Accuracy Improved Double-Talk Detector Based on State Transition Diagram

SangGyun Kim, Jong Uk Kim, and Chang D. Yoo

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology, Republic of Korea
zom@eeinfo.kaist.ac.kr, oribros@mail.kaist.ac.kr, and cdyoo@ee.kaist.ac.kr

Abstract

A double-talk detector (DTD) is generally used with an acoustic echo canceller (AEC) in pinpointing the region where far-end and near-end signal coexist. This region is called double-talk and during this region AEC usually freezes the adaptation. Decision variable used in DTD has a relatively longer transient time going from double-talk to single-talk than time going in opposite direction. Therefore, using a single threshold to pinpoint the location of double-talk region can be difficult. In this paper, a DTD based on a novel state transition diagram and a decision variable which requires minimal computational overhead is proposed to improve the accuracy of pinpointing the location. The use of different thresholds according to the state helps the DTD locate double-talk region more accurately. The proposed DTD algorithm is evaluated by obtaining a receiver operating characteristic (ROC) and is compared to that of Cho's DTD.

1. Introduction

An acoustic echo canceller (AEC) is used to eliminate acoustic feedback that is generated due to the coupling between a loudspeaker and a microphone as shown in Fig. 1. The far-end signal x goes through the echo path $\mathbf{h}$ and adds to the microphone signal y with the near-end signal v and noise n. The AEC adaptively estimates the echo path between the loudspeaker and the microphone [1]. When near-end signal is absent, the adaptive filter $\hat{\mathbf{h}}$ can converge to a good estimate of the echo path $\mathbf{h}$. However, when near-end signal and far-end signal coexist, the near-end signal acts as noise and hinders the estimation of $\mathbf{h}$. This signal may cause divergence. In order to prevent the adaptive filter from diverging during double-talk, a double-talk detector (DTD) is used to pinpoint double-talk regions where adaptation is stopped.

Traditionally, double-talk regions are located by comparing the decision variable of DTD to a preset threshold. To form a decision variable, several methods have been proposed [2]-[5]. In [2], a DTD algorithm based on the orthogonality theorem was proposed. In [3], the coherence between x and y was used as a decision variable. However, these algorithms require high computational complexity. Recently, a DTD algorithm based on cross-correlation was proposed [4], [5].

In general, a decision variable is calculated recursively with a forgetting factor in order to reduce the computational complexity and memory size. However, it has a relatively longer transient time going from double-talk to single-talk than time going in opposite direction. Therefore, using a single threshold to pinpoint the location of double-talk region can be difficult. In order to solve this problem, a DTD method using two decision variables, one of which has a relatively short transient time compared to the other, has been proposed in [4]. However, this method requires the calculation of two decision variables. In [6], a power that was necessary in calculating of the decision variable was updated periodically with a new average power. However, this method also requires additional computation for calculating the new average power periodically.

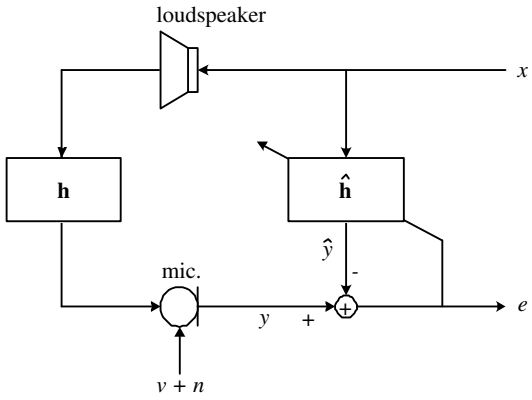

Figure 1: *Representation of AEC problem.*

In this paper, a DTD based on five-state transition diagram is proposed to solve the problem simply. The proposed DTD uses several thresholds according to the state. The state changes according to the value of decision variable. Therefore, we can use appropriate thresholds for the endpoint detection of double-talk. This can improve the accuracy of DTD. In this paper, the calculation of the decision variable proposed by Cho et al. is also simplified [5].

This paper is organized as follows. Section 2 presents simplified form for the decision variable proposed by Cho. Section 3 shows a DTD based on state transition diagram. Section 4 shows the simulation results and Section 5 concludes the paper.

