A NEW DOUBLE-TALK DETECTOR USING ECHO PATH ESTIMATION

Hae Kyung Jung, Nam Soo Kim, and Taejeong Kim

School of Electrical Engineering and the Institute of New Media and Communications
Seoul National University
Kwanak P. O. Box 34, Seoul 151-742, Korea
shizuka@infolab.snu.ac.kr, nkim@snu.ac.kr, tkim@snu.ac.kr

ABSTRACT

This paper presents a new double-talk detector (DTD) based on echo path estimation. The proposed algorithm consists of two stages. In the first stage, single-talk periods are distinguished from the double-talks or echo path changes based on the energy level of the echo path estimate. An accurate distinction between the double-talk and echo path change is made in the second stage based on the gradient of the energy level in the estimated echo path. By experiments, it is found that the proposed approach is effective in detecting double-talk periods. Moreover, the required decision delay is shorter than that of the conventional methods.

1. INTRODUCTION

Since the birth of telecommunication, the echo has been a long-standing problem hindering comfortable telephone conversations. Recently, a number of practical studies on echo cancelling can be found in the literature. Adaptive echo cancellers (AEC’s) of the finite impulse response (FIR) filter type have been dominantly among the echo cancellers. For the adaptive FIR system, the least mean square (LMS) and normalized LMS (NLMS) algorithms are prevailing owing to their simplicity and predictable behavior [1]. Fig.1 shows the basic structure of the AEC where the coefficients of the adaptive filter are updated so as to minimize the error $e(n)$ between the real echo signal and its replica generated by the adaptive filter.

One problem of the AEC’s is that the performance deteriorates drastically during the double-talk periods in which the signals from both the near-end and far-end speakers co-exist. In a common telephone call, the double-talk is found to occupy up to 20 percent of the whole period. During the double-talk periods, the echo is mixed into a large and strongly correlated near-end speech $u(n)$ and results in a large interference component in the error $e(n)$. Consequently, the AEC will quickly diverge from its converged state if the adaptive filter continues to update its coefficients during the double-talk periods.

For the above reason, a DTD is a very important part in practical AEC. According to the decision of the DTD, the adaptive filter updates its coefficients during the single-talk periods and freezes adaptation during the double-talk periods to avoid divergence. One of the problems in conventional DTD methods is that there is no way to distinguish the echo path change from the double-talk [2]-[4]. This distinction is important because the adaptive filter coefficients should be continuously updated during the echo path changes but not during the double-talk periods. Another problem is that the decision process requires a long delay.

In this paper, we introduce a new approach to double-talk detection based on echo path estimation. Three hypotheses exist in echo cancellation: the single-talk, the double-talk, and the echo path change. The proposed algorithm distinguishes the single-talk from the double-talk or echo path change by comparing the energy level of the echo path estimate. In addition, the double-talk is distinguished from the echo path change by the gradient of this energy level. Furthermore, it leads to a shorter decision delay than the

This research was supported by Korea Science and Engineering Foundation under Grant No. 98-0101-03-01-3 and by the Brain Korea 21 Project.
conventional method. In section 2, the echo path estimation is introduced and formulated. In section 3, we make the decision rules for this algorithm and a computer simulation is performed to verify the double-talk detection performance of the proposed algorithm.

2. DERIVATION OF THE ECHO PATH ESTIMATION

2.1. Single-talk case

Let us denote the far-end signal, near-end signal, channel noise, microphone signal, and echo path as \( x, u, w, y, \) and \( h, \) respectively.

In the single-talk case (when \( u \) is absent), the microphone signal \( y \) is described as

\[
y(n) = x(n) * h(n) + w(n) = x_n^T h_n + w(n)
\]

where

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

\[
h_n = [h(n) \ h(n-1) \ \cdots \ h(n-L+1)]^T,
\]

and \( L \) is the filter size. Multiplying both the sides of (1) by \( x_n \) and taking expectations, we obtain

\[
E \{ x_n y(n) \} = E \{ x_n x_n^T h_n \} + E \{ x_n w(n) \}.
\]

(2)

Here, \( E \{ x_n w(n) \} \approx 0 \) since the channel noise \( w(n) \) is assumed to be uncorrelated with the far-end signal \( x(n) \). Let \( \Phi_{xy}(n) = E \{ x_n y(n) \} \) and \( R_x = E \{ x_n x_n^T \} \). Then (2) becomes

\[
\Phi_{xy}(n) = R_x h_n.
\]

(3)

Multiplying both the sides of (3) by \( x_n^{-1} \), the echo path estimate \( \hat{h}_n \) is derived as follows:

\[
\hat{h}_n := R_x^{-1} \Phi_{xy}(n)
\]

(4)

where \( \hat{h}_n = [\hat{h}(n) \ \hat{h}(n-1) \ \cdots \ \hat{h}(n-L+1)]^T \).

2.2. Double-talk case

In the double-talk case (\( u \) is present), (3) becomes

\[
\Phi_{xy}(n) = R_x h_n + \Phi_{xu}(n).
\]

(5)

Therefore, the echo path estimate in the double-talk is obtained as follows:

\[
\hat{h}_n = R_x^{-1} \Phi_{xy}(n) = h_n + \varepsilon(x_n, u)
\]

(6)

where \( \varepsilon(x_n, u) = R_x^{-1} \Phi_{xu}(n) \), which is the error in the echo path estimate caused by the near-end signal.

In order to implement this idea, we use the sequential regression (SER) algorithm [5] for \( R_x^{-1} \) and a smoothed estimate for \( \Phi_{xy}(n) \) as follows:

\[
\Phi_{xy}(n) = (1 - \alpha) \Phi_{xy}(n-1) + \alpha x_n y(n)
\]

(7)

where \( \alpha \) is a smoothing factor which lies in \((0, 1)\).

