hsieh*[1] and Jang-Chyuan Jenq[2]

Department of Electrical Engineering[1]
National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China
Department of Computer and Communication Engineering[2]
China University of Technology[2], Hsinchu, Taiwan 300, Republic of China
E-mail: calvin0050@yahoo.com.tw, sfhsieh@mail.nctu.edu.tw, and jj.chyuan@msa.hinet.net

Abstract *Hand-free telephone system commonly drives the power-amplification and loudspeaker into saturated nonlinear region such that performance of conventional acoustic echo cancellation (AEC) is degraded. A computationally efficient cascade AEC model is used to include a memoryless piecewise linear (PWL) processor and a linear filter. In order to overcome the start-up divergence problem encountered in nonlinear AEC adaptation, a two-stage update scheme is adopted that starts with the linear filter update, and then joint adaptation of PWL and linear coefficients follows. Theoretical convergence analysis and stability criterion will be derived and validated by the computer simulation.*

Keywords*—acoustic echo cancellation, nonlinear filter, Hammerstein, LMS algorithm, convergence analysis*

1. INTRODUCTION

The nonlinear acoustic echo cancellation (AEC) system is shown in Fig.1. The signal *x(n)* from the far end goes through the nonlinear loudspeaker and the room impulse response and then is picked up by the microphone. The nonlinear AEC aims to cancel the nonlinear echo, but the design and analysis of nonlinear adaptive AEC is difficult [1]. Being a cascade of a memoryless nonlinear function and a linear filter, the simple Hammerstein model [2-5] has attracted most attention.

Unlike popular polynomial function, the piecewise linear (PWL) [6-8] processor is used to model the memoryless nonlinear function so that the computational cost of nonlinear coefficient update can be reduced. In the startup of updating nonlinear AEC, convergence cannot be guaranteed, since each filter (linear filter or PWL processor) behaves to compensate each other's misalignment, which can lead to a perpetual oscillating system.

In order to overcome this difficulty, a two-staged Hammerstein update strategy is adopted [1]. In the first stage, the linear filter has to adapt continuously so as to react to any change in the acoustic path, and the PWL filter must not adapt until the linear filter has sufficiently converged. In the second stage, joint update of both linear and nonlinear PWL coefficients will undergo.

Previous researches have studied both LMS and RLS nonlinear PWL adaptive algorithms without convergence analysis [7-8]. Although the statistical performance analysis of an adaptive Hammerstein filter has been proposed [2], a two-staged PWL Hammerstein model is used in this paper for theoretical convergence analysis.

Based on this two-staged strategy, we will derive 1st and 2nd staged convergence analysis from which the stability criterion for the step size can be obtained. Computer simulations will validate our analytical work.

2. ADAPTIVE NONLINEAR LMS AEC USING PWL STRUCTURE

In order to separate the identification of the nonlinear loudspeaker parameters and the tracking of the linear acoustic path changes, Fig. 1 shows a typical cascaded Hammerstein nonlinear AEC [3-4]. The far end signal $x(n)$ is fed into the PWL processor *f(x)* that approximates the nonlinear mapping function with one or more linear equations [7]. It has been exploited to compensate for the effect of nonlinear echo. The output $s(n)$ of the PWL processor passes through a linear filter ***h****(n)* to form a pseudo nonlinear echo $\hat{d}(n)$. Here we assume the nonlinear distortion is caused only by the loudspeaker and let $d(n)$ denote the desired signal.

The PWL function *f(x)* for the speech input range [-1 1] is assumed to be symmetric and its prototype is given by

$$f(x) = \begin{cases} m_1 x \\ m_2(x-\alpha_2) + m_1\alpha_2 \\ \vdots \\ m_N(x-\alpha_N) + \cdots m_1\alpha_1 \end{cases}$$

where $\alpha_N \le |x| < \alpha_{N+1}$, m_j and α_j account for the slope and partition parameters of each linear subregion, respectively, with $\alpha_1 = 0$ and $\alpha_{N+1} = 1$.