2. Simplification of the calculation of decision variable

Cho proposed a DTD method based on a normalized cross-correlation vector and his decision variable is given by

$$\begin{aligned} \xi_C &= \sqrt{\mathbf{r}_{\mathbf{x}y}^T(\sigma_y^2\mathbf{R}_{\mathbf{xx}})^{-1}\mathbf{r}_{\mathbf{x}y}} \\ &\approx \sqrt{\frac{\mathbf{r}_{\mathbf{x}y}^T\hat{\mathbf{h}}}{\sigma_y^2}} \end{aligned} \tag{1}$$

where $\mathbf{x}[n] = [x[n], x[n-1], ..., x[n-L+1]]^T$, $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}[n]\mathbf{x}[n]^T\}$, $\mathbf{r}_{\mathbf{x}y} = \mathbf{R}_{\mathbf{xx}}\mathbf{h}$, and σ_y^2 is the power of the microphone input y [5]. An approximation of (1) can be obtained

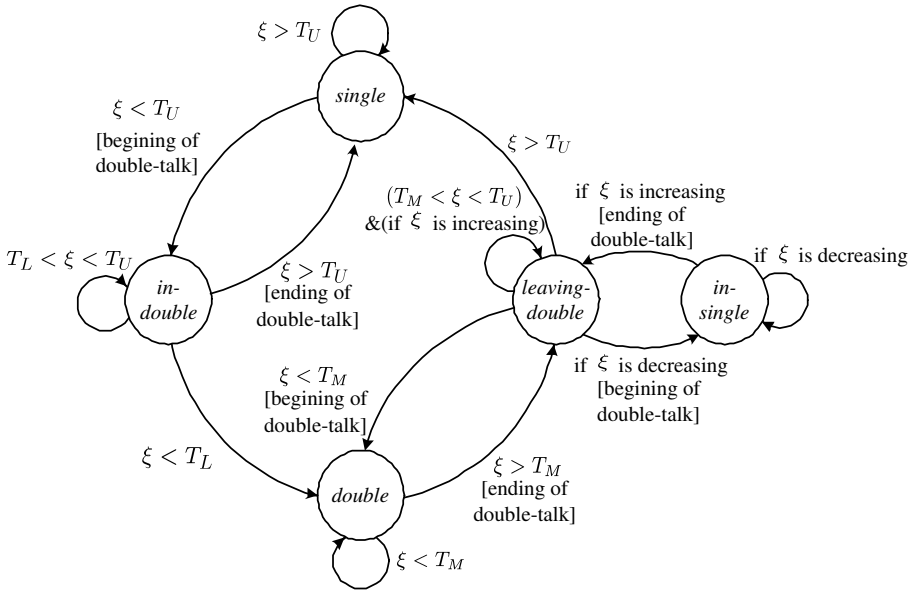

Figure 2: *State transition diagram for double-talk detection.*

by $\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{r}_{\mathbf{x}y} = \mathbf{h} \approx \hat{\mathbf{h}}$, assuming adaptive filter of the length L has converged. Parameter $\mathbf{r}_{\mathbf{x}y}$ at time n in (1) is estimated recursively as follow

$$\hat{\mathbf{r}}_{\mathbf{x}[n]y[n]} = (1-\alpha)\hat{\mathbf{r}}_{\mathbf{x}[n-1]y[n-1]} + \alpha\mathbf{x}[n]y[n] \tag{2}$$

where α is a forgetting factor. The recursive calculation is used in order to reduce computational complexity and memory size.

In this paper, the calculation of the above decision variable is simplified more. By setting $\mathbf{r}_{\mathbf{x}y} = E\{\mathbf{x}[n]y[n]\}$ into (1) and solving it, a simplified form of decision variable ξ is obtained as

$$\begin{aligned} \xi &= \sqrt{\frac{E\{\mathbf{x}^T[n]y[n]\}\hat{\mathbf{h}}}{\sigma_y^2}} \\ &= \sqrt{\frac{E\{\hat{y}[n]y[n]\}}{\sigma_y^2}} \\ &= \sqrt{\frac{r_{\hat{y}y}}{\sigma_y^2}} \end{aligned} \tag{3}$$

where $\hat{y}$ is an estimate of echo signal. Parameter $r_{\hat{y}y}$ is also estimated recursively as follow

$$\hat{r}_{\hat{y}[n]y[n]} = (1-\alpha)\hat{r}_{\hat{y}[n-1]y[n-1]} + \alpha\hat{y}[n]y[n]. \tag{4}$$

Equation (3) the value of ξ is close to 1 during single-talk since $\hat{y}$ is highly correlated to y and ξ is less than 1 during double-talk since cross-correlation between $\hat{y}$ and y is small. The use of (3) to calculate the decision variable instead of (1) leads to a high computational reduction: to compute $\mathbf{r}_{\mathbf{x}y}^T\hat{\mathbf{h}}$ requires ($3\times L$+1) multiplications per sample, however, to compute $r_{\hat{y}y}$, only 3 multiplications per sample is required. If the length of the adaptive filter L is 1000, the number of total multiplications to calculate the numerator part of ξ is approximately reduced by 1/1000.

