Nonlinear Acoustic Echo Cancellation Based on Higher Order Correlation

Guan-Yu Jiang[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: sjsmit63@yahoo.com.tw, sfhsieh@mail.nctu.edu.tw, and jj.chyuan@msa.hinet.net

Abstract *In order to compensate the nonlinearly distorted echo in hands-free telephone, a memoryless power-series polynomial nonlinear acoustic echo canceller (AEC) can be used. Conventional nonlinear AEC employs a NLMS adaptive algorithm which can fail to converge in case of doubletalk or a very noisy environment. In this paper we propose a higher order correlation algorithm using a white Gaussian training signal. With comparable complexity as NLMS, it is most effective at small SNR. Computer simulations demonstrate that the proposed algorithm has a smaller steady-state echo and it is also very robust to background noise.*

Keywords— *acoustic echo cancellation, nonlinear adaptive filter, correlation*

1. INTRODUCTION

Hands-free telephone usually suffers from the annoying acoustic echo problem. A linear adaptive filter is commonly used for acoustic echo cancellation (AEC). However, overdriving the power amplifier of loudspeaker will incur nonlinear distortion. Recently, several nonlinear AEC structures have been proposed to compensate this kind of distortion [1]-[3]. The cascaded nonlinear AEC structure has fewer coefficients than the Volterra filter [2] and has less computational complexity, if it is updated by the NLMS algorithm [3]. The cascade nonlinear AEC is shown in Fig 1 where the power amplifier before loudspeaker is modeled by the nonlinear processor (a power-series polynomial expansion) and the echo path is modeled by the linear FIR filter. It is well known that the nonlinear NLMS adaptive algorithm is based on a simple stochastic gradient and it has an error feedback structure with a low computational complexity. In general, this kind of adaptive algorithm has good performance in general. However, in case of doubletalk or a large background noise, the AEC coefficients may diverge due to the strongly perturbed error feedback signal [4].

In this paper we use a white Gaussian input signal to train both linear and nonlinear coefficients before transmission of a real speech signal. Our purpose is to aid the poor adaptive filter at low SNR. The training sequences, used in channel estimation, echo canceller, and nonlinear system identification, have been well studied in [4]-[7]. We will apply it to nonlinear AEC with a cascade structure and propose a higher order correlation (HOC) algorithm to find out both linear and nonlinear coefficients. Its low computational complexity, comparable to that of the NLMS algorithm, makes it very promising for nonlinear AEC at low SNR.

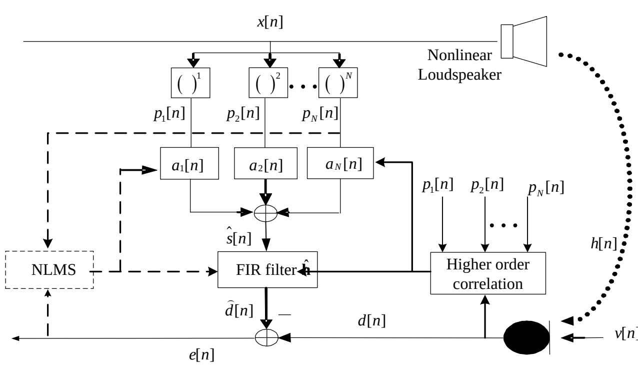

Fig 1. Nonlinear acoustic echo canceller

2. HIGHER ORDER CORRELATION FOR NONLINEAR AEC

In Fig 1, a white Gaussian training signal $x[n]$ is generated and played over a memoryless nonlinear loudspeaker modeled by power-series polynomial with coefficients a's, and through a linear time-invariant system with impulse response $h[n]$, it is then pickup by a microphone. By correlating the microphone signal $d[n]$ with the power signal $p_j[n]=x^j[n], 1\le j\le N$, a higher order correlation algorithm can be developed to solve for the nonlinear coefficients *a*'s and the linear FIR filter coefficients *h*'s.

