Probability of Detection of Residual Echo Based on Magnitude-Squared Coherence Estimate

Cliff Sui

Lucent Technologies, Inc.
Business Communications Systems Division
200 Laurel Avenue
Middletown, NJ 07748, U.S.A
Phone: (908)957-2373, 957-7488 (fax)
E-mail: csui@lucent.com

ABSTRACT

Residual echo is a distorted, partially-canceled, and transient echo of the near-end speech signal returned from the remote network. Previous work [2] has shown that residual echo can be detected and reduced by using a nonlinear processing technique based on frequency coherence. In this paper a mathematical approach to evaluate this nonlinear processor and the algorithm for computing the probability of detection using the magnitude-squared coherence estimate are presented.

1. INTRODUCTION

In many voice communication networks, the far-end returns to the local talker a delayed replica of his voice transmission because of impedance mismatch at the remote hybrid (H). This delayed replica is recognized as echo by the local party. When received with considerable delay, this echo becomes extremely bothersome and confusing in 2-way voice communication. A widely accepted practice is for the network to operate devices which suppress or cancel echoes near the source, e.g. echo canceler (EC) [1]. However, there are cases when the remote echo control may not be completely effective for removing echo. Thus, in many cases there are some residual echo returned to the local talker. A technique described in [2] to reduce or eliminate residual echo is to operate residual far-end echo control (RFEC) processing in the local network to improve the local talker's communication experience. This technique uses a multiband nonlinear processor and a coherence algorithm to detect and characterize the residual echo.

In general, a matched filter solution is preferred when the signal to be detected is of known form. In the residual echo case, however, the signals can be treated as short time stationary processes [3] and therefore a Wiener filter approach is applicable. The solution given by the Wiener theory shows that the performance of the optimal filter is determined by the magnitude of the coherence between the observed and the estimated signals [4]. By passing the observed signal through the optimal filter, an estimated signal with minimal error (in the mean square sense) can be produced. That is, the magnitude of coherence is unity when the two signals are completely linearly related. In other words, the magnitude of coherence between two arbitrary signals will be close or equal to unity if they are linearly related. On the other hand, if the two signals have no linear relationship the magnitude of

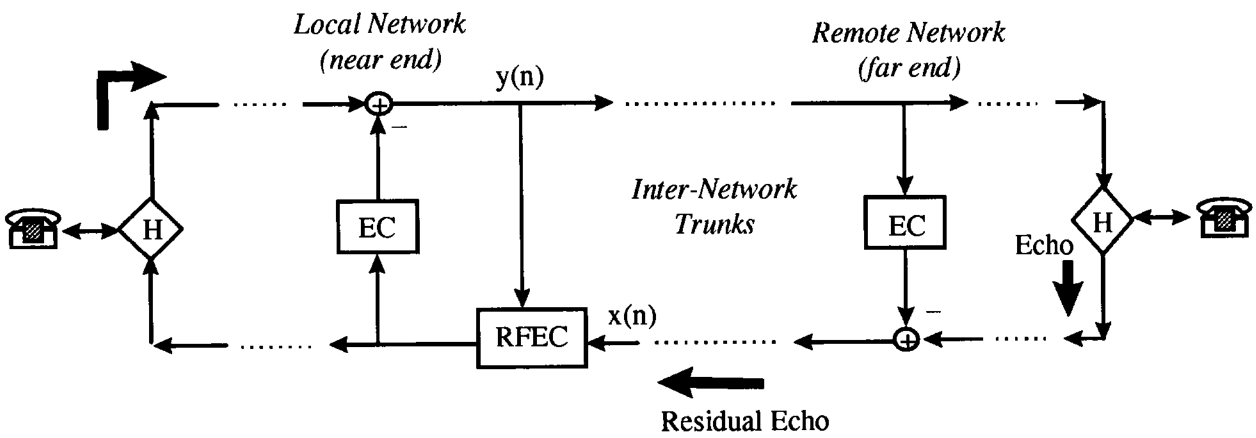

Figure 1. Network architecture utilizing residual far-end echo control (RFEC).

coherence will be zero. As a result, by measuring the linear relationship between the speech signals from the local network and the remote network by means of the frequency coherence function, residual echo within the speech signal from the remote network can be detected. An algorithm for computing the probability of detection can be expressed by utilizing the estimate of the magnitude-squared coherence (MSC) and its probability density function [5].

