Acoustic Echo Cancellation Using Iterative-Maximal-Length Correlation and Double-Talk Detection

Jang-Chyuan Jenq and Shih-Fu Hsieh

***Abstract*—The conventional maximal-length-correlation (MLC) algorithm to estimate room impulse response for adaptive echo cancellation (AEC) is disturbed by both far-end and near-end speeches. In this paper, a new iterative-maximal-length-correlation (IMLC) algorithm is proposed to reduce the far-end speech interference. To avoid the near-end interference, a new double-talk detection (DTD) method is proposed by tracking the squared coefficients errors of the AEC filter. This DTD method has well-separated detection margins among single-talk (ST), double-talk (DT), and echo path changes. Statistical analysis and computer simulations confirm that our proposed IMLC-DTD algorithm outperforms conventional methods.**

***Index Terms*—Acoustic echo cancellation, adaptive filters, double talk detection, echo suppression, maximal length sequence.**

I. Introduction

HANDS-FREE conversation is popular in various fields of communication such as teleconferencing, video conferencing and mobile radio telephone. However, in those applications the presence of coupling from the loudspeakers to the microphone would result in undesired acoustic echo and significantly degrade the speech quality. Therefore, an effective adaptive echo canceler (AEC) is required [1]. A typical AEC is shown in Fig. 1: where the coefficients $\hat{h}(n)$ of the AEC filter are used to model the room impulse response (RIR) $h(n)$ between the microphone and the loudspeaker. If a far-end speech $s(n)$ is sent into the near-end room, a synthesized echo replica speech $\hat{y}_s(n) = s(n) * \hat{h}(n)$ is generated by the AEC filter, ($*$ denotes linear convolution) and subtracted from the microphone signal $y(n)$. The residue echo signal $\varepsilon(n)$ in the return path is given by

$$\begin{aligned}\varepsilon(n) &= y(n) - \hat{y}_s(n)\\ &= s(n) * \left[h(n) - \hat{h}(n)\right] + u(n).\end{aligned} \tag{1}$$

Equation (1) shows that if the filter coefficients $\hat{h}(n)$ are the same as that of the RIR $h(n)$, then the echo will be canceled perfectly and there remains only the near-end signal $u(n)$, which includes a near-end speech $z(n)$ and a background noise $v(n)$, in the return path.

Manuscript received July 24, 2000; revised August 9, 2001. This work was supported in part by the National Science Council under Grant NSC 90-2213-E-009-112. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael S. Brandstein.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: ojjc@ms17.hinet.net; sfhsieh@cc.nctu.edu.tw).

Publisher Item Identifier S 1063-6676(01)09669-9.

Recently, several AEC filter techniques have been proposed [2]. They are typically implemented using a finite impulse response (FIR) filter for stability reason. The most popular and computationally efficient adaptive algorithms are the LMS algorithms [3]. However, all existing adaptive AEC filters share serious problems during "double-talk" (DT) when simultaneous talks occurs for both near-end and far-end speakers. In this situation, the microphone signal $y(n)$ includes near-end signal $u(n)$ which acts like a large disturbing noise to the residue echo signal $\varepsilon(n)$ and the filter coefficients will be greatly disturbed. Therefore, the echo cannot be canceled any more and will become intolerable echo.

To overcome the DT problem, almost all of current techniques attempt to effectively turn off adaptation during DT [6], [9], [10], [14]. However, a critical question is that merely measuring the residue echo cannot discriminate between DT and echo path changes. Furthermore, the detection and discrimination algorithm must be fast in order to prevent the adaptive filter from being misadjusted.

Another technique which is more robust to estimate RIR $h(n)$, during DT situation, is the maximal length correlation (MLC) algorithm [7], [8]. The basic MLC method is done by sending a periodic maximal-length sequence (MLS) $p(n)$ with period L to the loudspeaker [$s(n)$ is set to zero] and cross-correlate the microphone signal $y(n)$ with $p(n)$. We can estimate RIR $h(n)$ as follows:

$$\hat{h}(n) = \frac{1}{(L+1)}\, p(n) \copyright y(n), \qquad 1 \le n \le M \tag{2}$$

where © denotes *correlation*: $p(n) \copyright y(n) = \sum_{k=0}^{L-1} p(k)y(k-n)$; M is the order of the AEC filter and the microphone signal can be expressed as $y(n) = p(n) * h(n) + u(n)$. We can rewrite (2) as

$$\begin{aligned}\hat{h}(n) &= \frac{1}{(L+1)}\, p(n) \copyright [p(n) * h(n) + u(n)]\\ &= \frac{1}{(L+1)}\, [p(n) \copyright p(n)] * h(n) + \frac{1}{(L+1)}\, p(n) \copyright u(n).\end{aligned} \tag{3}$$

In (3), we have used the associative law between correlation © and convolution $*$: $p(n) \copyright [p(n) * h(n)] = [p(n) \copyright p(n)] * h(n)$. This is because of $p(n) \copyright y(n) = p(n) * y(-n)$ and the convolution operation satisfies the associative law. Since the MLS $p(n)$

1063–6676/01$10.00 © 2001 IEEE

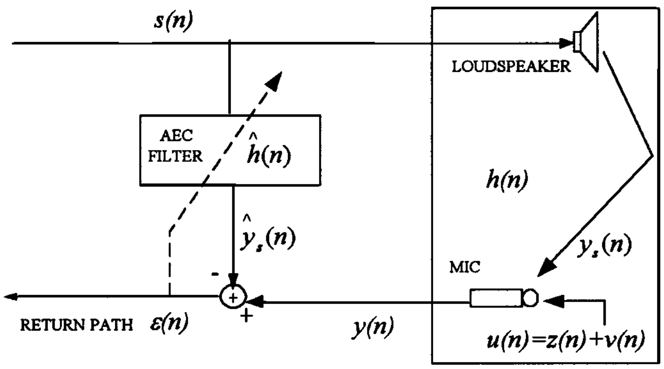

Fig. 1. Echo canceler.

is a pseudo random noise with magnitude ± 1, and its auto-correlation is nearly a delta function [8], we have $p(n)©p(n) = (L+1)\delta(n) - 1$. Equation (3) can be rewritten as

$$\begin{aligned}\hat{h}(n) &= \frac{1}{(L+1)}[(L+1)\delta(n) - 1] * h(n) \\ &\quad + \frac{1}{(L+1)} p(n)©u(n) \\ &= h(n) - \frac{1}{(L+1)}\sum_{k=1}^{M} h(k) + \frac{1}{(L+1)} p(n)©u(n) \\ &\approx h(n). \end{aligned} \tag{4}$$

In (4), the second term $(1/(L+1))\sum_{k=1}^{M} h(k)$ is the average of RIR $h(n)$ which can be neglected if $h(n)$ has no DC term, and the third term can be very small when L is large and $p(n)$ and $u(n)$ are uncorrelated. Thus we can find that during DT, the MLC method with large L is robust to estimate $h(n)$. In addition, because $p(n) \in \{1, -1\}$, we find the correlation in (2) only needs additions without multiplications. However, the conventional MLC method [7], [8] requires injecting the training sequence $p(n)$ to the near-end room before talking. This is too noisy for human hearing.

To overcome the noisy problem in MLC and maintain a robust estimation for RIR $h(n)$ during DT, Doherty, *et al.* [11] suggest adding a low level MLS to the far-end speech $s(n)$ before sending to the loudspeaker and then estimating $h(n)$ by the MLC method. According to the auditory masking effect [12], the MLS could be masked to the user if the power ratio of the far-end speech signal to the MLS is above 15 dB. This masked MLC structure is shown in Fig. 2. Notice that in the left top corner, MLS $p(n)$ with period L and magnitude controlled by gain G, is added to the far-end speech and the far-end signal becomes $x(n) = s(n) + Gp(n)$. Now the far-end speech $s(n)$ is equivalent to a disturbance noise when the MLC method is used to estimate RIR $h(n)$. Next, we want to find the estimation filter error of the MLC algorithm.