In vector form, output of the nonlinear processor $s(n)$ is

$$s = \mathbf{w}^T \cdot \mathbf{f}\,, \qquad (1)$$

where $\mathbf{w} = [w_1 \;\; w_2 \; ... \; w_N]^T$ and $\mathbf{f} = [\, f_1(x) \;\; f_2(x) \; ... \; f_N(x)]^T$ resemble a decomposition that maps real numbers into vectors using a set of predefined partition parameters $\{0,\; \alpha_2, \cdots \alpha_N, 1\}$.

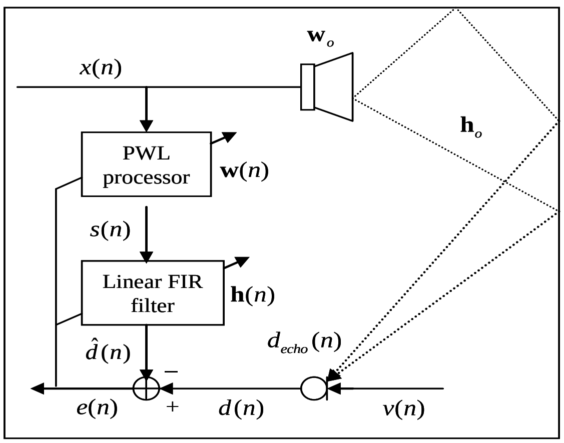

Fig. 1 Nonlinear acoustic echo canceller based on piecewise linear structure and a Hammerstein model

From Eq (1), the delay tap form of PWL processor can be expressed as

$$\mathbf{s}(n) = \left[s(n) \; s(n\text{-}1) \; ... \; s(n\text{-}M+1)\right]^T = \mathbf{F}(n) \cdot \mathbf{w} \qquad (2)$$

where

$$\mathbf{F}(n) = \left[\mathbf{f}(n) \;\; \mathbf{f}(n-1) \;\; \cdots \;\; \mathbf{f}(n-M+1)\right]^T \qquad (3)$$

is the delayed tap mapping matrix. Therefore, the nonlinear AEC output signal $d(n)$ can be written as

$$d(n) = \mathbf{s}^T(n) \cdot \mathbf{h}\,,$$

where $\mathbf{h} = [h_0 \; h_1 \; ... \; h_{M\text{-}1}]^T$ represents the estimated coefficients vector of the linear FIR filter with M being the length of the filter. The estimated output residual error is

$$e(n) = d(n) - d(n) = d(n) - \mathbf{s}^T(n) \cdot \mathbf{h} - v(n)$$

If the coefficients vectors are updated with step size μ_h and μ_w, a joint LMS adaptive algorithm according to the gradient of the cost function, $J(n) = e^2(n)$, is given by

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu_h \mathbf{s}(n) e(n) \qquad (4)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w \mathbf{F}^T(n) \cdot \mathbf{h}(n) e(n). \qquad (5)$$

3.TWO-STAGED ADAPTATION AND ITS CONVERGENCE ANALYSIS

Care should be taken in applying the nonlinear LMS algorithm in the beginning. Each filter of the nonlinear PWL AEC in the joint adaptation behaves to compensate each other's misalignment. This can lead to a perpetual oscillation Therefore, a two-staged strategy, as suggested in [7], was used to avoid this situation. This strategy is to start with a linear filter update in the first stage, and then joint update of both PWL and linear coefficients follows in the second stage.

