

Postprint of: Szwoch G., Czyżewski A., Kulesza M., A low complexity double-talk detector based on the signal envelope, Signal Processing, Vol. 88, Iss. 11 (2008), pp. 2856-2862, DOI: [10.1016/j.sigpro.2008.05.013](https://doi.org/10.1016/j.sigpro.2008.05.013)

© 2008. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

A low complexity double-talk detector based on the signal envelope

Grzegorz Szwoch<sup>1,2</sup>

Andrzej Czyżewski<sup>1</sup>

Maciej Kulesza<sup>1</sup>

<sup>1</sup> Gdansk University of Technology, Multimedia Systems Department, Gdansk, Poland

<sup>2</sup> Corresponding author; e-mail: [greg@sound.eti.pg.gda.pl](mailto:greg@sound.eti.pg.gda.pl), phone: +48 58 347 23 98

Gdansk University of Technology, Multimedia Systems Department, Narutowicza 11/12, 80-952 Gdansk, Poland

#### ABSTRACT

A new algorithm for double-talk detection, intended for use in the acoustic echo canceller for voice communication applications, is proposed. The communication system developed by the authors required the use of a double-talk detection algorithm with low complexity and good accuracy. The authors propose an approach to double-talk detection based on the signal envelopes. For each of three signals: the far-end speech, the microphone signal and the echo estimate, an envelope is detected. Next, using these envelopes, a detection function is determined and compared to the threshold. Additionally, a dynamic threshold is introduced in order to improve the accuracy of the algorithm. The results of the simulations presented in the paper proved that the accuracy of double-talk detection obtained using the proposed algorithm is higher than in the Geigel algorithm and comparable to the correlation-based methods, while the computational complexity of the proposed method remains at an acceptable level. The double-talk detection algorithm presented here may be used in voice communication systems having limited resources, allowing for accurate double-talk detection and, as a consequence, efficient acoustic echo cancellation.

#### KEYWORDS:

Double Talk Detection; Acoustic Echo Cancellation; Speech Communication; Speech Signal Processing

10 pages + 1 page with references

5 figures

1 table

No list of symbols necessary

1  
2  
3 INTRODUCTION  
4

5 The popularity of Internet-based voice communication systems has increased significantly during  
6 the recent years. Since many users of VoIP systems prefer to use a loudspeaker instead of  
7 headphones during the communication, developers of VoIP applications have to solve the problem  
8 of an acoustic echo, occurring when a speech signal flows from the loudspeaker to the microphone  
9 and back to the sender [1]. This type of the echo is not easy to remove due to a constantly changing  
10 properties of an acoustic feedback path and the reverberations added to the signal. The task of  
11 acoustic echo cancellation is usually performed by means of an adaptive filter that simulates the  
12 acoustic feedback path and subtracts an echo estimate from the microphone signal [2,3]. The main  
13 problem of this approach is that the adaptation of filter coefficients must be stopped when the ‘near  
14 speech’ is present, otherwise the filter diverges and the signal quality is deteriorated significantly.  
15 Therefore, a double-talk detector (DTD) is needed in order to check whether a double-talk is present  
16 or not and to switch the filter adaptation off and on accordingly [2,4].  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 The authors of this paper were involved in the development of a VoIP system with the focus on  
34 speech quality improvement. Achieving this goal required using complex and computationally  
35 expensive speech coding algorithms [5,6]. In order to maintain transmission delays at a reasonable  
36 level, a low-complexity acoustic echo canceller (AEC) was needed, providing good accuracy and  
37 efficiency at the same time. The complexity and accuracy of the AEC depends greatly on the  
38 implemented DTD algorithm. Since none of the currently available DTDs met the requirements of  
39 the system, the authors decided to propose their own algorithm, details of which are presented in the  
40 The AEC based on the adaptive filter proved to be the most efficient method of acoustic echo  
41 cancellation in voice communication systems [2]. The diagram of the AEC used in the system  
42 developed by the authors is presented in Fig. 1. The far-end speech signal  $x(n)$  is distorted and  
43 attenuated by the acoustic feedback path having an unknown impulse response. Next, this signal is  
44 mixed with the near-end speaker signal and the background noise, forming the microphone signal  
45  $d(n)$ . The adaptive filter processes the far-end signal in order to produce an estimate  $y(n)$  of the  
46  
47  
48  
49  
50  
51

