Echo Cancellation - A Likelihood Ratio Test for Double-talk Versus Channel Change

Neil J. Bershad, Jean-Yves Tourneret

Echo Cancellation—A Likelihood Ratio Test for Double-Talk Versus Channel Change

Neil J. Bershad, *Fellow, IEEE*, and Jean-Yves Tourneret, *Member, IEEE*

***Abstract*—Echo cancellers (ECs) are in wide use in both electrical (four-wire to two-wire mismatch) and acoustic (speaker–microphone coupling) applications. One of the main design problems is the control logic for adaptation. Basically, the algorithm weights should be frozen in the presence of double-talk and adapt quickly in the absence of double-talk. The control logic can be quite complicated since it is often not easy to discriminate between the echo signal and the near-end speaker. This paper derives a log-likelihood ratio test (LRT) for deciding between double-talk (freeze weights) and a channel change (adapt quickly) using a stationary Gaussian stochastic input signal model. The probability density function (pdf) of a sufficient statistic under each hypothesis is obtained, and the performance of the test is evaluated as a function of the system parameters. The receiver operating characteristics (ROCs) indicate that it is difficult to correctly decide between double-talk and a channel change based upon a single look. However, postdetection integration of approximately 100 sufficient statistic samples yields a detection probability close to unity (0.99) with a small false-alarm probability (0.01).**

***Index Terms*—Echo cancellation, channel change, double-talk, likelihood ratio test.**

I. Introduction

ECHO cancellers (ECs) have been used in networks for voice quality enhancement for several decades. There are two different kinds of applications for ECs. The network or hybrid echo on the public switched telephone network (PSTN) is caused by the four-wire to two-wire impedance mismatch. This mismatch results in unwanted reflection of transmitted energy back to the speaker or the source. Networks are equipped with ECs, known as network or line ECs, to remove these unwanted reflections. The International Telecommunication Union's (ITU's) Recommendation ITU-T G.168 2002 [1] specifies the minimum requirements and test conditions for performance of network ECs in the PSTN. Acoustic echo is another kind of echo which occurs widely in digital applications. Acoustic echo is the coupling of the received voice and the mouthpiece of a mobile handset or the coupling of the speaker and microphone of a hands-free mobile phone. Acoustic echo is typically more complex than the hybrid or network echo, and the echo delays are much longer.

N. J. Bershad is with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697 USA (e-mail: bershad@ece.uci.edu).

J.-Y. Tourneret is with IRIT-ENSEEIHT-TéSA, 31071 Toulouse Cedex 7, France (e-mail: Jean-Yves.Tourneret@enseeiht.fr).

The echo cancellation problem has been studied by many authors [2], [3] for more than 30 years. There are many similar design issues and parameters for acoustic and network echo cancellation. The two main design problems are 1) choice of adaptation algorithm(s) and 2) control logic for adaptation. The latter design problem is caused by double-talk. The EC observes the channel input vector and the scalar error signal. The error signal can consist of both double-talk (near-end speaker) and/or the uncancelled outgoing signal due to the far-end speaker. Specific control logic involves monitoring the error signal as well as the channel input vector (to handle nonstationary voice). Significant increases in the error signal power can be due to either double-talk or a channel change (ignoring voice nonstationarities). The algorithm weights should be frozen in the presence of double-talk and adapt quickly when there is a channel change.

The control logic can be quite complicated [2] since it is often not easy to discriminate between the echo signal and the near-end speaker. The primary problem is due to the nonstationarity of the channel input. There are many schemes described in both [2] and [3] for deciding when to adapt the adaptive filter weights [4]–[11]. Reference [4, p. 1717] states, "It should be noted that none of these detectors alone is yet sufficient to control the acoustic echo cancellation filter reliably," and "However, a combination of detectors is quite difficult and a lot of heuristics is involved." The details of these schemes will not be discussed here. Suffice it to say, to our knowledge, these or other schemes are not based on any optimum statistical tests such as a likelihood ratio test (LRT) [12, p. 34]. The principle reason for this lack is the difficulty modeling the nonstationarity of the voice data.

This paper derives a LRT for deciding between double-talk (freeze weights) and a channel change (adapt quickly) using a stationary Gaussian stochastic signal model. The LRT is then simplified to a sufficient statistic (a function of the observables that depends upon which hypothesis is true) to obtain an optimum test statistic. The probability density function (pdf) of the test statistic under each hypothesis is obtained and the performance of the test statistic is evaluated as a function of the system parameters. This performance is represented through receiver operating characteristics (ROCs) [12, p. 38]. These curves show the probability of detection (P_{D}) (deciding one hypothesis is true when it is actually true) versus probability of false alarm (P_{FA}) (deciding the same hypothesis is true when it is actually not true). The ROCs indicate that it is difficult to correctly decide between double-talk and a channel change based upon a single look. However, postdetection integration of about 100 successive LRT samples yields a P_{D} close to unity (0.99) with a small

P_{FA} (0.01). Note that the application of ROCs to the double-talk detection problem has been studied in [13]. The paper compares the performance of three different double-talk detectors using Monte Carlo simulations with real voice data and real channels. The simulations are required since no pdf's are available for these detectors. Our paper differs from [13] in that it considers channel changes and derives theoretical ROCs for the test.

The stationary signal model is not necessarily representative of speech since speech is highly nonstationary. However, as is usually the case with parametric signal models, the theoretical results are suggestive of good signal processing techniques. For example, the theoretical results for the optimum LRT provide upper bounds on the performance of any other test—i.e., one cannot do any better with any other test.

A particular EC structure (Fig. 1) is assumed in order to introduce the many parameters needed for the LRT. The EC consists of a nonadaptive main filter $\mathbf{H}_1$ and an adaptive shadow filter $\mathbf{H}_0$ [14]. The output of the main filter is subtracted from the echo to obtain the cancelled echo $e_{\mathrm{m}}(n)$. The shadow filter weights are adapted continuously and periodically transferred to the main filter using control logic based on measurements of various input parameters such as the far-end signal and received echo powers [5], for example. Consider the basic behavior of the EC when double-talk occurs or when a channel change occurs. Assume that the system is initially in steady state so that $\mathbf{H}_1 = \mathbf{H}_0$ and the two filter short-term time-averaged error powers $\widetilde{e_{\mathrm{s}}^2}(n)$ and $\widetilde{e_{\mathrm{m}}^2}(n)$ are small. Suppose double-talk occurs suddenly at time n_1. The two error powers now become large because of the double-talk. The shadow filter (incorrectly) adapts using this large error power and no longer matches the unknown channel. No transfer from the shadow filter to the main filter should occur. However, because the power of the double-talk is usually large compared with the error powers of the two filters prior to the appearance of the double-talk, $\widetilde{e_{\mathrm{s}}^2}(n)$ and $\widetilde{e_{\mathrm{m}}^2}(n)$ are primarily due to the double-talk. Thus, it is difficult to decide to transfer the weights from the shadow to the main filter using only the error powers of the two filters. On the other hand, suppose a channel change occurs at time n_1. The shadow filter now (correctly) adapts on this channel change. After some time, $\widetilde{e_{\mathrm{s}}^2}(n) < \widetilde{e_{\mathrm{m}}^2}(n)$ and a transfer from the shadow filter to the main filter should occur. This is an easy decision if one can wait long enough to detect the changes in the error powers. However, how can one determine the difference between the double-talk and a channel change when both events cause the shadow filter to immediately adapt? How should one make these decisions in an efficient manner based upon only the channel input and the outputs of the shadow and main filters? Some answers to these questions will be addressed in this paper.

Section II defines an hypothesis test based on the likelihood functions for double-talk versus a channel change. This hypothesis test yields a sufficient statistic for this problem. Section III derives the pdf of the sufficient statistics under both hypothesis. Section IV presents ROCs for different sets of parameters. A suboptimum postdetection integration procedure based on multiple samples of the sufficient statistic is proposed in Section V. The performance of this postdetection integrator is evaluated using Monte Carlo (MC) methods. Finally, Section VI applies this theory to full EC implementations for two cases: 1) a synthetically generated data model and 2) real voice data transmitted over a real channel. Some results and conclusions are reported in Section VII.