2.3. Echo path change case

Now we consider the case of an echo path change during the single-talk periods. Assume that the echo path changes from \( h_m \) to \( h'_{m} \) at time \( m \) and \( h'_{m} = h_m + \Delta h \). Since \( y(m) = x_n^T h_m + w(m) \), (7) becomes

\[
\Phi_{xy}'(m) = (1 - \alpha) \Phi_{xy}(m-1) + \alpha x_m x_n^T (h_m + \Delta h)
\]

(8)

\[
= \Phi_{xy}(m) + \alpha x_m x_n^T \Delta h.
\]

Note that \( w(m) \) does not exist in the single-talk periods and \( w(m) \) disappears during the correlation process. Therefore,

\[
\hat{h}_m' = \hat{h}_m + \alpha R_x^{-1} x_m x_n^T \Delta h
\]

(9)

\[
= h_m + \eta(x_m, \Delta h).
\]

\( \eta(x_m, \Delta h) \) is the error caused by the echo path change.

3. DECISION PROCEDURE

In order to perform double-talk detection, we define a decision parameter \( \xi_n \) computed at time \( n \) to be

\[
\xi_n := \frac{1}{L} ||\hat{h}_n||^2 = \frac{1}{L} \sum_{i=0}^{L-1} \hat{h}_i^2 (n - i).
\]

(10)

All the conditions encountered in echo cancellation, as identified by the following hypotheses, can be described in terms of this parameter.

- \( H_0 \): the single-talk periods without echo path change, \( \xi_n = \frac{1}{L} ||h_n||^2 \)
- \( H_1 \): the single-talk periods with echo path change, \( \xi_n = \frac{1}{L} ||h_n||^2 + \frac{1}{L} ||\eta(x_m, \Delta h)||^2 \)
- \( H_2 \): the double-talk periods without echo path change, \( \xi_n = \frac{1}{L} ||h_n||^2 + \frac{1}{L} ||\varepsilon(x_n, u)||^2 \)
- \( H_3 \): the double-talk periods with echo path change.

We are not concerned about \( H_3 \), the case of an echo path change during the double-talk periods, since the adaptation is frozen when the periods are decided as double-talk.
According to the above description, the first decision rule to distinguish $H_0$ from $H_1$ and $H_2$ is given as follows:

$$\xi_n = \frac{1}{L} \sum_{i=0}^{L-1} h^2(n-i) \begin{cases} H_1, & H_2 \end{cases} T_1 \quad (11)$$

where $T_1$ is a threshold. Next, let us compare $\eta(x_n, \Delta_h)$ in (9) with $\varepsilon(x_n, u)$ in (6) to distinguish $H_1$ from $H_2$.

- $H_1: \eta(x_n, \Delta_h) = \alpha R_{x}^{-1} x_n x_n^T \Delta_h$
- $H_2: \varepsilon(x_n, u) = R_{x}^{-1} \Phi_{xu}(n) = \alpha R_{x}^{-1} x_n u(n)$

In general, $\varepsilon$ is found small at the start of a double-talk period because $u(n)$, which usually represents a speech signal, grows slowly from a very small value. On the other hand, upon an echo path change, $\eta$ grows rapidly as the magnitude of $\Delta_h$ takes a relatively large value. As a result, $\xi_n$ increases more rapidly during the echo path change than during the double-talk periods. This leads us to propose an approach to detecting $H_1$ and $H_2$ based on the gradient of $\xi_n$ as follows:

$$\sum_{i=0}^{M-1} \beta_i (\xi_{n-i} - \xi_{n-i-1}) \begin{cases} H_1, & H_2 \end{cases} T_2 \quad (12)$$

where $\beta_i$ is a weighting factor and $T_2$ is a threshold.

### 4. RESULTS

To evaluate the proposed algorithm, we performed computer simulations. A sentence spoken by a female speaker was used to form the far-end signal $x(n)$ while another sentence from a male speaker was used for the near-end signal $u(n)$. White Gaussian noise with zero mean was used to simulate the channel noise $w(n)$ at the signal to noise ratio (SNR) of 30dB. In decision procedure, we let the output of the DTD in ‘H1:echo path change’ be the same as that in ‘H0:single-talk’ because the coefficients of the adaptive filter should also be updated during the echo path change.

First, the decision parameter $\xi_n$ in the cases of the double-talk and the echo path change is plotted in Fig.2. We find in the figure that, as pointed out earlier, $\xi_n$ at the starting point ($n = 3800$) of echo path change increases more rapidly than at the starting points ($n = 1267$ and $n = 4536$) of the double-talk periods. Based on these observations, we can ensure that (12) is a decision rule appropriate to detecting $H_1$ and $H_2$. Fig.3 shows a simulation result comparing the performance of the proposed method with that of the conventional method [3]. When echo path changes, the conventional method makes a wrong decision for the double-talk, whereas the proposed method makes the correct decision declaring the single-talk. Besides, we can also see that the starting points of the double-talk periods indicate a significant reduction of the decision delay in the proposed method as compared to the conventional method. Fig.4 shows the performance results under various near-end signal to far-end signal ratio (NFR) environments. From the results, it is shown that the proposed method noticeably shortens the delay required for the double-talk detection.

### 5. CONCLUSION

In this paper, we have presented a new DTD using echo path estimation. In an adaptive echo canceller, the distinction between the echo path change and the double-talk is very important. Using the fact that the echo path estimate increases more rapidly during the echo path changes than during the double-talk periods, we have proposed a gradient-based de-
Fig. 4. Comparison between the proposed method and the conventional method under various NFR environments.