3.1. First stage: Convergence analysis of linear adaptation

We denote the linear filter weight-error vector by

$$\boldsymbol{\varepsilon}_{\mathbf{h}}(n) = \mathbf{h}(n) - \mathbf{h}_o \qquad (6)$$

where $\mathbf{h}_o$ is the optimal linear filter. The estimation error produced by the nonlinear AEC filter can be expressed as

$$e(n) = d(n) - \hat{d}(n) = \mathbf{s}_o^T(n) \cdot \mathbf{h}_o + v(n) - \mathbf{s}^T(n) \cdot \mathbf{h}(n)$$
$$= v(n) - \mathbf{s}_e^T(n) \cdot \mathbf{h}_o - \mathbf{s}^T(n) \cdot \boldsymbol{\varepsilon}_{\mathbf{h}}(n) \qquad (7)$$

where $\mathbf{s}_e(n) = \mathbf{s}(n) - \mathbf{s}_o(n)$ is the error of PWL processor. Using Eq. (4),(6), and (7), we may rewrite $\boldsymbol{\varepsilon}_h(n+1)$ as

$$\boldsymbol{\varepsilon}_{\mathbf{h}}(n+1) = \mathbf{h}(n) + \mu_h \mathbf{s}(n) e(n) - \mathbf{h}_o$$
$$= \left[\mathbf{I} - \mu_h \mathbf{s}(n) \cdot \mathbf{s}^T(n)\right] \cdot \boldsymbol{\varepsilon}_{\mathbf{h}}(n) + \mu_h \left[v(n)\mathbf{s}(n) - \mathbf{s}(n) \cdot \mathbf{s}_e^T(n) \cdot \mathbf{h}_o\right]$$

The first moment of the linear weight error can be shown to be

$$E\{\boldsymbol{\varepsilon}_h(n)\} = -\frac{\sigma_{s,s_e}^2}{\sigma_s^2}\mathbf{h}_o + \left(\boldsymbol{\varepsilon}_h(0) + \frac{\sigma_{s,s_e}^2}{\sigma_s^2}\mathbf{h}_o\right)(1\text{-}\mu_h\sigma_s^2)^n \qquad (8)$$

from which we can get the upper bound of the step-size as

$$\mu_h < \frac{2}{M\sigma_s^2} \qquad (9)$$

Now we proceed to solve for the second moment $E\left\{\|\boldsymbol{\varepsilon}_h(n)\|_2^2\right\}$ of the linear filter weight error. With the approximations that, m_{s^3,s_e} and m_{s^2,s_e^2} are both comparable to $\sigma_s^2\sigma_{s,s_e}^2$ and $\sigma_s^2\sigma_{s_e}^2$, respectively,

$$E\left\{\mathbf{s}(n)\cdot\mathbf{s}^T(n)\mathbf{s}(n)\cdot\mathbf{s}^T(n)\right\}\approx M\sigma_s^4\mathbf{I} \text{ and } M>>1,$$

where $\sigma_{s_e}^2$ is the variance of $s_e(n)$, $m_{s^3,s_e}=E\left\{s_k^3 s_{e,k}\right\}$ and $m_{s^2,s_e^2}=E\left\{s_k^2 s_{e,k}^2\right\}$, $m_{s^4}=E\left\{s^4(n)\right\}$, we have

$$E\left\{\left\|\boldsymbol{\varepsilon}_h(n+1)\right\|_2^2\right\}\approx \left(1-2\mu_h\sigma_s^2+\mu_h^2 M\sigma_s^4\right)E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}+2E\left\{\boldsymbol{\varepsilon}_h(n)\right\}\times \mathbf{h}_o^T\mu_h\sigma_{s_e}^2\left(-1+\mu_h M\sigma_s^2\right)+\mu_h^2 M\sigma_s^2\left(\sigma_v^2+\sigma_{s_e}^2\left\|\mathbf{h}_o\right\|_2^2\right) \quad (10)$$