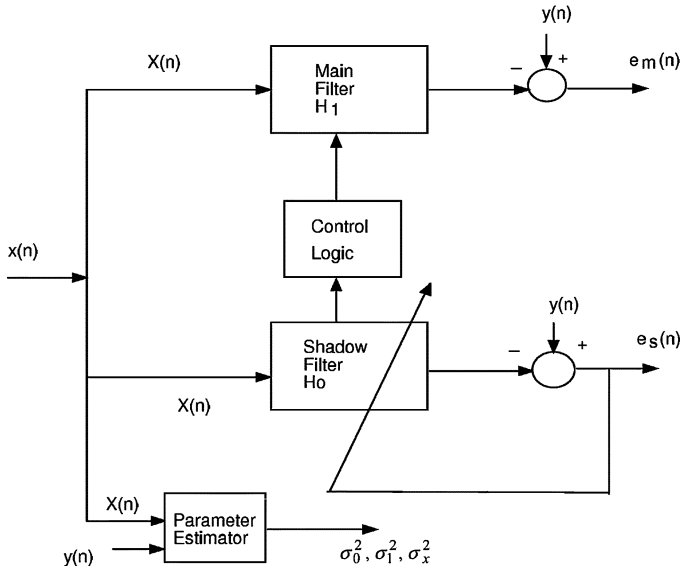


Fig. 1. Basic EC structure.

II. Hypothesis Test

Two of the primary signals that the EC uses for the control logic are the error signal $e_{\mathrm{m}}(n)$ (canceller output) and $e_{\mathrm{s}}(n)$ (shadow filter error signal). Whenever the powers of the error signals increase significantly over some quiescent level, the EC needs to decide whether the increase is due to double-talk or to a channel change. Either occurrence will cause a significant increase in the error powers. A statistical hypothesis testing problem is defined in what follows, which models these two possible events. It is assumed that the EC in Fig. 1 is able to accurately estimate the background noise power σ_0^2, the signal power σ_x^2, and the double-talk signal power σ_1^2. These powers are assumed to be time invariant, at least over the time interval of the data used in the hypothesis test.

A. Signal and Channel Models

The channel input vector $\mathbf{x}(n)$ is of dimension $N \times 1$ with $E[\mathbf{x}(n)\mathbf{x}^T(n)] = \sigma_x^2\mathbf{I}_N$ ($\mathbf{I}_N$ is the $N \times N$ identity matrix) and the channel output is a scalar $y(n)$. This paper assumes that $[y(n), \mathbf{x}^T(n)]^T$ is a zero-mean Gaussian vector. Let

$$\mathcal{H}_1 : y(n) \text{ is due to double-talk}$$
$$\mathcal{H}_0 : y(n) \text{ is due to a channel change.}$$

This choice is arbitrary; the reverse is also possible. Under $\mathcal{H}_1$

$$y(n) = \mathbf{x}^T(n)\mathbf{H}_1 + n_0(n) + n_1(n) \tag{1}$$

where $\mathbf{H}_1$ is an unknown channel that has been correctly identified prior to time n using the adaptive shadow filter and transferred to the main channel filter. The additive noise $n_0(n)$ is

stationary zero-mean white Gaussian, independent of $\mathbf{x}(n)$ with $E[n_0^2(n)] = \sigma_0^2$. The second additive noise $n_1(n)$, modeling the double-talk, is also zero-mean white Gaussian, and independent of both $\mathbf{x}(n)$ and $n_0(n)$ with $E[n_1^2(n)] = \sigma_1^2$. Under $\mathcal{H}_0$

$$y(n) = \mathbf{x}^T(n)\mathbf{H}_0 + n_0(n) \tag{2}$$

where $\mathbf{H}_0$ is a new unknown channel which is identified adaptively after time n using the shadow filter. It is assumed that no transfer from the shadow filter to the main filter occurs until after the hypothesis test has been performed. Thus, $\mathbf{H}_1$ is the main filter weights, and $\mathbf{H}_0$ is the shadow filter weights after convergence.[1]

Hence, all the parameters are known for the hypothesis test. Straightforward calculation yields

$$\begin{aligned} E\left[y^2(n)|\mathcal{H}_1\right] &= \sigma_x^2\mathbf{H}_1^T\mathbf{H}_1 + \sigma_0^2 + \sigma_1^2 \\ E\left[y^2(n)|\mathcal{H}_0\right] &= \sigma_x^2\mathbf{H}_0^T\mathbf{H}_0 + \sigma_0^2 \\ E\left[y(n)\mathbf{x}(n)|\mathcal{H}_1\right] &= \sigma_x^2\mathbf{H}_1 \\ E\left[y(n)\mathbf{x}(n)|\mathcal{H}_0\right] &= \sigma_x^2\mathbf{H}_0. \end{aligned} \tag{3}$$

Thus, the joint pdf of $\mathbf{v}(n) = [y(n), \mathbf{x}^T(n)]^T$ is Gaussian such that

$$p\left[\mathbf{v}(n)|\mathcal{H}_i\right] \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i), \quad i = 0, 1 \tag{4}$$

where $\mathbf{R}_i$ is the following $(N+1)\times(N+1)$ matrix:

$$\mathbf{R}_i = \sigma_x^2\mathbf{I}_{N+1} + \sigma_x^2\begin{pmatrix} a_i & \mathbf{H}_i^T \\ \mathbf{H}_i & \mathbf{0}_N \end{pmatrix} \tag{5}$$

where $\mathbf{0}_N$ is a $N \times N$ matrix of zeroes and where

$$a_0 = -1 + \mathbf{H}_0^T\mathbf{H}_0 + \frac{\sigma_0^2}{\sigma_x^2}, \quad a_1 = -1 + \mathbf{H}_1^T\mathbf{H}_1 + \frac{\sigma_0^2 + \sigma_1^2}{\sigma_x^2}.$$

B. Log-Likelihood Ratio Test

The log LRT for (4) accepts hypothesis $\mathcal{H}_1$ when

$$\ln\left[\frac{p\left(\mathbf{v}(n)|\mathcal{H}_1\right)}{p\left(\mathbf{v}(n)|\mathcal{H}_0\right)}\right] = \frac{1}{2}\mathbf{v}^T(n)\left(\mathbf{R}_0^{-1} - \mathbf{R}_1^{-1}\right)\mathbf{v}(n) + \frac{1}{2}\ln\left(\frac{|\mathbf{R}_0|}{|\mathbf{R}_1|}\right) \tag{6}$$

exceeds an appropriate threshold [12, p. 34]. Since the last term in this expression is not a function of the observables, the log LRT simplifies to

$$\mathbf{v}^T(n)\left(\mathbf{R}_0^{-1} - \mathbf{R}_1^{-1}\right)\mathbf{v}(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} T_1 \tag{7}$$

[1]If $\mathcal{H}_1$ is actually true (double-talk present), it will not be possible for the shadow filter to adapt and learn the true value of $\mathbf{H}_0$ needed for the test until the double-talk disappears. Instead, the output of the shadow filter can be used as if it has correctly estimated $\mathbf{H}_0$. The EC examples in Section VI function this way. The outputs of the main and shadow filters are used in the test statistic to decide whether or not to transfer the shadow filter to the main filter. As can be seen, the poor estimates of $\mathbf{H}_0$ during the double-talk period do not affect the transfer logic—no transfers occur during double-talk. Hence, it is not critical to the test not to have a good estimate of $\mathbf{H}_0$ during double-talk. To summarize, our mathematical model is robust with respect to this problem.

where T_1 is a threshold setting determined by P_{D} and P_{FA}. One decides $\mathcal{H}_1$ is true if the log LR exceeds the threshold T_1 and decides $\mathcal{H}_0$ otherwise. The threshold T_1 is selected so as to yield a given performance for the test. Equation (7) is a quadratic form in the observables whose matrix inverses need to be evaluated. Usually, this can be a formidable problem. However, the inverses can be evaluated here because $\mathbf{R}_0$ and $\mathbf{R}_1$ are each an identity matrix plus a rank-2 matrix. As a result, the inverse problem reduces to the following eigenvalue–eigenvector problem:

$$\mathbf{M}_k\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \tag{8}$$

where

$$\mathbf{M}_k = \begin{pmatrix} a_k & \mathbf{H}_k^T \\ \mathbf{H}_k & \mathbf{0}_N \end{pmatrix}. \tag{9}$$