2.1. Linear HOC coefficients estimation algorithm

First, some symbols are defined as follows,

$\mathbf{x}^t[n]=[x^t[n],x^t[n-1],...,x^t[n-M+1]]^T$, t^{th} order input data vector
$\mathbf{h}=[h_0,h_1,\cdots,h_{M-1}]^T$, room impulse response vector
$m_t=E[x[n]x^t[n]]$, t^{th} order moment
$r_{xd}[-k]=E[x[n-k]d[n]]$, cross correlation function
$r_{xx^t}[i-k]=E[x[n-k]x^t[n-i]]$, t^{th} order auto correlation
$\mathbf{r}_{xx^t}=[r_{xx^t}[0-k],\cdots,r_{xx^t}[M-1-k]]$, t^{th} auto correlation vector
$\mathbf{R}_{xx}=E[\mathbf{x}[n]\mathbf{x}^T[n]]$, auto correlation matrix
$\mathbf{R}_{xd}=E\left[\mathbf{x}[n]d[n]\right]$, cross correlation vector

This sequence cannot be a polar signal in order to have nontrivial higher order moments, which will be explained later [7]. The microphone signal $d[n]$ can be expressed as:

$$d[n]=\sum_{t=1}^{N}a_t\sum_{i=0}^{M-1}h_i x^t[n-i]+v[n] \tag{1}$$

M is the length of the linear filter, N is the nonlinear order, $v[n]$ is the independent background noise.

By defining the cost function as

$$J=E\{e^2[n]\}=E\{(d[n]-\hat{y}[n])^2\}$$

where

$$\hat{\mathbf{y}}[n]=\hat{\mathbf{h}}^T[n]\mathbf{X}[n]\hat{\mathbf{a}}[n]$$

and

$$\mathbf{X}[n]=[\mathbf{x}[n]\quad \mathbf{x}^2[n]\quad \ldots\quad \mathbf{x}^N[n]].$$

We can see that the optimum mean-square-error estimate for the nonlinear coefficient $\hat{\mathbf{a}}$ and linear FIR coefficients $\hat{\mathbf{h}}$ can be found by solving

$$\nabla_{\mathbf{h}}\ J=-2E\{(d[n]-\hat{\mathbf{h}}^T[n]\mathbf{X}[n]\hat{\mathbf{a}}[n])\mathbf{X}[n]\hat{\mathbf{a}}[n]\}=\mathbf{0}$$

$$\nabla_{\mathbf{a}}\ J=-2E\{(d[n]-\hat{\mathbf{h}}^T[n]\mathbf{X}[n]\hat{\mathbf{a}}[n])\mathbf{X}^T[n]\hat{\mathbf{h}}[n]\}=\mathbf{0}$$

which are unfortunately a set of nonlinear equations, unlike the well known normal equations solution for Volterra filter[7].

Motivated by the above cross correlation in the gradient vector, the cross correlation function between $x[n]$ and $d[n]$ can be written as:

$$\begin{aligned}r_{xd}[-k]&=\sum_{t=1}^{N}a_t\sum_{i=0}^{M-1}h_i r_{xx^t}[i-k]+E[x[n-k]v[n]]\\&=\sum_{t=1}^{N}a_t\mathbf{r}_{xx^t}\mathbf{h}\end{aligned} \tag{2}$$

Letting the lag index k go from 0 to M-1, we extend (2) in matrix form to obtain the cross correlation vector as:

$$\mathbf{R}_{xd}=a_1\mathbf{R}_{xx}\mathbf{h}+a_2\mathbf{R}_{xx^2}\mathbf{h}+\cdots+a_N\mathbf{R}_{xx^N}\mathbf{h} \tag{3}$$

where the autocorrelation matrices are

$$\mathbf{R}_{xx^t}=m_{t+1}\mathbf{I}_{M\times M} \tag{4}$$

for $1\le t\le N$ and t is odd. We note that $\mathbf{R}_{xx^2},\mathbf{R}_{xx^4},\cdots,\mathbf{R}_{xx^t}$ are zero matrices for $2\le t\le N$, t is even, since $x[n]$ is assumed to be a zero-mean white signal with a symmetric Gaussian probability density function. Here we let N be an odd integer and rewrite (3) as

$$\mathbf{R}_{xd}=(a_1m_2+a_3m_4+\cdots+a_Nm_{N+1})\mathbf{h} \tag{5}$$

Therefore, the column vector of $\mathbf{R}_{xd}$ is parallel to the vector of the room impulse response $\mathbf{h}$ and its vector length is composed of nonlinear coefficients. We assume that $\|\mathbf{h}\|_2^2$ equals to 1 thus the direction of $\mathbf{R}_{xd}$ is equal to $\mathbf{h}$, then the room impulse can be estimated by the normalized cross correlation vector

$$\mathbf{h}=\frac{\mathbf{R}_{xd}}{\|\mathbf{R}_{xd}\|_2} \tag{6}$$

2.2. Nonlinear HOC coefficients estimation algorithm

Because the input signal's Gaussian probability density function is symmetric, its odd-order moments are zeros. We divide the nonlinear HOC algorithm into two parts, odd- and even-ordered. First, we solve the odd-ordered nonlinear coefficients. Again, we multiply (1) by the odd-power signals $x^3[n-k],...,\ x^t[n-k]$ for $1\le t\le N$, t is odd, respectively, and take its expectation. Similar to (5), we can get the higher order crosscorrelation equations as:

$$\begin{aligned}\mathbf{R}_{x^3d}&=(a_1m_4+a_3m_6+\cdots+a_Nm_{N+3})\mathbf{h}\\&\vdots\\\mathbf{R}_{x^Nd}&=(a_1m_{N+1}+a_3m_{N+3}+\cdots+a_Nm_{2N})\mathbf{h}\end{aligned} \tag{7}$$

