ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

# A Double-talk resistant echo cancellation based on iterative maximal-length correlation

*J. C. Jenq and S. F. Hsieh*
Department of Communication Engineering
National Chiao Tung University
Hsinchu, Taiwan, Republic of China
E-mail:sfhsieh@cc.nctu.edu.tw

**ABSTRACT**

A novel acoustic echo canceller for speakerphone system based upon Iterative Maximal Length Correlation (IMLC) algorithm is proposed. This algorithm is robust to double talk because of the nature of the ML sequence; it also iteratively reduces the far-end speech's effects. This simple structure was found to perform very well. We also perform a rigorous convergence analysis and derive the lower bound for the coefficients error. Computer simulations also confirm our theoretical results.

## 1. INTRODUCTION

An adaptive acoustic echo canceller (AEC) is shown in Fig. 1, where an adaptive filter is used to model the room impulse response (RIR) between the microphone and the loudspeaker. A synthesized replica of the acoustic echo is generated, and then subtracted from the echo received by the microphone.

The adaptive filter is typically implemented using a finite impulse response (FIR) or infinite impulse response (IIR) filter to account for any changes in RIR. An FIR filter is generally favored because of its stable property.

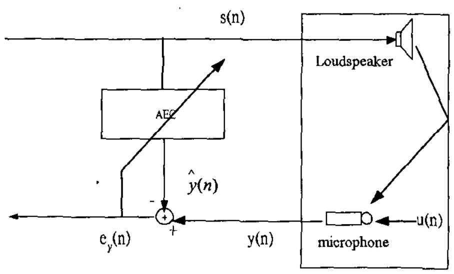


Fig. 1. An echo canceler

One of the simplest adaptation algorithms is the LMS algorithm [1]. The LMS coefficients are updated using the residue echo after cancellation. If the impulse response of the adaptive filter is the same as that of the RIR, then the echo will be canceled perfectly.

An open problem in AEC is to provide echo cancellation during double-talk. Double-talk occurs when a near-end speech source other than the loudspeaker is active within the room. In this case, the adaptive algorithm attempts to cancel both the echo and the near-end speech. This invariably leads to divergence of the adaptive filter from the optimum solution.

Current techniques attempt to effectively turn off the adaptation during double-talk [2][3][4][5]. A critical question is that merely measuring the prediction error will not discriminate between double talk and echo path changes. Thus, the AEC must react differently depending on whether double talk or a change in the echo path has occurred. Furthermore, the detection and discrimination algorithm must be fast in order to prevent the adaptive filter from being misadjusted.

Another technique that can be used to estimate RIR is the maximal length correlation (MLC) algorithm [6][7][9]. This is done by adding a maximal length sequence (MLS) $p(n)$ with length $L$ and low level $G$, to the far-end speech before it is sent into the room as in Fig.2 According to the auditory masking effect [8], The sequence $p(n)$ was masked to the user if the power ratio of the far-end speech signal to the maximum length sequence is above 15 dB.

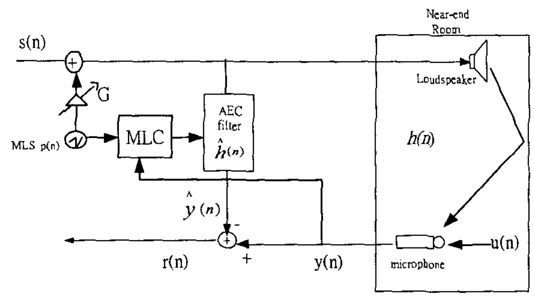


Fig.2 The MLC echo canceler

The microphone signal of the near-end room is modeled by

$$\begin{aligned} y(n) &= [s(n)+Gp(n)]*h(n)+u(n) \\ &= y_s(n)+Gy_p(n)+u(n) \qquad (1) \end{aligned}$$

where * denotes convolution, $y_s(n=s(n)*h(n)$ is the far-end echo speech, $y_p(n)=p(n)*h(n)$ is the output response of $p(n)$, $u(n)$ is the near-end signal (including the near-end speech and the near-end noise), and $G$ is a constant gain. The output of AEC filter cancels the real echo path signal $y_s(n)+Gy_p(n)$. The signal of echo path after echo cancellation is given by:

$$\begin{aligned} r(n) &= y(n)-\hat{y}(n) \\ &= y(n)-[s(n)+Gp(n)]*\hat{h}(n) \\ &= y_s(n)-\hat{y}_s(n)+Gy_p(n)-G\hat{y}_p(n)+u(n) \qquad (2) \\ &\approx u(n) \end{aligned}$$