By cascading up the first and second moment into one vector, we have the recursive vector equation as

$$\boldsymbol{\theta}(n+1)=\mathbf{A}\cdot\boldsymbol{\theta}(n)+\mathbf{b}, \quad (11)$$

where

$$\mathbf{b}=\left[\mu_h^2 M\sigma_s^2\left(\sigma_v^2+\sigma_{s_e}^2\left\|\mathbf{h}_o\right\|_2^2\right) \quad \mu_h\sigma_{s,s_e}^2\mathbf{h}_o^T\right]^T \quad (12)$$

$$\boldsymbol{\theta}(n)=\left[E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\} \quad E\left\{\boldsymbol{\varepsilon}_h(n)\right\}\right]^T \quad (13)$$

and

$$\mathbf{A}=\begin{bmatrix}\left(1-2\mu_h\sigma_s^2+\mu_h^2 M\sigma_s^4\right) & 2\mathbf{h}_o^T\mu_h\sigma_{s_e}^2\left(-1+\mu_h M\sigma_s^2\right)\\ \mathbf{0} & (1-\mu_h\sigma_s^2)\mathbf{I}_M\end{bmatrix} \quad (14)$$

The solution of Eq. (11) is given by

$$\boldsymbol{\theta}(n)=\left(\mathbf{I}-\mathbf{A}\right)^{-1}\cdot\mathbf{b}+\mathbf{A}^n\cdot\left(\boldsymbol{\theta}(0)-\left(\mathbf{I}-\mathbf{A}\right)^{-1}\cdot\mathbf{b}\right) \quad (15)$$

The steady state $\boldsymbol{\theta}(n)$ is equal to $\left(\mathbf{I}-\mathbf{A}\right)^{-1}\cdot\mathbf{b}$, from which the steady state of the second moment $E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}$ is given by

$$\lim_{n\to\infty}E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}=\frac{\mu_h M\left(\sigma_v^2+\sigma_{s_e}^2\left\|\mathbf{h}_o\right\|_2^2\right)+2\dfrac{\sigma_{s,s_e}^2\sigma_{s_e}^2}{\sigma_s^4}\left\|\mathbf{h}_o\right\|_2^2\left(1-\mu_h M\sigma_s^2\right)}{2-\mu_h M\sigma_s^2}. \quad (16)$$

Finally, after derivation of first and second moments of the linear coefficient weight error, we turn our attention to the residual output power

$$\begin{aligned}J_h(n)&=E\left\{\left|e(n)\right|^2\right\}\\&=\sigma_v^2+\mathbf{h}_o^T\mathrm{E}\left\{\mathbf{s}_e(n)\cdot\mathbf{s}_e^T(n)\right\}\mathbf{h}_o+E\left\{\boldsymbol{\varepsilon}_h^T(n)\cdot\mathbf{s}(n)\cdot\mathbf{s}^T(n)\cdot\boldsymbol{\varepsilon}_h(n)\right\}\\&\quad+2\mathbf{h}_o^T\cdot\mathrm{E}\left\{\mathbf{s}(n)\cdot\mathbf{s}_e^T(n)\right\}\cdot E\left\{\boldsymbol{\varepsilon}_{\mathbf{h}}(n)\right\}.\end{aligned} \quad (17)$$

Since the variation of $\boldsymbol{\varepsilon}_h(n)$ is slow compared to $\mathbf{s}(n)$, we have

$$E\left\{\boldsymbol{\varepsilon}_h^T(n)\cdot\mathbf{s}(n)\cdot\mathbf{s}^T(n)\cdot\boldsymbol{\varepsilon}_h(n)\right\}=\sigma_s^2 E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}$$

……………………………………………….(18)

The mean square error can be written as

$$J_h(n)=\sigma_v^2+\sigma_{s_e}^2\left\|\mathbf{h}_o\right\|_2^2+\sigma_s^2 E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}+2\sigma_{s,s_e}^2\mathbf{h}_o^T\cdot E\left\{\boldsymbol{\varepsilon}_{\mathbf{h}}(n)\right\} \quad (19)$$

which depends on $E\left\{\boldsymbol{\varepsilon}_{\mathbf{h}}(n)\right\}$ and $E\left\{\left\|\boldsymbol{\varepsilon}_h(n)\right\|_2^2\right\}$ derived earlier.