1  
2  
3 acoustic echo. After subtracting the echo estimate from  $d(n)$ , an error signal  $e(n)$  is created. The  
4  
5 coefficients of the adaptive filter are modified using the values of  $e(n)$  in order to obtain a more  
6  
7 accurate estimate of the acoustic feedback path [3]. A nonlinear processor (NLP) suppresses a  
8  
9 residual echo. The task of the DTD is to stop the adaptation process when the near-end speech is  
10  
11 present and restart it when there is no double-talk. A normalized least-mean square (NLS) algorithm  
12  
13 was selected for the filter adaptation, because it provides a good trade-off between the complexity  
14  
15 and the accuracy of the algorithm [3]. It was assumed that the AEC described above will fulfill the  
16  
17 low complexity and good accuracy requirements, provided that the DTD will also fulfill the same  
18  
19 requirements. Therefore, a proper choice of the DTD algorithm is crucial for achieving the desired  
20  
21 goal.  
22  
23  
24

25 Fig. 1  
26

## 27 DOUBLE-TALK DETECTOR

28  
29 Two main approaches are typically used for the task of double-talk detection in AEC systems [7].  
30  
31 The first approach is based on a comparison of energy of the far-end speech and the microphone  
32  
33 signal, while the second approach examines the correlation between the far-end speech and the  
34  
35 microphone signal. The most popular representative of energy-based DTDs is the Geigel algorithm  
36  
37 [8]. It is based on an observation that the energy of echo is typically much smaller than the energy  
38  
39 of far-end speech (due to attenuation of the signal in the acoustic feedback path). Therefore, if the  
40  
41 near-end speech is present, the energy of microphone signal increases. In the Geigel algorithm, a  
42  
43 double-talk detection function is calculated as:  
44  
45  
46  
47

$$48 \xi(n) = \frac{|d(n)|}{\max\{|x(n-1)|, \dots, |x(n-L)|\}} \quad (1)$$

49  
50 where  $x(n)$  is the far-end speech,  $d(n)$  is the microphone signal,  $L$  is a constant that determines a  
51  
number of past samples of the far-end speech signal that are used for the double-talk detection. A  
final decision is made by comparing  $\xi(n)$  with a constant threshold value  $T$ . If  $\xi(n) > T$ , the double-  
talk is declared. The Geigel DTD is a low complexity algorithm, but its accuracy in case of the

1  
2  
3 acoustic echo and unstable acoustic feedback path, resulting in introduction of reverberations to the  
4  
5 signal, was proved to be low [9]. The main problem of this approach is that it is not possible in the  
6  
7 general case to set the constant threshold  $T$  that results in a low number of both false-positive and  
8  
9 false-negative decisions. Moreover, the detection function  $\xi(n)$  changes rapidly, which makes the  
10  
11 detection task even more problematic.  
12  
13

14  
15 Some alternative approaches to the double-talk detection were proposed in the literature  
16  
17 [10,11,12,13]. For example, the algorithm based on a cross-correlation utilizes the observation that  
18  
19 the echo signal is highly correlated with the far-end speech while the correlation between the near-  
20  
21 speech and the far-speech is low. Therefore, the double-talk detection is possible by means of  
22  
23 calculation of a cross-correlation vector between the microphone signal and the far-end speech [12].  
24  
25 This approach fulfills one of the requirements (high accuracy), but it's computational cost is too  
26  
27 high to implement this method in the system developed by the authors of this paper. Therefore, a  
28  
29 novel algorithm for double-talk detection was proposed by the authors.  
30  
31  
32  
33  
34

### 35 DTD ALGORITHM BASED ON SIGNAL ENVELOPES

36  
37 The algorithm for the double-talk detection developed by the authors, shown in Fig. 2, is based on a  
38  
39 comparison of the microphone signal energy with the energy of far-end speech, similarly to the  
40  
41 Geigel algorithm. However, the Geigel algorithm is based on a comparison of the absolute sample  
42  
43 values which results in a detection function that changes rapidly. More accurate results may be  
44  
45 obtained by using the signal energy instead of the absolute sample values. The approach proposed  
46  
47 by the authors of this paper is based on a calculation of the signal envelope as a measure of signal  
48  
49 energy. This method is less computationally complex than the direct energy calculation. Various  
50  
51 methods of the envelope detection may be used, one example is a low-pass filtering of the signal. In  
the proposed DTD algorithm, the envelope  $v_x(n)$  is calculated from the absolute values of  $x(n)$   
samples, using the formula:

$$v_x(n) = \alpha v_x(n-1) + (1 - \alpha) |x(n)| \quad (2)$$

1  
2  
3 where  $\alpha$  is a ‘forgetting factor’ that defines how quickly the envelope detector reacts to rapid  
4  
5 changes in the signal amplitude. In order to obtain an accurate estimation of the signal energy, the  
6  
7 value of  $\alpha$  should be slightly less than one. In the experiments,  $\alpha = 0.99$  provided a sufficiently  
8  
9 accurate energy estimation.  
10