Since the MLC method estimates $h(n)$ from each period (length $=$ L) of the input MLS $p(n)$, we can represent the microphone signal by sums of shifted finite-length blocks of length L respectively: $y(n) = \sum_{m=0}^{\infty} y_m(n - mL)$, where $y_m(n) = y(n)|_{mL+1 \le n \le (m+1)L}$. The mth MLC estimate $\hat{h}_m(n)$ can be written as

$$\hat{h}_m(n) = \frac{1}{G(L+1)} p(n)©y_m(n), \qquad 1 \le n \le M. \tag{5}$$

Note that the mth block microphone signal $y_m(n)$ can be written by

$$y_m(n) = \{u(n) + [s(n) + Gp(n)] * h(n)\}|_{mL+1 \le n \le (m+1)L}. \tag{6}$$

Substituting (6) into (5), we get

$$\begin{aligned}\hat{h}_m(n) &= h(n) + \frac{1}{G(L+1)} p(n)©u(n) - \frac{1}{(L+1)}\sum_{k=1}^{M} h(k) \\ &\quad + \frac{1}{G(L+1)}[p(n)©s(n)] * h(n) \\ &= h(n) + I_{N,m}(n) + I_{F,m}(n) * h(n) \end{aligned} \tag{7}$$

where the near-end interference $I_{N,m}(n)$ and far-end interference $I_{F,m}(n)$ are defined as

$$\begin{aligned} I_N(n) &= \frac{1}{G(L+1)} p(n)©u(n) - \frac{1}{(L+1)}\sum_{k=1}^{M} h(k) \\ I_F(n) &= \frac{1}{G(L+1)} p(n)©s(n) \\ I_{N,m}(n) &= I_N(n)|_{mL+1 \le n \le (m+1)L} \\ I_{F,m}(n) &= I_F(n)|_{mL+1 \le n \le (m+1)L}. \end{aligned} \tag{8}$$

Notice that in (7) the near-end and far-end interferences are equivalent to the estimation errors. We can define the MLC coefficient error as

$$\begin{aligned} e_m(n) &\equiv \hat{h}_m(n) - h(n) \\ &= I_{N,m}(n) + I_{F,m}(n) * h(n). \end{aligned} \tag{9}$$

In (8), we find that better speech suppression can be achieved by increasing the magnitude G or the length L of the MLS; however, large G is too noisy for hearing and longer length of MLS needs longer time to converge and compute. Moreover, although the conventional MLC method [9] may suppress the DT effect, it still includes the far-end and near-end speeches disturbances that cannot be defeated and the echo cancellation performance is poor.

The main purpose in this paper is to find a way to overcome the MLC problem, i.e., to reduce the disturbances caused by the far-end and near-end speeches. In Section II, we propose an iterative MLC (IMLC) method [13] to reduce the far-end speech interference. This basic idea is similar in [17]. Moreover, because the near-end speech, in case of DT, cannot be removed, we will propose a new DT detection mechanism in Section III, to effectly distinguish between DT and echo path changes to prevent the AEC filter from misadjustment. This detection is based on comparison of the squared-coefficient error between two consecutive IMLC estimates. In Section IV, computer simulation and comparisons will be presented to support our analysis.

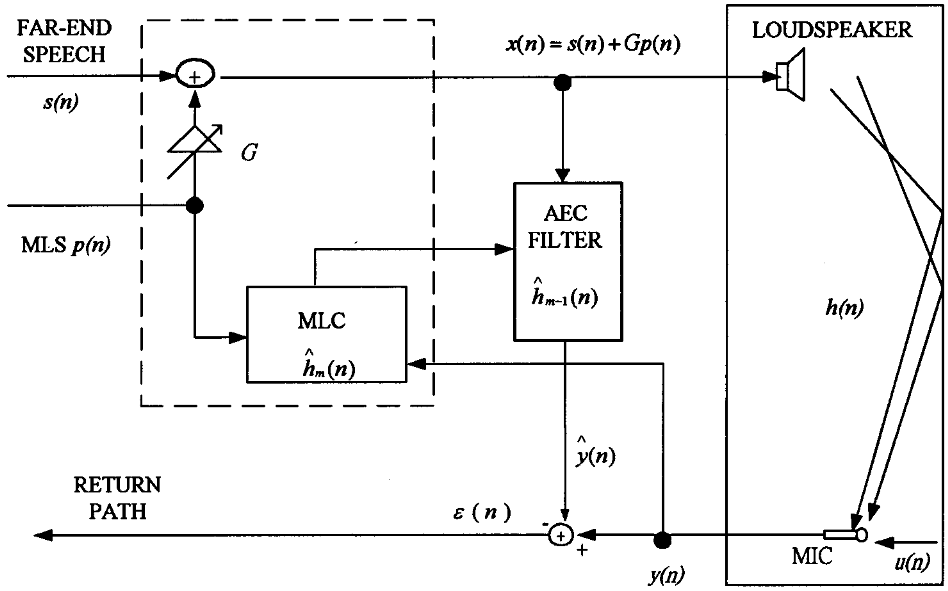

Fig. 2. AEC block digram with the masked MLC algorithm.

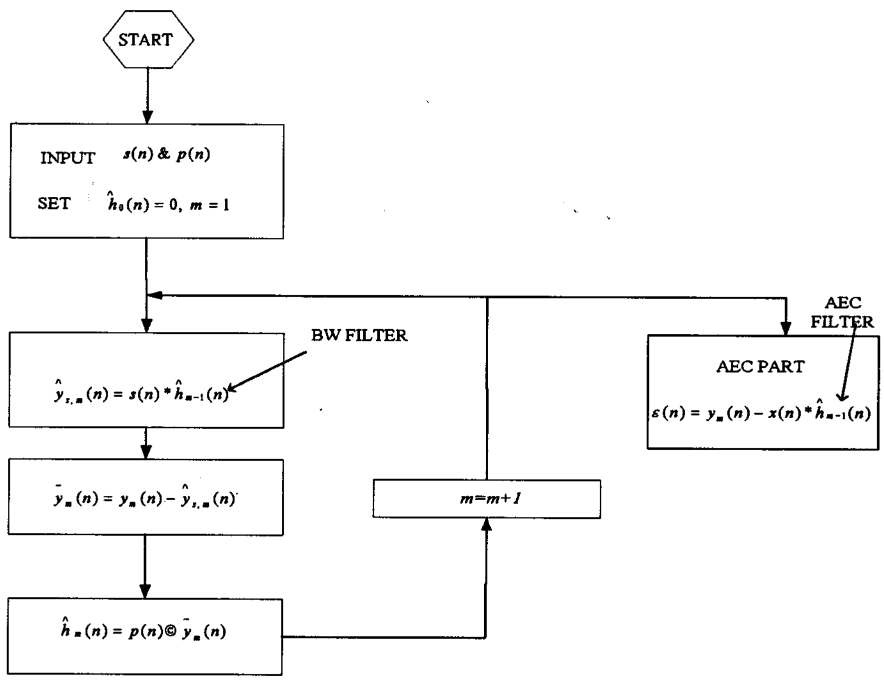

Fig. 3. Flowchart of the IMLC algorithm.

II. IMLC AEC Structure

A. IMLC Algorithm

In this section, we attempt to remove the MLC disturbance caused by the far-end speech to obtain a better estimate for the RIR $h(n)$. We propose an IMLC algorithm by inserting a backward (BW) filter $\hat{h}_{BW}(n)$ to estimate the far-end speech response $\hat{y}_s(n) = \hat{h}_{BW}(n) * s(n)$ and cancel its disturbance. This algorithm includes five steps:

1) set both BW and AEC filters coefficients $\{\hat{h}_0(n)\} = \{0\}$ and use the conventional MLC algorithm to obtain an initial estimate of the RIR $\hat{h}_1(n) = (1/(G(L+1)))p(n) © y_1(n)$, where $y_1(n)$ is the first block data of microphone output;
2) estimate the response of the (second block of) far-end speech by the BW filter: $\hat{y}_{s,2}(n) \approx s(n) * \hat{h}1(n)$;
3) cancel the $\hat{y}_{s,2}(n)$ signal from the second block data of the microphone output $y_2(n)$, thus $\tilde{y}_2(n) = y_2(n) - \hat{y}_{s,2}(n)$;
4) estimate $\hat{h}_2(n) = (1/(G(L+1)))p(n) © \tilde{y}_2(n)$ and copy it to BW and AEC filter;
5) the last step goes to the first step and recursively compute $\hat{h}_m(n) = (1/(G(L+1)))p(n) © \tilde{y}_m(n)$, $m = 2, 3, 4 \ldots$.

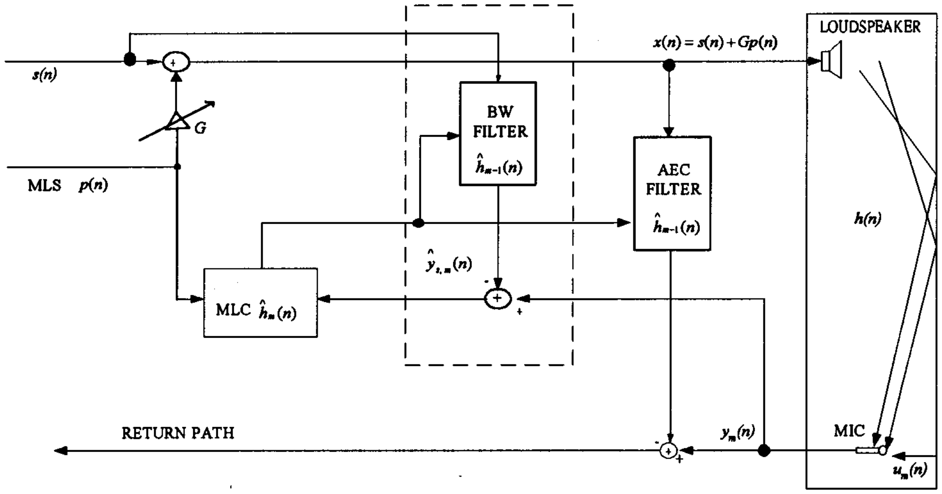

Fig. 4. IMLC AEC block diagram.