Multiplying (3) and (7) by $\mathbf{h}^{\mathrm{T}}$, we have

$$\begin{bmatrix}\mathbf{h}^T\mathbf{R}_{xd}\\\mathbf{h}^T\mathbf{R}_{x^3d}\\\vdots\\\mathbf{h}^T\mathbf{R}_{x^Nd}\end{bmatrix}=\underbrace{\begin{bmatrix}m_2&m_4&\cdots&m_{N+1}\\m_4&m_6&\cdots&m_{N+3}\\\vdots&\vdots&\ddots&\vdots\\m_{N+1}&m_{N+3}&\cdots&m_{2N}\end{bmatrix}}_{\mathbf{G}_{odd}}\begin{bmatrix}a_1\\a_3\\\vdots\\a_N\end{bmatrix} \tag{8}$$

which can be solved to find the odd-ordered nonlinear coefficients $a_1,\ a_3,\cdots,a_N$ as:

$$\begin{bmatrix} a_1 \\ a_3 \\ \vdots \\ a_N \end{bmatrix} = \mathbf{G}_{odd}^{-1} \begin{bmatrix} \mathbf{h}^T\mathbf{R}_{xd} \\ \mathbf{h}^T\mathbf{R}_{x^3d} \\ \vdots \\ \mathbf{h}^T\mathbf{R}_{x^Nd} \end{bmatrix} \tag{9}$$

We note that the matrix $\mathbf{G}_{\text{odd}}$ in (9) has to be nonsingular in order to have a unique solution for the nonlinear coefficients [7]. It is obvious that the input training sequence $x[n]$ cannot be a polar ±1 signal.

Next we will proceed to find the even-ordered nonlinear coefficients. Similarly, we start with (1) but there are some differences from previous procedure. Although the input signal is white with zero mean but its even-ordered moments are not equal to zero. In order to get a diagonal correlation matrix we multiply (1) by the moment-reduced power signals $x^2[n-k]-m_2$ and take expectation to get

$$\mathbf{R}_{(x^2-m_2)d} = a_2(m_4-m_2^2)\mathbf{h} + a_4(m_6-m_2m_4)\mathbf{h} + \cdots + a_4(m_{N+2}-m_2m_N)\mathbf{h}.$$

In matrix form, we have

$$\begin{bmatrix} \mathbf{h}^T\mathbf{R}_{(x^2-m_2)d} \\ \mathbf{h}^T\mathbf{R}_{(x^4-m_4)d} \\ \vdots \\ \mathbf{h}^T\mathbf{R}_{(x^{N-1}-m_{N-1})d} \end{bmatrix} = \underbrace{\begin{bmatrix} m_4-m_2^2 & m_6-m_2m_4 & \cdots & m_{N+2}-m_2m_N \\ m_6-m_2m_4 & m_8-m_4^2 & \cdots & m_{N+3}-m_4m_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N+1}-m_2m_{N-1} & m_{N+3}-m_4m_{N-1} & \cdots & m_{2N-2}-m_{N-1}^2 \end{bmatrix}}_{\mathbf{G}_{even}} \begin{bmatrix} a_2 \\ a_4 \\ \vdots \\ a_{N-1} \end{bmatrix}$$

With $\mathbf{G}_{even}$ being assumed to be nonsingular, the even-numbered nonlinear coefficients a_t can be solved by

$$\begin{bmatrix} a_2 \\ a_4 \\ \vdots \\ a_{N-1} \end{bmatrix} = \mathbf{G}_{even}^{-1}\mathbf{h}^T \begin{bmatrix} \mathbf{R}_{(x^2-m_2)d} \\ \mathbf{R}_{(x^4-m_4)d} \\ \vdots \\ \mathbf{R}_{(x^{N-1}-m_{N-1})d} \end{bmatrix} \tag{10}$$