2. COHERENCE FUNCTION

A block diagram of the speech and residual echo path is shown in Figure 1. The received signal from the remote network is denoted by x(n). The speech signal from the local talker is y(n). The RFEC processing uses the two signals x(n) and y(n) to make decision whether x(n) contains residual echo or not. Both x(n) and y(n) are assumed to be wide-sense stationary real zero-mean processes. The optimal Wiener filter $H_o(f)$ for obtaining the estimate $x_o(n)$ from y(n) is expressed by

$$H_o(f) = \frac{S_{xy}(f)}{S_y(f)} \qquad (1)$$

In [4], the minimum mean-squared-error (MSE) of the optimal filter performance can be expressed as

$$MSE = \int_{-\infty}^{\infty} S_x(f)\, df - \int_{-\infty}^{\infty} H_o(f)\, S^*_{xy}(f)\, df \qquad (2)$$

Substitution of (1) into (2) yields

$$MSE = \int_{-\infty}^{\infty} S_x(f)\ [1 - |\gamma_{xy}(f)|^2]\, df \qquad (3)$$

where

$$|\gamma_{xy}(f)|^2 = \frac{|S_{xy}(f)|^2}{S_x(f)\, S_y(f)} \qquad (4)$$

which is the magnitude-squared coherence function between x(n) and y(n). $S_x(f)$ and $S_y(f)$ are the (auto) spectral densities and $S_{xy}(f)$ is the cross spectral density. If $\gamma^2_{xy}(f) = 0$, then $H_o(f) = 0$. That is, x(n) and y(n) have no linear relationship and $x_o(n)$ cannot be estimated from y(n). If $\gamma^2_{xy}(f) = 1$, then minimum MSE = 0. Thus x(n) and y(n) are completely linearly related in the mean-square sense. A decision variable α which is used as a parameter for residual echo detection is defined by

$$\alpha = \frac{1}{K} \sum_{i=0}^{K-1} |\gamma_{xy}(f_i)|^2 \qquad (5)$$

where $f_i \in [f_1, f_2]$ represents the i-th frequency in the interval in which the coherence function is of interest and $f_i = f_1 + i(f_2 - f_1)/K$, $i = 0,1,2,..K$. By comparing α to a pre-defined threshold level T, the residual echo is present in y(n) if $\alpha \geq T$, not present if $\alpha < T$.

3. ALGORITHM DESCRIPTION

This section presents an estimation procedure for the magnitude-squared coherence and the algorithm for computing the probability of detection of residual echo. The first step of the estimation procedure is to obtain estimates of the spectral densities $S_x(f)$, $S_y(f)$, and $S_{xy}(f)$. The signal inputs x(n) and y(n) are digital, telephone bandwidth signals from the remote network and the local network. Each signal is segmented into data frames of equal length for frame-by-frame processing. Samples of each data frame are weighted (e.g. Hanning window) to achieve good sidelobe reduction. In order to obtain an estimate with low variance and good spectral resolution with efficient use of limited data, a segment overlapping scheme [5] can be used to optimize the conflicting requirements on the number of segments N and the segment length. Then the spectral density estimates can be obtained by

$$\hat{S}_x(f) = \frac{1}{N} \sum_{k=1}^{N} |X_k(f)|^2 \qquad (6)$$

$$\hat{S}_y(f) = \frac{1}{N} \sum_{k=1}^{N} |Y_k(f)|^2$$

$$\hat{S}_{xy}(f) = \frac{1}{N} \sum_{k=1}^{N} X_k(f)\, Y_k^*(f)$$

where $X_k(f)$ and $Y_k(f)$ are the FFT of the k-th windowed segment of x(n) and y(n) respectively. $Y_k^*(f)$ denotes the complex conjugate of $Y_k(f)$. Then, the estimate of the magnitude-squared coherence is

$$|\hat{\gamma}_{xy}(f)|^2 = \frac{|\hat{S}_{xy}(f)|^2}{\hat{S}_x(f)\, \hat{S}_y(f)} \qquad (7)$$