11 Fig. 2

12 In order to form the decision function, two envelope detectors are needed:  $v_x(n)$  for the far-end  
13  
14 speech signal  $x(n)$  and  $v_d(n)$  for the microphone signal  $d(n)$ . The detection function is given by the  
15  
16 formula:  
17

$$18 \xi(n) = \frac{v_d(n)}{v_x(n) + \gamma} \quad (3)$$

19  
20 The  $\gamma$  parameter is used in order to limit the values of detection function during parts of the signal  
21  
22 containing only the noise, when values of both envelopes are low. The  $\gamma$  value should be small, in  
23  
24 the experiments the highest detection accuracy was obtained for  $\gamma = 0.05$ .  
25

26  
27 It is assumed that the echo energy is low compared to the far-end speech energy while the near-  
28  
29 speech energy is significantly higher than the echo energy. Therefore, the value of the detection  
30  
31 function will increase significantly when the double-talk occurs and it will decrease when the  
32  
33 double-talk ends. This concept is similar to the one utilized in the Geigel algorithm. However, the  
34  
35 detection function in the proposed method is much more smooth and slowly changing, which allows  
36  
37 for easier detection of the signal parts that contain the double-talk. If the value of the detection  
38  
39 function exceeds the selected threshold value  $T$ , the double-talk is detected.  
40  
41  
42  
43  
44  
45

46  
47 The algorithm described above suffers from the same drawback as the Geigel algorithm: a choice of  
48  
49 the threshold value allowing for an accurate double-talk detection for a wide range of signals having  
50  
51 different energy is problematic. If the threshold is set too low, it will result in many false-positive  
decisions, if the threshold is set too high, it will cause many false-negative results. In order to avoid  
the dependency of the DTD accuracy on the threshold selection, the authors proposed a method of  
dynamic threshold setting, at the cost of introducing one additional envelope detector to the  
algorithm. The concept of the DTD with the dynamic threshold is based on an observation that in

1  
2  
3 the presence of double-talk, the energy of microphone signal is considerably larger than the echo  
4  
5 energy. In order to make use of this observation, the estimated echo signal produced by the  
6  
7 converged adaptive filter in the AEC may be used as an approximation of a real echo. The envelope  
8  
9 of the echo estimate  $y(n)$  is compared to the envelope of the microphone signal  $d(n)$ . If the envelope  
10  
11 of  $d(n)$  is significantly higher than envelope of  $y(n)$ , it indicates the presence of the double-talk.  
12  
13  
14 However, the DTD based directly on this idea proved to be inaccurate in the experiments. A  
15  
16 modified approach that provided much more accurate double-talk detection was based on using the  
17  
18 estimated echo signal envelope  $v_y(n)$  in the calculation of the dynamic detection threshold,  
19  
20 according to the formula:  
21  
22

23  
24  
25 
$$T(n) = \frac{v_y(n)}{v_x(n) + \gamma} + \beta \quad (4)$$
  
26  
27

28 which is almost identical to the formula (3), except that the envelope of the echo estimate replaces  
29  
30 the envelope of microphone signal. The parameter  $\beta$  is used in order to leave some margin for the  
31  
32 detection error and it should be a small positive value.  
33  
34

35 The modification described above removes the need of setting the constant threshold value for the  
36  
37 detection. The dynamic threshold adapts to the changes in echo signal envelope in real time. If the  
38  
39 double-talk is present, the envelope of microphone signal is greater than the envelope of estimated  
40  
41 echo. Therefore, the double-talk may be detected by comparing the current value of detection  
42  
43 function to the current dynamic threshold value. The far-end speech envelope  $v_x(n)$  may be treated  
44  
45 as a normalization term that improves the detection accuracy comparing to the original version of  
46  
47 the algorithm described earlier.  
48  
49

50  
51 It has to be noted that the accuracy of the presented DTD depends on the accuracy of echo  
estimation produced by the adaptive filter, therefore the calculation of dynamic threshold should not  
be performed until the filter convergence is finished. Therefore, the threshold is initially set to a  
constant value  $T_{init}$ . All three envelopes are calculated starting from the first sample. When the filter  
adaptation is finished, the dynamic threshold is calculated and used instead of  $T_{init}$ . Additionally, it

1  
2  
3 is convenient to set lower and upper bounds on the dynamic threshold ( $T_{min}$  and  $T_{max}$ ) that help to  
4  
5 avoid detection errors in parts of the signal containing only noise. To summarize the proposed  
6  
7 algorithm, a double-talk is detected if the initial filter convergence period has passed and:  
8  
9

$$\xi(n) > T, \text{ where } T = \begin{cases} T(n), & T_{min} < T(n) < T_{max} \\ T_{min}, & T(n) < T_{min} \\ T_{max}, & T(n) > T_{max} \end{cases} \quad (5)$$

