Decision of Double-Talk and Time-Variant Echo Path for Acoustic Echo Cancelation

J. C. Jenq and S. F. Hsieh

***Abstract*—A new double-talk detection (DTD) method is proposed for acoustic echo cancelation (AEC) using the iterative maximal-length-correlation (IMLC) algorithm. Based on the assumption of uniformly distributed time-variant echo-path-change, we develop a simplified likelihood ratio test and plot the detection performance by a receiver operating characteristic. Computer simulation shows the proposed DTD method is effective in discriminating double-talk from time-variant echo-path-change.**

***Index Terms*—Acoustic echo cancelation, double-talk detection.**

I. Introduction

HANDS-FREE conversation is popular in various fields of communication such as teleconferencing, video conferencing, and mobile radiotelephone. However, in those applications, the presence of coupling from the far-end signal (loudspeaker) to the microphone would result in undesired acoustic echo and significantly degrade the speech quality. Therefore, an effective acoustic echo canceler (AEC) is required. AEC is usually implemented by an adaptive finite-impulse response filter, and various adaptation algorithms have been suggested by many researchers [1]. However, all existing adaptive AEC filters share serious problems during "double-talk" when simultaneous speech occurs for both near-end and far-end speakers. In this situation, the microphone signal includes the near-end signal, which acts like a large disturbing noise to the residue echo signal. If the adaptive filter continues to adjust, its coefficients will be greatly disturbed and result in intolerable echo.

To overcome the double-talk problem, almost all current techniques use double-talk detectors (DTDs) and attempt to turn off adaptation during this situation. Many detection methods [2]–[4] are accomplished by observing the far-end signal, the microphone signal, and the residue error, but a critical question is that merely measuring these signals cannot discriminate between double-talk and echo-path-change. If echo path change is mislabeled as double-talk, AEC performance degrades. Therefore, we have proposed the iterative maximal-length correlation (IMLC) DTD method. In our previous paper [5], the IMLC AEC structure is introduced, and the probability density functions (pdfs) of the filter coefficient squared errors are derived under double-talk and path-change. This DTD method has a well-separated detection margin between double-talk and echo-path-change if the change is assumed to be large and deterministic. However, this assumption is not valid in the case of slowly changing random echo path, because the decision margins are time-variant and generally unknown in practice.

Manuscript received October 10, 2002; revised November 14, 2002. This work was supported in part by the National Science Council under Grant NSC 92-2213-E-009-129. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark Hasegawa-Johnson.

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@mail.nctu.edu.tw).

Digital Object Identifier 10.1109/LSP.2003.817169

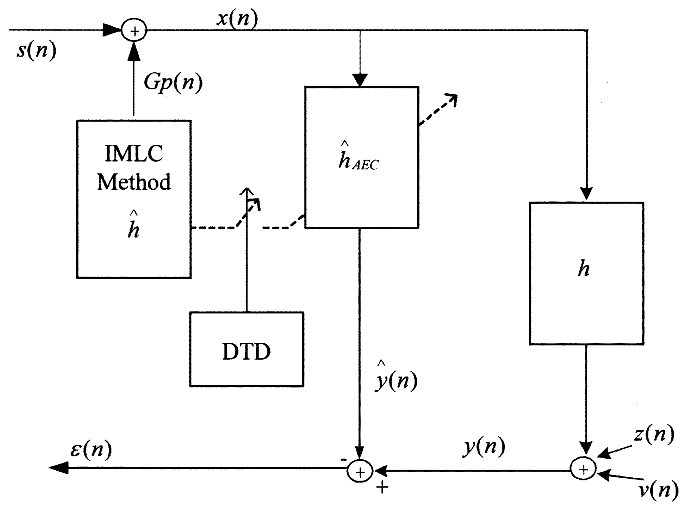

Fig. 1. IMLC-DTD AEC structure.