3. A DTD based on state transition diagram

A decision variable which is obtained from a recursive estimation with a forgetting factor α has a relatively long transient time going from double-talk to single-talk. Therefore, it is difficult to select a proper threshold and for this reason DTD can not accurately pinpoint the double-talk region. A double-talk region that is longer than reality is often detected. Usually the end of the double-talk is inaccurately detected since the decision variable is less sensitive to the change as it leaves the double-talk region. We can increase the sensitivity by increasing α; however, this may lead to unreliable ξ during single-talk region. In order to solve this problem, a DTD method based on two cross-correlations or periodically updated correlation was reported in [4], [6], respectively. However, these methods require far great computational overhead.

In this paper, a DTD method based on five-state transition diagram is proposed to improve the accuracy of detecting double-talk period. Fig. 2 shows the proposed state transition diagram for double-talk detection. The five states are *in-single*, *single*, *in-double*, *double*, and *leaving-double*. It is assumed that the *single* state is the initial state. The input is the decision variable and the output is the location of double-talk region. T_U, T_M, and T_L are three thresholds with $T_U > T_M > T_L$. State transition is made by comparing the value of the decision variable to the threshold of each state.

The state stays in the *single* state until $\xi < T_U$ at which point near-end signal enters the microphone. The actions are to give a starting point of double-talk for freezing the adaptation of the adaptive filter and to move to the *in-double* state. It stays in the *in-double* state until $\xi < T_L$ or $\xi > T_U$. In the case of $\xi < T_L$, the cross-correlation between the estimated echo and the microphone input signal becomes small and the state moves to the *double* state. In the case of $\xi > T_U$, near-end signal becomes small and thus the endpoint of double-talk is declared. The state moves back to the *single* state. The state diagram stays in the *double* state until $\xi > T_M$ that means the end of the double-talk. The *double* state enables a more accurate detection of the end of double-talk because T_M is selected less than T_U.

Then, the state moves to the *leaving-double* state. It stays in the *leaving-double* state while $T_M < \xi < T_U$ and ξ is increasing. In the *leaving-double* state, when $\xi > T_U$, the state moves to the *single* state and when ξ is decreasing, meaning near-end signal is present, it moves to the *in-single* state and the beginning of double-talk is declared. The *in-single* state is introduced in order to distinguish the long transition situation from the reentrance situation of near-end signal during the transition. The state stays in the *in-single* state while ξ is decreasing. When ξ is increasing again, the end of double-talk is declared and the state moves to the *leaving-double* state. And then, if $\xi < T_M$, it moves back to the *double* state and the beginning of double-talk is declared.

The proposed method based on the five-state transition diagram accurately locates endpoint of double-talk regions with minimal calculation overhead. The method based on state transition diagram enables that several thresholds can be used according to the state for an accurate double-talk detection. It introduces the *in-single* state to consider the situation when near-end signal enters the microphone during the transition from double-talk to single-talk. In this DTD method, if ξ does not becomes less than T_L during double-talk, the proposed method will operate similarly as general DTD.

4. Simulations

In this section, simulation conducted to evaluate the performance of the proposed DTD method is presented. In DTD, the role of a threshold is very important to the performance. However, the thresholds are generally selected using a heuristic method, therefore, it is difficult to evaluate and compare the performance of DTD. In order to objectively evaluate the performance of DTD, Benesty et al. proposed an objective evaluation method based on the receiver operating characteristic (ROC) [7]. The double-talk detection problem was regarded as a binary detection problem: the probability of false alarm (P_f^s) is defined as a probability of declaring detection when double-talk is not present and the probability of miss (P_m) is defined as a probability of detection failure when double-talk is present. The performance of DTD is evaluated in terms of P_m as a function of near-end and far-end signal ratio (NFR, σ_v/σ_x) under constrained P_f^s. In this paper, another evaluation factor is added. Although the threshold is selected under a given probability of the false alarm P_f^s when v=0, it does not guarantee the restriction of the probability of the false alarm (P_f) when $v \neq 0$ happens sporadically to the predetermined value: especially, DTD suffers a wrong detection at the endpoint of double-talk. Therefore, the performance of DTD is also evaluated in terms of P_f as a function of NFR.