0-7803-5482-6/99/$10.00 ©2000 IEEE

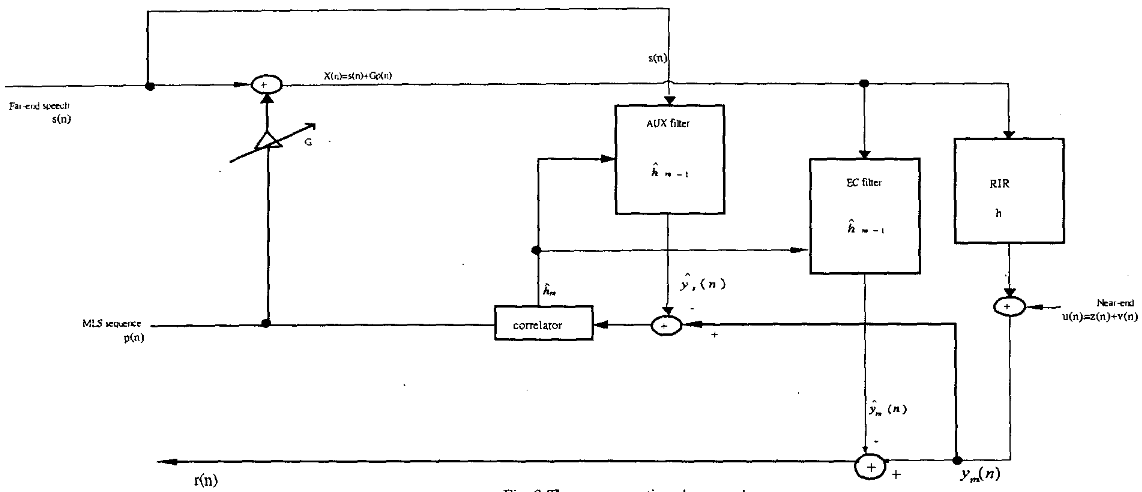


Fig. 3 The new acoustic echo canceler

where $\hat{y}_s(n)=s(n)*\hat{h}(n)$ and $\hat{y}_p(n)=p(n)*\hat{h}(n)$.

Since the MLS is continuous and independent of either the near-end or far-end speech, the RIR estimate can be obtained by cross-correlating $p(n)$ with the microphone output $y_m(n)$:

$$\hat{h}_{MLC}(n)=\frac{1}{G(L+1)}p(n)\Theta y_m(n) \qquad (3)$$

where the subscript $m$ denotes the m-th block data, e.g., $y_m(n)=y(n)$, $(m-1)L+1\le n\le mL$ represents the m-th block of $y(n)$, and so on. We will show later that the conventional MLC includes the disturbance caused by the far-end and near-end speech. Thus the MLC technique can't be used in AEC alone, even in the single-talk (far-end speech) condition.

In this paper we attempt to remove the MLC disturbance caused by the far-end speech to obtain a better estimate for the RIR $h$. Based on this idea, we propose an iterative MLC (IMLC) algorithm without the need for double-talk detection. The main idea of IMLC is that we try to cancel the far-end speech first before the RIR $h$ is estimated. We extend this idea and develop an iterative algorithm to update the AEC filter. This new simple structure is found to give excellent performance. We carry out a rigorous convergence analysis and derive the lower bound for the coefficients error. Simulation results also confirm our theoretical derivation.

## 2. NEW ACOUSTIC ECHO CANCELER

From Eq. (3) the micrphone output $y_m(n)$ can be expressed as $y_m(n)=[s_m(n)+Gp(n)]*h(n)+u_m(n)$. We can show that the conventional MLC filter coefficients error is

$$e_{MLC}(n)=h(n)-\hat{h}_{MLC}(n)$$
$$=\frac{1}{(L+1)}\sum_k h(k)-\frac{1}{G(L+1)}p(n)\Theta u_m(n)-\frac{1}{G(L+1)}p(n)\Theta h(n)\otimes s_m(n) \quad (4)$$

where $\Theta$ and $\otimes$ denote *circular correlation and convolution*, respectively:

$$p(n)\Theta y(n)=\sum_{k=1}^{L}p(n)y((k+n)_{\mathrm{mod}\ L})$$
$$p(n)\otimes y(n)=\sum_{k=1}^{L}p(n)y((n-k)_{\mathrm{mod}\ L})$$

where $(n)_{\mathrm{mod}\ L}$ is n (modulo) L. In Eq. (4), we have used an MLS property [6] :

$$p(n)\Theta p(n)=(L+1)\delta(n)-1 \qquad (5)$$

Although from Eq.(4) the MLC technique mitigates the effects of double talk when $L$ and $G$ are large enough, it still contains disturbance caused by the far-end and near-end speech signals.