In this letter, we aim to obtain a new double-talk detection method based on the IMLC structure. By the assumption of uniformly distributed time-variant echo-path-change, we develop a simplified likelihood ratio test by setting a decision threshold and plot the detection performance by a receiver operating characteristic (ROC). Computer simulations confirm that our proposed algorithm outperforms previous DTD methods.

II. IMLC-DTD AEC Structure

A typical IMLC-DTD AEC structure is shown in Fig. 1, where the AEC filter's coefficients $\hat{h}_{\mathrm{AEC}}(n)$, estimated from an IMLC method [5], are used to model the room impulse response $h(n)$ between the microphone and the loudspeaker. The basic IMLC method is done by adding a periodic maximal-length sequence (MLS) [8] $p(n)$ with period L and magnitude G to the far-end speech $s(n)$, so that the far-end signal becomes $x(n) = s(n) + Gp(n)$ which is then fed into the loudspeaker. Once per L samples, $\hat{h}(n)$ is estimated by cross-correlating the microphone signal $y(n)$ with $p(n)$ [6]. Since $\hat{h}(n)$ is disturbed by the far-end speech $s(n)$, the near-end speech $z(n)$, and the near-end noise $v(n)$, the IMLC method predicts $s(n)$ and iteratively reduces its interference, while DTD tracks the sum-squared filter coefficient change so that interferences due

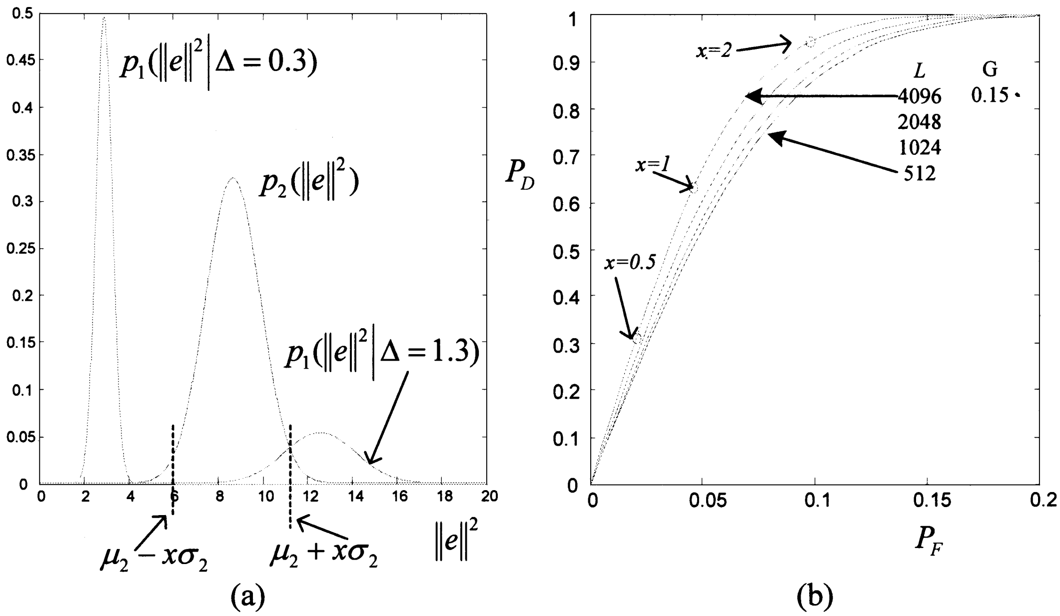

Fig. 2. (a) PDFs of $\|\mathrm{e}\|^2$. (b) ROC curves with different sizes x and MLS lengths L.

to $z(n)$ can be avoided. If no double-talk is detected, $\hat{h}_{\mathrm{AEC}}(n)$ is set equal to $\hat{h}(n)$.

Next, based on this structure, we develop a new double-talk detection that is a simplified likelihood ratio test and is valid in the case of random echo-path-change.

III. Simplified Likelihood Test and ROC

Suppose the IMLC algorithm begins from single-talk and has converged; then we have $\hat{h}_{\mathrm{AEC}}(n) \approx h(n)$. Denote H_1 and H_2 as echo-path-change and double-talk hypotheses, respectively.