The simulation is conducted with 8 kHz sampled near-end and far-end signals and for better statistical significance, the P_m and P_f are obtained by averaging over 12 different conditions: four different 0.8-s, 0.9-s, and 1-s near-end signals at three different positions within the 3-s far-end signal. The above simulation is repeated over a range of NFR values from -20 dB to 20 dB. The characteristics of P_m and P_f are measured under the constraints $P_f^s = 0.1$ and $P_f^s = 0.3$. Once the double-talk or single-talk is declared, the detection is held for 15 ms in order to suppress detection dropouts due to the noisy behavior of the decision variable. The probability of false alarm when v=0 is calculated as

$$P_f^s = \frac{\Sigma D}{X} \tag{5}$$

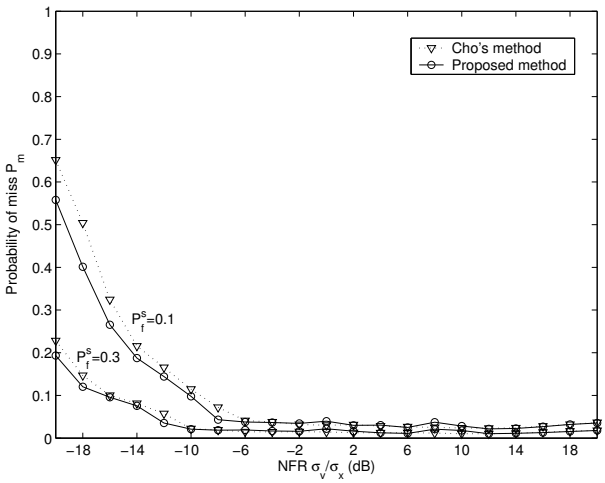

Figure 3: P_m *characteristics versus NFR* σ_v/σ_x.

where D is the DTD output and X is the length of the entire far-end signal. In the proposed DTD, the thresholds T_L, T_M, and T_U are selected as 0.2, 0.5, and 0.98 for $P_f^s = 0.1$ and 0.2, 0.5, and 0.9865 for $P_f^s = 0.3$, respectively. For comparison, the performance of the DTD proposed by Cho is also evaluated with same data. In the method, the threshold is selected as 0.9785 for P_f^s=0.1 and 0.986 for P_f^s=0.3, respectively. Once the thresholds for a given P_f^s constraint are determined, the performance of DTD is evaluated with near-end signal at different power. The probability of miss is calculated as

$$P_m = 1 - \frac{\Sigma D \cdot V}{\Sigma V} \tag{6}$$

where V is the activity detector outputs by using NFR. The logical AND is represented as $(\cdot)$ operator. Cho counted the probability of miss only when both x and v are active. However, in this paper, it is counted as double-talk region only when v is relatively larger than x because the adaptive filter can be adapted when x is relatively larger than v although v is active. The probability of false alarm when $v \neq 0$ happens sporadically is calculated as

$$P_f = \frac{\Sigma(D - D \cdot V)}{\Sigma D}. \tag{7}$$

The simulation is conducted with the element of $\hat{\mathbf{h}}$ given by $\hat{h}_i = (1 + \epsilon_i)h_i$ where h_i is the echo path and ϵ_i is an uncorrelated Gaussian noise of power -30 dB. The adaptive filter has a length of $L = 1000$ and uses a normalized least mean-square (NLMS) algorithm [8].

The P_m characteristics of the proposed and Cho's DTD are shown in Fig. 3. It is observed that the P_m decreases with increasing NFR and converges to a value for NFR above -8 dB for both $P_f^s = 0.1$ and 0.3. Both DTD methods have similar P_m characteristics, however, the decision variable of the proposed DTD is calculated with lower computational cost.

The P_f characteristics of both DTD methods are also evaluated and shown in Fig. 4 when $P_f^s = 0.3$. The same results can also be obtained when $P_f^s = 0.1$. The proposed method has

