International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan

A METHOD FOR DETECTING ECHO PATH VARIATION

Jiaquan Huo,Sven Nordholm, Zhuquan Zang

Western Australian Telecommunications Research Institute
Australian Telecommunications Cooperative Research Center
Curtin University of Technology, Perth, Australia

ABSTRACT

A low cost method for detecting echo path variation is proposed. A detection parameter is constructed by averaging the coherence function between the synthetic echo and the error signal over time and frequency. The detection delay caused by the time averaging is accounted for by setting different detection thresholds for the rising and dropping edges of the detection parameter. Numerical examples show that the proposed method can reliably differentiate between echo path variation and double talk.

1. INTRODUCTION

Acoustic echo cancellers are used to suppress acoustic echoes [1] in modern communication systems. An acoustic echo canceller (AEC) is generally implemented by an adaptive finite impulse response (FIR) filter as illustrated in Fig. 1. The presence of the near-end speech makes the adaptation of the echo cancellation filter problematic. A strong near-end speech acts as a large disturbance to the adaptive filtering algorithm and may cause the echo cancellation filter to diverge [2]. It is for this reason that the update of the echo cancellation filter should be turned off when near-end speech is detected. An algorithm that detects the presence of near-end speech is called a double talk detector (DTD).

The common procedure of detecting double talk involves the forming of a detection variable which, ideally, is sensitive to the increase of near-end signal power but insensitive to the change of echo path characteristics. Forming such a detection variable is generally computationally complicated. For example, the normalized cross-correlation method [3] requires the use of an auxiliary filter. Alternatively, multiple detection variables can be used in a fuzzy logic type of system [1] so that detection variables that are computationally much less demanding to calculate can be used in combination to deliver reliable detections.

In this paper, we propose an algorithm that differentiates echo path variation from double talk.The detection of echo path variation is done by comparing the frequency properties of the synthetic echo $\hat{y}(n)$ and the error signal $e(n)$ as indicated in Fig. 1. The comparison is made by means of the coherence function. A time-frequency average of the estimated coherence function is used as the detection variable. A method for setting the threshold against which the decision statistics is compared is proposed to account for the delay caused by the time average. Numerical examples show that the proposed algorithm can reliably discriminate between echo path variation and double talk.

2. DETECTION PRINCIPLES

The basic idea of an acoustic echo canceller is depicted in Fig. 1. The microphone signal $d(n)$ consists of the echo of the far-end signal $y(n)$ and the near-end signal $v(n)$

$$d(n) = y(n) + v(n). \tag{1}$$

It is assumed that the echo path can be modelled by an N tap time varying FIR filter $\mathbf{h}(n)$,

$$y(n) = \mathbf{h}^T(n)\mathbf{x}(n) \tag{2}$$

where $\mathbf{x}(n)$ is a N element vector with its lth element as

$$\mathbf{x}_l(n) = x(n-l) \qquad l \in [0, N-1]. \tag{3}$$

A synthetic echo is generated with an N tap FIR filter $\hat{\mathbf{h}}(n)$ as

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n)\mathbf{x}(n), \quad (4)$$

and subtracted from the microphone signal. The resulting error signal

$$e(n) = d(n) - \hat{y}(n) \quad (5)$$

is transmitted to the far-end.

The proposed algorithm exploits the coherence function between the synthetic echo and the error signal, defined as

$$\gamma_{\hat{y}e}(\omega) = \frac{|S_{\hat{y}e}(\omega)|^2}{S_{\hat{y}\hat{y}}(\omega)S_{ee}(\omega)} \quad (6)$$

where $S_{x_1x_2}(\omega)$ is the cross-spectrum of the signals $x_1(n)$ and $x_2(n)$. The error signal $e(n)$ consists the residual echo $\epsilon(n)$ and the near-end signal

$$e(n) = \epsilon(n) + v(n) \quad (7)$$

$$\epsilon(n) = y(n) - \hat{y}(n). \quad (8)$$

Assuming that $\epsilon(n)$ is uncorrelated with $v(n)$, one has

$$\gamma_{\hat{y}e}(\omega) = \frac{S_{\epsilon\epsilon}(\omega)}{S_{ee}(\omega)} = 1 - \frac{1}{1 + \frac{S_{\epsilon\epsilon}(\omega)}{S_{vv}(\omega)}}. \quad (9)$$