The filter coefficent error $e(n) \equiv \hat{h}(n) - h(n) \approx \hat{h}(n) - \hat{h}_{\mathrm{AEC}}(n)$ can be expressed as [5]

$$\begin{aligned} H_1 : e(n) =& [f(n) - h(n)] + \frac{1}{\mathrm{GL}} p(n) \copyright v(n) \\ &+ \frac{1}{\mathrm{GL}} p(n) \copyright s(n) * [f(n) - h(n)] \\ H_2 : e(n) =& \frac{1}{\mathrm{GL}} p(n) \copyright [z(n) + v(n)] \end{aligned} \quad (1)$$

where $1 \leq n \leq M$, M is the filter order, and © and $*$ denote linear correlation and convolution, respectively.

The room impulse response is assumed to change from $h(n)$ to $f(n)$. Assume $v(n)$, $z(n)$, and $s(n)$ are Gaussian distributed and mutually independent, with zero means and variances σ_v^2, σ_z^2, and σ_s^2, respectively. The filter error vector $\mathbf{e} = [e(1), e(2), \ldots, e(M)]^t$ under each hypothesis can be shown to be Gaussian with pdf as

$$\begin{aligned} H_i : p_i(\mathbf{e}) =& \frac{1}{(2\pi)^{M/2} |\mathbf{K}_i|^{1/2}} \\ &\cdot \exp\left[-\frac{1}{2} (\mathbf{e}^t - \mathbf{m}_i^t) \mathbf{K}_i^{-1} (\mathbf{e} - \mathbf{m}_i) \right], \quad i = 1, 2 \end{aligned} \quad (2)$$

where mean vectors are $\mathbf{m}_1 = \mathbf{f} - \mathbf{h} = [f(1) - h(1), \ldots, f(M) - h(M)]$ and $\mathbf{m}_2 = \mathbf{0}$ and covariance matrices are $\mathbf{K}_1 = \left((\Delta\sigma_s^2 + \sigma_v^2)/G^2L\right)\mathbf{I}_{M\times M} \equiv \varsigma_1^2 \mathbf{I}_{M\times M}$ and $\mathbf{K}_2 = \left((\sigma_z^2 + \sigma_v^2)/G^2L\right)\mathbf{I}_{M\times M} \equiv \varsigma_2^2 \mathbf{I}_{M\times M}$, where the room impulse responses are normalized to $\|\mathbf{f}\|^2 = \|\mathbf{h}\|^2 = 1$ and the difference of impulse responses (DIR) is defined as $\Delta \equiv \|\mathbf{f} - \mathbf{h}\|^2$.

Obviously, in (2), the pdfs of the binary hypotheses are quite different. In general, the likelihood ratio decision rule [7] can be used for DTD as follows:

$$\frac{p_1(\mathbf{e})}{p_2(\mathbf{e})} \underset{H_2}{\overset{H_1}{\gtrless}} \Lambda$$

where Λ denotes some decision threshold. Therefore, in our case, the DTD can be written as

$$\begin{aligned} \|\mathbf{e}\|^2 \underset{H_2}{\overset{H_1}{\gtrless}} \gamma, \quad \gamma = & \frac{\varsigma_1^2 \varsigma_2^2}{\varsigma_1^2 - \varsigma_2^2} \\ & \cdot \left(2 \ln\left(\frac{\varsigma_1}{\varsigma_2} \Lambda \right) + \frac{1}{\varsigma_1^2} \mathbf{m}_1^t \mathbf{m}_1 - \frac{2}{\varsigma_1^2} \mathbf{e}^t \mathbf{m}_1 \right). \end{aligned} \quad (3)$$

