A SURVEY OF DOUBLE-TALK DETECTION SCHEMES FOR ECHO CANCELLATION APPLICATIONS

Thien-An Vu[1], Heping Ding[2], and Martin Bouchard[1]
[1]School of Information Technology and Engineering, University of Ottawa, Ontario, Canada
[2]Acoustics and Signal Processing, IMS, National Research Council, Ottawa, Ontario, Canada

1. INTRODUCTION

An echo canceller removes undesired echo in full-duplex speech communication. The cancellation is done by modeling the echo path impulse response with an adaptive finite impulse response filter and subtracting the echo estimate from the received signal. A typical diagram of an echo canceller is depicted in **Figure 1**. The signal x(n) and v(n) represent the far-end and near-end speeches respectively. The signal s(n) and y(n) represent the echo signal generated by the actual echo path **h** and the echo estimate produced by the adaptive filter. The signal e(n) denotes the residual error signal, which is transmitted to the far-end side and is used to update the coefficient **w** of the adaptive filter.

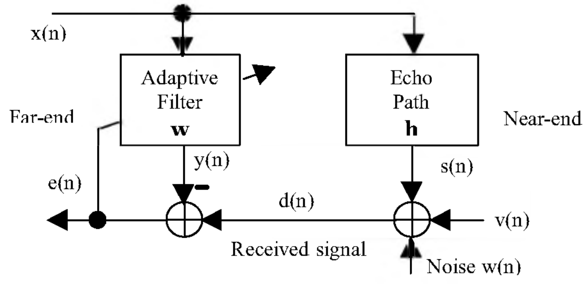

Figure 1: A diagram of an echo canceller.

When the speech signal v(n) is zero and the near-end noise w(n) is assumed to be insignificant, the adaptive filter, **w** can converge to a good estimate of the echo path, **h** and successfully cancel the echo. However, when both v(n) and x(n) are not zero, i.e. double-talk (DT) situation, the near-end speech v(n), which acts as an uncorrelated noise to the adaptive algorithm, may cause the adaptive filter to diverge and allow excessive un-cancelled echo to pass through to the far-end. The common solution to this problem is to slow down or completely stop the filter adaptation when the presence of the near-end speech is detected. This is the role of a double-talk detector.

A special case that some DT detection algorithms seem to have problem with is when there is a change in the echo path, for example in acoustic environment. This can often be falsely detected as a DT condition by those DT detection algorithms. As this is a case the adaptive filter really needs to adapt to the change in the echo path, it is not desired that the adaptation is unnecessarily turned off because DT has falsely been detected. Furthermore, the background noise w(n) at the near-end should not be detected as double-talk.

2. DOUBLE-TALK DETECTION SCHEMES

2.1 Basics

A common basic for most DT detection schemes involves computation of a detection variable from the available data such as the near-end, far-end and/or residue error signal, and comparison of the detection variable with a preset constant threshold. Depending on whether the detection variable is above or below the threshold, a decision is made on whether double-talk condition is present or not. If DT is declared, the filter adaptation is stopped or slowed down for a minimum period of hold time. When the non-DT condition lasts consecutively over the hold time, the adaptation can be resumed until the next DT condition occurs. The hold time is necessary to suppress detection dropouts because of the noisy behavior of the detection variable [1].

There are many different methods existing in the literature on how to form a detection variable for a DT detector. In this paper, attention is paid to some better-known algorithms, which are based on energy comparison and cross-correlation. These algorithms are briefly summarized in the following sub-sections.

2.2 Energy-based algorithms

A simple approach in this category is the well-known Geigel algorithm [2]. The Geigel algorithm compares the magnitude of the near-end received signal d(n) with the maximum magnitude of L most recent samples of the far-end signal x(n), where L is the adaptive filter's length. L past samples are used because of the possible end delay of x(n) through the echo path. The echo path typically dampens the far-end signal x(n), and as a result the magnitude of the received signal d(n) containing only the echo s(n) will be smaller than the received signal d(n) containing both the echo s(n) and the near-end speech signal v(n). The Geigel algorithm computes its detection variable ξ and makes decision as

$$\xi = \frac{|d(n)|}{\max\{|x(n)|,\ldots,|x(n-L+1)|\}} > T$$

If ξ is larger than the threshold T, DT is declared otherwise it is not. The choice of T needs to be made with care, and will strongly affect the performance of the detector. For line echo cancellers, T is set to 0.5 because the hybrid attenuation is assumed to be 6dB. However, for an acoustic echo cancellation environment, not only the background noise level but also the echo path characteristics are time varying. Therefore, it is not easy to decide a proper value for the threshold T. In particular, for the time-varying echo path, the Geigel algorithm can falsely regard a change of the echo path as a DT situation. As a result, the adaptive filter stops updating the coefficients when the coefficient update is actually needed.

Another energy-based DT detection scheme, proposed in [5], exploits the idea of difference in bandwidths of the echo signal s(n) (300-3400 Hz) and the near-end speech v(n) (wider band). Please refer to [5] for more details of the approach.