It is seen from (9) that the coherence function $\gamma_{\hat{y}e}(\omega)$ is governed by the residual echo to near-end signal ratio $S_{\epsilon\epsilon}(\omega)/S_{vv}(\omega)$. The properties of the coherence function is summarized in Table 1. The situations in which the AEC is operating are denoted by a three digit binary number, with the first two digits representing the activities of the far- and near-end users (0='inactive',1='active') respectively and the third digit representing the sufficiency of adaptation (0='sufficient',1='insufficient'). The situations 00x (idle) and 01x (near-end single talk) can easily be recognized by using a far-end voice activity detector. There are existing algorithms that reliably detect the situation 100. In this work, we use the algorithm presented in [4]. The situation 101 (echo path variation) and 110 (double talk) can be differentiated by looking at the coherence function $\gamma_{\hat{y}e}(\omega)$ because $\gamma_{\hat{y}e}(\omega)$ changes in opposite directions in these two situations. In the situation 111 (double talk plus echo path variation) $\gamma(\omega)$ can change in either direction because both $S_{\epsilon\epsilon}(\omega)$ and $S_{vv}(\omega)$ increase. Nonetheless, it is reasonable to expect an adaptive filtering algorithm being capable of adapting the echo cancellation filter toward the echo path when the residual echo $\epsilon(n)$ is stronger than the near-end signal $v(n)$. Therefore, the situation 111 can be treated as 101 or 110 depending on the ratio $S_{\epsilon\epsilon}(\omega)/S_{vv}(\omega)$.

Situation	$S_{\epsilon\epsilon}(\omega)$	$S_{vv}(\omega)$	$\gamma_{\hat{y}e}(\omega)$
00x	≈ 0	≈ 0	?
01x	≈ 0	$\neq 0$	≈ 0
101	$\uparrow$	unchanged	$\uparrow$
100	$\leq S_{vv}(\omega)$	$\geq S_{\epsilon\epsilon}(\omega)$	≤ 0.5
110	unchanged	$\uparrow$	$\downarrow$
111	$\uparrow$	$\uparrow$	?

Table 1: Properties of $\gamma_{\hat{y}e}(\omega)$ in different situations.

3. ECHO PATH VARIATION DETECTION

To detect echo path variation, a detection variable $\xi(k)$ is formed based on the estimated coherence function $\hat{\gamma}_{\hat{y}e}(\omega, k)$ at iteration k and compared to a threshold T_γ. If $\xi(k) \geq T_\gamma$, echo path variation is detected.

The detection variable $\xi(n)$ is constructed by averaging the coherence estimate $\hat{\gamma}_{\hat{y}e}(\omega, k)$ over time and frequency as

$$\xi(k) = \lambda\xi(k-1) + \frac{1-\lambda}{I}\sum_{i=0}^{I-1}\hat{\gamma}_{\hat{y}e}(\omega_i, k). \quad (10)$$

The frequencies over which the coherence function is averaged are chosen as follows. A set of frequencies in the range $300 \leq \omega f_s/(2\pi) \leq 1800$ Hz such that every pair of the frequencies in the set are separated by a minimum of $W f_s/(2\pi)$ Hz. The coherence function is then averaged over the I frequencies at which the estimated error signal spectrum $\hat{S}_{ee}(\omega)$ takes its I largest values. The frequencies are constrained to between 300-1800 Hz because this is the frequency region where the dominating part of speech energy falls in. The minimum separation between the frequency samples is imposed so that the frequency samples can be considered independent of each other. By picking peaks of the error signal spectrum, we avoid averaging the coherence estimate at those frequencies where the excitation is poor.

