ECHO CANCELLATION - A LIKELIHOOD RATIO TEST FOR DOUBLE-TALK VS. CHANNEL CHANGE

Neil J. Bershad and Jean-Yves Tourneret†*

†IRIT/ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
*University of California, Irvine, 1621 Santiago Drive, Newport Beach, CA 92660, USA
bershad@ece.uci.edu, Jean-Yves.Tourneret@enseeiht.fr

ABSTRACT

Echo cancellers 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 [1] 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 for deciding between double-talk (freeze weights) and a channel change (adapt quickly) using a stationary Gaussian stochastic input signal model. The probablity density function 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 indicate that it is difficult to correctly decide between double-talk and a channel change based upon a single look. However, post-detection integration of approximately one hundred sufficient statistic samples yields a detection probability close to unity with a small false alarm probability.

1. INTRODUCTION AND PROBLEM FORMULATION

The echo cancellation problem has been studied by many authors [1, 2, 3] for more than 30 years. 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 echo canceller (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 signal due to the far-end speaker. Specific control logic involves monitoring the error signal as well as the channel input vector (to handle nonstationarities of the voice signal). 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 [1] since it is often difficult 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 [1] and [3] for deciding when to adapt the adaptive filter weights [4, 5]. Suffice to say, to our knowledge, these or other schemes are not based on any optimum statistical test such as a Likelihood Ratio Test (LRT) [6].

This paper derives a LRT for deciding between doubletalk (freeze weights) and a channel change (adapt quickly) using a stationary Gaussian stochastic signal model. The LRT is then simplified to a sufficient statistic 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 thru Receiver Operating Characteristic (ROC) curves [7]. These curves show the Probability of Detection (P_D) (deciding one hypothesis is true when it is actually true) vs. Probability of False Alarm (PFA) (deciding the same hypothesis is true when it is actually not true). The ROC's indicate that it is difficult to correctly decide between doubletalk and a channel change based upon a single look. However, post-detection integration of about one hundred successive LRT samples yields a PD close to unity (.99) with a small PFA (.01).

The stationary signal model is not necessarily representative of speech since speech is highly non-stationary. 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 (Figure 1) is assumed in order to obtain good estimates of the many parameters needed for the LRT. The EC consists of a non-adaptive main filter and an adaptive shadow filter [4]. The output of the main filter is subtracted from the echo to obtain the cancelled echo. 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 [8].

2. THE HYPOTHESIS TEST

Two of the primary signals that the EC uses for the control logic are the error signal $e_m(n)$ (canceller output) and $e_s(n)$ (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 doubletalk or to a channel change. Either occurence will cause a significant increase in the error powers. A statistical test is defined in what follows which models these two possible events. It is assumed that the EC in Figure 1 is able to accurately estimate the powers of the background noise, signal, and double-talk. These powers are assumed time-invariant for the data in the hypothesis test.

2.1. Signal and channel models

The channel input vector is of dimension $N \times 1$, with covariance matrix $E[X(n)X^T(n)] = \sigma_x^2 I_N$ (I_N is the $N \times N$ identity matrix). The channel output is a scalar $y(n)$. $X(n)$ and $y(n)$ are zero mean jointly Gaussian vectors. Let

$$\begin{aligned} \mathcal{H}_1 &: \ y(n) \text{ is due to doubletalk} \\ \mathcal{H}_0 &: \ y(n) \text{ is due to a channel change} \end{aligned} \tag{1}$$

Under $\mathcal{H}_1$,

$$y(n) = X^T(n)H_1 + n_0(n) + n_1(n),$$

where H_1 is an unknown channel which has been correctly identified prior to time n using the adaptive shadow filter and transferred to the main channel filter. $n_0(n)$ is the additive stationary zero mean white Gaussian noise, independent of $X(n)$ with $E[n_0^2(n)] = \sigma_0^2$. $n_1(n)$ is a second additive zero mean white Gaussian noise modeling double-talk, independent of both $X(n)$ and $n_0(n)$ with $E[n_1^2(n)] = \sigma_1^2$. Under $\mathcal{H}_0$,

$$y(n) = X^T(n)H_0 + n_0(n),$$

where 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, H_1 is the main filter weight vector and H_0 is the shadow filter weight vector after convergence. Hence, all the parameters are known for the hypothesis test. Straightforward computations allow to obtain

$$\begin{aligned} &E[y^2(n)|\mathcal{H}_1] = \sigma_x^2 H_1^T H_1 + \sigma_0^2 + \sigma_1^2, \\ &E[y(n)X(n)|\mathcal{H}_1] = \sigma_x^2 H_1, \\ &E[y^2(n)|\mathcal{H}_0] = \sigma_x^2 H_0^T H_0 + \sigma_0^2, \\ &E[y(n)X(n)|\mathcal{H}_0] = \sigma_x^2 H_0. \end{aligned} \tag{2}$$