2.3 Correlation-based algorithms:

A correlation-based DT detector proposed in [3] makes use of the orthogonality principle. When the adaptive filter converges to its optimal solution, the orthogonality principle between the far-end vector $\underline{\mathbf{x}}(n) = [x(n)\ x(n-1)\ \ldots\ x(n-L+1)]$ and the residual error e(n) is satisfied, i.e. $E[e(n)\underline{\mathbf{x}}(n)] = 0$ (E[.] denotes the statistical expectation). In DT situation, the received signal d(n), and therefore the residue echo gets larger abruptly because of the presence of the near-end speech v(n). However, as long as the near end signal v(n) is uncorrelated with the far-end signal x(n), which is usually the case in practice, the orthogonality principle still holds. On the other hand, when the echo path changes, the orthogonality principle cannot be satisfied anymore. The cross-correlation vector between $\underline{\mathbf{x}}(n)$ and e(n) is defined as $\underline{\mathbf{c}}_{xe} = [c_{xe,0}\ \ c_{xe,1}\ \ldots\ c_{xe,L-1}]^T$

Where $c_{xe,i} = \dfrac{E[x(n-i)e(n)]}{\sqrt{E[x^2(n-i)][e^2(n)]}}$

The detection variable is defined as $\xi = \frac{1}{L}\sum_{i=0}^{L-1}\left|c_{xe,i}\right|$. When $\xi \leq T$, a properly chosen threshold, the adaptive filter has converged; otherwise, the adaptive has not converged or the echo path has changed. In general, this algorithm does not detect DT condition explicitly. Instead, it decides whether the adaptive filter has converged or not. If the adaptive filter has converged, the adaptation is stopped to protect the filter from being disturbed by DT interference. On the other hand, if the adaptive filter has not converged or the echo path has changed, the adaptive filter will keep adapting.

The algorithm in [3] defines a cross-correlation vector, which is not well normalized. The amount of cross-correlation depends on the statistics of the signals and of the echo path. As a result, the appropriate value for threshold T can vary from one experiment to another [1]. A similar idea that uses the cross correlation vector between $\underline{\mathbf{x}}(n)$ and d(n) is the normalized cross-correlation algorithm, introduced in [4]. The algorithm normalizes the cross-correlation vector in the sense that the detection variable is equal to one when the near-end signal v(n) is zero and less than one when v(n) is not. The normalized cross-correlation vector is defined as $\underline{\mathbf{c}}_{xd} = (\sigma_d^2 R_x)^{-1/2}\, \underline{r}_{xd}$ [1], where $\sigma_d^2 = E[d^2(n)]$ is the variance of d(n), $\mathbf{R}_x = E[\underline{\mathbf{x}}(n)\underline{\mathbf{x}}(n)^T]$ is the auto-correlation matrix of $\underline{\mathbf{x}}(n)$, and $\underline{\mathbf{r}}_{xd}$ is the cross-correlation vector between vector $\underline{\mathbf{x}}(n)$ and a scalar d(n). The detection variable is therefore

$$\xi = \sqrt{\underline{r}_{xd}^T (\sigma_d^2 R_x)^{-1}\, \underline{r}_{xd}} = \frac{\sqrt{\underline{\mathbf{h}}^T \mathbf{R}_x \underline{\mathbf{h}}}}{\sqrt{\underline{\mathbf{h}}^T \mathbf{R}_x \underline{\mathbf{h}} + \sigma_v^2(n)}}$$

When $\xi < T$, DT is declared, and when $\xi \geq T$, DT is not present. The threshold T is selected between 0 and 1.

Another DT algorithm, proposed in [5], is based on the orthogonality between e(n) and y(n) and can distinguish between double-talk and echo path change with a low complexity. For more details, please refer to [5].

3. SUMMARY

This paper reviews some typical DT detection schemes existing in the literature, which are based on energy comparison and cross-correlation algorithms. The energy-based algorithms, in general, have the benefit of being computationally simple, needing very little memory, and have been successfully used in line echo cancellation; however, they do not always provide reliable performance in an acoustic echo path environment. On the other hand, the correlation-based algorithms show improved detection performance in such time-varying environment but they would require relatively higher memory storage and computational complexity due to vector or matrix-based operations.

REFERENCES

[1] S. L. Gay, and J. Benesty, *Acoustic Signal Processing for Telecommunication*, Kluwer Academic Publishers, 2000.

[2] D. L. Duttweiller, "A twelve-channel digital echo canceler," IEEE Trans. Comm., vol. 26, pp. 647-653, May 1978.

[3] H. Ye, and B. X. Wu, "A new double-talk detection algorithm based on orthogonality theorem," IEEE Trans. Comm., vol. 39, pp. 1542-1545, November 1991.

[4] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," IEEE Trans. Speech Audio Processing, vol. 8, pp. 168-172, March 2000.

[5] H. Ding, and Frank Lau, "Double-talk detection schemes for echo cancellation," Acoustics Week in Canada - Canadian Acoustical Association, October 2004.