3.2 Second stage: Convergence analysis of joint adaptation of linear and PWL coefficients

After the initial convergence of linear coefficients in the first stage, we switch to the second stage in which both linear and PWL coefficients will be updated jointly. Now, the residual error is given by

$$e(n)=v(n)-\mathbf{w}_o^T\cdot\mathbf{F}^T(n)\cdot\boldsymbol{\varepsilon}_h(n)-\boldsymbol{\varepsilon}_w^T(n)\cdot\mathbf{F}^T(n)\cdot\mathbf{h}_o-\boldsymbol{\varepsilon}_w^T(n)\mathbf{F}^T(n)\boldsymbol{\varepsilon}_h(n) \quad (20)$$

The coupled linear and nonlinear weight error in the fourth term renders difficulty in convergence analysis. However, with wideband signal like speech, loudspeaker nonlinearities are much less dominant than the linear components in general. Therefore, we can assume initial PWL weight error is much smaller than optimal PWL coefficients. Moreover, the converged linear coefficients are also very close to the optimal linear filter, namely,

$$\boldsymbol{\varepsilon}_h(n)<<\mathbf{h}_o,\ \boldsymbol{\varepsilon}_w(n)<<\mathbf{w}_o, \quad (21)$$

where $\boldsymbol{\varepsilon}_w(n)=\mathbf{w}(n)-\mathbf{w}_o$. With this assumption of sufficiently small perturbation errors in linear and nonlinear coefficients, the second order perturbation term can be discarded and the estimation error becomes

$$e(n)\approx v(n)-\mathbf{w}_o^T\cdot\mathbf{F}^T(n)\cdot\boldsymbol{\varepsilon}_h(n)-\boldsymbol{\varepsilon}_w^T(n)\cdot\mathbf{F}^T(n)\cdot\mathbf{h}_o \quad (22)$$

Denote the combined linear and PWL coefficient weight error as

$$\boldsymbol{\varepsilon}(n)=\left[\boldsymbol{\varepsilon}_h(n) \quad \boldsymbol{\varepsilon}_w(n)\right]^T$$

and let

$$\mathbf{G}(n)=\left[\mathbf{w}_o^T\cdot\mathbf{F}^T(n) \quad \mathbf{F}^T(n)\cdot\mathbf{h}_o\right]^T \quad (23)$$

then Eq. (22) becomes:

$$e(n)\approx v(n)-\mathbf{G}^T(n)\cdot\boldsymbol{\varepsilon}(n) \quad (24)$$

Denote the combined coefficient weight error as

$$\boldsymbol{\varepsilon}(n+1)=\begin{bmatrix}\mathbf{h}(n+1)\\ \mathbf{w}(n+1)\end{bmatrix}-\begin{bmatrix}\mathbf{h}_o\\ \mathbf{w}_o\end{bmatrix} \quad (25)$$

we have

$$\boldsymbol{\varepsilon}(n+1)=\boldsymbol{\varepsilon}(n)+\mathbf{T}\cdot\begin{bmatrix}\mathbf{F}(n)\cdot\left(\boldsymbol{\varepsilon}_w(n)-\mathbf{w}_o\right)\\ \mathbf{F}^T(n)\cdot\left(\mathbf{h}_o-\boldsymbol{\varepsilon}_h(n)\right)\end{bmatrix}\left(v(n)-\mathbf{G}^T(n)\cdot\boldsymbol{\varepsilon}(n)\right)$$

…………………………………………………..(26)

where

$$\mathbf{T}=\begin{bmatrix}\mu_h\mathbf{I}_M & \mathbf{O}\\ \mathbf{O} & \mu_w\mathbf{I}_N\end{bmatrix}.$$