The joint pdf of $y(n)$ and $X(n)$ is Gaussian such that

$$\begin{aligned} p[y(n), X^T(n)|\mathcal{H}_1] \sim \mathcal{N}(\underline{0}, R_1), \\ p[y(n), X^T(n)|\mathcal{H}_0] \sim \mathcal{N}(\underline{0}, R_0), \end{aligned} \tag{3}$$

where R_1 and R_0 can be written

$$R_1 = \sigma_x^2 I_{N+1} + \sigma_x^2 \begin{pmatrix} H_1^T H_1 + \frac{\sigma_0^2+\sigma_1^2}{\sigma_x^2} - 1 & H_1^T \\ H_1 & 0 \end{pmatrix}, \tag{4}$$

$$R_0 = \sigma_x^2 I_{N+1} + \sigma_x^2 \begin{pmatrix} H_0^T H_0 + \frac{\sigma_0^2}{\sigma_x^2} - 1 & H_0^T \\ H_0 & 0 \end{pmatrix}. \tag{5}$$

The second matrices in (4) and (5) are of the form

$$M_k = \begin{pmatrix} a_k & H_k^T \\ H_k & 0 \end{pmatrix},$$

where $k = 0, 1$ and

$$\begin{aligned} a_1 &= H_1^T H_1 + \tfrac{\sigma_0^2+\sigma_1^2}{\sigma_x^2} - 1, \\ a_0 &= H_0^T H_0 + \tfrac{\sigma_0^2}{\sigma_x^2} - 1. \end{aligned}$$

2.2. The Log LRT

By denoting $v(n) = [y(n), X^T(n)]$, the log LRT for (3) rejects hypothesis $\mathcal{H}_1$ when

$$\ln\left(\frac{p(v(n)|\mathcal{H}_1)}{p(v(n)|\mathcal{H}_0)}\right) = \frac{1}{2} v(n) \left(R_0^{-1} - R_1^{-1}\right) v(n)^T + \frac{1}{2}\ln\left(\frac{|R_0|}{|R_1|}\right)$$

exceeds an appropriate threshold ([6], chapter 2). Since the last term in this expression is not a function of the observables, the log LRT simplifies to

$$(y(n), X^T(n))^T \{R_0^{-1} - R_1^{-1}\} \begin{pmatrix} y(n) \\ X(n) \end{pmatrix} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} T_1, \tag{6}$$

where T_1 is a threshold setting determined by the probability of detection P_D and the probability of false alarm P_{FA}. The detection strategy (6) can be simplified as follows [7]

$$\gamma\left(y(n), X^T(n)\right) = y(n)z(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} T, \tag{7}$$

where γ is the sufficient statistic of the test (6), T is the new test threshold and

$$\begin{aligned} &z(n) = Ky(n) + 2\alpha X^T(n)H_1 - 2\beta X^T(n)H_0, \\ &K = k_1 + k_2 - k_2 - k_4, \alpha = \tfrac{k_1}{\lambda_{11}} + \tfrac{k_2}{\lambda_{21}}, \beta = \tfrac{k_3}{\lambda_{10}} + \tfrac{k_4}{\lambda_{20}}, \\ &\lambda_{1,k} = \tfrac{1}{2}(a_k + \sqrt{a_k^2 + 4H_k^T H_k}), \\ &\lambda_{2,k} = \tfrac{1}{2}(a_k - \sqrt{a_k^2 + 4H_k^T H_k}), \\ &k_1 = \tfrac{\lambda_{11}}{1+\lambda_{11}} \tfrac{\lambda_{11}^2}{\lambda_{11}^2 + H_1^T H_1}, k_2 = \tfrac{\lambda_{21}}{1+\lambda_{21}} \tfrac{\lambda_{21}^2}{\lambda_{21}^2 + H_1^T H_1}, \\ &k_3 = \tfrac{\lambda_{10}}{1+\lambda_{10}} \tfrac{\lambda_{10}^2}{\lambda_{10}^2 + H_0^T H_0}, k_4 = \tfrac{\lambda_{20}}{1+\lambda_{20}} \tfrac{\lambda_{20}^2}{\lambda_{20}^2 + H_0^T H_0}. \end{aligned} \tag{8}$$

Thus, the sufficient statistics is the product of two zero mean correlated Gaussian variates. $y(n)$ is the channel output at time n. $z(n)$ is a linear combination of the scaled channel output, the scaled output of the shadow filter and the scaled output of the main filter.