These steps are expressed by the flow chart shown in Fig. 3. We extend this algorithm to develop an IMLC AEC structure shown in Fig. 4.

The new AEC structure has two identical filters: backward (BW) and AEC filters. Because the far-end speech is readily available, we can subtract the BW filter output $\hat{y}_{s,m}(n)$ from $y_m(n)$ before using the MLC algorithm to estimate $h(n)$. The mth estimated coefficients $\hat{h}_m(n)$ by the IMLC method is expressed as

$$\hat{h}_m(n) = \frac{1}{G(L+1)} p(n) \copyright [y_m(n) - \hat{y}_{s,m}(n)]. \tag{10}$$

Comparing (5) and (10), because the far-end speech disturbance $\hat{y}_s(n) = \hat{h}_{BW}(n) * s(n)$ is estimated by the BW filter and canceled from $y_m(n)$, we can obtain a better estimate for the RIR $h(n)$. Next we give convergence analysis for squared coefficient error of IMLC algorithm.

B. Estimation Error

Now we will show that the IMLC estimation coefficient error can be reduced by further iterations. Follow the same process of MLC method in (6)–(8)

$$\hat{h}_m(n) = h(n) + I_{N,m}(n) - I_{F,m}(n) * [\hat{h}_{m-1}(n) - h(n)]. \tag{11}$$

In (11), the third term is the disturbance due to the far-end speech. So long as the BW filter's coefficients $\hat{h}_{m-1}(n)$ are close to $h(n)$, the far-end speech will be canceled more perfectly and further iterations can improve estimation even better. By definition in (9), the IMLC coefficient error becomes

$$e_m(n) = I_{N,m}(n) - I_{F,m}(n) * e_{m-1}(n) \tag{12}$$

and we have the initial state $e_0(n) \equiv \hat{h}_0(n) - h(n)$ where $\{\hat{h}_0(n)\} = \{0\}$. Now, we are interested in the convergence issue of squared coefficient error $\sum_{n=1}^{M} e_m^2(n) = \|\mathbf{e}_m\|^2$.

In (12), from the convolution property we have

$$\begin{aligned}
&e_{m-1}(1)\\
&\quad = I_{N,m}(1) - [I_{F,m}(0)e_{m-1}(1) + I_{F,m}(-1)e_{m-1}(2)\\
&\qquad\qquad + \cdots I_{F,m}(1-M)e_{m-1}(M)\\
&e_{m-1}(2)\\
&\quad = I_{N,m}(2) - [I_{F,m}(1)e_{m-1}(1) + I_{F,m}(2-2)e_{m-1}(2)\\
&\qquad\qquad + \cdots I_{F,m}(2-M)e_{m-1}(M)\\
&\vdots\\
&e_{m-1}(M)\\
&\quad = I_{N,m}(M) - [I_{F,m}(M-1)e_{m-1}(1)\\
&\qquad\qquad + I_{F,m}(M-2)e_{m-1}(2)\\
&\qquad\qquad + \cdots I_{F,m}(0)e_{m-1}(M).
\end{aligned} \tag{13}$$

By definition, in (8) we have $I_{F,m}(0) = I_{F,m-1}(L) \cdots I_{F,m}(-k) = I_{F,m-1}(L-k)$. For simplicity, (13) is written in a vector form as

$$\mathbf{e}_m = \mathbf{I}_{N,m} - \boldsymbol{\Psi}_m \mathbf{e}_{m-1} \tag{14}$$

where the coefficients error and the near-end interference vector are defined as

$$\mathbf{e}_m = \begin{bmatrix} e_m(1) \\ e_m(2) \\ \vdots \\ e_m(M) \end{bmatrix} \qquad \mathbf{I}_{N,m} = \begin{bmatrix} I_{N,m}(1) \\ I_{N,m}(2) \\ \vdots \\ I_{N,m}(M) \end{bmatrix}$$

and the far-end interference matrix is defined as

$$\boldsymbol{\Psi}_m = \begin{bmatrix} I_{F,m-1}(L) & I_{F,m-1}(L-1) & \cdots & I_{F,m-1}(L-M+1) \\ I_{F,m}(1) & I_{F,m-1}(L) & \cdots & I_{F,m-1}(L-M+2) \\ \vdots & \vdots & \cdots & \vdots \\ I_{F,m}(M-1) & I_{F,m}(M-2) & \cdots & I_{F,m-1}(L) \end{bmatrix}.$$

Now, by recursive substitution, in (14) and proper expansion, we may express (14) as follows:

$$\begin{aligned}\mathbf{e}_m =& \mathbf{I}_{N,m} + (-1)^1\mathbf{\Psi}_m\mathbf{I}_{N,m-1} + \cdots \\ &+ (-1)^{m-1}(\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\cdots\mathbf{\Psi}_2)\mathbf{I}_{N,1} \\ &+ (-1)^m(\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\cdots\mathbf{\Psi}_1)\mathbf{e}_0 \end{aligned} \tag{15}$$

and $\|\mathbf{e}_m\|^2$ can be expressed as

$$\begin{aligned}\|\mathbf{e}_m\|^2 =& \mathbf{I}^t_{N,m}\mathbf{I}_{N,m} + \mathbf{I}^t_{N,m-1}(\mathbf{\Psi}^t_m\mathbf{\Psi}_m)\mathbf{I}_{N,m-1} \\ &+ \mathbf{I}^t_{N,m-2}(\mathbf{\Psi}^t_{m-1}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{\Psi}_{m-1})\mathbf{I}_{N,m-2} + \cdots \\ &+ \mathbf{I}^t_{N,1}(\mathbf{\Psi}^t_2\cdots\mathbf{\Psi}^t_{m-1}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\cdots\mathbf{\Psi}_2)\mathbf{I}_{N,1} \\ &+ \mathbf{e}^t_0(\mathbf{\Psi}^t_1\cdots\mathbf{\Psi}^t_{m-1}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\cdots\mathbf{\Psi}_1)\mathbf{e}_0. \end{aligned} \tag{16}$$

In (16), we have assumed that different block of near-end interference vectors satisfy $\mathbf{I}^t_{N,i}\mathbf{D}\mathbf{I}_{N,j} = 0,\ i \neq j$ where $\mathbf{D}$ is a square matrix obtained by the product of the far-end interference matrices $\{\cdots\mathbf{\Psi}^t_{k-1}\mathbf{\Psi}^t_k\mathbf{\Psi}_k\mathbf{\Psi}_{k-1}\cdots\}$. This is because $\mathbf{I}_{N,j}$ (or $\mathbf{D}$) is regarded as the vector of the near-end signal $u(n)$ [or the matrix of the far-end signal $s(n)$] spread by the MLS $p(n)$. Note that in (16), $\mathbf{\Psi}^t_k\mathbf{\Psi}_k$ is a symmetric matrix and assume the smallest eigenvalue $\lambda_{\min} = \min_k \lambda(\mathbf{\Psi}^t_k\mathbf{\Psi}_k)$ and largest eigenvalue $\lambda_{\max} = \max_k \lambda(\mathbf{\Psi}^t_k\mathbf{\Psi}_k)$. From the quadratic form property we have

$$\begin{aligned}\lambda_{\min}\|\mathbf{I}_{N,i}\|^2 \leq& \mathbf{I}^t_{N,i}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{I}_{N,i} \leq \lambda_{\max}\|\mathbf{I}_{N,i}\|^2 \\ \lambda^2_{\min}\|\mathbf{I}_{N,i}\|^2 \leq& \mathbf{I}^t_{N,i}\mathbf{\Psi}^t_{m-1}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\mathbf{I}_{N,i} \\ \leq& \lambda^2_{\max}\|\mathbf{I}_{N,i}\|^2 \\ &\vdots \\ \lambda^{m-1}_{\min}\|\mathbf{I}_{N,i}\|^2 \leq& \mathbf{I}^t_{N,i}\mathbf{\Psi}^t_2\cdots\mathbf{\Psi}^t_{m-1}\mathbf{\Psi}^t_m\mathbf{\Psi}_m\mathbf{\Psi}_{m-1}\cdots\mathbf{\Psi}_2\mathbf{I}_{N,i} \\ \leq& \lambda^{m-1}_{\max}\|\mathbf{I}_{N,i}\|^2 \end{aligned} \tag{17}$$

where $\mathbf{I}_{N,i}$ defined in (8) is regarded as the near-end signal $u(n)$ spread by $p(n)$ and suppressed by the factor $1/(G(L+1))$. When $G(L+1)$ is large, difference of $\|\mathbf{I}_{N,i}\|^2$ in consecutive blocks is reduced, too. In our analysis we assume it is the same for all i. From (16) in (17) we have

$$\begin{aligned}&\sum_{k=0}^{m-1}\lambda^k_{\min}\|\mathbf{I}_{N,m}\|^2 + \lambda^m_{\min}\|\mathbf{e}_0\|^2 \\ &\quad \leq \|\mathbf{e}_m\|^2 \leq \sum_{k=0}^{m-1}\lambda^k_{\max}\|\mathbf{I}_{N,m}\|^2 + \lambda^m_{\max}\|\mathbf{e}_0\|^2. \end{aligned} \tag{18}$$