According to the small perturbation assumption, it can be approximated as

$$\boldsymbol{\varepsilon}(n+1) \approx \left(\mathbf{I}-\mathbf{T}\cdot\mathbf{G}(n)\cdot\mathbf{G}^T(n)\right)\cdot\boldsymbol{\varepsilon}(n)-v(n)\mathbf{T}\cdot\mathbf{G}(n) \quad (27)$$

The first moment of $\boldsymbol{\varepsilon}(n+1)$ can be shown to be

$$E\{\boldsymbol{\varepsilon}(n)\} \approx \left(\mathbf{I}-\mathbf{T}\cdot\mathbf{R_G}\right)^N \cdot E\{\boldsymbol{\varepsilon}(0)\} \quad (28)$$

where $\mathbf{R_G}$ is the correlation matrix of $\mathbf{G}(n)$. With a suitable step-size, the magnitude of this geometric ratio must be less than unity for all *n*, we can see that the estimate is unbiased.

Similarly, the second moment of $\boldsymbol{\varepsilon}(n)$ is given by

$$E\left\{\|\boldsymbol{\varepsilon}(n+1)\|_2^2\right\} = E\left\{\|v(n)\mathbf{T}\cdot\mathbf{G}(n)\|_2^2\right\} +$$

$$E\left\{\boldsymbol{\varepsilon}_h^T(n)\left(1-2\mathbf{T}\cdot\mathbf{R_G}+\mathbf{T}^2 E\left\{\mathbf{G}(n)\cdot\mathbf{G}^T(n)\cdot\mathbf{G}(n)\cdot\mathbf{G}^T(n)\right\}\right)\boldsymbol{\varepsilon}_h(n)\right\}$$

…………………………………………………..(29)

Here, we assume the term $E\left\{\mathbf{G}(n)\cdot\mathbf{G}^T(n)\cdot\mathbf{G}(n)\cdot\mathbf{G}^T(n)\right\}$ can be approximated as $\mathbf{R}_G^2$. By applying the similarity transformation, $\mathbf{R}_G$ is transformed into a simpler form:

$$\mathbf{Q}^T\cdot\mathbf{R}_G\cdot\mathbf{Q}=\mathbf{D},$$

where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{D}$ is a diagonal matrix consisting of the eigenvalues λ_i of $\mathbf{R}_G$. Letting

$$\mathbf{K}(n)=\mathbf{Q}^T\cdot\boldsymbol{\varepsilon}(n),$$

we can deduce the second moment of $\boldsymbol{\varepsilon}(n)$ and mean square error $J(n)$ as follows:

$$E\left\{\|\boldsymbol{\varepsilon}(n)\|_2^2\right\} = \sum_{i=1}^{M+N}\frac{\sigma_v^2 T_i\lambda_i}{2-T_i\lambda_i}+\left[\left|k_i(0)\right|^2-\frac{T_i\sigma_v^2}{2-T_i\lambda_i}\right](1-T_i\lambda_i)^{2n}$$

…………………………………………………..(30)

$$J(n)=\sigma_v^2+\sum_{i=1}^{M+N}\frac{\sigma_v^2 T_i\lambda_i}{2-T_i\lambda_i}+\lambda_i\left[\left|k_i(0)\right|^2-\frac{T_i\sigma_v^2}{2-T_i\lambda_i}\right](1-T_i\lambda_i)^{2n}$$

…………………………………………………..(31)

where T_i is the *i*-th diagonal entry of $\mathbf{T}$ and $k_i(0)$ is the initial value of *i*-th entry of $\mathbf{K}(n)$ at *n=0*.