Following the techniques in [15], solving (8) yields

$$\begin{aligned} \lambda_{1,k} &= \frac{a_k + \sqrt{a_k^2 + 4\mathbf{H}_k^T\mathbf{H}_k}}{2} \\ \lambda_{2,k} &= \frac{a_k - \sqrt{a_k^2 + 4\mathbf{H}_k^T\mathbf{H}_k}}{2}, \quad k = 0, 1. \end{aligned}$$

The eigenvectors $\boldsymbol{\phi}_{i,j}$, $i = 1,2$ and $j = 0,1$, are given by

$$\boldsymbol{\phi}_{i,j} = \frac{1}{\sqrt{\lambda_{i,j}^2 + \mathbf{H}_j^T\mathbf{H}_j}}\begin{pmatrix} \lambda_{i,j}^2 \\ \mathbf{H}_j \end{pmatrix}. \tag{10}$$

Using (8) and (10), the following result can be obtained:

$$\sigma_x^2\mathbf{R}_j^{-1} = \mathbf{I}_{N+1} - \sum_{i=1}^{2}\frac{\lambda_{i,j}}{1+\lambda_{i,j}}\boldsymbol{\phi}_{i,j}\boldsymbol{\phi}_{i,j}^T, \quad j = 0, 1. \tag{11}$$

Inserting (11) in (7) yields

$$\frac{1}{\sigma_x^2}\mathbf{v}^T(n)\left[\sum_{i=1}^{2}\sum_{j=0}^{1}\frac{(-1)^{j+1}\lambda_{i,j}}{1+\lambda_{i,j}}\boldsymbol{\phi}_{i,j}\boldsymbol{\phi}_{i,j}^T\right]\mathbf{v}(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} T_1. \tag{12}$$

Hence, inserting (10) in (12) and performing the matrix multiplications yields the test statistic

$$\sum_{i=1}^{2}\sum_{j=0}^{1}(-1)^{j+1}k_{ij}\left[y(n) + \frac{\mathbf{x}^T(n)\mathbf{H}_j}{\lambda_{i,j}}\right]^2 \tag{13}$$

where

$$k_{i,j} = \frac{\lambda_{i,j}}{1+\lambda_{i,j}} \cdot \frac{\lambda_{i,j}^2}{\lambda_{i,j}^2 + \mathbf{H}_j^T\mathbf{H}_j}.$$

Expanding (13) and, ignoring terms that do not change under either hypothesis yields the following sufficient statistic for the test:

$$Ky^2(n) + 2\sum_{i=1}^{2}\sum_{j=0}^{1}\frac{(-1)^{j+1}k_{ij}}{\lambda_{i,j}}y(n)\mathbf{x}^T(n)\mathbf{H}_j$$

where

$$K = \sum_{i=1}^{2}\sum_{j=0}^{1}(-1)^{j+1}k_{i,j}.$$

By dividing by K and noting that $\lambda_{1,k}\lambda_{2,k} = -\mathbf{H}_k^T\mathbf{H}_k$, some algebra leads to the following test

$$\gamma(n) = y(n)z(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} T_2, \tag{14}$$

where

$$z(n) = y(n) + 2\left(\frac{\sigma_0^2}{\sigma_1^2}\right)\mathbf{x}^T(n)\mathbf{H}_1 - 2\left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_1^2}\right)\mathbf{x}^T(n)\mathbf{H}_0. \tag{15}$$

Thus, the sufficient statistic is the product of two zero-mean correlated Gaussian variates. Here, $y(n)$ is the channel output at time n and $z(n)$ is a linear combination of the channel output, the scaled output of the main filter, and the scaled output of the shadow filter. The scalings are simply the ratio of the additive channel noise power to the double-talk signal power, and 1 plus this ratio. Note that the test statistic does not depend on the input signal power but that the performance of the test does. Thus, one observes two interesting situations as limiting cases. When the double-talk is large in comparison to the background noise and the channel input, i.e., $\sigma_0^2/\sigma_1^2 \to 0$ and $\sigma_x^2/\sigma_1^2 \to 0$, we obtain

$$\lim_{\sigma_1^2/\sigma_0^2\to\infty}\gamma(n) = y(n)\left[y(n) - 2\mathbf{x}^T(n)\mathbf{H}_0\right] \approx y^2(n).$$

Hence, one just measures the power in the channel output, agreeing with intuition. On the other hand, if the double-talk is small in comparison to the background noise and the channel input, i.e., $\sigma_1^2/\sigma_0^2 \to 0$ and $\sigma_1^2/\sigma_x^2 \to 0$, we obtain

$$\gamma(n) \approx 2\frac{\sigma_0^2}{\sigma_1^2}y(n)\mathbf{x}^T(n)[\mathbf{H}_1 - \mathbf{H}_0].$$

Thus, one cross-correlates the channel output with the channel input vector and weights the resultant with the difference between the two channel vectors. The use of cross-correlation is well known [16]. In the general case, the test is a combination of a power measurement of $y(n)$ and the weighted cross-correlation vector $y(n)\mathbf{x}(n)$. The nice feature of $\gamma(n)$ in the general case is that it indicates how to optimally combine these two measurements. The next section derives the pdf of $\gamma(n)$ under either hypothesis.

III. Pdf of the Sufficient Statistic

Since $\mathbf{v}(n) = [y(n), \mathbf{x}^T(n)]^T$ is a zero-mean Gaussian vector, it follows that $y(n)$ is a zero-mean scalar Gaussian variate with variance given by (3) under the two hypothesis. Here, $z(n)$ is also a zero-mean scalar Gaussian variate with a variance that can be computed from (15), and $[y(n), z(n)]^T$ is linearly related to $\mathbf{v}(n) = [y(n), \mathbf{x}^T(n)]^T$ through the matrix relation

$$\begin{bmatrix} y(n) \\ z(n) \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ 1 & 2\alpha\mathbf{H}_1^T - 2\beta\mathbf{H}_0^T \end{bmatrix}\begin{bmatrix} y(n) \\ \mathbf{x}(n) \end{bmatrix} \tag{16}$$

where $\alpha = \sigma_0^2/\sigma_1^2$ and $\beta = (\sigma_0^2+\sigma_1^2)/\sigma_1^2$. Thus, $[y(n), z(n)]^T$ is a Gaussian vector with mean $[0,0]^T$ and covariance matrix $\mathbf{\Sigma}_i$, $i = 0{,}1$ under hypothesis $\mathcal{H}_i$ ($i = 0{,}1$), where

$$\begin{aligned}\mathbf{\Sigma}_i &= \begin{bmatrix} 1 & \mathbf{0}^T \\ 1 & 2\alpha\mathbf{H}_1^T - 2\beta\mathbf{H}_0^T \end{bmatrix}\mathbf{R}_i\begin{bmatrix} 1 & 1 \\ \mathbf{0} & 2\alpha\mathbf{H}_1 - 2\beta\mathbf{H}_0 \end{bmatrix} \\ &= \sigma_x^2\begin{bmatrix} m_{11}^i & m_{12}^i \\ m_{21}^i & m_{22}^i \end{bmatrix}\end{aligned} \tag{17}$$

with

$$\begin{aligned} m_{11}^i &= 1 + a_i \\ m_{12}^i &= 1 + a_i + 2\mathbf{H}_i^T(\alpha\mathbf{H}_1 - \beta\mathbf{H}_0) = m_{21}^i \\ m_{22}^i &= 1 + a_i + 4\left(\alpha\mathbf{H}_i^T\mathbf{H}_1 - \beta\mathbf{H}_i^T\mathbf{H}_0\right) \\ &\quad + 4\left(\alpha\mathbf{H}_1^T - \beta\mathbf{H}_0^T\right)(\alpha\mathbf{H}_1 - \beta\mathbf{H}_0)\end{aligned}$$

where a_i has been defined below (5). The joint pdf of $y(n)$ and $z(n)$ under hypothesis $\mathcal{H}_i$ can be written

$$p_i(y,z) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}_i|}}\exp\left[-\frac{1}{2}(y,z)\mathbf{\Sigma}_i^{-1}(y,z)^T\right] \tag{18}$$

where

$$\mathbf{\Sigma}_i^{-1} = \frac{1}{\sigma_x^2\left(m_{11}^i m_{22}^i - m_{12}^i m_{21}^i\right)}\begin{bmatrix} m_{22}^i & -m_{12}^i \\ -m_{21}^i & m_{11}^i \end{bmatrix}. \tag{19}$$