When (18) satisfies the convergence condition, $|\lambda_{\max}| < 1$, we may express it as

$$\frac{1-\lambda^m_{\min}}{1-\lambda_{\min}}\|\mathbf{I}_{N,m}\|^2 \leq \|\mathbf{e}_m\|^2 \leq \frac{1-\lambda^m_{\max}}{1-\lambda_{\max}}\|\mathbf{I}_{N,m}\|^2 \tag{19}$$

and we find that the error $\mathbf{e}_0$ due to the first iteration decays after few iterations. Thus, the error propagation is not a serious problem and will be confirmed by our simulations. Because elements of the near-end interference matrix $\mathbf{\Psi}_m$ are $I_{F,m}(n) = (1/(G(L+1)))p(n)©s(n)$, the maximum eigenvalue $\lambda_{\max}$ of $\mathbf{\Psi}^t_m\mathbf{\Psi}_m$ has the factor $1/(G^2(L+1)^2)$. When $G(L+1)$ is large, we may well have $\lambda_{\max} \ll 1$ and (19) becomes

$$\|\mathbf{e}_m\|^2 \approx \|\mathbf{I}_{N,m}\|^2. \tag{20}$$

In (20), we find that the estimation coefficient error of the IMLC method is merely caused by the near-end signal. In (8), because the near-end signal $u(n)$ includes the near-end speech $z(n)$ and the background noise $v(n)$, we have $u(n) = z(n) + v(n)$. To compare the coefficient error between MLC in (7), and the IMLC method in (20), let H_{st} and H_{dt} represent the hypotheses of single-talk (ST) $(z(n) = 0)$ and DT, respectively, we have

$$H_{dt}: \begin{cases} e_{MLC,m}(n) = \dfrac{1}{G(L+1)}p(n)©[z(n)+v(n)] \\ \quad -\dfrac{1}{(L+1)}\displaystyle\sum_{k=1}^{M}h(k) + \dfrac{1}{G(L+1)}p(n)©s(n)*h(n) \\ e_{IMLC,m}(n) = \dfrac{1}{G(L+1)}p(n)©[z(n)+v(n)] \\ \quad -\dfrac{1}{(L+1)}\displaystyle\sum_{k=1}^{M}h(k) \end{cases} \tag{21}$$

$$H_{st}: \begin{cases} e_{MLC,m}(n) = \dfrac{1}{G(L+1)}p(n)©v(n) \\ \quad -\dfrac{1}{(L+1)}\displaystyle\sum_{k=1}^{M}h(k) + \dfrac{1}{G(L+1)}p(n)©s(n)*h(n) \\ e_{IMLC,m}(n) = \dfrac{1}{G(L+1)}p(n)©v(n) \\ \quad -\dfrac{1}{(L+1)}\displaystyle\sum_{k=1}^{M}h(k). \end{cases} \tag{22}$$

From (21) and (22), we find that when the number of iterations m is large enough, (normally, $m > 5$), the IMLC method is no longer disturbed by the far-end speech. In another word, during DT situation, the MLC coefficients errors are disturbed by both near-end and far-end interferences. However, the IMLC method only has near-end interferences. In a ST situation, the MLC method still has far-end interference, but the IMLC method could estimate coefficients more perfectly. This is the main reason why we propose this IMLC method.

Although IMLC outperforms MLC during ST by effectively removing the far-end interference, the near-end signal $u(n)$ during DT is still troublesome. In next section, based on the IMLC structure, we propose a DT detection mechanism, which compares the squared errors of the filter coefficients to effectively distinguish between DT and echo path changes and control AEC filter updates properly.

III. AEC Structure With Doubtalk Detection

A. Conventional DTD AEC Structure

There are a number of DT detectors. One is the cross-correlation method [15], [16]. If the filter has converged to its optimal solution, the detection is accomplished by observing whether the input signal $x(n)$ is orthogonal to the residue error

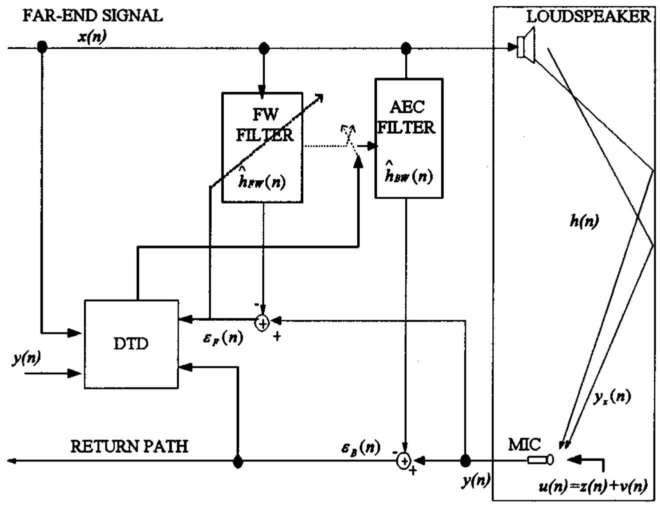

Fig. 5. Conventional DTD AEC structure.

$\varepsilon(n)$ or not. By the orthogonal principle they will be orthogonal to each other in case of ST. However, in case of DT, the residue error $\varepsilon(n)$ gets larger abruptly and they are not orthogonal. But if the echo path changes, they are also neither orthogonal and detection error may arise. Another one is the level-comparison method [4]–[7], [14] by observing the power levels of the far-end signal $x(n)$, the microphone signal $y(n)$, and the residue error $\varepsilon(n)$. One of the popular structures of AEC, with level-comparison DT detection, is depicted in Fig. 5.

The structure has two separate FIR filters, one (forward FW) for adaptively identifying RIR $h(n)$ and the other (AEC) for synthesizing the echo replica to cancel out the echo. The coefficients of the AEC filter are refreshed by the adaptive FW filter, only when the latter gives a better estimation of RIR $h(n)$ than the former. This is done by the following procedures. First, we observe the input data $x(n)$ block by block, with $x_m(n) = x(n)|_{mL+1 \le n \le (m+1)L}$ denoting the mth block of $x(n)$, and calculate the echo return loss (ERL) [6] as $ERL_{F,m} = ||\mathbf{x}_m||^2/||\boldsymbol{\varepsilon}_F||^2$ and $ERL_{B,m} = ||\mathbf{x}_m||^2/||\boldsymbol{\varepsilon}_B||^2$ where $\boldsymbol{\varepsilon}_F$ and $\boldsymbol{\varepsilon}_B$ denote the residue error signals from FW and AEC filters, respectively. Second, if $ERL_{F,m} > ERL_{B,m}$, which is the converging situation during ST, then $\hat{h}_{FW,m}(n)$ is much closer to RIR $h(n)$, and the AEC filter coefficients $\hat{h}_{BW,m}(n)$ are replaced by $\hat{h}_{FW,m}(n)$. Thus, the AEC filter coefficients $\hat{h}_{BW,m}(n)$ will be always close to $h(n)$.

If $ERL_{F,m} \le ERL_{B,m}$ which can be the case of DT, echo path changes, or converged ST, the DTD method compares the power levels of the input signal $x(n)$ and microphone output signal $y(n)$. If the difference between them is small, $\hat{h}_{BW,m}(n)$ are updated and replaced by $\hat{h}_{FW,m}(n)$, because the power level of the microphone signal seldom changes in case of RIR changes. On the other hand, if the difference is large, $\hat{h}_{BW,m}(n)$ are freezed and will not be replaced, because the power level tends to increase in case of DT.

Although this conventional DTD method is able to discriminate between DT and echo path change, its discriminating performance is poor, especially when the far-end speech $x(n)$ is correlated with the near-end speech $z(n)$, which is usually the case in practice. To clarify and compare the detection performance between this conventional and our proposed methods, we view this problem as a hypothesis test. First, we determine the conditional probability density function (pdf) of the microphone signal power $||\mathbf{y}||^2 = \sum_{n=1}^{L} y^2(n)$ in case of ST, DT, and echo path change. Then, we will pinpoint the detection problem.