4. COMPUTER SIMULATIONS

In our simulations, unless otherwise stated, the far end signal is a uniformly distributed white noise. The nonlinear I/O mapping curve is a raised-cosine function. The room impulse response of length 128 is generated by a random number generator with an exponential damping factor. In the first stage, only linear filter is undated, and we let the step size $\mu_h=0.01$ which is securely below the upper bound. SNR=30 dB, the nonlinear filter order is 3 with a uniform partition on [0 0.33 0.66 1], the PWL coefficients of the loudspeaker is $w_{o,1}=1.9050$ $w_{o,2}=$ -0.8886 $w_{o,3}=$ -0.8875 to approach a raised-cosine curve and the initial values of the PWL processor are $w_1=1.905$ $w_2=0$ $w_3=0$, which is equivalent to a straight line with slope 1.905. Fig. 2 compares the simulated and theoretical residual error powers in the first stage. If the step size is chosen to be 0.013, exceeding the bound 0.012, divergence does happen in our simulations.

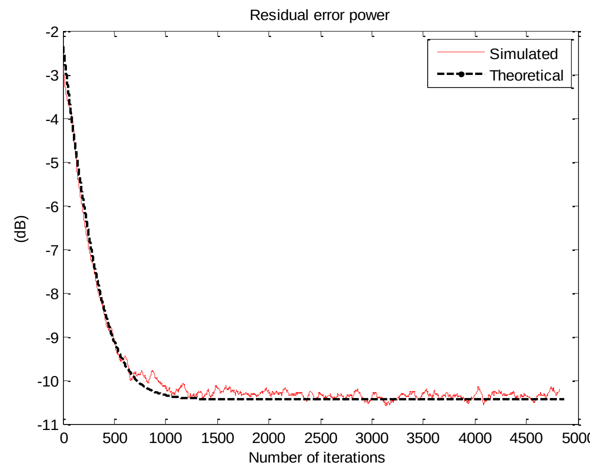

Fig. 2 Nonlinear PWL AEC with step size 0.01 at first stage.

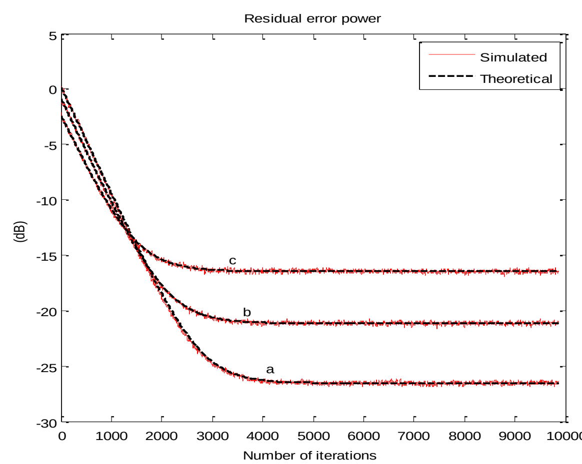

Fig. 3 Residual echo power comparison with different degrees of nonlinearity at first stage: (a) slight (b) moderate (c) severe.

One important factor is the nonlinearity of loudspeaker. If it is highly nonlinear, the covariance of $\mathbf{s}(n)$ and $\mathbf{s}_e(n)$ and the variance of $s_e(n)$ in Eq. (16) are larger. In Fig. 3, three nonlinear loudspeakers are used with (a) slight, (b) moderate, (c) severe nonlinearities. It shows that the steady-state error is a monotonically increasing function of the nonlinearity factors.

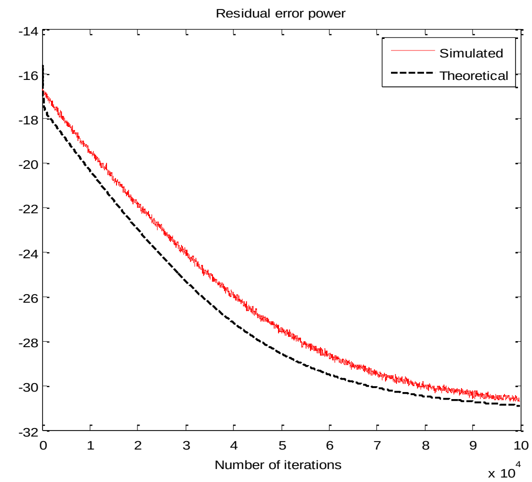

Fig. 4 Comparison of theoretical and simulated residual error powers during joint adaptation of the second stage.