For a well designed adaptive filtering algorithm, it is reasonable to expect that the steady state residual echo is slightly weaker than the near-end signal due to the average involved in the adaptation procedure. Therefore, a natural setting of the threshold T_γ would be

$$T_\gamma = 0.5. \tag{11}$$

However, the time average in the construction of the detection parameter introduces a delay in the detection. At the rising edge of $\xi_(k)$, this delay slows down the tracking of echo path variation. The detection delay is particular harmful in the situation 111. During the near-end speech pause, $S_{vv}(\omega)$ falls low, and $S_{\epsilon\epsilon}(\omega)$ remains high because of the mismatch between the echo cancellation filter and the echo path remains high. The detection variable $\xi(k)$ would rise and may exceed the threshold in this circumstance. After the speech pause, $S_{vv}(\omega)$ rises and $\xi(k)$ may not be able to follow the rising fast enough. This allows the filter to be adapted when the near-end speech signal is actually strong enough to cause divergence. To account for the detection delay, we set two different thresholds, one for the rising edge of $\xi(k)$, the other for the dropping edge, as

$$T_\gamma = \begin{cases} 0.5 - \phi & \xi(k) \geq \xi(k-1) \\ 0.5 + \varphi & \xi(k) < \xi(k-1). \end{cases} \tag{12}$$

where $0 < \phi, \varphi \leq 0.1$. By setting the thresholds in this way, we can advance the threshold crossing and mitigate the ill-effects of the detection delay.

4. SIMULATION RESULTS

Numerical examples are presented in this section to demonstrate the performance of the proposed echo path variation detector.

The first example concerns the capability to detect echo path variation without any near-end speech. A male speech is broadcasted into an office. At sample 100000, the microphone is displaced by 4cm to simulate echo path variation. A 1024 tap adaptive filter is used to cancel the echo. The adaptive filter is adapted in 128 subbands every $D = 64$ samples using the subband robust algorithm [5]. In Fig. 2(a), the adaptation is controlled using the angle parameter [4]

$$\rho(k) = \frac{\sum_{l=0}^{L_\rho - 1} \hat{y}(kD-l)d(kD-l)}{\sqrt{\sum_{l=0}^{L_\rho - 1} \hat{y}^2(kD-l) \sum_{l=0}^{L_\rho - 1} d^2(kD-l)}}, \tag{13}$$

with $L_\rho = 64$ and a threshold $T_\rho = \sqrt{0.86}$. $\xi(k)$ is also calculated with $I = 3$, $\lambda = 0.9$. The coherence function is estimated with 128 point FFT using the Welch method. The detection variables $\rho(k)$ and $\xi(k)$ are plotted. It can be seen that $\rho(k)$ falls below the threshold T_ρ when echo path varies, and the echo canceller is not able to track the echo path variation. This is confirmed by the convergence curve (dotted line) depicted in Fig. 2(b). The parameter $\xi(k)$, on the other hand, rises beyond the threshold T_γ and remain above it because the adaptation is frozen, giving clear indication of echo path variation. We update the echo cancellation filter when either $\rho(k) \geq T_\rho$ or $\xi(k) \geq T_\gamma$. The result is shown in Fig. 2(b) by the solid line. It is clear that the AEC is capable of tracking echo path variation and this capability comes from the incorporation of the proposed method for detecting echo path variation.

In the second example, echo path variation is introduced at sample 100000 by displacing the microphone by 4 cm as in the first example. A female speech signal is added to the microphone signal from sample 90000 to 110000 to simulate the near-end speech. The adaptation of the echo cancellation filter is controlled as such. When both $\rho(k)$ and $\xi(k)$ is below their thresholds, double talk is declared and the adaptation is frozen. The detection of double talk is held for 4 iterations, i.e., the adaptation is suspended in the immediate following 4 iterations regardless of the result of the detection. When $\xi(k)$ exceeds T_γ, echo path variation is detected and the echo cancellation filter is updated. The scale factors for the robust adaptation are updated with a forgetting factor of 0.75 in this case. When $\xi(k) < T_\gamma$ and $\rho \geq T_\rho$, far-end single talk is detected and the echo cancellation filter is adapted. The scale parameters are updated with a forgetting factor 0.95. The threshold settings are the same as in the first example. The convergence curves with different levels of near-end speech (relative to the echo level) are depicted in Fig. 3. It can be seen from the figure that echo path variation and double talk is properly discriminated.