Assume that in case of echo path changes, RIR changes from $h(n)$ to $f(n)$ and the norm of RIR $||\mathbf{h}|| = ||\mathbf{f}|| = 1$, for simplicity. In fact, the gains of the loudspeaker and the echo path can both be normalized. If $||\mathbf{h}|| \neq 1$, we can change the input signal $x(n)$ by $x(n)/||\mathbf{h}||$. We also assume that the system RIR input $x(n)$ and the output signal $y_{xh}(n) = x(n) * h(n)$ or $y_{xf}(n) = x(n) * f(n)$, have equal power. For convenience of analysis, assume that $y_{xh}(n)$, $y_{xf}(n)$, $z(n)$ and $v(n)$ are independent sequences of Gaussian distribution with zero mean and variances σ_x^2, σ_x^2, σ_z^2, and σ_v^2, respectively. Therefore, $y(n)$ is an iid (identical and independent) sequence of Gaussian distribution with zero mean and its variances under three hypotheses of ST (H_{st}), DT (H_{dt}), and echo path change (H_{hv}) are

$$
\begin{aligned}
H_{st}&: \sigma_y^2 = \sigma_x^2 + \sigma_v^2 \\
H_{dt}&: \sigma_y^2 = \sigma_x^2 + \sigma_z^2 + \sigma_v^2 + E[y_{xh}(n)z(n)] \\
H_{hv}&: \sigma_y^2 = \sigma_x^2 + \sigma_v^2.
\end{aligned} \tag{23}
$$

Thus, $||\mathbf{y}||^2 = \sum_{n=1}^{L} y^2(n)$ is a Gamma distribution with two parameters denoted by $\Gamma((L/2), (1/(2\sigma_y^2)))$. When L is large enough, by the central limit theorem, the distribution is approximately a normal distribution with mean $L\sigma_y^2$ and variance $2L(\sigma_y^2)^2$, denoted by $p(||\mathbf{y}||^2) \approx N(L\sigma_y^2, 2L(\sigma_y^2)^2)$. We have pdfs of $||\mathbf{y}||^2$ as

$$
\begin{aligned}
H_{st}: p(||\mathbf{y}||^2) &= N(\mu_{st}, \sigma_{st}^2) \\
&= N(L(\sigma_x^2 + \sigma_v^2), 2L(\sigma_x^2 + \sigma_v^2)^2) \\
H_{dt}: p(||\mathbf{y}||^2) &= N(\mu_{dt}, \sigma_{dt}^2) \\
&= N(L(\sigma_x^2 + \sigma_z^2 + \sigma_v^2 + 2E[y_{xh}(n)z(n)]), \\
&\qquad 2L(\sigma_x^2 + \sigma_z^2 + \sigma_v^2 + 2E[y_{xh}(n)z(n)])^2) \\
H_{hv}: p(||\mathbf{y}||^2) &= N(\mu_{hv}, \sigma_{hv}^2) \\
&= N(L(\sigma_x^2 + \sigma_v^2), 2L(\sigma_x^2 + \sigma_v^2)^2)
\end{aligned} \tag{24}
$$

where μ_{st}, μ_{dt}, μ_{hv}, σ_{st}^2, σ_{dt}^2 and σ_{hv}^2 are the mean and variance under different hypotheses. Notice that in (24), the pdf under ST is the same as that of the echo path change. Fig. 6 shows typical pdfs of $||\mathbf{y}||^2$ by assuming $L = 4095$ and ignoring $v(n)$. The cross-correlation value of $E[y_{xh}(n)z(n)]$ is assumed to be -0.45, so that the pdfs are partially overlapped.

In Fig. 6, to distinguish between DT from echo path change (or the ST), a threshold γ that has the minimum detection error, will be determined. We use the maximum *a posteriori* (MAP) detection rule

$$
\frac{p(||\mathbf{y}||^2|H_{dt})}{p(||\mathbf{y}||^2|H_{hv})} \underset{H_{hv}}{\overset{H_{dt}}{\gtrless}} \frac{p(H_{hv})}{p(H_{dt})} \tag{25}
$$

where $p(H_{hv})$ and $p(H_{dt})$ are the probabilities of echo path change and DT occurrence. Equation (25) states that we should choose hypothesis H_{dt} if the ratio of the conditional pdfs $p(||\mathbf{y}||^2|H_{dt})/p(||\mathbf{y}||^2|H_{hv})$ is greater than $p(H_{hv})/p(H_{dt})$;

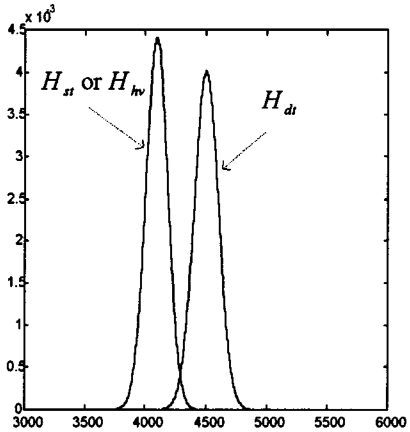

Fig. 6. Conditional pdf of $||\mathbf{y}||^2$.

otherwise, we should choose hypothesis H_{hv}. Because $p(||\mathbf{y}||^2)$ under different hypotheses are approximately of Gaussian distribution, (25) can be written as

$$\Lambda(y) \underset{H_{hv}}{\overset{H_{dt}}{\gtrless}} \gamma \tag{26}$$

where

$$\Lambda(y) = \frac{(||\mathbf{y}||^2 - \mu_{dt})^2}{\sigma_{dt}^2} - \frac{(||\mathbf{y}||^2 - \mu_{hv})^2}{\sigma_{hv}^2}$$

and the threshold

$$\gamma = 2\ln\sqrt{\frac{\sigma_{hv}^2}{\sigma_{dt}^2}\,\frac{p(H_{dt})}{p(H_{hv})}}.$$

The flowchart of DT detection is depicted in Fig. 7.

However, there exists a serious problem to determine the threshold γ. Because $p(||\mathbf{y}||^2)$ is a function of $E[y_{xh}(n)z(n)]$ that can not be known exactly, thus, the threshold γ can not be determined exactly and a risk of wrong detection is very high.

In the next section, we propose an AEC structure which incorporates the IMLC method and a new DT detection mechanism. This mechanism monitors the variation of the filter coefficients to offer a well-separated detection margin to effectively discriminate among ST, DT, and echo path change.

B. IMLC DTD AEC Structure

The new AEC structure is depicted in Fig. 8 and is basically similar to the one in Fig. 5. It also has FW and AEC filters for adaptively identifying echo path transfer characteristics and synthesizing the echo replica. The coefficients of the AEC filter $\hat{h}_{BW}(n)$ are refreshed and copied from the FW filter's coefficients $\hat{h}_{FW}(n)$ only when the FW is found to give a better echo path transfer characteristics than the AEC filter. However, in the new structure the FW filter coefficients are updated by the IMLC method described in Section II. Because (21) shows that the IMLC method is more robust to estimate $h(n)$ during DT, the new DT detection mechanism is to track the variation of the squared coefficients error $||\mathbf{e}||^2 = ||\hat{\mathbf{h}} - \mathbf{h}||^2$ to effectively

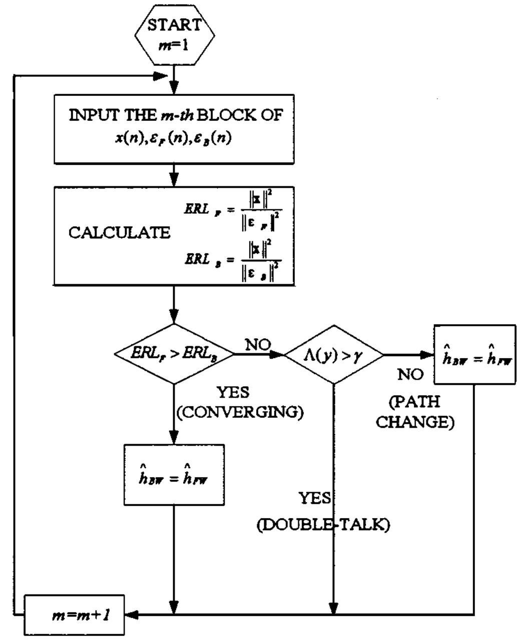

Fig. 7. Flowchart of DT detection based on the MAP rule.

discriminate among the ST, DT, and echo path change. This differs significantly from the conventional DTD method by tracking $||\mathbf{y}||^2$.