In [6], the probability density function of the estimate is given as

$$p(\hat{\gamma}|n, \gamma) = (n-1)(1-\gamma)^n (1-\hat{\gamma})^{n-2} F(n,n;1;\hat{\gamma}\gamma) \quad (8)$$

where $F(\cdot)$ is the hypergeometric function defined in [6] and n = N given that the estimate is formed from N segments. Note that the frequency dependency and subscripts are dropped for notational simplicity. In [7], it is shown that the function $F(\cdot)$ can be expressed by a simple Legendre polynomial as follows:

$$F(n,n;1,\gamma\hat{\gamma}) = (1-\gamma\hat{\gamma})^n\, P_{n-1}\left(\frac{1+\gamma\hat{\gamma}}{1-\gamma\hat{\gamma}}\right) \quad (9)$$

where $P_n(\cdot)$ is the n-th degree Legendre polynomial and the probability density function becomes

$$p(\hat{\gamma}|n, \gamma) = \frac{(n-1)(1-\gamma)^n(1-\hat{\gamma})^{n-2}}{(1-\gamma\hat{\gamma})^n}\, P_{n-1}\left(\frac{1+\gamma\hat{\gamma}}{1-\gamma\hat{\gamma}}\right) \qquad (10)$$

To compute the probability density function, a recursive procedure with the following formulas can be used

$$\frac{p(\hat{\gamma}|n+1,\gamma)}{p(\hat{\gamma}|n,\gamma)} = \frac{n}{n-1}\, \frac{(1-\gamma)(1-\hat{\gamma})}{(1-\gamma\hat{\gamma})} \cdot \frac{P_n[(1+\gamma\hat{\gamma})/(1-\gamma\hat{\gamma})]}{P_{n-1}[(1+\gamma\hat{\gamma})/(1-\gamma\hat{\gamma})]} \qquad (11)$$

and

$$\frac{P_n(x)}{P_{n-1}(x)} = \frac{(2n-1)x}{n} - \frac{n-1}{n} \Big/ \frac{P_{n-1}(x)}{P_{n-2}(x)} \quad (12)$$

Begin with n=2, the initial condition is

$$p(\hat{\gamma}|n=2,\gamma) = \frac{(1-\gamma)^2}{(1-\gamma\hat{\gamma})^2} P_1\left(\frac{1+\gamma\hat{\gamma}}{1-\hat{\gamma}\gamma}\right) \quad (13)$$

and $P_0(x) = 1$, $P_1(x) = x$.

As shown in (5), the decision parameter α is the mean value of the coherence estimates over K different frequencies. By convolving the function in (8) K times, a probability density distribution of α, denoted by $p(\alpha)$ can be obtained.

$$p(\alpha) = p(\hat{\gamma}_0) \otimes p(\hat{\gamma}_1) \otimes \ldots \otimes p(\hat{\gamma}_{K-1})$$

Then the probability of detection of residual echo, p_{rfec}, can be calculated by

$$p_{rfec} = 1 - \int_0^T p(\alpha \mid \alpha \geq T)\, d\alpha$$

where T is the threshold level for residual echo detection. The probability of false alarm, p_{false}, is

$$p_{false} = \int_0^T p(\alpha \mid \alpha < T)\, d\alpha$$

4. SUMMARY

A mathematical expression to evaluate the RFEC processing has been shown by finding the probability of detection of residual echo based on the estimate of the magnitude-squared coherence. A recursive procedure using Legendre polynomial expression can be used for computing the probability values. An optimal value for the threshold level T can be chosen to give acceptable values for the probability of detection and the probability of false alarm. The performance of the RFEC processing can be evaluated by using these probability values.

5. REFERENCES

[1] M. Sondhi, W. Kellermann, eds., "Adaptive Echo Cancellation for Speech Signals", Advances in Speech Signal Processing, Marcel-Dekker, 1992.

[2] P. M. Velardo, W. D. Wynn, "Method and apparatus for reducing residual far-end echo in voice communication networks", U.S. Patent 5587998.

[3] L. R. Rabiner, R. W. Schafer, Digital Processing of Speech Signals, Prentice Hall, 1978.

[4] W. A. Gardner, Introduction to Random Processes, Macmillan, 1986.

[5] G. C. Carter, C. H. Knapp, and A. H. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast fourier Transform processing", IEEE Trans. Audio Electroacoust., vol. 21, Aug. 1973.

[6] M. Abranowitz, I. A. Stegun, Eds., Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, Washington, D.C.: U.S. Gov. Printing Off. 1964.

[7] T. E. Barnard, "Legendre polynomial expressions for the probability density function of magnitude-squared coherence estimates", IEEE Trans. Acoust., Speech, and Signal Proc., vol. ASSP-29, Feb. 1981.