The new AEC is shown in Fig.3; we try to cancel the far-end speech first before the RIR $h$ is estimated. In case of a single talk period, the IMLC method includes two steps. The first step is to get a rough estimation of the RIR $\hat{h}_1(n)$ by cross-correlation of $p(n)$ and the microphone output $y_1(n)$. The second step is to recursively compute $\hat{h}_m(n)$ as follows:

$$\hat{h}_m(n)=\frac{1}{G(L+1)}p(n)\Theta[y_m(n)-\hat{y}_{s(m-1)}(n)]$$
$$=\frac{1}{G(L+1)}p(n)\Theta[y_m(n)-s_m(n)\otimes\hat{h}_{m-1}(n)] \quad (6)$$

Based on the IMLC method, the new AEC structure has two identical filters: auxiliary (AUX) filter and the AEC filter. Because the far-end speech is readily available, we can subtract the AUX filter output $\hat{y}_{sm}$ from $y_m(n)$, so long as the AUX filter's coefficients $\hat{}$ is close to $h$, then the far-end speech will be canceled more perfectly and further iterations can improve estimation even more.

Next we will derive the convergent coefficients error $e_m(n)$ between the AEC filter and the RIR $h$, which is related to the ERLE

2

performance.

## 3. COEFFICIENTS ERROR ANALYSIS

In this section, AEC performance analysis is carried out, from which we can see the influence of each parameter. The echo return loss enhancement (ERLE) is usually used as an AEC performance measure defined as follows

$$\begin{aligned} ERLE\,(dB) &\equiv 10\log_{10}\frac{E[y(n)^2]}{E[|\,y(n)-\hat{y}(n)\,|^2]} \\ &= 10\log_{10}\frac{E([h(n)*s(n)]^2)}{E[|\,[h(n)-\hat{h}(n)]*s(n)\,|^2]} \\ &= 10\log_{10}\frac{E([h(n)*s(n)]^2)}{E[|\,[e(n)*s(n)\,|^2]} \end{aligned} \qquad (7)$$

Because the filter coefficients error *e(n)* plays an important role, we derive the IMLC filter coefficients error $e_m(n)$, after the m-th iteration. From Eq. (6), we can show that

$$\begin{aligned} &h(n)-\hat{h}_m(n) \\ &=\frac{1}{(L+1)}\sum_k h(k)-\frac{1}{G(L+1)}u_{pm}(n)-\frac{1}{G(L+1)}s_{pm}(n)\otimes[h(n)-\hat{h}_{m-1}(n)] \end{aligned} \quad (8)$$

where

$$s_{pm}(n) \equiv s_m(n)\,\Theta\,p(n) \qquad (9)$$

$$u_{pm}(n) \equiv u_m(n)\,\Theta\,p(n) \qquad (10)$$

In Eq.(8) the far end signal *s(n)* can be neglected if $\hat{h}_{m-1}$ is very close to RIR *h*. We define the filter coefficients error as

$$e_m(n) \equiv h(n) - \hat{h}_m(n) \qquad (11)$$

and

$$\xi_m(n) \equiv \frac{1}{(L+1)}\sum_k h(k) - \frac{1}{G(L+1)}u_{pm}(n) \qquad (12)$$

Using Eq. (11)(12), Eq. (8) can be written as

$$e_m(n) = \xi_m(n) - \frac{1}{G(L+1)}s_{pm}(n)\otimes e_{m-1}(n) \qquad (13)$$

We are interested in the convergence issue of $e_m(n)$. To this end, we note that the estimation error of RIR decreases monotonically to a lower bound $\xi_m(n)$. The initial state of Eq.(13) can be given as

$$\begin{aligned} e_1(n) &= \xi_1(n) - \frac{1}{G(L+1)}s_{p1}(n)\otimes[h(n)-\hat{h}_0(n)] \\ &= \xi_1(n) - \frac{1}{G(L+1)}s_{p1}(n)\otimes h(n) \end{aligned} \qquad (14)$$

where $\hat{h}_0[n] = 0$ .