The new detection method can be expressed as follows: If $ERL_{F,m} > ERL_{B,m}$, the AEC filter coefficients $\hat{h}_{BW}(n)$ are replaced by $\hat{h}_{FW}(n)$. If $ERL_{F,m} \le ERL_{B,m}$, we need to inspect the squared coefficients error, $||\mathbf{e}_m||^2$, at the mth IMLC iteration. We find that the coefficients error $e_m(n) = \hat{h}_m(n) - h(n)$ can not be measured, in practice. However, in the structure of Fig. 8, when the IMLC algorithm begins from ST, after a few iterations, we have $\hat{h}_{BF}(n) \approx h(n)$ and because of $\hat{h}_{FW}(n) = \hat{h}_m(n)$, we can estimate $e_m(n)$ as

$$e_m(n) \approx \hat{h}_{FW}(n) - \hat{h}_{BW}(n). \tag{27}$$

Now, we are interested in determining pdfs of $||\mathbf{e}_m||^2$ under three different hypotheses: $||\mathbf{e}||_{st}^2 ||\mathbf{e}||_{dt}^2$ and $||\mathbf{e}||_{hv}^2$. We have computed the filter coefficients errors, $e_m(n)$, of IMLC during DT and ST. As before, we assume that the near-end speech $z(n)$ and the near-end noise $v(n)$ are both of white Gaussian distribution with zero mean. By (21) and (22), the coefficients errors are modeled as a linear combination of the near-end signal and assume that $h(n)$ has no DC term, $(1/(L+1))\sum_{k=1}^{M} h(k)$ can be neglected. Thus, $e_{st,m}(n)$ and $e_{dt,m}(n)$ are also of Gaussian distribution with mean $\mu_{st} = \mu_{dt} = 0$ and variances $\sigma_{st}^2 = \sigma_v^2/(G^2(L+1))$ and $\sigma_{dt}^2 = (\sigma_z^2 + \sigma_v^2)/(G^2(L+1))$. Following the same procedure in (24), we have the conditional pdfs of $||\mathbf{e}_m||^2$ in ST and DT as

$$H_{st}\colon p(||\mathbf{e}_m||^2) = N\left(M\,\frac{\sigma_v^2}{G^2(L+1)},\, 2M\left(\frac{\sigma_v^2}{G^2(L+1)}\right)^2\right)$$

$$H_{dt}\colon p(||\mathbf{e}_m||^2) = N\left(M\,\frac{\sigma_z^2+\sigma_v^2}{G^2(L+1)},\, 2M\left(\frac{\sigma_z^2+\sigma_v^2}{G^2(L+1)}\right)^2\right). \tag{28}$$

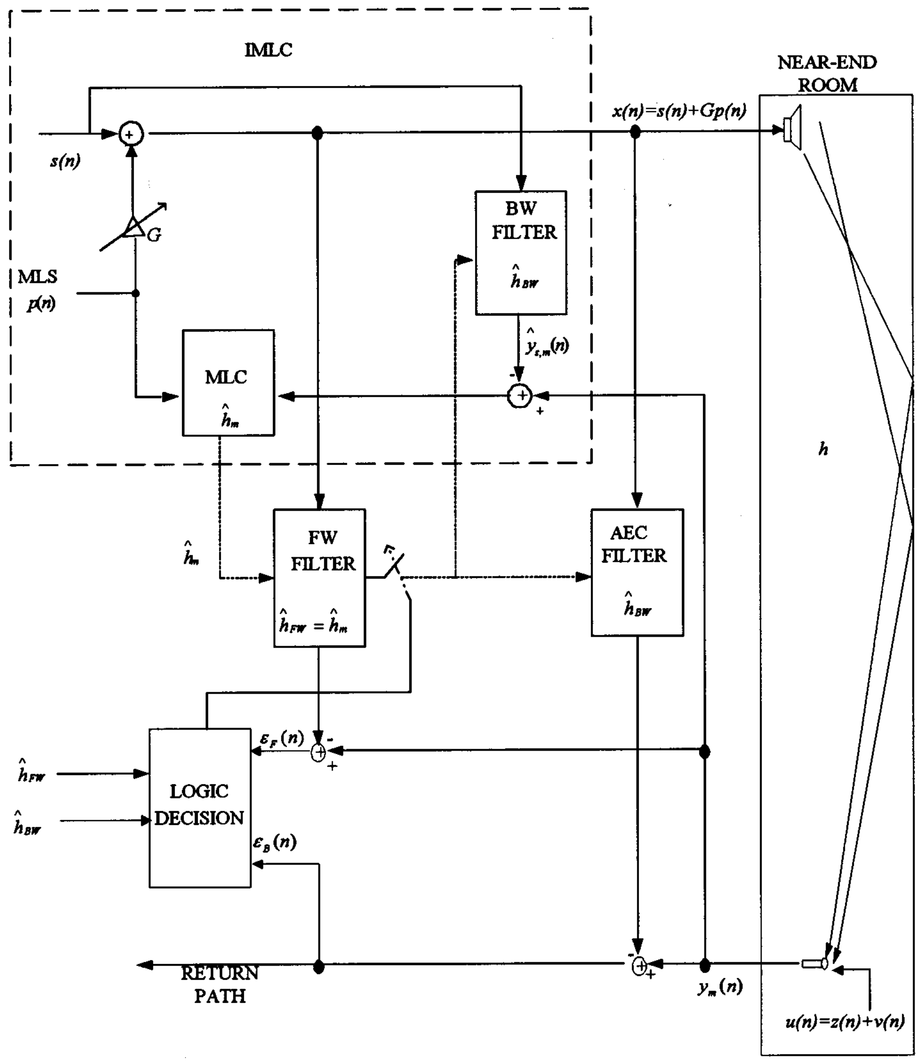

Fig. 8. Proposed IMLC AEC with DTD structure.

Next, we will consider the pdf of $||\mathbf{e}||^2_{hv}$, in the hypotheses H_{hv} of echo path change. Suppose the IMLC algorithm begins from ST and the echo path changes from $h(n)$ to $f(n)$ at some time instant. In this case, according to (27), we have the coefficients of the FW filter $\hat{h}_{FW}(n) = \hat{f}(n)$, and the estimated coefficients error $e_{hv}(n)$ will become

$$\begin{aligned} e_{hv}(n) &= \hat{h}_{FW}(n) - \hat{h}_{BW}(n) \approx \hat{f}(n) - h(n) \\ &= [f(n) - h(n)] + e_f(n) \end{aligned} \tag{29}$$

where $e_f(n) = \hat{f}(n) - f(n)$ is the new coefficient error. Assume that $[f(n) - h(n)]$ and $e_f(n)$ are uncorrelated, then the squared coefficients error $||\mathbf{e}||^2_{hv}$ can be approximately expressed as

$$||\mathbf{e}||^2_{hv} \approx ||\mathbf{f} + \mathbf{e}_f - \mathbf{h}||^2 = ||\mathbf{f} - \mathbf{h}||^2 + ||\mathbf{e}_f||^2. \tag{30}$$

From (11), the newly estimated IMLC coefficients can be rewritten as

$$\hat{f}(n) = f(n) + I_{N,m}(n) + I_{F,m}(n) * [f(n) - \hat{h}_{m-1}(n)]. \tag{31}$$

Because $\hat{h}_{m-1}(n) \approx h(n)$ and in the absence of the near-end speech [$I_{N,m}(n)$ is ignored], the new coefficient error $e_f(n)$ can be rewritten as

$$e_f(n) = I_{F,m}(n) * [f(n) - h(n)].$$

Using (8), we have

$$\begin{aligned} e_f(n) &= \frac{1}{G(L+1)} p(n) © [f(n) - h(n)] * s(n) \\ &= \frac{1}{G(L+1)} \Delta_p(n) © s(-n) \end{aligned} \tag{32}$$

where $\Delta_p(n) \equiv p(n) © [h(n) - f(n)]$. Notice that (32) is similar to (22) and assume $s(n)$ is an iid sequence of Gaussian distribution with zero mean, we may derive the pdf of $e_f(n)$ as previously discussed in (28). The variance of $e_f(n)$ is $\sigma_f^2 = (\sigma_s^2/(G^2(L+1)^2)) \sum_k^L \Delta_p^2(k)$ where $\Delta_p^2(n)$, by the definition and property of MLS $p(n)$, can be written as: $\Delta_p^2(n) = ||\mathbf{h} - \mathbf{f}||^2$

and σ_f^2 becomes $\sigma_f^2 = (\sigma_s^2(\|\mathbf{f}-\mathbf{h}\|^2))/(G^2(L+1))$. The pdf of $\|\mathbf{e}_f\|^2$ can be expressed as

$$p(\|\mathbf{e}_f\|^2) = N\left(M\,\frac{\sigma_s^2(\|\mathbf{f}-\mathbf{h}\|^2)}{G^2(L+1)},\, 2M\left[\frac{\sigma_s^2(\|\mathbf{f}-\mathbf{h}\|^2)}{G^2(L+1)}\right]^2\right). \tag{33}$$

In (30), we have known the relation between $\|\mathbf{e}\|_{hv}^2$ and $\|\mathbf{e}_f\|^2$, the pdf of $\|\mathbf{e}\|_{hv}^2$ now becomes

$$p(\|\mathbf{e}\|_{hv}^2) = N\left(\|\mathbf{f}-\mathbf{h}\|^2 + M\,\frac{\sigma_s^2(\|\mathbf{f}-\mathbf{h}\|^2)}{G^2(L+1)},\; 2M\left[\frac{\sigma_s^2(\|\mathbf{f}-\mathbf{h}\|^2)}{G^2(L+1)}\right]^2\right). \tag{34}$$