3. PDF OF THE SUFFICIENT STATISTIC

Since $v(n) = [y(n), X^T(n)]$ is a zero mean Gaussian vector, it follows that $y(n)$ is a zero mean scalar Gaussian variate with variance given by (2) under the two hypotheses. $z(n)$ is is also a zero mean scalar Gaussian variate with a variance that can be computed from its expression in (8). $[y(n), z(n)]$ is linearly related to $[y(n), X^T(n)]$ thru the matrix relation

$$\begin{pmatrix} y(n) \\ z(n) \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ K & 2\alpha H_1^T - 2\beta H_0^T \end{bmatrix} \begin{pmatrix} y(n) \\ X(n) \end{pmatrix}.$$

Thus, $[y(n), z(n)]$ is a Gaussian vector with mean $[0, 0]$ and covariance matrix Σ_i under hypothesis $\mathcal{H}_i$ $(i = 0, 1)$. More precisely

$$\begin{aligned} \Sigma_i &= \begin{bmatrix} 1 & 0 \\ K & 2\alpha H_1^T - 2\beta H_0^T \end{bmatrix} \sigma_x^2 M_i \begin{bmatrix} 1 & K \\ 0 & 2\alpha H_1 - 2\beta H_0 \end{bmatrix}, \\ &= \sigma_x^2 \begin{bmatrix} m_{11}^i & m_{12}^i \\ m_{21}^i & m_{22}^i \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} m_{11}^i &= a_1, m_{12}^i = Ka_i + 2H_i^T(\alpha H_1 - \beta H_0) = m_{21}^i \\ m_{22}^i &= K^2 a_i + 4K(\alpha H_i^T H_1 - \beta H_i^T H_0) + \\ &\quad 4(\alpha H_1^T - \beta H_0^T)(\alpha H_1 - \beta H_0) \end{aligned}$$

The joint pdf of $[y(n), z(n)]$ under hypothesis $\mathcal{H}_i$ can be written

$$p_i(y, z) = \frac{1}{2\pi\sqrt{|\Sigma_i|}} \exp\left(-\frac{1}{2}(y, z)\Sigma_i^{-1} \begin{pmatrix} y \\ z \end{pmatrix}\right),$$

where

$$\Sigma_i^{-1} = \frac{1}{\sigma_x^2(m_{11}^i m_{22}^i - m_{12}^i m_{21}^i)} \begin{bmatrix} m_{22}^i & -m_{12}^i \\ -m_{21}^i & m_{11}^i \end{bmatrix}.$$

Since y and z are jointly Gaussian with zero means, the pdf of the product $u = yz$ is given by [9, p. 45]

$$p_i(u) = \frac{\sqrt{|\Sigma_i^{-1}|}}{\pi} \exp\left[-u(\Sigma_i^{-1})_{12}\right] K_0\left(|u|\sqrt{(\Sigma_i^{-1})_{11}(\Sigma_i^{-1})_{22}}\right),$$

where K_0 is the modified Bessel function of the second kind and of zero order.

4. PERFORMANCE CURVES

4.1. Theoretical curves

The performance of the sufficient statistic can be defined by the two following probabilities [6, p. 38]

$$P_D = P\left[\text{accepting } \mathcal{H}_1 \; | \mathcal{H}_1 \text{ is true}\right] = \int_T^\infty p_1(u)du, \quad (9)$$

$$P_{FA} = P\left[\text{accepting } \mathcal{H}_1 \; | \mathcal{H}_0 \text{ is true}\right] = \int_T^\infty p_0(u)du. \quad (10)$$

Thus, for each value of T, there exists a pair (P_{FA}, P_D). The curves of P_D as a function of P_{FA} are called Receiver Operating Characteristics (ROC curves) [6, p. 38].