Since y and z are jointly Gaussian with zero means, the pdf of the product $u = yz$ is given by [17, p. 45]

$$p_i(u) = \frac{\exp\left[-u(\mathbf{\Sigma}_i^{-1})_{12}\right]}{\pi\sqrt{|\mathbf{\Sigma}_i|}}K_0\left[|u|\sqrt{\left(\mathbf{\Sigma}_i^{-1}\right)_{11}\left(\mathbf{\Sigma}_i^{-1}\right)_{22}}\right] \tag{20}$$

where K_0 is the modified Bessel function of the second kind and of zero order. It is interesting to note that the two channels are related to the pdf $p_i(u)$ through $\mathbf{H}_1^T\mathbf{H}_0$, $\mathbf{H}_0^T\mathbf{H}_0$, and $\mathbf{H}_1^T\mathbf{H}_1$. Consequently, any pair of channels with the same values for these three parameters will yield the same detection performance for a given value of $(\sigma_x^2, \sigma_0^2, \sigma_1^2)$. Note also that when $\sigma_0^2 = 0$, the covariance matrix $\mathbf{\Sigma}_0$ is singular. In this case, the pdf of the product $u = yz$ under hypothesis $\mathcal{H}_0$ reduces to $p_0(u) = [-2\pi\sigma_x^2(\mathbf{H}_0^T\mathbf{H}_0)u]^{-1/2}\exp\{u/[2\sigma_x^2(\mathbf{H}_0^T\mathbf{H}_0)]\}$, i.e., $u/[-(\mathbf{H}_0^T\mathbf{H}_0)\sigma_x^2]$ is distributed according to a χ^2 distribution with one degree of freedom.

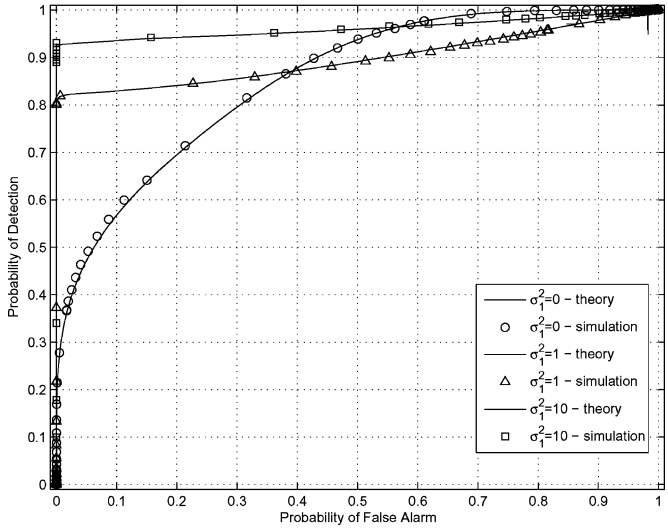


Fig. 2. Comparison of theory and MC simulations. $P_{\rm D}$ versus $P_{\rm FA}$ for different double-talk power levels σ_1^2 with $\sigma_x^2 = 1, \sigma_0^2 = 0, N = 1024$, and orthogonal channels.

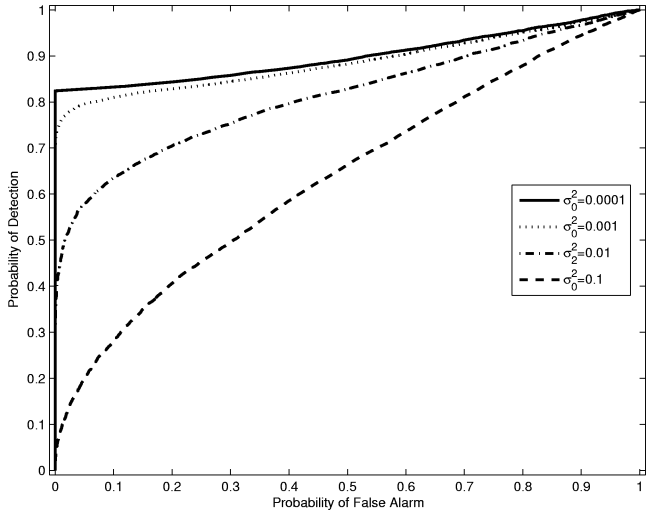


Fig. 3. $P_{\rm D}$ versus $P_{\rm FA}$ (MC simulations) for different values of σ_0^2 with $\sigma_x^2 = 1, \sigma_1^2 = 1, N = 1024$, and orthogonal channels.

IV. Performance Curves

A. Theoretical Curves

The performance of the sufficient statistic can be defined by the two following probabilities [12, p. 34]:

$$P_{\rm D} = P[\text{accepting } \mathcal{H}_1 | \mathcal{H}_1 \text{ is true}] = \int_T^\infty p_1(u)du \quad (21)$$

$$P_{\rm FA} = P[\text{accepting } \mathcal{H}_1 | \mathcal{H}_0 \text{ is true}] = \int_T^\infty p_0(u)du. \quad (22)$$

Alternatively, if $\mathcal{H}_0$ is critical

$$1 - P_{\rm FA} = P[\text{accepting } \mathcal{H}_0 | \mathcal{H}_0 \text{ is true}],$$
$$1 - P_{\rm D} = P[\text{accepting } \mathcal{H}_0 | \mathcal{H}_1 \text{ is true}].$$

Thus, for each value of T, there exists a pair $(P_{\rm FA}, P_{\rm D})$. The curves of P_D as a function of P_{FA} are called ROCs [12, p. 38].

B. Monte Carlo Simulations

A set of 10 000 MC simulations has been run for the sufficient statistic in (13) as a check on the theoretical results obtained in (20)–(22). Fig. 2 shows some typical ROCs for $N = 1024$ and different parameter selections. Here, $\mathbf{H}_0$ and $\mathbf{H}_1$ are two one-sided exponential channels with attenuation between successive taps

$$\mathbf{H}_i(j) = \begin{cases} c(0.95)^{j-\Delta_i}, & j \geq \Delta_i \\ 0, & \text{otherwise} \end{cases}$$

where Δ_i is a relative delay of the individual channel $\mathbf{H}_i$ and the parameter c is defined by the filter gain, which is $G = \mathbf{H}_1^T\mathbf{H}_1 = \mathbf{H}_0^T\mathbf{H}_0$. Two cases will be considered here: the first one is defined by $G = 0.1$, corresponding to a -10-dB channel gain (typical in electrical applications); the second case is defined by $G = 4$, corresponding to an acoustic channel (with a 6-dB gain). Each filter is effectively about 80 taps. The two filters differ only in a bulk delay (a difference of more than 200 taps for the orthogonal case).

Excellent agreement between the theory and MC simulations was obtained over all values of $P_{\rm D}$ and $P_{\rm FA}$. Fig. 2 shows the ROCs for different double-talk powers, no additive noise, and orthogonal channels ($\mathbf{H}_1^T\mathbf{H}_0 = 0$) with $G = 0.1$. It is seen that a $P_{\rm D}$ approaching unity requires a fairly large $P_{\rm FA}$, even with no additive background noise. Fig. 2 displays the relatively poor behavior with no background noise because 1) the sufficient statistic is noncoherent (quadratic in the data rather than linear) and 2) only one time sample of the data vector is used in the decision.