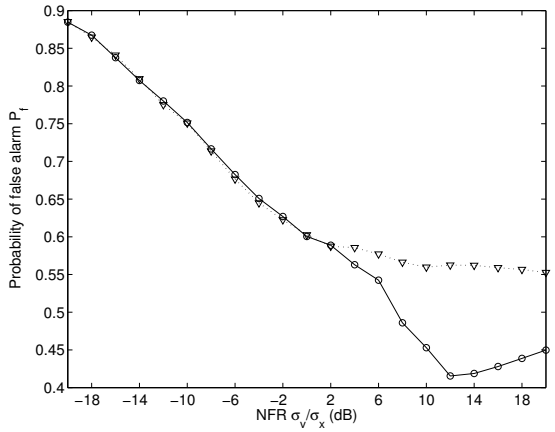

Figure 4: P_f *characteristics versus NFR* σ_v/σ_x *when* P_f^s*=0.3.*

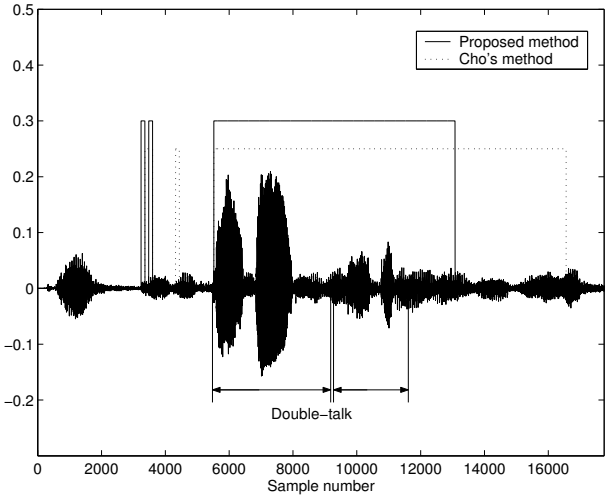

Figure 5: *DTD results of both methods.*

a smaller P_f than the Cho's method for NFR above 4 dB. In a relatively small NFR from -20 dB to 2 dB, both methods have a similar P_f characteristics since the decision variable of the proposed does not go below T_L. However, as NFR gets larger and thus the decision variable becomes less than T_L, the proposed method operates based on the state transition diagram. This enables a more accurate double-talk detection, therefore, the probability of false alarm becomes smaller than the case of Cho's method.

Fig. 5 shows an example of double-talk detection with far-end and near-end signal using both DTD methods. The double-talk exists from sample number 5490 to 9190 and from 9250 to 11610 and NFR is 10 dB. The proposed DTD method gives a more accurate detection than the DTD method proposed by Cho. Both DTD methods give inaccurate result when NFR is small. This is because ξ is affected by noise n. We can achieve better performance using the proposed DTD method than using the DTD method proposed by Cho.

5. Conclusions

A DTD method based on five-state transition diagram is proposed. The state transition diagram is introduced to alleviate the problem that a decision variable of DTD has long transient time going from double-talk to single-talk. Different thresholds and state transition conditions are applied to each state for accurate double-talk detection. In this paper, the calculation of the decision variable proposed by Cho is also simplified approximately from $3\times L$ multiplications to 3 multiplications where L is the length of the adaptive filter. The simulation results show that the proposed DTD has a similar P_m characteristics to Cho's and better P_f characteristics than Cho's when NFR is relatively high. This is due to a more accurate detection of the end of double-talk. The proposed DTD method gives more accurate double-talk detection than the conventional DTD method with a lower computational cost.

6. References

[1] M. M. Sondhi, "An adaptive echo canceler", Bell Syst. Tech. J., Vol. 46, Mar. 1967.

[2] Hua Ye and Bo-Xiu Wu, "A new double-talk detection algorithm based on the orthogonality theorem", IEEE Trans. Comm., Vol. 39, No. 11, Nov. 1991.

[3] Tomas Gansler and Goram Salomonsson, "A double-talk detector based on coherence", IEEE Trans. Comm., Vol. 44, No. 11, Dec. 1996.

[4] Seon Joon Park and Dae Hee Youn, "Integrated echo and noise canceller for hands-free applicaions", IEEE Trans. Circ. and Syst., Vol. 49, No. 3, Mar. 2002.

[5] Jacob Benesty and Jun H. Cho, "A new class of double-talk detectos based on cross-correlation", IEEE Trans. Speech and Audio Proc., Vol. 8, No. 2, Mar. 2000.

[6] S. H. Kim, H. S. Kwon, K. S. Bae, K. J. Byun, and K. S. Kim, "Performance improvement of double-talk detection algorithm in the acoustic echo canceller", IEEE Int. Conf. Acoustics, Speech, and Signal Proc., Vol. 5, 2001

[7] Jun H. Cho, Dennis R. Morgan, and Jacob Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers", IEEE Trans. Speech and Audio Proc., Vol. 7, No. 6, Nov. 1999.

[8] Simon Haykin, Adaptive filter theory, Englewood Cliffs, NJ: Prentice-Hall, 1996.