4.2. Monte Carlo simulations

Ten thousand Monte Carlo simulations have been run for the sufficient statistic in (7) as a check on the theoretical results. Figure 2 shows some typical ROC curves for $N = 1024$ and different parameter selections. H_0 and H_1 are two sided exponential channels with $H_i(j) = (0.5)^{|j|}c, j = 0, \pm 1, \ldots$. The parameter c is defined by the filter gain which is here $H_1^T H_1 = H_0^T H_0 = 0.1$. The filters differ only in a bulk delay of 200 taps.

Excellent agreement between the theory and MC simulations was obtained over all values of P_D and P_{FA}. Figure 2 shows the ROC curves for different double talk powers, no additive noise and $H_1^T H_0 = 0$. A P_D approaching unity results in a fairly large P_{FA}. The poor behavior is because 1) the sufficient statistic is non-coherent (quadratic in the data) and 2) only one time sample of the data vector is used in making the decision.

4.3. Using the MC simulations to predict the theory

Some numerical integration problems were encountered using (9) and (10). Thus, because of excellent agreement between theory and MC simulations, the ROC curves were generated from the MC simulations instead. Figure 3 shows the effect of decreasing the background noise power. The performance improvement approaches the top curve as the background noise power approaches zero. Hence, the hypothesis test in (1) is not noise limited. Figure 4 shows that the performance does not increase monotonically with increasing levels of double-talk. This effect occurs because of the non-coherent nature of the sufficient statistic.

5. POST-DETECTION INTEGRATION

The previous ROC curves suggest that one time sample is not enough to make a reliable decision. Thus, one would like to derive the sufficient statistic for p time samples of the vector $[y(j), X^T(j)]$ for $j = n-p+1$ to n. The problem is that inversion of the covariance matrix is extremely difficult. A way to avoid this statistical problem is to use the MC simulations. Consider the time averaged sufficient statistic

$$\Gamma(n) = \frac{1}{p} \sum_{m=n-p+1}^{n} y(m)z(m). \quad (11)$$

Ten thousands MC simulations of (11) were run for different values of p and $\sigma_0^2 = \sigma_1^2 = 1$. Figure 5 shows that $p = 100$ yields an excellent ROC curve.

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 to set $H_0 = H_1$ in our model and proceed to generate ROC curves etc ...

6. RESULTS AND CONCLUSIONS

This paper has derived the 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 ROC's indicate that it is difficult to correctly decide between double-talk and a channel change based upon a single look. However, Monte Carlo simulations of the post-detection integration of approximately one hundred sufficient statistic samples yields a detection probability close to unity (.99) with a small false alarm probability (.01). Thus, use of an LRT based test to decide between a channel change or double-talk offers a significant improvement in EC performance.

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 Echo Canceller 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 here can be of great value to an EC designer even though they may not match precisely the parameters of the environment.

7. REFERENCES

[1] C. Breining, P. Dreiscitel, E. Hnsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control," *IEEE Signal Processing Magazine*, vol. 16, pp. 42–69, July 1999.

[2] "Digital network echo cancellers," in *ITU-T Recommendation G. 168*, 2002.

[3] W. Kellermann, "Current topics in adaptive filtering for hands-free-acoustic communication and beyond," *Signal Processing (EURASIP)*, vol. 80, pp. 1695–1696, Sept. 2000.

[4] A. Mader, H. Puder, and G. Schmidt, “Step-size control for echo cancellation filters - an overview,” *Signal Processing (EURASIP)*, vol. 80, pp. 1697–1719, Sept. 2000.

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

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

[7] N. J. Bershad and J.-Y. Tourneret, “Echo cancellation: a likelihood ratio test for double talk versus channel change,” *IEEE Trans. Signal Processing*, submitted in August 2005.

[8] P. Heitkamper, “An adaptation control for acoustic echo cancellers,” *IEEE Signal Processing Letters*, vol. 4, pp. 170–172, June 1997.

[9] K. S. Miller, *Multidimensional Gaussian distributions*. New York: Wiley, 1964.