10  
11  
12  
13  
14  
15 where  $T(n)$  is the dynamic threshold given by (4).  
16  
17  
18  
19

## 20 EVALUATION OF THE DOUBLE-TALK DETECTOR PERFORMANCE

21  
22 The proposed DTD based on signal envelopes, described in the previous section, was implemented  
23  
24 in the AEC based on the NLS algorithm, and its performance was evaluated. For the assessment of  
25  
26 the DTD accuracy, a method proposed by Cho et al [9] was used. However, Cho et al tested various  
27  
28 DTD algorithms using a set of studio recordings with artificially added reverberation, delay and  
29  
30 noise. The authors of this paper used speech signals recorded in an actual Internet-based speech  
31  
32 communication system, in order to measure the DTD performance in conditions similar to those  
33  
34 existing in real communication systems. The delay between the echo and the far-end speech in the  
35  
36 recordings was about 3.375 ms and the delay on the far-end of the system was about 273.25 ms,  
37  
38 which resulted in an audible echo effect. For the double-talk detection, the near-end speech was  
39  
40 artificially added to the recorded echo signal.  
41  
42  
43  
44

45 Fig. 3 shows an example of double-talk detection results obtained using the DTD presented in this  
46  
47 paper. It can be seen that identification of the double-talk section in the signal was correct. There is  
48  
49 a noticeable delay between the beginning of the actual double-talk section and the beginning of the  
50  
51 detected double-talk. However, this delay is not long enough to cause a divergence of the adaptive  
filter. The dynamic threshold function, shown in Fig. 3, allows for an accurate detection of both the  
beginning and the end of the double-talk. The experiment was repeated for other test signals and  
similar results were obtained.

Fig. 3

1  
2  
3 A quantitative assessment of the envelope-based DTD was performed using the method proposed  
4  
5 by Cho et al. [9], with some modifications. Two main parameters that describe the accuracy of DTD  
6  
7 are a probability of miss  $P_m$  (a false negative decision) and a probability of false alarm  $P_f$  (a false  
8  
9 positive decision). Of these two parameters, a low value of  $P_m$  is much more crucial, because a  
10  
11 missed detection of the double-talk causes a divergence of the adaptive filter and a deterioration of  
12  
13 the signal quality. In order to measure the DTD performance, averaged values of  $P_m$  and  $P_f$  were  
14  
15 calculated for different ratios of the near-end speech energy to the far-end speech energy (near-to-  
16  
17 far ratio, NFR) and for different test signals. Additionally, the same procedure was applied to the  
18  
19 standard Geigel DTD and to the DTD based on the normalized cross-correlation (NCC) [9,12] in  
20  
21 order to do a comparative analysis of all three DTD algorithms. For each DTD algorithm, ‘optimal’  
22  
23 values of the parameters were found during the experiments so that the best trade-off between  $P_m$   
24  
25 and  $P_f$  is obtained. For the envelope-based DTD, value  $\beta = 0.02$  was selected. The results of the  
26  
27 assessment of all DTDs are plotted in Figures 4 and 5, showing probability of miss and probability  
28  
29 of false alarm, respectively, as a function of NFR.

30  
31  
32  
33  
34  
35 In order to obtain more detailed information on the performance of the envelope-based algorithm,  
36  
37 three parameters were calculated for each tested DTD, as well as for the AEC without a DTD and  
38  
39 for a signal not processed by the AEC. The test set contained four studio recordings of Polish  
40  
41 sentences, two by a male speaker and another two by a female speaker, having length of 80 to 120  
42  
43 seconds. The averaged results of this experiment, shown in Table 1, allow for comparison of the  
44  
45 proposed algorithm performance with other DTDs. A signal to noise ratio (SNR) was calculated as a  
46  
47 ratio of the power of the near-end speech signal section to the power of the signal part containing  
48  
49 only the echo. An echo return loss enhancement (ERLE) was calculated as a ratio of the power of  
50  
51 the unprocessed signal (containing the echo) to the power of the signal processed by the AEC (after  
the echo cancellation). A PESQ (Perceptual Evaluation of Speech Quality) score is a measure of  
quality of the double-talk part of the signal, ranging from 4.5 (the highest possible quality) to 0 (the  
worst quality). Signal distortions caused by an incorrect double-talk detection decrease the PESQ

Fig. 4



1  
2  
3 value. This measure was computed using the Opticom OPERA software, a clean near-end speech  
4  
5 was used as a reference signal.  
6  
7  
8  
9