However, in (3), we find that $\mathbf{m}_1 = \mathbf{f} - \mathbf{h}$ cannot be known in practice. Thus, we are still unable to determine the decision regions from the observed test statistics $\|\mathbf{e}\|^2$ to discriminate echo-path-changes (H_1) from double-talk (H_2). Next, we will examine the pdfs $p(\|\mathbf{e}\|^2)$ and find a practical decision rule. When the order M is large, by the central limit theorem, the sufficient statistics $\|\mathbf{e}\|^2$ can be approximated as a Gaussian random variable

$$H_i : p_i(\|\mathbf{e}\|^2) \approx \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left[-\frac{1}{2\sigma_i^2} (\|\mathbf{e}\|^2 - \mu_i)^2 \right] \quad (4)$$

where $\mu_1 = \Delta + M\left((\Delta\sigma_s^2 + \sigma_v^2)/G^2L\right)$, $\sigma_1^2 = 2M\left[(\Delta\sigma_s^2 + \sigma_v^2)/G^2L\right]^2$, $\mu_2 = M\left((\sigma_z^2 + \sigma_v^2)/G^2L\right)$, and $\sigma_2^2 = 2M\left[(\sigma_z^2 + \sigma_v^2)/G^2L\right]^2$.

We observe that μ_2 and σ_2^2 can be determined from its parameters (G, L, and M, etc.), but μ_1 and σ_1^2 are functions of the DIR $\Delta \equiv \|\mathbf{f} - \mathbf{h}\|^2$, which is unknown. Fig. 2(a) shows an example of pdfs $p_1(\|\mathbf{e}\|^2)$ and $p_2(\|\mathbf{e}\|^2)$ with $G = 0.15, L = 512$, $M = 100$, $\sigma_z^2 = \sigma_s^2 = 1$, and $\sigma_v^2 = 0.0001$. Measured room responses with reverberation time150 ms are used in our simulations. We notice that when DIR $\Delta = 1.3$ or 0.3, i.e., in the case of significant or slight echo-path-change, μ_1 and σ_1^2 will be large or small. Thus, $p_1\left(\|\mathbf{e}\|^2 \middle| \Delta\right)$ may be located to the left or right of $p_2(\|\mathbf{e}\|^2)$ as seen in this plot.

In our previous work [5], we assumed that $\Delta = 2$, and therefore $p_1(\|\mathbf{e}\|^2)$ was always located to the far right-hand side of $p_2(\|\mathbf{e}\|^2)$. We now assume $\|\mathbf{f} - \mathbf{h}\|$ is uniformly distributed over [0, 2]. Thus, the pdf of $\Delta = \|\mathbf{f} - \mathbf{h}\|^2$ becomes $f(\Delta) = 1/(4\sqrt{\Delta})$ over [0,4].

Our strategy is to decide H_2 when $\|\mathbf{e}\|^2 \in R_2 = [\mu_2 - x\sigma_2, \mu_2 + x\sigma_2]$ where x is the size controlling the decision interval. Define the probability of detection $p_D = \text{prob}(\|\mathbf{e}\|^2 \in R_2 | H_2)$ (if double-talk is true and double-talk is decided) and the false alarm probability $p_F = \text{prob}(\|\mathbf{e}\|^2 \in R_2 | H_1)$ (if echo-path-change is true but double-talk is decided). We have

$$p_D = \int_{\mu_2 - x\sigma_2}^{\mu_2 + x\sigma_2} p_2(\|\mathbf{e}\|^2) d\,\|\mathbf{e}\|^2$$

$$p_F = \int_0^4 p(\Delta) \int_{\mu_2 - x\sigma_2}^{\mu_2 + x\sigma_2} p_1(\|\mathbf{e}\|^2) d\,\|\mathbf{e}\|^2\, d\Delta \qquad (5)$$

from which the receiver operating characteristic curve (ROC) relating p_D and p_F can be obtained by choosing different sizes x.