5. CONCLUSIONS

In this paper, we propose a method for detecting echo path variation. In the proposed method, the detection parameter is constructed by averaging the coherence function between the synthetic echo and the error signal over time and frequency. The detection delay caused by the time averaging is accounted for by setting different detection thresholds for the rising and dropping edges of the detection parameter. It is shown by numerical examples that the proposed method can reliably differentiate between echo path variation and double talk.

6. REFERENCES

[1] C. Breining et. al., "Acoustic echo control: an application of very-high-order adaptive filters," *IEEE SP Mag.*, pp. 42–69, July 1999.

[2] T. Gänsler, J. Benesty, and S.L. Gay, "Double-talk dection schemes for acoustic echo cancellation," in *Acoustic signal processing for telecommunication*, S.L. Gay and J. Benesty, Eds., chapter 5, pp. 81–97. Kluwer Academic Publishers, 2000.

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

[4] K. Ghose and V.U. Reddy, "A double-talk detector for acoustic echo cancellation applications," *Signal Processing*, vol. 80, pp. 1459–1467, 2000.

[5] J. Huo, S. Nordholm, , K.F.C. Yiu, and K.L. Teo, "Sub-band and frequency domain double talk robust algorithms," in *Proc. RVK'2002*, 2002.

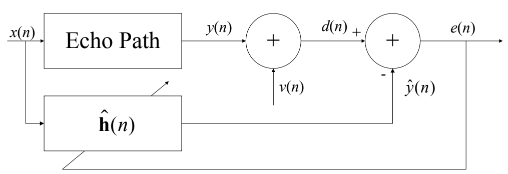
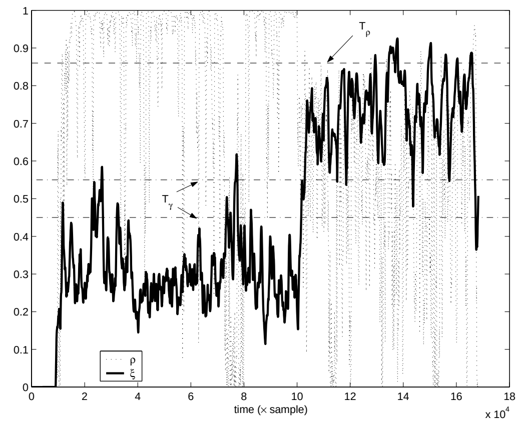
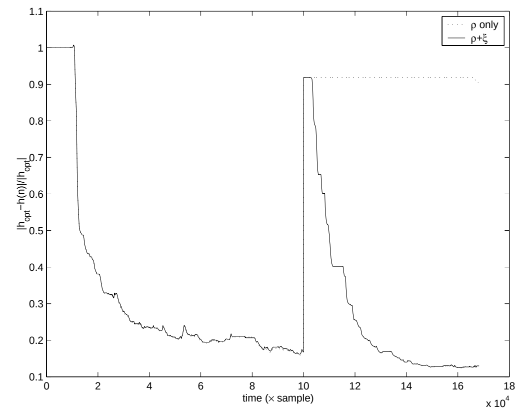


Figure 1: Basic idea of acoustic echo cancellation.



(a) Detection parameter behavior during echo path variation.



(b) Tracking of echo path variation.

Figure 2: Tracking of echo path variation.

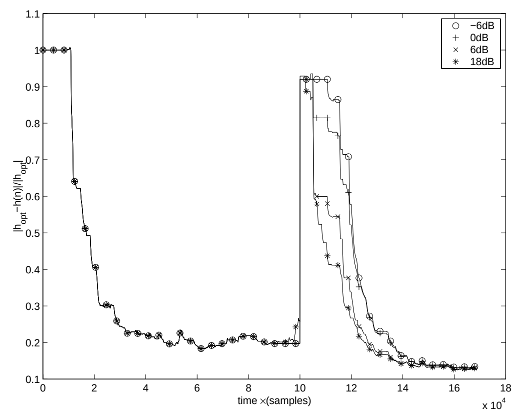


Figure 3: Differentiation between echo path variation and double talk.