## 10 DISCUSSION

11  
12 The accuracy of the DTD is measured in terms of the probability of miss and the probability of false  
13  
14 detection. Both values should be as low as possible. The results of the simulations show that the  
15  
16 proposed algorithm yields a significant improvement in the double-talk detection accuracy over the  
17  
18 Geigel DTD. Introduction of the dynamic threshold that adapts to the signal removes the most  
19  
20 important problem of the Geigel algorithm – the dependency of the DTD accuracy on the threshold  
21  
22 setting. Using slowly changing envelopes instead of actual sample values in the calculation of  
23  
24 detection functions simplifies the task of automatic selection of the dynamic threshold. Two  
25  
26 parameters,  $\beta$  and  $\gamma$ , have to be set in the envelope-based DTD. They can be seen as tune-up settings  
27  
28 that allow for decreasing  $P_f$  at the cost of some increase in  $P_m$ .  
29  
30  
31

32  
33 From Figs 4 and 5 it can be clearly seen that the accuracy of the envelope-based DTD for all NFR  
34  
35 values is higher than in the Geigel algorithm and comparable to the NCC method. Although  $P_f$   
36  
37 increases with the NFR in the envelope DTD, it still remains at a sufficiently low level. Moreover,  
38  
39 higher  $P_f$  values do not have as significant impact on DTD performance as higher  $P_m$  values. The  
40  
41 values of the parameters shown in Table 1 indicate that the quality of the signal processed with the  
42  
43 AEC and the envelope DTD is comparable to the quality obtained using the NCC method and  
44  
45 higher than in both the Geigel algorithm and the unprocessed signal.  
46  
47

48  
49 Regarding the computational complexity, the envelope-based DTD with the dynamic threshold  
50  
51 requires a calculation of three envelopes, three absolute values, the detection function and the  
dynamic threshold, as well as a comparison of the detection function with the threshold. Each  
envelope detector requires two multiplications and two addition/subtraction operations. In total, the  
proposed DTD requires eight multiply/divide operations and nine add/subtract operations.

Therefore, the computational complexity of the envelope-based DTD is higher than in the Geigel

1  
2  
3 DTD, but it is still acceptable. In contrast to that, the original NCC method requires the computation  
4  
5 of two correlation vectors and an autocorrelation matrix that has to be inverted [9], therefore this  
6  
7 method is impractical due to a high computational cost. A simplified method [11] was used in the  
8  
9 experiments described in this paper, however, this ‘fast’ method still requires  $(5L + 4)$   
10  
11 multiplications, where  $L$  is the length of the analysis window ( $L = 1024$  in the simulations).  
12  
13 Therefore, the computational cost of the proposed algorithm is considerably lower than in the NCC  
14  
15 method.  
16  
17

18  
19 From Fig. 3 it can be seen that there is a delay between the end of the actual double-talk section and  
20  
21 the end of the double-talk detected by the envelope-based DTD. This effect is caused by the  
22  
23 envelope detector that is not able to react to rapid decreases in signal amplitude without a delay.  
24  
25 However, this effect does not necessarily decrease the accuracy of echo cancellation, because in  
26  
27 most AEC systems, a period of latency is introduced so that the filter adaptation is restarted only if  
28  
29 the double-talk was not found during this period. In the proposed algorithm, there is no need to  
30  
31 introduce this additional latency, because the delay caused by the long decay time of the envelope  
32  
33 detector takes care of this problem. Therefore, the slow reaction time for a rapid decrease of the  
34  
35 signal amplitude in the proposed DTD does not reduce the accuracy of double-talk detection in a  
36  
37 significant way.  
38  
39

40  
41 One more important thing to note is that the accuracy of double-talk detection in the proposed  
42  
43 algorithm depends on the accuracy of echo estimation in the adaptive filter. If the echo estimate is  
44  
45 inaccurate, the dynamic threshold value will be inappropriate, which may result in wrong DTD  
46  
47 decision. In return, a wrong DTD decision may cause an unneeded switching of the adaptation filter  
48  
49 state, which results in wrong echo estimation. Consequently, this feedback loop may cause the  
50  
51 divergence of the filter and the deterioration of the signal quality. However, this effect was not  
observed during the experiments. The reaction time of the DTD on the double-talk occurrence is  
short enough and the risk of filter divergence due to a wrong DTD decision is low as long as the  $\beta$   
parameter is set properly.