2.3. Recursive HOC algorithm

In order to compute the linear and nonlinear coefficients in (9) and (10), we need to compute the autocorrelation function and cross correlation function. In practice, we replace the expectation operation with the sample mean. Thus, the estimated cross correlation function $\widehat{R}_{xd}[n]$ can be expressed as:

$$\begin{aligned} \widehat{R}_{xd}[n] &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}[i]d[i] \\ &= \frac{n-1}{n}\widehat{R}_{xd}[n-1] + \frac{1}{n}\mathbf{x}[n]d[n] \end{aligned} \tag{11}$$

In (6), the estimated room impulse $\hat{h}[n]$ becomes

$$\hat{h}[n] = \frac{n-1}{n}\hat{h}[n-1] + \frac{1}{n\left\|\widehat{\mathrm{R}}_{xd}[n]\right\|_2}\mathbf{x}[n]d[n] \tag{12}$$

Similarity, in (6), (9) and (10), the correlation matrix of the nonlinear HOC algorithms is replaced by the sample mean and its recursive form can be rewritten as follows:

$$\begin{bmatrix} \hat{a}_1[n] \\ \hat{a}_3[n] \\ \vdots \\ \hat{a}_N[n] \end{bmatrix} = \frac{n-1}{n}\begin{bmatrix} \hat{a}_1[n-1] \\ \hat{a}_3[n-1] \\ \vdots \\ \hat{a}_N[n-1] \end{bmatrix} + \frac{d[n]\mathbf{G}_{odd}^{-1}\hat{h}^T[n]}{n}\begin{bmatrix} \mathbf{x}^1[n] \\ \mathbf{x}^3[n] \\ \vdots \\ \mathbf{x}^N[n] \end{bmatrix}$$

$$\begin{bmatrix} \hat{a}_2[n] \\ \hat{a}_4[n] \\ \vdots \\ \hat{a}_N[n] \end{bmatrix} = \frac{n-1}{n}\begin{bmatrix} \hat{a}_2[n-1] \\ \hat{a}_4[n-1] \\ \vdots \\ \hat{a}_N[n-1] \end{bmatrix} + \frac{d[n]\mathbf{G}_{even}^{-1}\hat{h}^T[n]}{n}\begin{bmatrix} (\mathbf{x}^2[n]-m_2) \\ (\mathbf{x}^4[n]-m_4) \\ \vdots \\ (\mathbf{x}^N[n]-m_N) \end{bmatrix}$$

Here we have assumed that the moments of input signals are known *a priori* so that the matrix inverse $\mathbf{G}_{odd}^{-1}$ and $\mathbf{G}_{even}^{-1}$ can be pre-computed. The steady state of linear and nonlinear HOC coefficients will achieve the optimum Wiener solution because the correlation estimation can smooth out the perturbing noise *v[n]* when the iteration number $n \to \infty$. By contrast, the coefficients computed by the NLMS algorithm have excess mean square error resulting from gradient noise at small step size.

In view of computational complexity, HOC needs about $2M+N$ and $MN+2N+N^2$ multiplications per iteration to solve the linear and nonlinear coefficients, respectively. Here, we do not take into account the computational cost of the matrix inverse $\mathbf{G}^{-1}$, since it can be computed *a priori.* To generate the nonlinear components, it needs about *N-1* multiplications. Table 1 compares the computational complexities between HOC and NLMS adaptive filter [3]. The computational complexity of the HOC method is almost the same as the NLMS adaptive filter, as the nonlinear order *N* is generally much less than the FIR filter order *M*. Both methods can be used to find the AEC coefficients. In case of a noisy environment, the HOC method is more attractive.

Table 1 Comparison of computational complexity

Nonlinear algorithms	Number of multiplication
HOC	$MN+2M+4N-1+N^2$
NLMS	$MN+2M+4N-1$

3. SIMULATION RESULTS

In the our simulations, the length of the room impulse response is set to be 128, which is identical to the number of taps of the room impulse response; the nonlinear coefficients are $a_1 = 2.5967$, $a_3 = -3.3283$, $a_5 = 1.7833$, $a_2 = a_4 = 0$. The signal to noise ratio is defined as the power ratio of echo and background noise. The residual echo is shown in Fig 2, when SNR is equal to 10 dB. The convergence rate of the NLMS adaptive filter is dependent on the step size, thus we use two sets of step size, [$\mu_a = 0.05$, $\mu_h = 0.05$], and [$\mu_a = 0.25$, $\mu_h = 0.1$].

The HOC method has better steady state performance than the NLMS method since performance of NLMS is limited by the background noise due to its error feedback structure. Although the adaptive filter has faster convergence rate at the beginning of iteration, but its steady state performance is poorer as a tradeoff. In Fig 3, comparison is made when SNR is equal 0 dB to account for doubletalk. Convergence rate of HOC is relatively insensitive to SNR variations, except for a bias due to the noise. This is the main advantage of the HOC method that makes it preferable than NLMS adaptive filter at low SNRs.

