A novel normalized cross-correlation based echo-path change detector

M. A. Izbal

Steven L. Grant
Missouri University of Science and Technology, sgrant@mst.edu
A NOVEL NORMALIZED CROSS-CORRELATION BASED ECHO-PATH CHANGE DETECTOR

Mohammad Asif Iqbal and Steven L. Grant
University of Missouri-Rolla, Rolla, MO 65409. Email: {ammq2,sgrant}@umr.edu

Abstract—A double-talk detector is used to freeze acoustic echo canceller’s (AEC) filter adaptation during periods of near-end speech. Increased sensitivity towards double-talk results in declaring echo-path changes as double-talk which adversely effects the performance of an AEC as we freeze adaptation when we really need to adapt. Thus, we need an efficient and simple echo-path change detector so as to differentiate any echo-path variations from double-talk condition. In this paper, we derive a novel test statistic for echo-path change detection. The proposed decision statistic detects any echo-path variations, is normalized properly and is computationally very efficient as compared to existing techniques. Simulations demonstrate the efficiency of the proposed algorithm.

I. INTRODUCTION

Most teleconferencing conversations are conducted in the presence of acoustic echoes [1]; if the delay between the speech and its echo is more than a few tens of milliseconds, the echo is distinctly noticeable. An acoustic echo canceller (AEC) is used to remove the echo created due to the loudspeaker-microphone environment (h) [3]. In an AEC the echo-path (loudspeaker-microphone path h) is adaptively modelled using a filter (\(\hat{h}\)), which is then used to synthesize a replica of the echo (\(\hat{e}\)). This synthesized replica of the echo is subtracted from the echo-corrupted microphone signal (m) to get an echo-free signal (e). When the near-end talker (v) is active or when the speech comes from both the far-end (x) and near-end (y), the adaptive filter coefficients diverge from the true echo path impulse response if the adaptation is not halted. A double-talk detector is used to stop the AEC’s filter adaptation during periods of near-end speech [3]. A double-talk detector should be able to detect a double-talk condition quickly and accurately so as to freeze adaptation as soon as possible; at the same time it should be able to track any echo-path changes and should be able to distinguish the double-talk from the echo-path variations [5]. Typically, better immunity towards double-talk results in declaring echo-path changes as double-talk which adversely effects the performance of an AEC as we freeze adaptation when we really need to adapt. Thus, we need an efficient and simple echo-path change detector so as to differentiate any echo-path variations from double-talk. An optimum decision variable for echo-path change detection should behave as follows:

1) If no echo-path variations i.e. when the adaptive filter is converged \(\xi_{EP} < T_{EP}\).
2) During echo-path variations i.e. when the adaptive filter is not converged \(\xi_{EP} \geq T_{EP}\) and
3) \(\xi_{EP}\) is insensitive to near-end speech v.

Figure 1 shows the basic structure of the adaptive acoustic echo canceller. The far-end signal x is filtered through the room impulse response h to get the echo signal

\[
y(n) = h^T x
\]

where

\[
h = [h_0, h_1, \ldots, h_{L-1}]^T,
\]

\[
x = [x(n), x(n-1), \ldots, x(n-L+1)]^T,
\]

and L is the length of the echo-path. This echo signal is added to the near-end speech signal v to get the microphone signal

\[
m(n) = y(n) + v(n)
\]

We define the error signal at time n as

\[
e(n) = m(n) - \hat{h}^T x
\]

This error signal is used to adapt the L taps of the adaptive AEC filter \(\hat{h}\).
This paper is structured as follows: In section 2 the proposed echo-path change detection statistic is derived. We do a comprehensive study on the proposed algorithm for echo-path change detection in section 3 which is followed by a summary and conclusion in section 4.

II. ECHO-PATH CHANGE DETECTION ALGORITHM

In this section, we derive a novel normalized cross-correlation based echo-path change detector which detects any echo-path variations. Referring to figure 1, the cross-correlation between the microphone signal \( m \), and the cancellation error \( e \) is given by:

\[
\begin{align*}
\mathbf{r}_{em} & = E[em^T] \\
& = E[(y+v-\mathbf{h}^T\mathbf{x})(y+v)^T] \\
& = E[(\mathbf{h}^T\mathbf{x}-\mathbf{h}^T\mathbf{x}+v)(\mathbf{h}^T\mathbf{x}+v)^T] \\
& = (\mathbf{h} - \hat{\mathbf{h}})^T E[\mathbf{x}\mathbf{x}^T]\mathbf{h} + E[\mathbf{v}\mathbf{v}^T] \\
& = (\mathbf{h} - \hat{\mathbf{h}})^T R_{xx}\mathbf{h} + \sigma_v^2
\end{align*}
\]

where \( \sigma_v^2 \) is the variance of the near-end speech, the far-end speech vector \( \mathbf{x} \) the near-end signal \( \mathbf{v} \) are independent and are assumed to be of zero mean. Variance of the microphone signal is given by:

\[
\begin{align*}
\sigma_m^2 & = E[mm^T] = E[(y+v)(y+v)^T] \\
& = E[yy^T] + E[\mathbf{v}\mathbf{v}^T] = E[\mathbf{h}^T\mathbf{x}(\mathbf{h}^T\mathbf{x})^T] + \sigma_v^2 \\
& = \mathbf{h}^T R_{xx}\mathbf{h} + \sigma_v^2
\end{align*}
\]

