# DOUBLE-TALK DETECTION SCHEMES FOR ECHO CANCELLATION

**Heping Ding[1] and Frank Lau[2]**
[1]Acoustics and Signal Processing, IMS, National Research Council, 1200 Montreal Rd., Ottawa, Ontario K1A 0R6
*heping.ding@nrc-cnrc.gc.ca*
[2]Nortel Networks, Westwinds Innovation Center, 5050 40th St. NE, Calgary, Alberta T3J 4P8
*franklau@nortelnetworks.com*

## 1. Introduction

Widely used in telecommunications, an acoustic or network echo cancellation system, as Fig. 1 shows, subtracts an echo estimate $y(n)$, made by an adaptive filter inputting the far-end signal $x(n)$, from the desired input $d(n)$ to reduce the echo $u(n)$ therein while leaving the near-end signal $s(n)$ intact. In order for $y(n)$ to approximate $u(n)$, the filter coefficients are updated by an adaptation algorithm seeking the minimum of the mean of $e^2(n)$ so that the adaptive filter converges to mimic the echo path. Fundamentals of adaptive filtering and echo cancellation can be found in a text book, such as [1].

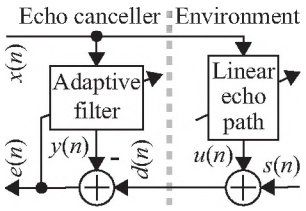


Fig. 1. Echo cancellation

In a practical echo canceller, the adaptation should be a) active when there is little $s(n)$ - so that the filter can converge; and b) halted if $s(n)$ is significant - to prevent the filter from diverging. The key element in controlling this is a so-called double-talk (DT) detector, which detects conditions where $d(n)$ contains a significant near-end signal $s(n)$ mixed with a much stronger $u(n)$ (e.g., 25 dB stronger in a typical speakerphone application), echo of the far-end signal $x(n)$. It is a challenge in academia and industry to find DT detectors that do this job with few misses and false alarms.

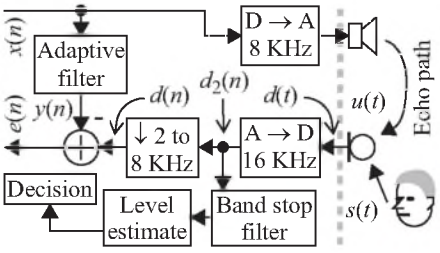


Fig. 2. Proposed Scheme 1 for DT detection

Typical existing DT detection schemes are reviewed in a separate paper [2] and will not be discussed here. This paper presents two new schemes, see for example [3] and [4]†. These schemes have been demonstrated to be pretty reliable, robust, and simple.

## 2. Scheme 1: Out-of-Telephony-Band Level

In telephony, $u(n)$, linear echo of $x(n)$, has a frequency bandwidth of about 300-3400 Hz, the same as that of $x(n)$. On the other hand, a voice $s(t)$ usually has a wider band, i.e., containing energy outside of this range, and, in a speaker phone, $d(t)$ with a wider band can be readily available since its acoustic source is local. Thus, a DT condition can be claimed if $d(t)$ contains significant energy outside of the telephony band.

Fig. 2 shows the proposed scheme, where $d_2(n)$ is the analog input $d(t)$ sampled at 16 KHz, double the telephony system's sampling rate. In addition to being decimated to form the needed $d(n)$, $d_2(n)$ is band-stop filtered, ridding components within 300-3400 Hz and retaining those within 50-300 Hz and 3400-7000 Hz, for level estimation and decision making.

Simulations with various talkers and speech samples were performed. Fig. 3 shows a typical example, where $d_2(n)$ contains an $s(n)$ 25 dB weaker than $u(n)$ so that it looks almost the same as $u(n)$. This big level difference makes DT detection difficult for some conventional schemes, but the proposed scheme correctly identifies places of significant $s(n)$.

Scheme 1 effectively detects $s(n)$ masked by a much stronger $u(n)$ if a wider band $d(t)$ is made available by the electronics; therefore, it is suitable for speakerphone applications.

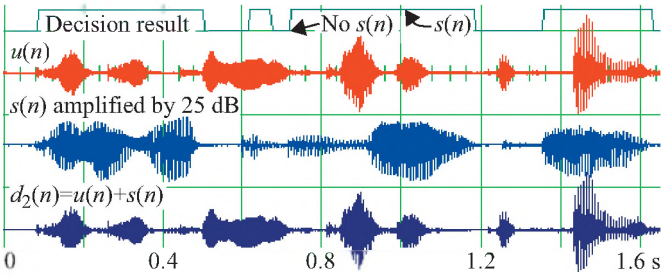


Fig. 3. Simulation result for Scheme 1

† Heping Ding was employed by Nortel Networks during development of the work presented in this article. Nortel Networks owns intellectual property rights associated with the work and this article.

## 3. Scheme 2: Orthogonality Principle

Comparing levels of various signals in Fig. 1, an energy comparison scheme, discussed in [2], can detect the presence of $s(n)$ *or* a non-convergence echo (uncanceled echo due to non-convergence of the adaptive filter), in $e(n)$, but it cannot distinguish between the two, because either of them results in the same fluctuations in signal levels. Assuming that $s(n)$ is uncorrelated with $x(n)$, the proposed Scheme 2 does the distinction by evaluating the inner product $E[e(n)y(n)]$, where $E[\cdot]$ stands for expectation.