1  
2  
3 The results of the comparative analysis presented in Table 1 show that the accuracy of echo  
4  
5 cancellation in the AEC utilizing the proposed DTD algorithm is comparable to the cross-  
6  
7 correlation algorithm and significantly higher than in the Geigel algorithm. In terms of quality of  
8  
9 the processed signal, the results of the PESQ test prove that the proposed algorithm provides an  
10  
11 acceptable trade-off between the low complexity of the Geigel algorithm and the high accuracy of  
12  
13 the cross-correlation method. Hence, the aim of the research was successfully achieved.  
14  
15  
16  
17  
18

## 19 CONCLUSIONS

20  
21 The goal of the work was to develop a double-talk detector suitable for use in the AEC in speech  
22  
23 communication systems, that provides both low computational complexity and good accuracy of the  
24  
25 double-talk detection. The approach presented in this paper is based on the detection of signal  
26  
27 envelopes of the far-end speech, microphone and echo estimate signals. It can be seen as an  
28  
29 expansion of the original Geigel algorithm. However, using a signal envelope instead of absolute  
30  
31 values of signal samples, together with the introduction of the dynamic threshold, resulted in an  
32  
33 improved accuracy of double-talk detection.  
34  
35  
36

37 The simulations of the acoustic echo cancellation, performed using the test signals recorded in the  
38  
39 real speech communication system, proved that the proposed DTD algorithm provides good  
40  
41 accuracy of double-talk detection, comparable to the correlation-based method and higher than in  
42  
43 the Geigel algorithm. It is also possible to tune up the performance of the algorithm by changing the  
44  
45 value of  $\beta$  parameter. The computational complexity of the proposed approach, although higher than  
46  
47 in the Geigel DTD, is still low, allowing for implementation of this algorithm in the developed  
48  
49 communication system. The complexity of the presented method is considerably lower than in the  
50  
51 NCC algorithm, while the accuracy of both methods and the processed signal quality are similar. As  
a conclusion, it may be stated that the DTD algorithm presented here fulfills the requirements of the  
voice communication system that is being developed by the authors. Further experiments will focus

1  
2  
3 on the implementation and testing of the proposed DTD algorithm in the physical system and also  
4  
5 on further modifications of the algorithm, improving its performance.  
6  
7  
8  
9

10 **ACKNOWLEDGEMENTS**

11  
12 Research funded by the Polish Ministry of Education and Science within the Grant No. 3 T11D 004  
13

14  
15 28  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

1  
2  
3 REFERENCES  
4

5 [1] International Telecommunication Union, Acoustic echo controllers, ITU-T Recommendation  
6 G.167, 1993.  
7

8  
9  
10 [2] S.M. Kuo, B.H. Lee, W. Tian, Real-Time Digital Signal Processing: Implementations and  
11 Applications, John Willey & Sons, 2006, Chapter 10: Adaptive Echo Cancellation.  
12

13  
14 [3] S. Haykin, Adaptive filter theory, 4th edition, Prentice Hall Inc, 2002.  
15

16  
17 [4] S.V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, John Willey & Sons,  
18 2000, Chapter 14: Echo cancellation.  
19

20  
21 [5] M. Kulesza, A. Czyżewski, Speech Codec Enhancements Utilizing Time Compression and  
22 Perceptual Coding, Proc. 112th Audio Eng. Conv, Vienna, 5-8 May 2007, preprint No. P-3.  
23

24  
25 [6] G. Szwoch, M. Kulesza, A. Czyżewski, Transient detection for speech coding applications, Int.  
26 Journ. of Computer Science and Network Security 6 (12) (2006) 320-325.  
27

28  
29 [7] T. Vu, H. Ding, M. Bouchard, A Survey of Double-talk Detection Schemes for Echo  
30 Cancellation Applications, Canadian Acoustics 32 (3) (2004) 144-145.  
31

32  
33 [8] A. Adrian, Voice over Internet Acoustic Echo Cancellation, DFS Deutsche Flugsicherung, 2004,  
34 Internet: [http://home.arcor.de/andreadrian/echo\\_cancel/](http://home.arcor.de/andreadrian/echo_cancel/).  
35

36  
37 [9] J. Cho, D. Morgan, An Objective Technique for Evaluating Doubletalk Detectors in Acoustic  
38 Echo Cancelers, IEEE Trans. Speech Audio Processing 7 (6) (1999) 718-724.  
39

40  
41 [10] T. Gaensler, M. Hansson et al., A double-talk detector based on coherence, IEEE Trans. on  
42 Communications, 44 (11) (1996) 1421-1427.  
43

44  
45 [11] T. Gaensler, J. Benesty, The fast normalized cross-correlation double-talk detector, Signal  
46 Processing, 86 (6) (2006) 1124-1139.  
47

48  
49 [12] J. Benesty, D.R. Morgan, J.H. Cho, A New Class of Doubletalk Detectors Based on Cross-  
50 correlation, IEEE Trans. Speech Audio Processing 8 (3) (2000) 168-172.  
51

[13] H. Ye, B. Vu, A new double-talk detection algorithm based on the orthogonality theorem, IEEE  
Trans. on Communications 39 (11) (1991) 1542-1545.