C. Using the MC Simulations to Validate the Theory

Some numerical integration problems were encountered using (21) and (22), as the tails of the density functions are not particularly well behaved. Because Fig. 2 showed excellent agreement between the theory and MC simulations, it was decided to display the ROC curves generated from the MC simulations instead. Thus, the ROC curves in the subsequent figures were obtained using 10 000 MC simulations rather than by direct integration. This approach was also useful when obtaining ROCs for a postdetection integration scheme presented in Section V. Fig. 3 shows the effect of decreasing the background noise power on the ROC curves. The improvement in performance asymptotically approaches the top curve as the background noise power approaches zero. Hence, the hypothesis test defined by (14) is not noise-limited. Fig. 4 shows that the performance of the sufficient statistic does not increase monotonically with increasing levels of double-talk. This agrees with physical intuition. At very low levels of double-talk, the double-talk is buried in the background noise. Thus, the channel output dominates the test statistic. As the double-talk power level increases, the channel output is somewhat obscured by the double-talk and the performance of the

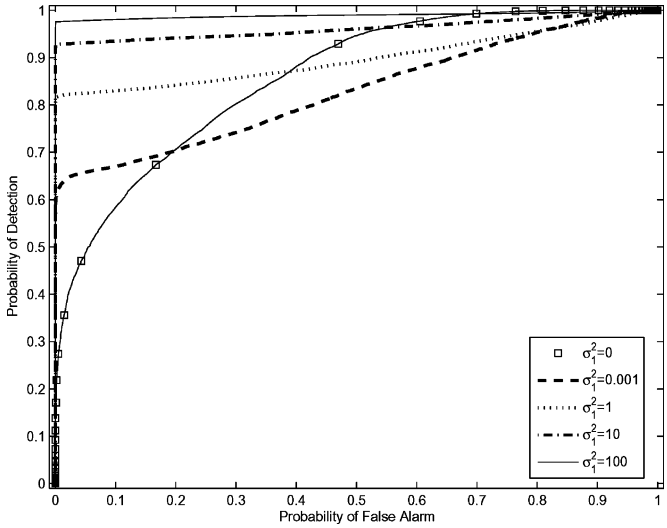


Fig. 4. $P_{\rm D}$ versus $P_{\rm FA}$ (MC simulations) for different values of σ_1^2, $\sigma_x^2 = 1$, $\sigma_0^2 = 0.001$, $N = 1024$, and orthogonal channels.

test statistic decreases. Eventually, the double-talk power dominates, and the performance again improves. This effect occurs because of the noncoherent nature of the sufficient statistic.

V. Postdetection Integration

The previous ROCs suggest that one time sample of the sufficient statistic is not enough to make a reliable decision. Thus, one would like to derive the sufficient statistic for p time samples of the vector $\mathbf{v}(j) = [y(j), \mathbf{x}^T(j)]^T$ for $j = n - p + 1, \ldots, n$. The problem with this approach is that inversion of the covariance matrix of the p data vectors is extremely difficult unless it is assumed that successive time samples are independent. This is not a viable or useful assumption because both the sequences $y(j)$ and $\mathbf{x}(j)$ are strongly correlated for different j, $y(j)$ through the memory of the channel and $\mathbf{x}(j)$ through the tapped delay line structure of the adaptive filter in the EC.

A way to get around this statistical problem is to use the MC simulation approach. Consider the time averaged sufficient statistic

$$\Gamma(n) = \frac{1}{p} \sum_{m=n-p+1}^{n} y(m)z(m). \tag{23}$$

A total of 10 000 MC simulations of (23) were run for orthogonal channels and channels whose differential delay is $\Delta = \Delta_1 - \Delta_0 = 1$. It is straightforward to show that a differential delay of 200 taps yields essentially orthogonal channels (for the one-sided exponential channel used in this paper). Figs. 5(a) and (b) and 6(a) and (b) show the resulting ROCs for different values of p and $\sigma_0^2 = \sigma_1^2 = 1$. It can be seen that $p = 100$ yields excellent ROCs ($P_{\rm D}$ approaches unity with a small $P_{\rm FA}$) for all cases except for the nonorthogonal 6-dB case. For the good cases, the $p = 100$ curves are in the extreme upper left hand corner of the figures and are difficult to discern. However, it is clear from the $p = 50$ curves that changes in delays of one tap can be detected with $P_{\rm D} \geq 0.9$ and $P_{\rm FA} \leq 0.01$. Fig. 6(b) indicates that the nonorthogonal 6-dB case would require $p > 1000$ to obtain good performance. This result suggests that it will be very difficult to differentiate double-talk and channel change due to a loss of synchronization (defined by $\Delta = \Delta_1 - \Delta_0 = 1$) for a 6-dB channel gain.

VI. Application to ECs

The LRT theory derived in this paper has been tested for two distinct examples in full EC implementations of Fig. 1 with transfer logic between the shadow and main filters modified to use the postdetection test statistic (23). The first example consists of a synthetically generated data set whose channel change and double-talk parameters are assumed known to the EC. The second example consists of real voice data transmitted over a real channel.

The first EC uses a partial Haar adaptive filter to estimate the bulk channel delay for sparse channels. It consists of a main filter (128 taps), an adaptive shadow filter (128 taps), and a second adaptive filter (256 taps) to handle sparse channels as described in [18]. The second adaptive filter operates on Haar transformed inputs to estimate the channel bulk delay. An overall channel delay of 1024 taps can be accommodated in this way. The 128-tap adaptive filter uses the affine projection (AP) algorithm of order 2. The 256-tap Haar adaptive filter uses the NLMS algorithm.

The second bench-tested EC uses a time-domain sub-sampling adaptive filter scheme (Duttweiler filter) as described in [19], instead of the Haar-based adaptive filter used in the first example. The EC structure consists of a main filter (158 taps), an adaptive shadow filter (158 taps), and a second adaptive filter (108 taps) to handle sparse channels. The second adaptive filter operates on sub-sampled inputs to estimate the channel bulk delay. An overall channel delay of 1024 taps can be accommodated in this way. The 158-tap adaptive filter uses the AP algorithm of order 2. The 108-tap adaptive filter also uses the AP algorithm of order 2.

A. Synthetic Data

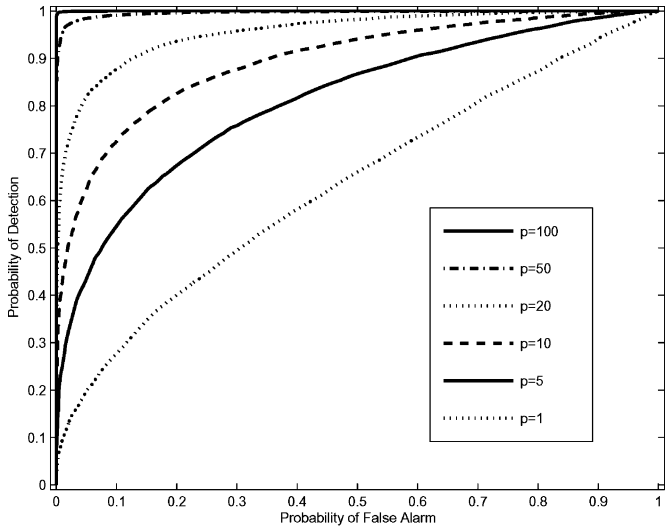
The input to the canceller and the unknown channel output was synthetically generated. The channel input $\mathbf{x}(n)$ consisted of four 1-s (8000 samples/s) sets of zero-mean white Gaussian variates with unit variance. The unknown channel output $y(n)$ consisted of four 1-s segments generated as follows:

$$y(n) = \begin{cases} \mathbf{x}^T(n)\mathbf{H}_1 + n_0(n), & n = 1, \ldots, 8000 \\ \mathbf{x}^T(n)\mathbf{H}_0 + n_0(n), & n = 8001, \ldots, 16000 \\ \mathbf{x}^T(n)\mathbf{H}_0 + n_0(n) + n_1(n), & n = 16001, \ldots, 24000 \\ \mathbf{x}^T(n)\mathbf{H}_0 + n_0(n), & n = 24001, \ldots, 32000 \end{cases}$$

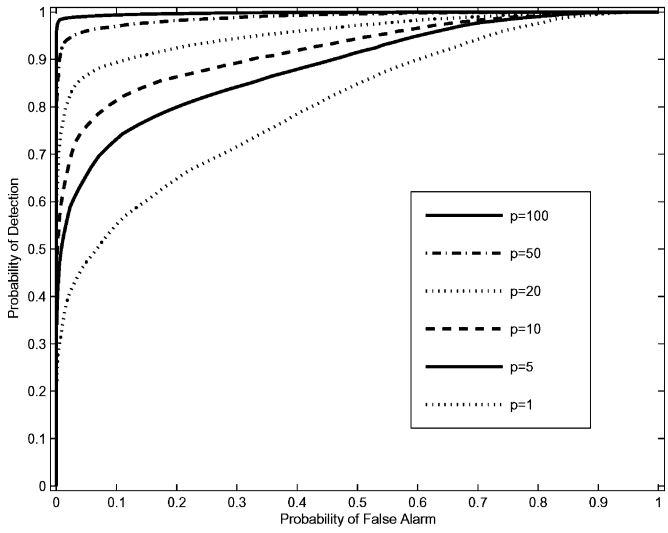
where $E[n_1^2(n)] = 1$ and $E[n_0^2(n)] = 10^{-3}$. Also, $\mathbf{H}_0$ corresponds to a time-delayed version of $\mathbf{H}_1$ as given in Section IV-B. Thus, $y(n)$ consist of channel changes at $n = 1$ and $n = 8001$, double-talk but no channel change at $n = 16001$, and the double-talk disappears at $n = 24001$ without another channel change. The parameters needed in (23) were set *a priori*, $y(n)$ was generated as above, and $\mathbf{x}^T(n)\mathbf{H}_1$ and $\mathbf{x}^T(n)\mathbf{H}_0$ were replaced by the outputs of the main and shadow filters, respectively.