For convenience, Eq.(13) is written in a vector form:

$$\mathbf{e}_m = \boldsymbol{\xi}_m - \frac{1}{G(L+1)}\mathbf{S}_{pm}\mathbf{e}_{m-1} \qquad (15)$$

where

$$\mathbf{e}_m = \begin{bmatrix} e_m(1) \\ e_m(2) \\ \cdot \\ \cdot \\ e_m(M) \end{bmatrix} \qquad \boldsymbol{\xi}_m = \begin{bmatrix} \xi_m(1) \\ \xi_m(2) \\ \cdot \\ \cdot \\ \xi_m(M) \end{bmatrix}$$

$$\mathbf{S}_{pm} = \begin{bmatrix} s_{pm}(1) & s_{pm}(0) & s_{pm}(-1) & \cdot\ \cdot & s_{pm}(-M+2) \\ s_{pm}(2) & s_{pm}(1) & s_{pm}(0) & \cdot\ \cdot & s_{pm}(-M+3) \\ s_{pm}(3) & s_{pm}(2) & s_{pm}(1) & \cdot\ \cdot & s_{pm}(-M+4) \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ s_{pm}(M) & s_{pm}(M-1) & s_{pm}(M-2) & \cdot\ \cdot & s_{pm}(1) \end{bmatrix}$$

and $M$ is the order of the AEC filter. Now, by recursive substitution in Eq.(15) and proper expansion, we may express Eq.(15) as follows

$$\begin{aligned} \mathbf{e}_m = [\mathbf{I} &+ \frac{-1}{G(L+1)}\mathbf{S}_{pm} + (\frac{-1}{G(L+1)}\mathbf{S}_{pm})^2 + (\frac{-1}{G(L+1)}\mathbf{S}_{pm})^3 - \cdots \\ &+ (\frac{-1}{G(L+1)}\mathbf{S}_{pm})^{m-2}]\boldsymbol{\xi}_m + (\frac{-1}{G(L+1)}\mathbf{S}_{pm})^{m-1}\mathbf{e}_1 \end{aligned} \qquad (16)$$

A well known lemma from Linear Algebra is quoted as follows:

[Lemma]: If the eigenvalues $\lambda_1$, $\lambda_2 \ldots$ $\lambda_n$ of a matrix **A** satisfy the conditions $|\lambda_j| < 1$ for j=1,2,...,n, then we deduce the binomial theorem that

$$\begin{aligned} \sum_{p=0}^{m}\mathbf{A}^p &= [\mathbf{I}+\mathbf{A}]^{-1}[\mathbf{I}-\mathbf{A}^{m+1}] \\ &= [\mathbf{I}+\mathbf{A}]^{-1},\ m \to \infty \end{aligned} \qquad (17)$$

According to this lemma, we may rewrite Eq. (16) in the desired form:

$$\mathbf{e}_m = [\mathbf{I}+\frac{1}{G(L+1)}\mathbf{S}_{pm}]^{-1}[\mathbf{I}-(\frac{-1}{G(L+1)}\mathbf{S}_{pm})^{m-1}]\boldsymbol{\xi}_m + (\frac{-1}{G(L+1)}\mathbf{S}_{pm})^{m-1}\mathbf{e}_1 \quad (18)$$

If the eigenvalues of $\mathbf{S}_{pm}$ are $\lambda_{s1}$, $\lambda_{s2} \ldots$ $\lambda_{sn}$, it follows that the necessary and sufficient condition for $\mathbf{e}_m$ to converge is that

$$\left|\frac{\lambda_{max}}{G(L+1)}\right| < 1 \qquad (19)$$

where $\lambda_{max}$ is the largest eigenvalue of the matrix $\mathbf{S}_{pm}$ (for all m) When Eq. (18) converges, that is, $\lambda_{max}$ is within the bounds defined by Eq. (19), we finally have, irrespective of the initial conditions,

$$\mathbf{e}_m = [\mathbf{I} + \frac{1}{G(L+1)}\mathbf{s}_{pm}]^{-1}\boldsymbol{\xi}_m \qquad (20)$$

For large values of *G(L+1)*, the coefficients error $\mathbf{e}_m$ will converge quickly to the lower bound

$$\mathbf{e}_{\infty} \approx \xi_{\infty} \qquad (21)$$

From Eq.(12), we have

$$e_{\infty}(n) = \xi_{\infty}(n)$$
$$= \frac{1}{L+1}\sum_{k} h(k) - \frac{1}{G(L+1)} u_{p\infty}[n] \quad (22)$$

From Eq.(22) we find two interesting properties of the IMLC structure. The first one, when the number of iterations $m$ is large enough the filter coefficient are no longer disturbed by the far-end speech. In another word, the IMLC structure cancels the acoustic echo perfectly during the single-talk period. The second one, better double-talk suppression can be achieved by increasing the magnitude $G$ or the length L of the MLS; however large $G$ is too noisy in auditory and longer length of MLS needs longer time to converge and compute.