and finally the variance of the cancellation error \( e \) is given by:

\[
\begin{align*}
\sigma_e^2 & = E[ee^T] \\
& = E[((\mathbf{h} - \hat{\mathbf{h}})^T\mathbf{x} + v)((\mathbf{h} - \hat{\mathbf{h}})^T\mathbf{x} + v)^T] \\
& = (\mathbf{h} - \hat{\mathbf{h}})^T E[\mathbf{x}\mathbf{x}^T](\mathbf{h} - \hat{\mathbf{h}}) + E[\mathbf{v}\mathbf{v}^T] \\
& = (\mathbf{h} - \hat{\mathbf{h}})^T R_{xx}(\mathbf{h} - \hat{\mathbf{h}}) + \sigma_v^2
\end{align*}
\]

We define our new normalized decision statistic to be

\[
\xi_{Asif} = \frac{\mathbf{r}_{em} - \sigma_e^2}{\sigma_m^2 - \mathbf{r}_{em}}
\]

substituting equations 4, 5 and 6 in 7 we get:

\[
\xi_{Asif} = \frac{(\mathbf{h} - \hat{\mathbf{h}})^T R_{xx}\hat{\mathbf{h}}}{\mathbf{h}^T R_{xx}\mathbf{h}}
\]

we observe from equation 8, for \( \mathbf{h} \approx \hat{\mathbf{h}} \), \( \xi_{Asif} \approx 0 \) and for \( \mathbf{h} \neq \hat{\mathbf{h}} \), \( \xi_{Asif} > 0 \). Thus the proposed echo-path change detector meets the needs of an optimal echo-path change detector.

III. EXPERIMENTS AND RESULTS

Our algorithm is computationally very efficient, we only need to do 3 multiplications, 3 additions, 2 subtractions and a division to compute the decision statistic at each sample (i.e. 9 operations per sample) as compared to 2l + 2 multiplications, l + 1 divisions and 3l + 1 additions (i.e. 6l+4 operations) per sample are required for the decision statistic proposed in [5]. Further, our decision statistic is normalized appropriately i.e. it is approximately zero in the absence of echo-path variations and is greater than zero during echo-path variations.

The values of \( \mathbf{r}_{em} \), \( \sigma_m^2 \) and \( \sigma_e^2 \) in 7 are exact and not available in practice. As a result, the final decision statistic is given by:

\[
\xi_{Asif} = \frac{\bar{\mathbf{r}}_{em} - \sigma_e^2}{\sigma_m^2 - \bar{\mathbf{r}}_{em}}
\]
where the estimates denoted by a hat are obtained using the exponential recursive weighting algorithm, [4] [2]:

\[
\hat{r}_{em}(t) = \lambda \hat{r}_{em}(t-1) + (1-\lambda)e(t)m^T(t) \\
\hat{\sigma}^2_m(t) = \lambda \hat{\sigma}^2_m(t-1) + (1-\lambda)m(t)m^T(t) \\
\hat{\sigma}^2_v(t) = \lambda \hat{\sigma}^2_v(t-1) + (1-\lambda)e(t)e^T(t)
\]

The echo-path change detector works as follows:

1) When \(\hat{\xi}_{Asif} > T\) (\(T\) is a properly chosen detection threshold), we declare that the echo-canceller has not converged i.e. the echo-path has changed, the adaptation is enabled even if the double-talk detector declares a double-talk.

2) Whenever \(\hat{\xi}_{Asif} < T\), the detector decides that the echo-canceller has converged i.e. there are no echo-path variations.

Detection threshold \(T\) is chosen to be slightly greater than the steady state value (the value of \(\hat{\xi}_{Asif}\) in the absence of any echo-path variations) as shown in figure 2. The recorded digital speech sampled at 16 KHz is used as far-end speech \(x\) and near-end speech \(v\) and a measured \(L = 8000\) sample \((500\text{ ms})\) room impulse response of a \(10' \times 10' \times 8'\) room is used as the loudspeaker-microphone environment \(h\). The room response was collected using a stereo system.

To create echo-path variations we moved from the collected left channel impulse response to the right channel response after 320 frames, as can be seen in figure 3 these changes in echo-path were detected. The echo-path change statistic goes above the detection threshold \(T\) as observed, and hence the variations in echo-path are detected. Next, we increased the echo-path gain by 2 dB after 320 frames. Simulations demonstrate the efficiency of the algorithm, even variations in filter coefficients by 2 dB are detected as shown in figure 4.

IV. CONCLUSIONS

We have proposed a novel normalized sample by sample echo-path change detector. To summarize we list the major advantages of the proposed echo-path change detector:

- Detects any echo-path variations and is normalized appropriately i.e. the detection statistic is greater than zero only for echo-path variations.
- Low added complexity when implemented with any adaptive algorithm.
- Independent of the near-end speech/doubletalk.

The proposed echo-path change detector can also serve as a good download test for a two-path AEC, and when used with a good double-talk detector makes the complete system (AEC) very robust.

REFERENCES