Next, we consider $\|\mathbf{f}-\mathbf{h}\|^2$ in (34). When the relative location between the microphone and the loudspeaker varies slightly and the echo path changes from $\mathbf{h}$ to $\mathbf{f}$, our experiment indicates that $\mathbf{f}^T\mathbf{h} \approx 0$ and $\|\mathbf{h}\| \approx \|\mathbf{f}\|$, therefore, we assume

$$\|\mathbf{f}-\mathbf{h}\|^2 = \|\mathbf{f}\|^2 + \|\mathbf{h}\|^2 - 2\mathbf{f}^T\mathbf{h} \approx 2\|\mathbf{h}\|^2. \tag{35}$$

In summary, the pdfs of $\|\mathbf{e}\|^2$ under three different hypotheses are

$$\begin{aligned} p(\|\mathbf{e}\|_{st}^2) &= N\left(\frac{M\sigma_v^2}{G^2(L+1)},\, 2M\left[\frac{\sigma_v^2}{G^2(L+1)}\right]^2\right)\\ p(\|\mathbf{e}\|_{dt}^2) &= N\left(\frac{M(\sigma_z^2+\sigma_v^2)}{G^2(L+1)},\, 2M\left[\frac{\sigma_z^2+\sigma_v^2}{G^2(L+1)}\right]^2\right)\\ p(\|\mathbf{e}\|_{hv}^2) &= N\left(2\|\mathbf{h}\|^2 + \frac{2M\|\mathbf{h}\|^2\sigma_s^2}{G^2(L+1)},\, 2M\left[\frac{2\|\mathbf{h}\|^2\sigma_s^2}{G^2(L+1)}\right]^2\right). \end{aligned} \tag{36}$$

Fig. 9 shows the pdfs of $\|\mathbf{e}\|^2$ by assuming that $\sigma_z^2 = \sigma_s^2 = 1$, $\|\mathbf{h}\|^2 = 1, L = 4095, G = 0.15$ and the near-end SNR is 10 dB.

Comparing the new DTD in (36) and conventional one in (24), we find some interesting properties. First, the new method can easily distinguish among ST, DT, and echo path change because the parameters of its pdfs are known and the detection margins are well separated, but the conventional method cannot. Second, because the detection margins are well separated, the decision rule can be simply expressed as

$$\begin{aligned} &H_{st}\colon \|\mathbf{e}\|^2 < \gamma_d\\ &H_{dt}\colon \gamma_d \le \|\mathbf{e}\|^2 < \gamma_h\\ &H_{hv}\colon \|\mathbf{e}\|^2 \ge \gamma_h \end{aligned} \tag{37}$$

where the threshold γ_d can be simply chosen by averaging the means of $p(\|\mathbf{e}\|_{st}^2)$ and $p(\|\mathbf{e}\|_{dt}^2)$. Similarly, γ_h can be chosen by averaging the means of $p(\|\mathbf{e}\|_{dt}^2)$ and $p(\|\mathbf{e}\|_{hv}^2)$. The complete flow chart of the proposed DTD is depicted in Fig. 10.

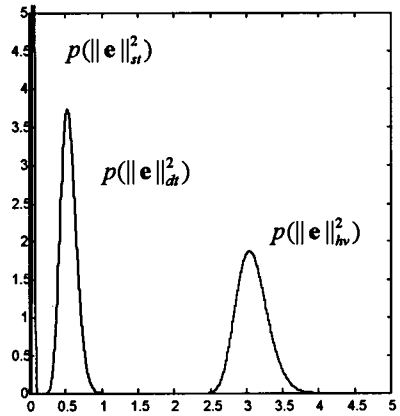

Fig. 9. Probability density functions of $\|\mathbf{e}\|^2$ under ST, DT, and path change.

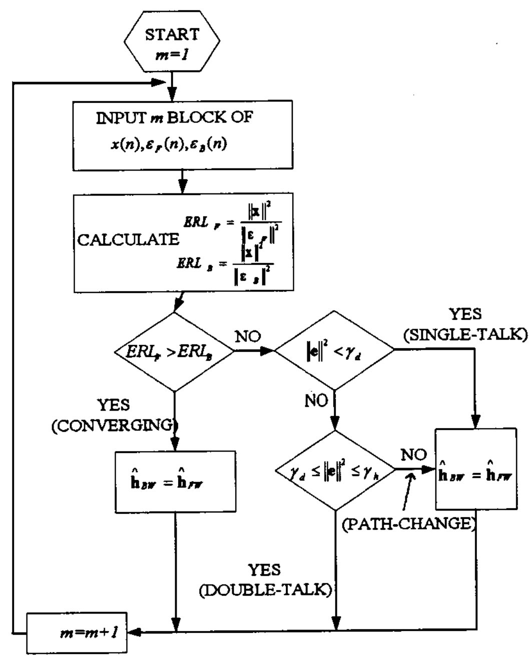

Fig. 10. Flowchart of proposed DT detection.

IV. Computer Simulation and Comparisons

The performance of the algorithms are verified and compared by extensive computer simulation. We use $ERLE$ as the criterion, defined as

$$ERLE_m\ (\mathrm{dB}) \equiv 10\log_{10}\frac{\sum_{n=1}^{L} y_m(n)^2}{\sum_{n=1}^{L}[y_m(n)-\hat{y}_m(n)]^2} = 10\log_{10}\frac{\sum_{n=1}^{L} y_m(n)^2}{\sum_{n=1}^{L}[e_m(n)*s(n)]^2}. \tag{38}$$

Fig. 11 shows RIR $h(n)$ which is measured from a real room. The sampling frequency is 10k, and down sampled by 100. The

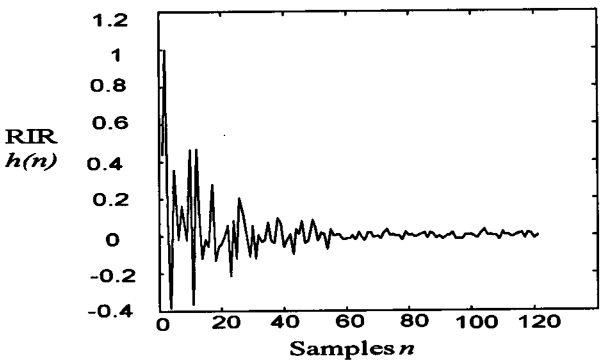

Fig. 11. RIR used for simulations.

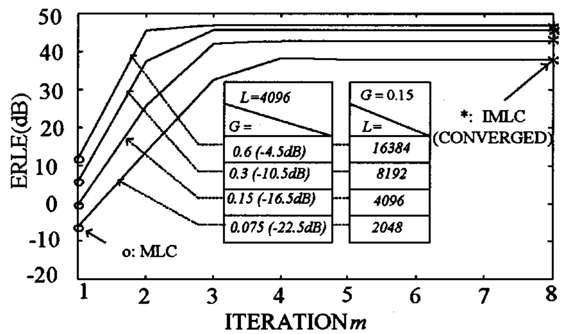

Fig. 12. $ERLE$ comparison of MLC and IMLC algorithms in ST.

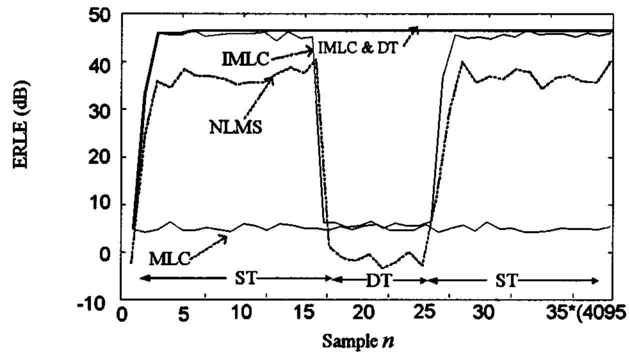

Fig. 13. $ERLE$ comparisons of MLC, IMLC, and IMLC&DT algorithms in ST and DT.

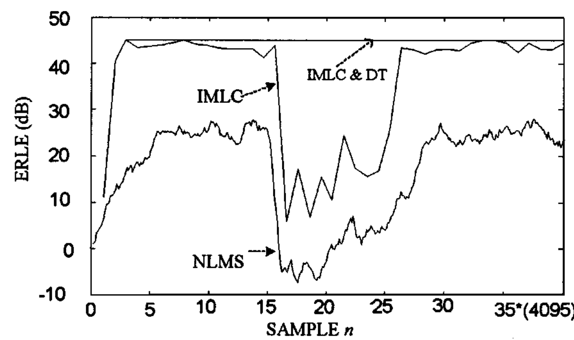

Fig. 14. $ERLE$ comparisons using real speech signals.

far-end $s(n)$, near-end $z(n)$ speeches and background noise $v(n)$ are white Gaussian with $\sigma_z^2 = \sigma_s^2 = 1$, and $\sigma_v^2 = 0.0001$. The MLS period length L is 4096.