The threshold setting for (23) was set at

$$T_2 = E\left[\gamma(n)|\mathcal{H}_0\right] + b\left\{E\left[\gamma(n)|\mathcal{H}_1\right] - E\left[\gamma(n)|\mathcal{H}_0\right]\right\} \tag{24}$$

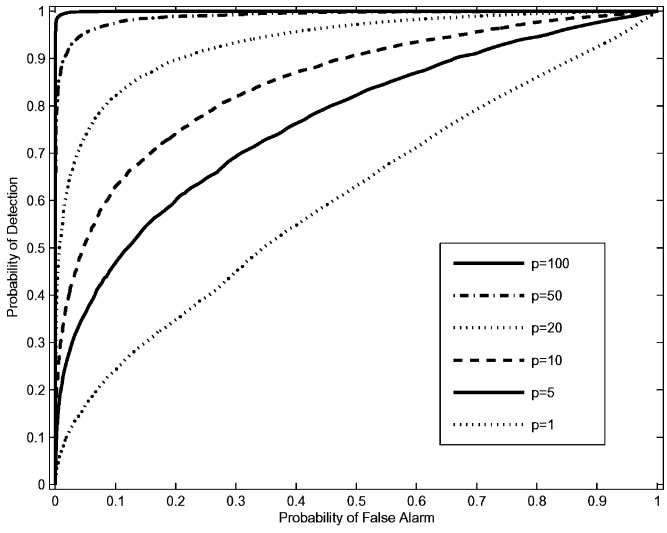


(a) $G = -10$ dB

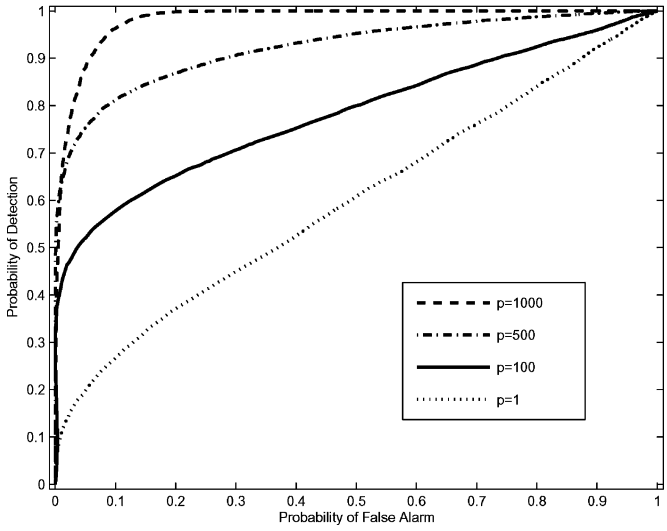


(b) $G = 6$ dB

Fig. 5. $P_{\rm D}$ versus $P_{\rm FA}$ (MC simulations) for postdetection integration of p samples of the LRT for (a) $G = -10$ dB and (b) $G = 6$ dB (orthogonal channels, $\sigma_x^2 = \sigma_1^2 = \sigma_0^2 = 1$, $N = 1024$).



(a) $G = -10$ dB



(b) $G = 6$ dB

Fig. 6. $P_{\rm D}$ versus $P_{\rm FA}$ (MC simulations) for postdetection integration of p samples of the LRT for (a) $G = -10$ dB and (b) $G = 6$ dB ($\Delta = 1$, $\sigma_x^2 = \sigma_1^2 = \sigma_0^2 = 1$, $N = 1024$).

where

$$E\left[\gamma(n)|\mathcal{H}_1\right] = \sigma_x^2\left(1 + 2\frac{\sigma_0^2}{\sigma_1^2}\right)\mathbf{H}_1^T\mathbf{H}_1 - 2\sigma_x^2\left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_1^2}\right)\mathbf{H}_1^T\mathbf{H}_0 + \sigma_0^2 + \sigma_1^2$$

and

$$E\left[\gamma(n)|\mathcal{H}_0\right] = \sigma_x^2\left(1 - 2\frac{\sigma_0^2+\sigma_1^2}{\sigma_1^2}\right)\mathbf{H}_0^T\mathbf{H}_0 + 2\sigma_x^2\left(\frac{\sigma_0^2}{\sigma_1^2}\right)\mathbf{H}_1^T\mathbf{H}_0 + \sigma_0^2.$$

Here, b is a scalar that controls controls the location of the threshold with respect to the means under the two hypothesis. When $b = 0$, $T_2 = E[\gamma(n)|\mathcal{H}_0]$ (at the mean under $\mathcal{H}_0$); when $b = 1$, $T_2 = E[\gamma(n)|\mathcal{H}_1]$ (at the mean under $\mathcal{H}_1$) and when $b = 0.5$, $T_2 = 0.5\{E[\gamma(n)|\mathcal{H}_1] + E[\gamma(n)|\mathcal{H}_0]\}$ (halfway between the two means). Here, b was chosen equal to 0.2 for the subsequent figures. Threshold T_2 can be related to $P_{\rm D}$ and $P_{\rm FA}$ through (22) for $p = 1$. This information is of limited value here since $p \geq 1$ and the ROCs are obtained from the MC simulations.

Figs. 7–9 display the mean-square error (MSE) (top curves), the number of transfers from the shadow filter to the main filter every 200 samples (middle curves), and the average sufficient statistic $\Gamma(n)$ and threshold T_2 (bottom curves) for $p = 100$, $p = 10$, and $p = 1$, respectively, as a function of the number of algorithm iterations. The MSE is defined as the uniformly weighted time average of the squared error over 100 adjacent time samples. The MSE begins at about 90 dB ($\sigma_x^2\mathbf{H}_1^T\mathbf{H}_1 = 0.1$ implies MSE $=$ 90 dB) and converges to about 70 dB ($\sigma_0^2 = 10^{-3}$ implies MSE $=$ 70 dB) for the first two channel changes. The channel changes are correctly detected in all cases. It takes about 400 ms for the shadow filter to adapt to the unknown channel and transfer this information to the main filter when the shadow filter is initialized at zero. It takes about 750 ms for the shadow filter to change from $\mathbf{H}_1$ to $\mathbf{H}_0$. The reason

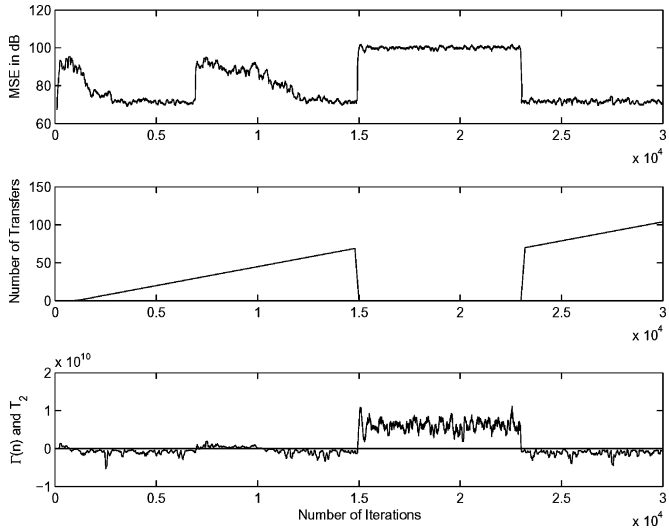


Fig. 7. EC performance for $p = 100$. (Top) MSE. (Middle) Number of transfers from the shadow filter to the main filter. (Bottom) Time-average sufficient statistic $\Gamma(n)$ and threshold T_2.