### 4. COMPUTER SIMULATION

The performance of the new AEC algorithm is investigated by extensive computer simulations. The ERLE defined in Eq. (7), is used as the criterion. We also investigate the effects of different parameters on the convergence behavior as well as on the tracking characteristics of the new algorithm.

For simplicity, assume the RIR $h$ to be identified is a 100-tap FIR filter obtained by down sampling and truncating an impulse response measured in a real room. The near-end signal is modeled by $u(n)=z(n)+v(n)$ where $z(n)$ is the near-end speech and $v(n)$ is the near-end background white Gaussian noise with unit-variance (SNR 40dB). The far-end signal $s(n)$ and the near-end signal $z(n)$ use real speech data, $p(n)$ is MLS with amplitude $-1$ to $+1$.

In Fig.4, under single-talk, the simulation results confirm our convergence analysis from Eq.(22) (to estimate the lower bound of converged coefficients error denoted by *), when the number of iterations $m$ is more than 4. We also compare the conventional MLC (denoted by o) with our proposed IMLC algorithms using $m$ iterations and different gains $G$. It is expected that with same length $L$, increasing the magnitude $G$ achieves better ERLE performance.

In Fig.5, we compare the performance and converging rate of the NLMS and IMLC algorithms in the cases of single-talk and double-talk, which shows that the IMLC algorithm provides good echo cancellation without double-talk detection. In the period of double-talk (from n=3000 to n=7000), both NLMS and IMLC spend about 8000 samples to converge. As to computation complexity, the NLMS needs about 5x100 additions and 3x100 multiplications and the IMLC needs about 4096x100 additions and 100 multiplications

### 5. CONCLUSION

We propose the IMLC algorithm without double-talk detection to track and estimate the echo path. Our preliminary work indicates that the AEC algorithm can maintain a superior ERLE value even during double-talk periods. A convergence analysis and the lower bound of the error coefficients are also given. Although a few iterations are required to reduce the far-end speech effects, further investigation is still necessary for real time applications. Performance of the structure is justified by extensive computer simulations.

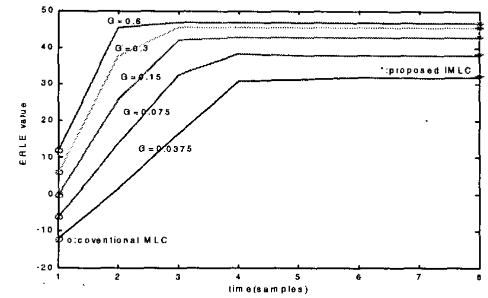


Fig.4. Convergence characteristic of different gains G (MLS p(n) has length 4095).

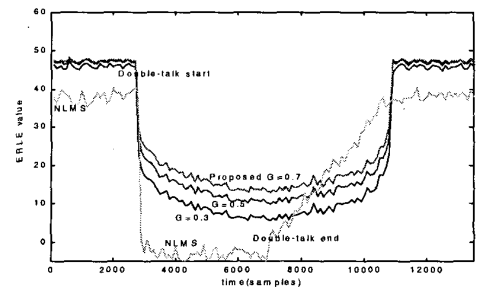


Fig.5 ERLE comparisons of NLMS and IMLC algorithms with MLS length 4095 in single-talk and double-talk.

### REFERENCES


[1] S. Haykin, *Adaptive filter theory*, Printice-Hall, Englewood Cliffs, 1986

[2] O. A. Horna, "Echo canceller with adaptive transversal filter utilizing pseudo-logarithmic coding", *COMSAT Tech. Rev.*, vol. 7, no. 2, pp. 393-427, 1977.

[3] N. Furuya *et al.*, "High performance custom VLSI echo canceler," in *ICC'85 Rec.*, 1985, pp 1470-1476.

[4] S. Minami, *et al.*, "A double-talk detection method for an echo canceller," in *ICC'85 Rec.*, 1985, pp.1492-1497.

[5] K. Ochiai *et al.*, "Echo canceller with two echo path model," *IEEE Trans. Commun.*, vol. COM-25, June 1977.

[6] M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Am.*, vol. 66 (1979 Aug).

[7] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc.*, vol. 31, pp. 478-488 (1983 July Aug).

[8] M J. Crocker, *Handbook of Acoustics*, Wiley & Sons, INC, 1998.

[9] J. F. Doherty and R Porayath., "A robust echo canceler for acoustic environments," *IEEE Trans on circuit and systems-11.*, vol 44. no. 5, may 1997, pp389-39.