In Fig. 12, assume the filter length $M = 100$ and RIR $h(n)$ also has an order of 100 during ST situation, we compare the $ERLE$ performance between MLC and IMLC methods. In these simulations, two groups of parameters, L and G of MLS, are considered. 1) Fixed $L = 4096$ and change G from 0.075 (-22.5 dB) to 0.6 (-4.5 dB). It is expected that increasing G will improve the $ERLE$ performance. 2) Fixed $G = 0.15$ (-16 dB) and change L from 2048 to 16 384. Likewise, increasing L will improve the $ERLE$ performance. We find that if $G(L+1)$ is fixed as a constant, the performances of $ERLE$ are identical. For example, if G $= 0.3$ (-10 dB) and $L = 4095$ are chosen, the $ERLE$ is the same with that of G $= 0.15$ (-16 dB) and $L = 8192$. This result is confirmed by (21) and (22). To keep the same $ERLE$ performance, we may either choose larger G and shorter L (more noise but less computation complexity) or smaller G and longer L (less noise but more computation complexity). "$\circ$" is the result of MLC and "$*$" is the converged results of IMLC. Note that in the beginning, the AEC filter's coefficients $\hat{h}_{BW}(n)$ may diverge due to the far-end speech, especially when the products GL are small. When the coefficients $\hat{h}_{BW}(n)$ are close to $h(n)$, the far-end speech will be canceled more perfectly and further iterations can improve estimation even better. We find that after three or four iterations, IMLC improves 40 dB compared to the conventional MLC method.

Fig. 13 compares the $ERLE$ performance of NLMS, MLC, IMLC, and IMLC&DT (IMLC with DT detection) algorithms in cases of ST and DT using white Gaussian noise as input. In this simulation, we select $G = 0.3$, $L = 4095$, filter taps $M = 100$ and RIR orders $= 100$, the $ERLE$ of MLC is only 6 dB in either cases of ST and DT. The converged $ERLE$ of IMLC in ST is about 45 dB and degrades to 6 dB in DT. The $ERLE$ of IMLC&DT always maintains about 45 dB.

Fig. 14 compares the $ERLE$ performances using speech signals. In IMLC and MLC simulations, we choose $G = 0.3$, $L = 4096$, filter taps $M = 100$ and RIR orders $= 100$, the converged $ERLE$ of IMLC in ST is also about 45 dB and degrades to 6 dB minimum in DT. The $ERLE$ of IMLC&DT always maintains about 45 dB. In NLMS simulations, we choose step size $= 1$, filter length $= 100$. The converged $ERLE$ of NLMS in ST is about 27 dB and degrades to -5 dB minimum in DT.

Fig. 15 shows the squared coefficients errors $||\mathbf{e}||^2$ of IMLC, using white Gaussian noise and real speech signals, in cases of ST, DT, and echo path change (when positions of the microphone and the loudspeaker vary slightly). In this simulation, we choose G $= 0.3$ (the mask level of the far-end speech to MLS is about 10 dB); the results are close to theoretical values: $E(||\mathbf{e}||_{dt}^2) = -9$ dB, $E(||\mathbf{e}||_{st}^2) = -48$ dB. Notice that the detection margin between ST and DT is about 30 dB and between DT and echo path change (HV) is about 10 dB. Although there are more fluctuations for speech signals in case of DT, our detection algorithm still works very well.

Next, we compare the computation complexities of MLC, IMLC, and NLMS algorithms.

In Table I, we find that the computation complexities of IMLC are more than MLC but equal or less than NLMS. (Note that $*$ indicates the computation complexities of the NLMS algorithm

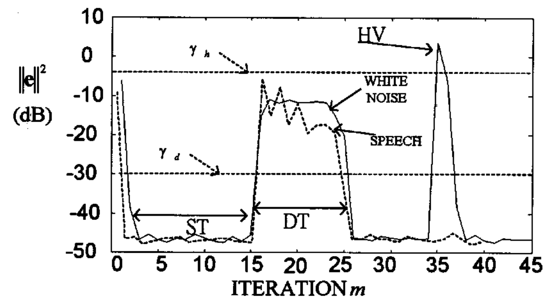

Fig. 15. Squared coefficients errors $\|\mathrm{e}\|^2$ in ST, DT, and echo path change (HV).

TABLE I
COMPARISON OF COMPUTATION COMPLEXITIES

ALGORITHM / ITEM	MLC	IMLC	NLMS
$\hat{y}_{s,m}(n) = s(n) * \hat{h}_{m-1}$		Add: $L(M-1)$ Mul: LM	
$\hat{h}_m(n) = \frac{1}{G(L+1)} p(n) © (y_m(n) \text{ or } \tilde{y}_m(n))$	Add: $(L-1)M$ Mul: M	Add: $(L-1)M$ Mul: M	
$\varepsilon(n) = y_m(n) - x(n) * \hat{h}_{m-1}$	Add: LM Mul: LM	Add: LM Mul: LM	Add: LM Mul: LM
$\hat{\mathbf{h}}_m = \hat{\mathbf{h}}_{m-1} - \frac{\mu}{\\|\mathbf{x}\\|^2} \mathbf{x}\varepsilon(n)$			Add: $L(2M-1)$ Mul: $2L(M+1)$
TOTAL	Add: $2LM$ Mul: LM	Add: $3LM$ Mul: $2LM$	Add*: $3LM$ Mul*: $3LM$

can be reduced to approximately $2LM$ due to the fact that the norm of the input signal is calculated recursively.)

V. Conclusion

In this paper, we investigate the known MLC method for acoustic echo cancellation. By using a BW filter to estimate the response of the far-end speech, an improved iterative MLC algorithm is proposed to reduce the coefficient estimation error due to the far-end speech interference. A new DT detection based on hypothesis test is also proposed to monitor the squared coefficient errors so that the near-end speech interference can be avoided. Theoretical analysis and computer simulation demonstrates that the IMLC is efficient and effective for acoustic echo cancellation.

References

[1] N. A. M. Verhoeckx, "Digital echo cancellation for base-band data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, no. 6, pp. 768–781, 1979.
[2] E. Hansler, "The hands-free telephone problem; An annotated bibliography," *Signal Process.*, vol. 27, pp. 259–271, 1992.
[3] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[4] O. A. Horna, "Echo canceller with adaptive transversal filter utilizing pseudo-logarithmic coding," *COMSAT Tech. Rev.*, vol. 7, no. 2, pp. 393–427, 1977.
[5] N. Furuya, *et al.*, "High performance custom VLSI echo canceller," in *ICC'85 Conf. Rec.*, 1985, pp. 1470–1476.
[6] S. Minami, *et al.*, "A double-talk detection method for an echo canceller," in *ICC'85 Conf. Rec.*, 1985, pp. 1492–1497.
[7] D. D. Rife and J. Vanderkooy, "Transfer function measurement with maximum length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–443, June 1989.
[8] 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, July–Aug. 1983.
[9] C. Breining, "Applying a neural network for step-size control in echo cancellation," in *Proc. 5th Int. Workshop Acoustic Echo Noise Control*, London, U.K., 1997, pp. 441–444.
[10] P. Heitkamper, "An adaptation control for acoustic echo cancellers," *IEEE Signal Processing Lett.*, vol. , pp. 170–172, June 1997.
[11] J. F. Doherty, *et al.*, "A robust echo canceler for acoustic environments," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 389–396, May 1997.
[12] M. J. Crocker, *Handbook of Acoustics*: Wiley, 1998.
[13] J. C. Jenq and S. F. Hsieh, "A double-talk resistant echo cancellation based on iterative maximal-length correlation," in *IEEE Symposium on Circuits and Systems*, vol. 5, May 2000, pp. 237–241.
[14] K. Ochiai, *et al.*, "Echo canceller with two echo path models," *IEEE Trans. Commun.*, vol. COM-25, pp. 589–595, June 1977.
[15] T. Gilloire and M. Hansson, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, vol. 44, pp. 1421–1427, Nov. 1996.
[16] H. Ya and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Commun.*, vol. 39, pp. 1542–1545, Nov. 1991.
[17] C. Antweiler and M. Dörbecker, "Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control," *Annales des Telecommunications*, no. 7–8, pp. 386–397, July–August 1994.

Jang-Chyuan Jenq was born in Taoyuan, Taiwan, R.O.C. He received the B.S. degree in electrical engineering from the Chung-Cheng Institute of Technology, Taoyuan, in 1989 and the M.S. degree in communication engineering from the National Chiao-Tung University, Hsinchu, Taiwan, in 1993, where he is currently pursuing the Ph.D. degree.

His main research interests are in acoustic echo cancellation, data detection, and active noise control.

Shih-Fu Hsieh received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1984, and the M.S. and Ph.D. degrees from the University of California, Los Angeles, in 1987 and 1990, respectively, all in electrical engineering.

From 1984 to 1986, he was an Ensign Instructor at the Naval Communication and Electronics School, Taiwan. Since 1990, he has been with the National Chiao Tung University, Hsinchu, Taiwan, where he is currently an Associate Professor of communication engineering. His current research interests include acoustical signal processing, digital communication systems, and parallel algorithms and architectures.

Dr. Hsieh received the IEEE Signal Processing Society's 1993 Senior Award in VLSI Signal Processing.