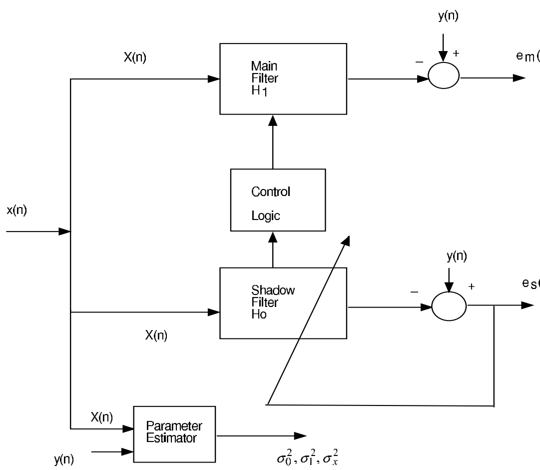

Fig. 1. Basic Echo Canceller Structure.

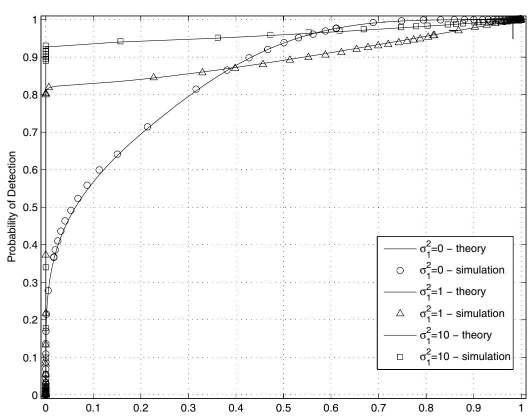

Fig. 2. Comparison of Theory and Monte Carlo Simulations. P_D versus P_{FA} for different doubletalk power levels σ_1^2 for $\sigma_0^2 = 0, N = 1024$ and orthogonal channels.

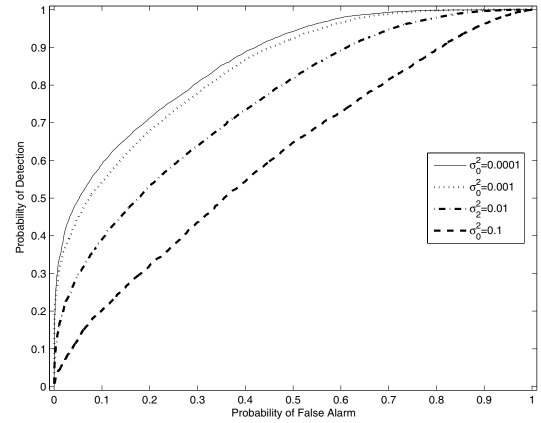

Fig. 3. P_D versus P_{FA} (MC simulations) for different values of $\sigma_0^2, \sigma_1^2 = 1, N = 1024$ and orthogonal channels.

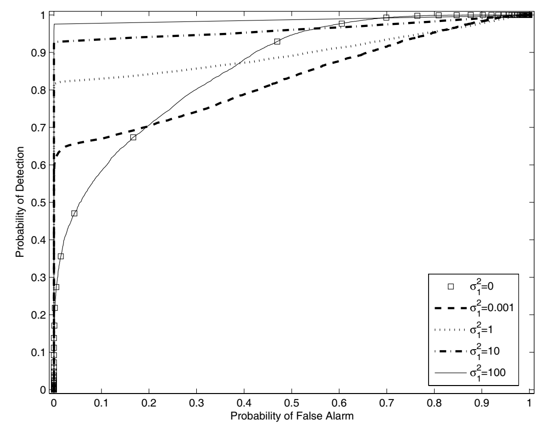

Fig. 4. P_D versus P_{FA} (MC simulations) for different values of $\sigma_1^2, \sigma_0^2 = 0.001, N = 1024$ and orthogonal channels.

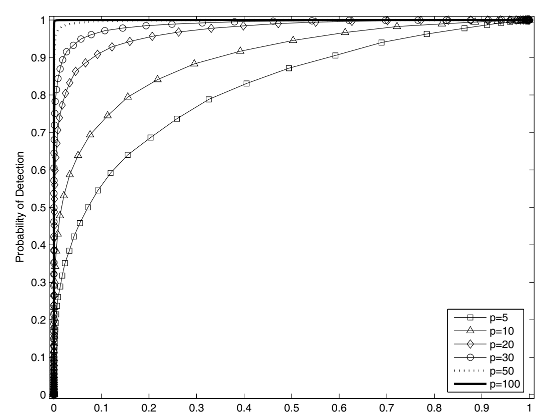

Fig. 5. P_D versus P_{FA} (MC simulations) for post-detection integration of p samples of the LRT for $\sigma_1^2 = \sigma_0^2 = 1, N = 1024$ and orthogonal channels.