1  
2  
3 TABLE CAPTIONS  
4  
5  
6

7 Table 1. Comparison of the averaged quality measures of the signal processed by the AEC based on  
8 the NLS adaptive filter, with various DTD algorithms  
9

10  
11 FIGURE CAPTIONS  
12

13 Fig. 1. Block diagram of the acoustic echo canceller with the double-talk detector (DTD)  
14

15 Fig. 2. Block diagram of the acoustic echo canceller with the proposed double-talk detector based  
16 on signal envelopes  
17

18  
19 Fig. 3. The example of the double-talk detection obtained in the simulations using the proposed  
20 algorithm The dashed vertical lines mark the beginning and the end of the inserted double-talk  
21 signal. The light gray background marks the signal part detected as a double-talk by the proposed  
22 algorithm. The thin line is the detection function, the thick line is the dynamic threshold function  
23  
24

25 Fig. 4. The measurement results of the probability of miss ( $P_m$ ) as a function of NFR, for the  
26 proposed (EnvDTD), Geigel and normalized cross-correlation (NCC) algorithms  
27

28 Fig. 5. The measurement results of the probability of false alarm ( $P_f$ ) as a function of NFR, for the  
29 proposed (EnvDTD), Geigel and normalized cross-correlation (NCC) algorithms  
30  
31  
32  
33  
34  
35  
36  
37  
38

Table 1. Comparison of the averaged quality measures of the signal processed by the AEC based on the NLS adaptive filter, with various DTD algorithms

DTD algorithm	SNR [dB]	ERLE [dB]	PESQ score
Proposed algorithm (envelope)	68.12	32.26	3.71
Normalized cross-correlation (NCC)	69.98	34.02	3.86
Geigel	44.59	8.38	3.63
No double-talk detection	36.63	7.06	2.91
No echo cancellation	36.36	0	3.29