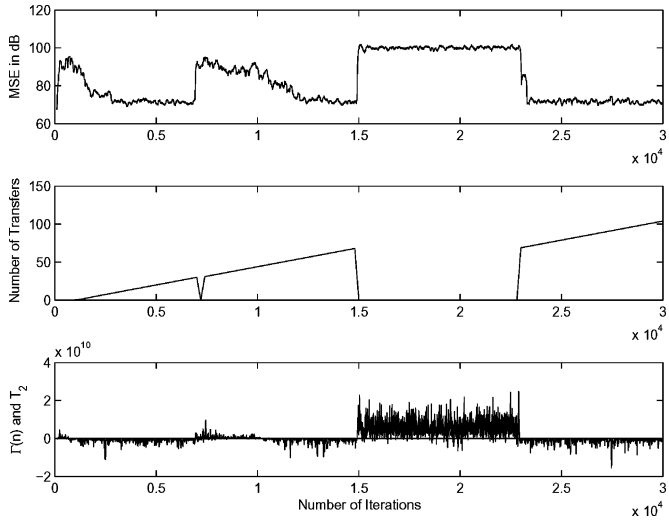


Fig. 8. EC performance for $p = 10$. (Top) MSE. (Middle) Number of transfers from the shadow filter to the main filter. (Bottom) Time-average sufficient statistic $\Gamma(n)$ and threshold T_2.

for this longer convergence time is due to the convergence time of the Haar adaptive filter. In the third phase, the MSE is determined by the double-talk, which is 30 dB above the noise floor (at 100 dB). The third phase is the most interesting. It demonstrates the sensitivity to double-talk as p varies. This is shown in the middle curves of Figs. 7–9. Fig. 7 shows that no transfers occur during double-talk when $p = 100$, whereas Fig. 9 shows that numerous transfers occur for $p = 1$. This behavior is supported by the MSE curves. Further support for this behavior is provided by the bottom curves of Figs. 7–9, which compare $\Gamma(n)$ and T_2. The fluctuations in $\Gamma(n)$ decrease as p increases, reflecting the change in the time averaging of $\Gamma(n)$. The case $p = 10$ displays mixed behavior with some sensitivity to double-talk. Note that the middle curves of Figs. 7–9 show $\Gamma(n) < T_2$ for some values of n, but no transfers occur. This is because of the *ad hoc* requirement for a transfer that all previous $p-1$ samples of $y(m)z(m)$ be less than T_2 (for additional double-talk protection). This prevents a transfer when previous

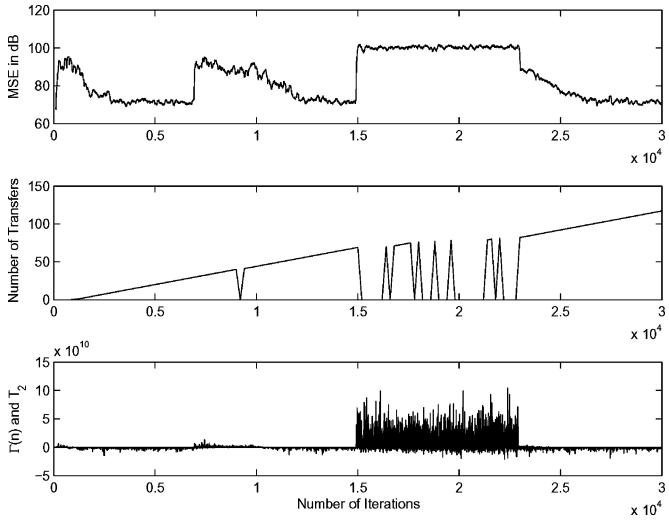


Fig. 9. EC performance for $p = 1$. (Top) MSE. (Middle) Number of transfers from the shadow filter to the main filter. (Bottom) Time-average sufficient statistic $\Gamma(n)$ and threshold T_2.

double-talk has badly affected the shadow filter weights before they have received a transfer back from the main filter. The bottom curves of Figs. 7–9 show that lowering T_2 (reducing b) will increase both P_{D} and P_{FA}. This will change the operating point $(P_{\mathrm{FA}}, P_{\mathrm{D}})$ to a different place on the ROC. The behavior displayed in Figs. 7–9 is in agreement with the ROCs shown in Figs. 5(a) and (b) and 6(a) and (b). When p is increased, no transfer from the shadow filter to the main filter occurs in the presence of double-talk, as is observed in the middle curves of Figs. 7–9.

B. Voice Data Over a Real Channel

The EC structure used in this example has been described previously. For comparison purposes, the LRT-based EC was obtained from the conventional EC with only one modification. The logic for transfer from the shadow to main filters was changed in accordance with (23) and (24). The various parameters in (24) were replaced by estimates obtained from other portions of the EC.

The voice data file is approximately 114-K-samples long. The language is Swedish. The channel output consists of a far-end speaker (0–27 K) during which time a channel change occurs at 20 K, double-talk (27–93 K), a second channel change (93 K), and far-end speaker (93–114 K). Thus, the file consists of an initial training period of 20 K, a channel change with only 7 K for training, a long period of high-level double-talk, a second channel change, and 21 K for training after the second channel change. Thus, this file tests three properties of an EC: learning speed, double-talk sensitivity, and response to channel changes. It should be noted that the first channel change does not involve a significant change in echo return loss (ERL) or delay. Hence, the estimate of the bulk delay should not change. The second channel change (93 K) involves a change in channel delay of about 300 samples. For this channel change, the estimate of the bulk delay changes significantly. The real channels were unknown. Hence, the adaptive filter weights, after convergence, provided the following information about the unknown impulse responses:

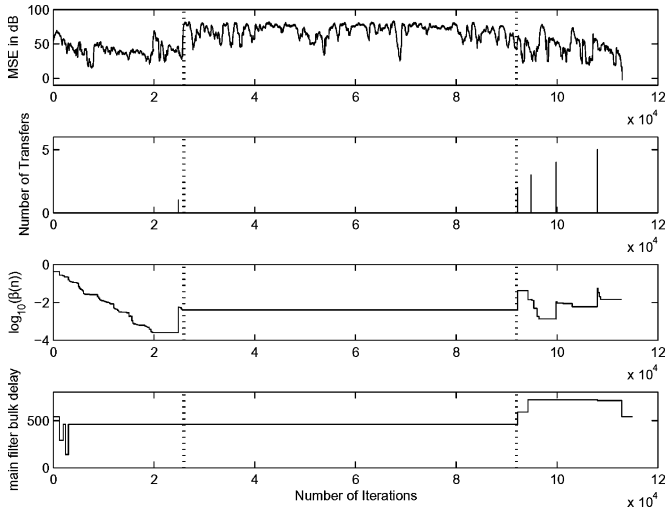


Fig. 10. Performance of conventional EC. (Top) MSE. (Second) Number of transfers from the shadow filter to the main filter. (Third) $10\ln_{10}(\beta)$. (Bottom) Main filter bulk delay.

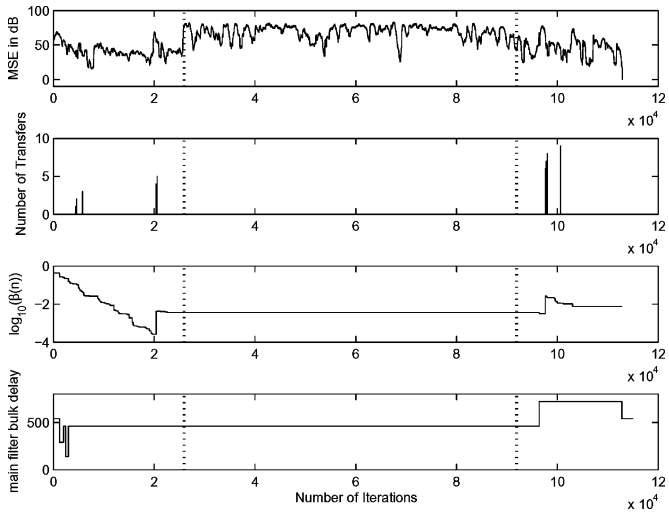


Fig. 11. Performance of the LRT based EC. (Top) MSE. (Second) Number of transfers from the shadow filter to the main filter. (Third) $10\ln_{10}(\beta)$. (Bottom) Main filter bulk delay.

- prior to first channel change—a highly oscillatory of length about 80 taps with six positive and five negative discernible peaks (of positive amplitudes about 0.12, 0.08, 0.05, 0.04, 0.01, 0.01 and negative amplitudes of 0.11, 0.09, 0.07, 0.015, 0.01);
- after first channel change—the same filter shape but a change in delay of one tap;
- second channel change—the same filter shape but a very large change in delay (300 taps).