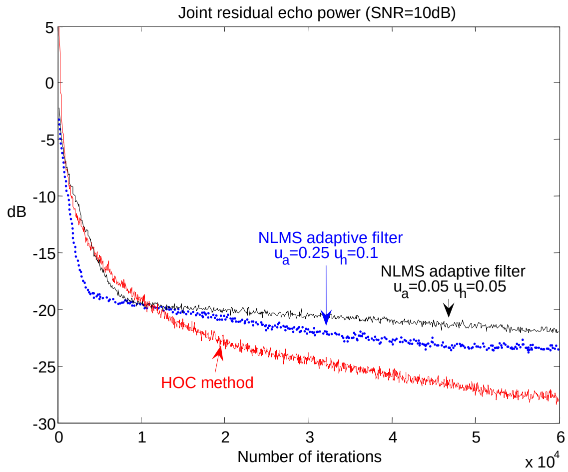

Fig 2 Residual echo power comparison of NLMS and HOC (SNR=10 dB)

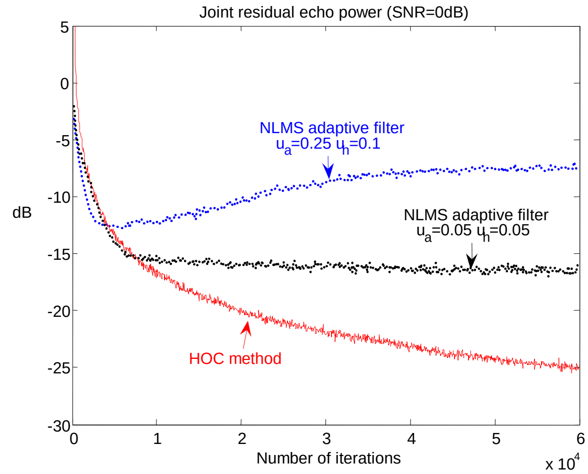

Fig 3 Residual echo power comparison of NLMS and HOC under noisy environment (SNR=0 dB)

For a real low-cost loudspeaker of 2.5-inch diameter and a Creative-MC1000 microphone, placed 4 inches away, a nonlinear polynomial of 5^{th} and an FIR of 256^{th} order are used to model the loudspeaker-to-microphone channel. We add a near-end speech signal to account for large background noise. Comparison of ERLE is shown in Table 2 that again justifies the advantages of the robust HOC method.

Table 2 Average ERLE comparison at different SNRs

Avg ERLE / SNR	NLMS adaptive filter	HOC method
-3 dB	-4.2dB	3.8dB
3 dB	3.8dB	6dB
6 dB	8.6dB	6.5dB
12 dB	13.5dB	6.9dB

4. SUMMARY

We proposed a higher order correlation estimation algorithm based on a training white signal. The recursive HOC algorithm has comparable computational cost but better echo cancellation performance than NLMS, especially in noisy environment. From the experiment of a real speech into a real echo path, we know that in case of a large noise or doubletalk, when NLMS adaptive filter diverges, the HOC algorithm still retains its robust nonlinear echo estimation capability.

ACKNOWLEDGEMENT

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

REFERENCES

[1] J. P. Costa, A. Lagrange, and A. Arliaud, "Acoustic echo cancellation using nonlinear cascade filters," *ICASSP*, vol. 5, pp. 389-392, Apr. 2003.

[2] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," *ICASSP*, vol. 2, pp. 877–880, Nov. 1999.

[3] A. Stenger and W. Kelermann, "Nonlinear acoustic echo cancellation with fast converging memoryless preprocessor," *ICASSP*, vol.2, pp. 805-808, June 2000.

[4] J.C. Jenq and S.F. Hsieh, "Acoustic echo cancellation using iterative maximal length correlation and double-talk detection," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 9, no. 8, pp. 932-942, Nov. 2001.

[5] A. Tchamkerten and I. E. Telatar, "On the use of training sequences for channel estimation," *IEEE Trans. Information Theory*, vol. 52, no. 3, Mar. 2006.

[6] G. Long and F. Ling, "Fast initialization of data-driven Nyquist in-band echo cancellers," *IEEE Trans. Commun.* vol.41, no. 6, pp. 893-904, June 1993.

[7] R. D. Nowak and B. D. Van Veen, "Random and pseudorandom inputs for Volterra filter identification," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2124-35, Aug. 1994.