Considering the adaptive filter to be $L$-tap FIR and with

$$\underline{X}(n) \equiv \left[x(n)\ x(n-1)\ \ldots\ x(n-L+1)\right]^T,\ \boldsymbol{R}(n) \equiv E[\underline{X}(n)\underline{X}^T(n)],$$
$$\underline{W} \equiv \left[w_0\ w_1\ \ldots\ w_{L-1}\right]^T,\ \text{and}\ y(n) = \underline{X}^T(n)\underline{W}, \tag{1}$$

the expectation of interest can be found as

$$E[e(n)y(n)] = (\underline{W}^{\text{opt}} - \underline{W})^T \boldsymbol{R}(n)\underline{W}, \tag{2}$$

where $\underline{W}^{\text{opt}}$ is the optimal value of the filter coefficient vector $\underline{W}$. Once the adaptive filter has converged, i.e., $\underline{W} = \underline{W}^{\text{opt}}$, Eq. (2) vanishes - whether there is an uncorrelated $s(n)$ or not. If the adaptive filter has not converged, Eq. (2) will in general be non-zero - although not guaranteed, it is so in practice. This can also be understood intuitively in an infinite-dimensional Hilbert space, in which $x(n)$, ..., $x(-\infty)$, $y(n)$, $d(n)$, and $e(n)$ are vectors. $d(n)$ consists of: $u(n)$ being a linear combination of $\{x(n), \ldots, x(-\infty)\}$, and $s(n)$ being not. A linear combination of $\{x(n), \ldots, x(n\text{-}L\text{+}1)\}$, the optimum $y(n)$, $y^{\text{opt}}(n) = \underline{X}^T(n)\underline{W}^{\text{opt}}$, can resemble part of $u(n)$, but not the uncorrelated $s(n)$ or the residual echo due to under modeling. $y^{\text{opt}}(n)$ is the projection of $d(n)$ onto the sub-space occupied by $\{x(n), \ldots, x(n\text{-}L\text{+}1)\}$ and is orthogonal to the projection error $e(n)$.

To summarize, $E[e(n)y(n)]$ becomes small with a converged adaptive filter, regardless of $s(n)$, and is large in magnitude if the filter has not converged.

With ensemble means replaced by time averages, the scheme is implemented as

$$\begin{aligned} R_{ey}(n) &= \beta R_{ey}(n-1) + (1-\beta)e(n)y(n) \\ A_{ey}(n) &= \beta A_{ey}(n-1) + (1-\beta)|e(n)y(n)| \end{aligned}, \tag{3}$$

where $0 < \beta < 1$ is a forgetting factor to provide a low-pass smoothing effect, and $A_{ey}(n)$ gives a reference for evaluating the magnitude of $R_{ey}(n)$. When the energy comparison part claims a candidate for the presence of either $s(n)$ or a non-convergence echo in $e(n)$, a distinction is made as per

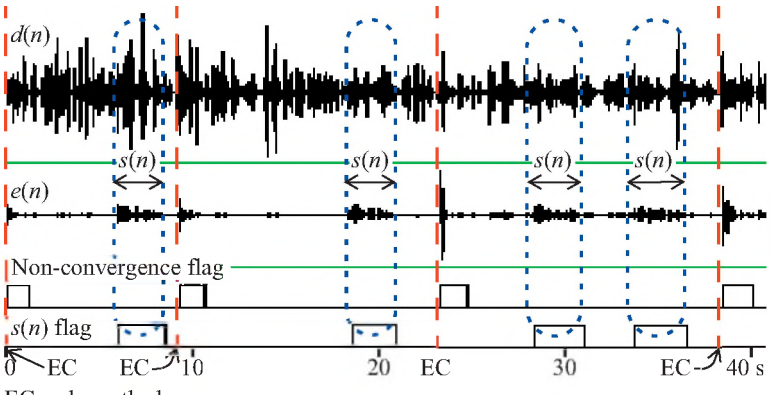
$$|R_{ey}(n)| > T_{\text{echo}} A_{ey}(n) \quad \Rightarrow \text{Candidate is non-conv. echo} \tag{4}$$

$$|R_{ey}(n)| < T_{s(n)} A_{ey}(n) \quad \Rightarrow \text{Candidate is } s(n), \tag{5}$$

where the thresholds $0 < T_{s(n)} \le T_{\text{echo}} < 1$ are experimentally determined.

Unlike Scheme 1, this scheme does not need a wider band $d(t)$, and therefore is suitable for general echo cancellation.

Simulations with similar conditions as that with Scheme 1 were performed, with an example shown in Fig. 4. It is seen that $e(n)$ contains both $s(n)$ and non-convergence echoes, as results of a non-converged adaptive filter after echo path changes. An energy comparison based DT detection scheme is only able to spot these $e(n)$ level increases as candidates for either of the two events, while the proposed Scheme 2 clearly distinguishes between them by raising the non-convergence flag if Eq. (4) is satisfied, or the $s(n)$ flag if Eq. (5) is met.



EC: echo path change

Fig. 4. Simulation result for Scheme 2

## 4. Summary

This paper presents two DT detection schemes which, compared to other existing ones reviewed in [2], are reliable, robust, and relatively simple in terms of implementation. Both schemes have been verified in simulations and Scheme 2 has further been incorporated into acoustic and network echo cancellation products.

## References


[1] Simon Haykin, *Adaptive Filter Theory*, 4th Edition, Prentice Hall, Sept. 2001.

[2] Thien-An Vu, Heping Ding, and Maritn Bouchard, "A survey of double-talk detection schemes for echo cancellation applications," *Acoustics Week in Canada* - Canadian Acoustical Association, Oct. 6-8, 2004, Ottawa, Canada.

[3] Heping Ding and Frank Lau, "Circuit and Method of Double Talk Detection for Use in Handsfree Telephony Terminals," U.S. Patent 6,049,606, issued April 11, 2000, assigned to Nortel Networks.

[4] Heping Ding, "Method of Distinguishing between Echo Path Change and Double Talk Conditions in an Echo Canceller," U.S. Patent 6,226,380, issued May 1, 2001, assigned to Nortel Networks.