IV. Computer Simulations

Fig. 2(b) shows the curves of p_D versus p_F in (5) with different sizes x. The pdf of Δ is $f(\Delta) = 1/(4\sqrt{\Delta})$ over [0, 4]. Signals $s(n)$, $z(n)$, and $v(n)$ are white Gaussian, and their parameters used are as the same as in Fig. 2(a). The MLS magnitude $G = 0.15$ (-16 dB) is fixed as a constant, and its length L is changed from 512 to 4096. Notice that when the decision interval $[\mu_2 - x\sigma_2, \mu_2 + x\sigma_2]$ is enlarged by increasing x, p_D and p_F increase too. From this ROC curve, the decision probability p_D is readily available by specifying a tolerable p_F. For example, if $L = 4096$, and $p_F = 0.1$ is permitted, then the size $x = 2$, and p_D can be assured to be 0.95. It is seen that as L is increased, i.e., more data are used, the detection performance also improves. To simplify our simulations, all signals are white. The detection algorithm will still work well with real speech signals. The observed test statistics $\|\mathbf{e}\|^2$ to discriminate echo-path-changes (H_1) from double-talk (H_2) are still quite different, as discussed in [5].

Now we compare our ROC curve with other methods: the Geigel algorithm [2]–[4]. In Fig. 3, for comparison purpose, we assume $\Delta = 0$ because the other methods assume the echo path is time-invariant. In this figure " false alarms" and "detection" refer only to double-talk detection, and that there is no path change. There are $L = 512$ data samples and $G = 0.15$; the powers of the far-end and the near-end signals are equal. This figure shows that our proposed algorithm outperforms other methods because the pdfs $p_1\left(\|e\|^2 \middle| \Delta = 0\right)$ and $p_2(\|e\|^2)$ will be well separated as seen in Fig. 2(a). The reason for the perfect performance of the IMLC method is that IMLC method takes advantage of an auxiliary MLS signal in order to measure the room response.

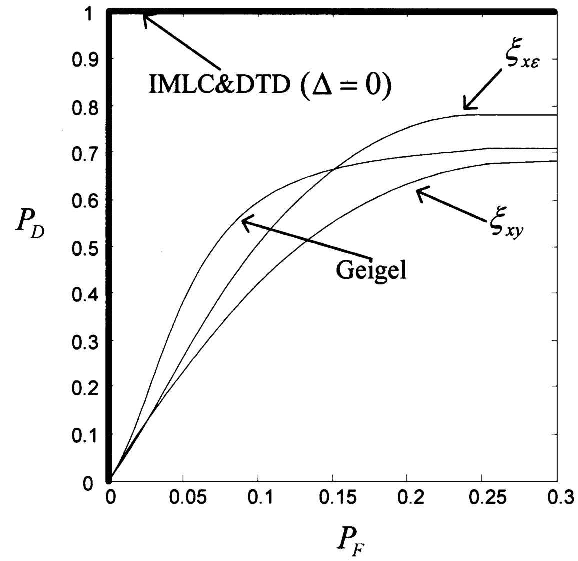

Fig. 3. Comparison of the ROC curves with other decision methods.

V. Conclusion

We have presented a new DTD method for the IMLC AEC structure. Following a likelihood ratio test of the coefficient error and the assumption of uniformly distributed echo-path-change, an ROC curve is plotted and shows the tradeoff of double-talk detection probability and false-alarm probability. From this ROC curve, a decision rule can be used to find a proper size x for DTD. Computer simulations show the DTD method is very effective.

References

[1] M. J. Crocker, *Handbook of Acoustics*. New York: Wiley, 1998.
[2] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. COMM-26, pp. 647–653, May 1978.
[3] H. Ye 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.
[4] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancellers," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 718–724, Nov. 1999.
[5] J. C. Jenq and S. F. Hsieh, "Acoustic echo cancellation using iterative-maximal-length correlation and double-talk detection," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 932–942, Nov. 2001.
[6] J. F. Doherty *et al.*, "A robust echo canceler for acoustic environments," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 389–396, May 1997.
[7] L. L. Scharf, *Statistical Signal Processing Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
[8] M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Amer.*, vol. 66, no. 2, pp. 497–500, 1979.