After the convergence of the first stage, we can see in Fig. 4 that the simulated residual error power is slightly larger than the theoretical curve.

Finally a true speech 8KHz experiment is performed. A low-cost 2.5 inch diameter desktop loudspeaker is placed 6 inches above the recording Creative-MC1000 microphone. Fig. 5 shows the nonlinear PWL AEC has 3 dB ERLE gain over linear AEC. It is also better than the commonly used polynomial-nonlinear adaptive filter[3].

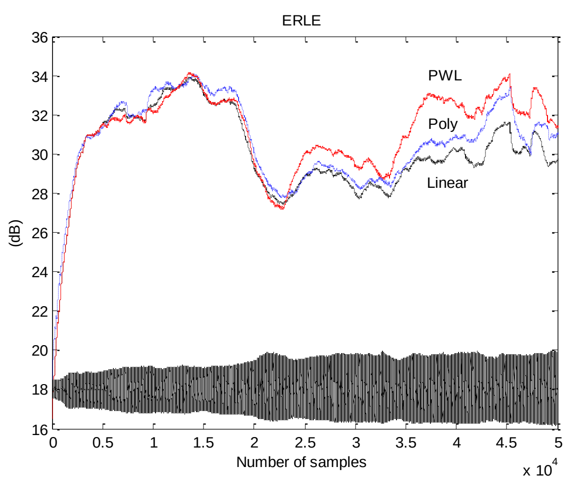

Fig. 5 ERLE comparison for a true speech and loudspeaker system.

5. CONCLUSIONS

We have developed the nonlinear AEC system where the scheme is a cascade model of a memoryless PWL processor and linear FIR filter. The advantage of this cascade model is its joint LMS adaptive algorithm has lower computation load than conventional nonlinear AEC based on polynomial function.

Since each filter (linear filter or PWL processor) behaves to compensate each other's misalignment, which can lead to a perpetual oscillating phenomena. In order to overcome this difficulty, we have adopted two-staged algorithm, starting with a linear filter, and then joint PWL and linear coefficients update follow with a steady state linear filter and derived its convergence analysis.

ACKNOWLEDGEMENT

This work was supported by the NSC and MediaTek research center at National Chiao Tung University.

REFERENCES

[1] A. Guerin, G.. Faucon and R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters,"*IEEE Trans. Speech and Audio Processing*, vol. 11, no. 6, pp.672-683, Nov. 2003.

[2] J. Jeraj and V.J. Mathews, "Stochastic mean-square performance analysis of an adaptive Hammerstein filter," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2168-2177, June 2006.

[3] G. Y. Jiang and S. F. Hsieh, "Nonlinear acoustic echo cancellation using orthogonal polynomial," *ICASSP*, vol.5, pp. 273-276, 2006.

[4] J. P. Costa, A. Lagrange, and A. Arliaud, " Acoustic echo cancellation using nonlinear cascade filters," *ICASSP*, pp.v389-v392, 2003.

[5] K. Shi; X. Ma, and G. T. Zhou, **"**Adaptive nonlinearity identification in a Hammerstein system using a pseudo coherence function**,"** *IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 745-748, Aug. 2007.

[6] L.O. Chua and S. M. Kang, "Section-wise piecewise-linear functions: Canonical representation, properties, and applications," *Proceedings of the IEEE*, vol. 65, no. 6, pp.915-929. Jun. 1977.

[7] J. N. Lin and R. Unbehauen, "Adaptive nonlinear digital filter with canonical piecewise-linear structure," *IEEE Tran. Circuits and Systems*, vol. 37, no. 3, pp. 347-353. Mar. 1990.

[8] V. Vörös, "Modeling and parameter identification of systems with multisegment piecewise-linear characteristics," *IEEE Trans. Automatic Control*, vol.47, no. 1, pp. 184-188, Jan. 2002.