The first channel change did not require a new estimate of delay by the partial Haar filter, in contrast to the second channel change. Figs. 10 and 11 show four curves for each of the two EC results: the smoothed MSE of the main filter in decibels (top), the number of transfers from shadow to main (second), a measure of the adaptive filter weight errors (third) (a small value of the norm of the "delay coefficients" [20], denoted as β, means that the adaptive filter is at or near convergence), and the bulk delay of the main filter (bottom). Note that the beginning and the end of double-talk are indicated by vertical dotted lines on all figures. These two sets of curves can be interpreted as follows.

- Both ECs undergo a learning phase from 0 to 20 K. The transfers from the shadow to the main filter occur under the control of some transient learning logic. This logic is the same for both ECs and is not related to the channel change versus double-talk logic. Hence, the β curves are identical during this phase.
- The first channel change at 20 K is detected at 20.4 K (within 400 samples) for the LRT-based EC (Fig. 11, middle two curves) but is detected at 24.8 K (within 4800 samples) for the conventional EC (Fig. 10, middle two curves). Note that the parameters for the LRT-based EC are $p = 500$ in (23) and $b = 0$ in (24).
- The curves for the second channel change at 93 K are more difficult to interpret because of the effects of the Duttweiler filter and the transient learning logic. Fig. 10 has the following interpretation: the first change in the third figure (at about 92 K) is an incorrect estimate of the bulk delay. A correct estimate occurs at about 94.2 K. Numerous transfers from the shadow filter to the main filter occur due to the transient training logic. However, the conventional EC transfers the channel change to the main filter at 94.8 K (in 600 samples). The jump in β at 99.6 K is due to a small change in the Duttweiler filter estimate of the delay (not shown here) at 99.5 K. This changes the delay of the shadow filter and causes it to re-adapt. This is interpreted by the conventional EC as another channel change and, hence, a transfer at 99.6 K. A similar comment applies to the transfer at about 108 K. Fig. 11 has the following interpretation: the bottom curve indicates that the first change in the bulk delay is correct (at 96.2 K). The middle two curves indicate that the LRT-based EC first transfers the shadow filter to the main filter at 97.6 K (in 1400 samples). Then, the transient training logic takes over, causing β to decrease. Note that a significant portion of the total delay (from the channel change at 93 K) is due to the Duttweiler filter estimating the new bulk delay and transferring this to the main filter.
- Both cancellers are insensitive to the heavy double-talk during 27–93 K, as shown in the second and third curves of Figs. 10 and 11.

To summarize for this particular example, both ECs are not sensitive to double-talk. The LRT-based EC yields a much faster transfer from the shadow to the main filter for the first channel change, whereas the conventional EC is somewhat faster for the second channel change. The latter result is somewhat clouded by the effects of the Duttweiler adaptive filter.

VII. Results and Conclusion

This paper has derived a LRT for deciding between double-talk (freeze weights) and a channel change (adapt quickly) for a stationary Gaussian stochastic input signal model. The pdf of the sufficient statistic under each hypothesis was obtained and the performance of the sufficient statistic was evaluated as a function of the system parameters. The ROCs indicate that it is difficult to correctly decide between double-talk and a channel

change based upon a single look. However, MC simulations of the postdetection integration of approximately 100 sufficient statistic samples yields a detection probability close to unity (0.99) with a small false alarm probability (0.01). Thus, use of an LRT to decide between a channel change or double-talk offers a significant improvement in EC performance. It should be noted that the simpler problem of detecting double-talk only is a special case of what has been studied here. One need only set $\mathbf{H}_1 = \mathbf{H}_0$ in (1) and proceed to generate ROCs, etc.[2]

The LRT is highly parametric and requires detailed statistical information about the input under both hypotheses. This will not be the case in a real echo cancellation environment. Thus, any practical application of the LRT to an EC will suffer performance degradation as compared to the ROC curves presented here. These degradations are due to the difficulty of the EC to accurately estimate these parameters in an actual voice signal environment. However, the real value of the ROC curves is to upper bound the performance of any less-than-optimum system. Thus, the ROC curves presented in this paper (or others derived using the theory in this paper) can be of great value to an EC designer even though they may not match precisely the parameters of the environment.

The effects of parameter estimation errors can be studied through the use of the generalized-likelihood ratio test (GLRT) when the parameters are assumed unknown and are simultaneously estimated while deciding which hypothesis is true. Of course, the GLRT will not perform as well as the LRT.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments and A. Bentahir for help with some simulation results.

REFERENCES

[1] *Digital Network Echo Cancellers*, ITU-T Recommendation G. 168, 2002.

[2] C. Breining, P. Dreiscitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.

[3] W. Kellermann, "Current topics in adaptive filtering for hands-free-acoustic communication and beyond," *Signal Process.*, vol. 80, no. 9, pp. 1695–1696, Sep. 2000.

[4] A. Mader, H. Puder, and G. Schmidt, "Step-size control for echo cancellation filters—An overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, Sep. 2000.

[5] P. Heitkamper, "An adaptation control for acoustic echo cancellers," *IEEE Signal Process. Lett.*, vol. 4, no. 6, pp. 170–172, Jun. 1997.

[6] R. Frenzel and M. E. Hennecke, "Using prewhitening and stepsize control to improve the performance of the LMS algorithm for acoustic echo cancellation," in *Proc. Int. Symp. Circuits. Systems (ISCAS)*, San Diego, CA, May 1992, pp. 1930–1932.

[7] H. Ezzaidi, I. Bourmeyster, and J. Rouat, "A new algorithm for double-talk detection and separation in the context of digital mobile radio," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Munich, Germany, Apr. 1997, pp. 1897–1900.

[2]It should also be noted that the rare case of the simultaneous occurrence of both double-talk and a channel change can also be handled within the present framework. It is only necessary to reverse the definitions of the main filter and shadow filters. In this case, (2) represents the output of the main filter (whose transfer function is $\mathbf{H}_0$) and (1) represents the output of the shadow filter ($\mathbf{H}_1$ is due to the channel change and $n_1(n)$ is due to double-talk). The two hypotheses are then $\mathcal{H}_0$: no channel change and no double-talk, $\mathcal{H}_1$: $y(n)$ is due to a channel change and double-talk.

[8] T. Gansler, M. Hansson, C. Ivarson, and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, vol. 44, no. 11, pp. 1421–1427, Sep. 1996.

[9] T. Gansler, "A double-talk resistant subband echo canceller," *Signal Process.*, vol. 65, no. 1, pp. 89–101, Jan. 1998.

[10] T. Gansler, S. Gay, M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[11] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. New York: Springer-Verlag, 2001.

[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: Wiley, 1968.

[13] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[14] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceller with two echo path models," *IEEE Trans. Commun.*, vol. 25, no. 6, pp. 589–595, Jun. 1977.

[15] R. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[16] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

[17] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: Wiley, 1964.

[18] N. J. Bershad and A. Bist, "Fast coupled adaptation for sparse impulse responses using a partial haar transform," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 966–976, Mar. 2005.

[19] D. Duttweiler, "Subsampling to estimate delay with application to echo cancelling," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 5, pp. 1090–1099, Oct. 1983.

[20] E. Haensler and G. Schmidt, *Acoustic Echo and Noise Control*. New York: Wiley, 2004.

Neil J. Bershad (S'60–M'62–SM'81–F'88) received the B.E.E. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1958, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1960, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute in 1962.

He joined the Faculty of the Henry Samueli School of Engineering, University of California, Irvine, in 1966 and is currently an Emeritus Professor of Electrical Engineering and Computer Science. His research interests have involved stochastic systems modeling and analysis. His recent interests are in the area of stochastic analysis of adaptive filters. He has published a significant number of papers on the analysis of the stochastic behavior of various configurations of the LMS adaptive filter. His present research interests include the statistical learning behavior of adaptive filter structures for nonlinear signal processing and electronic and acoustic echo cancellation.

Dr. Bershad has served as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of phase-locked loops and synchronization. More recently, he was an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING in the area of adaptive filtering.

Jean-Yves Tourneret (M'94) received the Ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique, Toulouse (ENSEEIHT), France, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, France, in 1992.

He is currently a Professor in ENSEEIHT. He is a member of the IRIT Laboratory (UMR 5505 of the CNRS), where his research activity is centered around estimation, detection, and classification of non-Gaussian and nonstationary processes.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse, France, in 2002. He was also member of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2006 organizing committee. He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society.