Figure 1

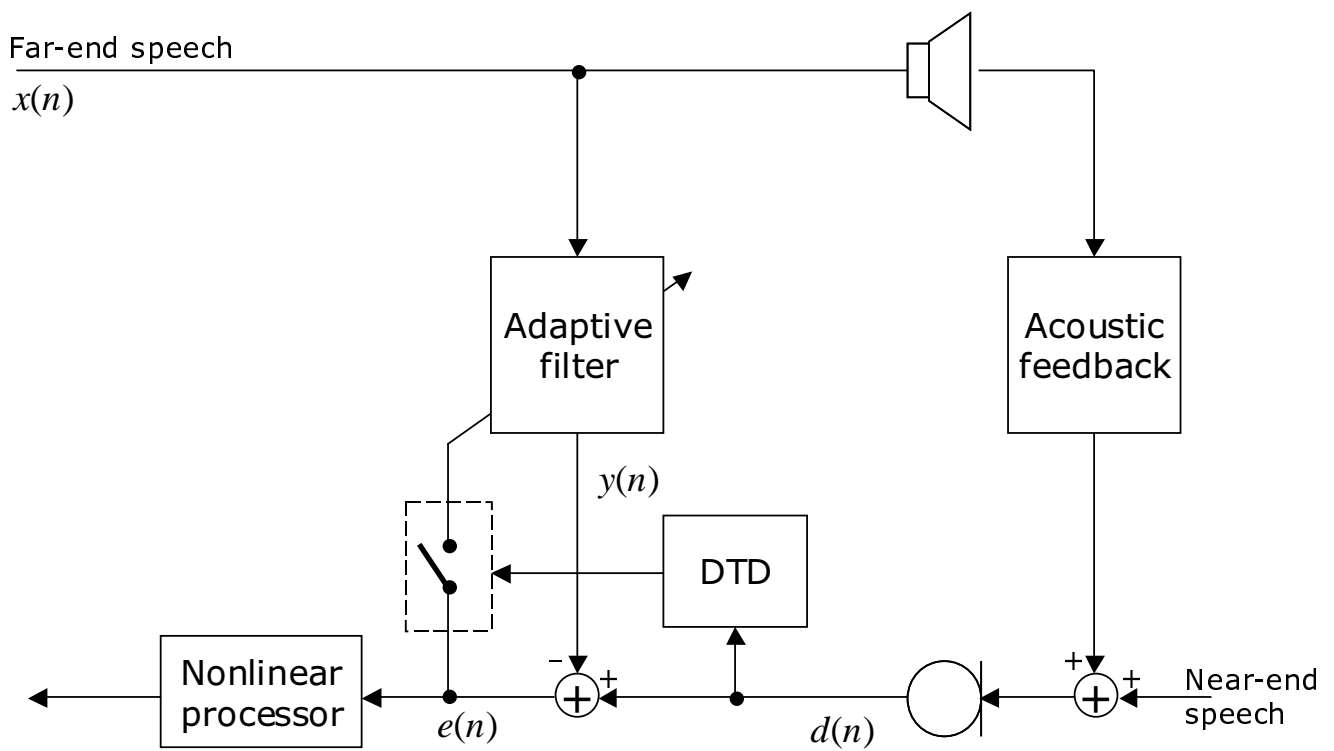




Figure 2

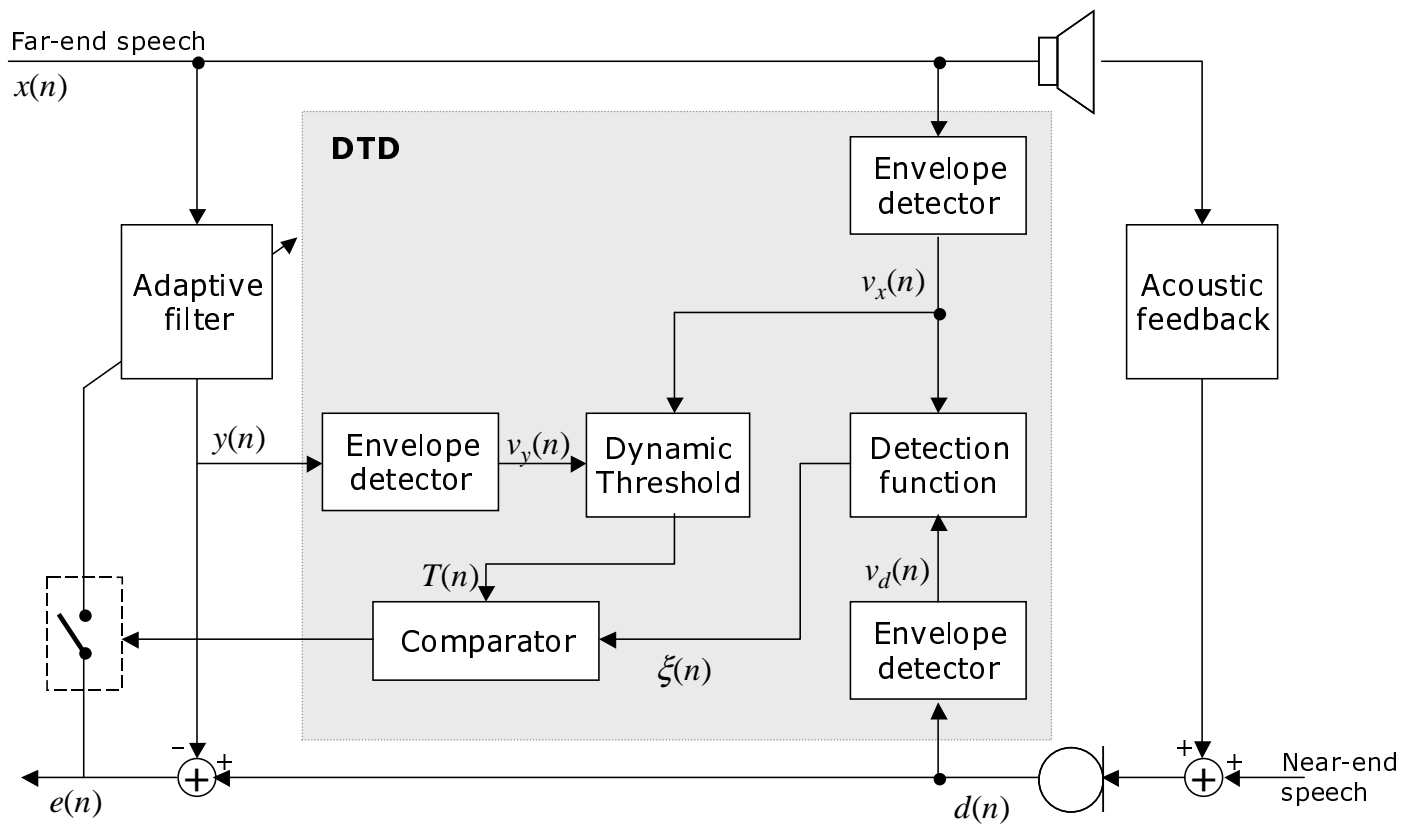


Figure 3

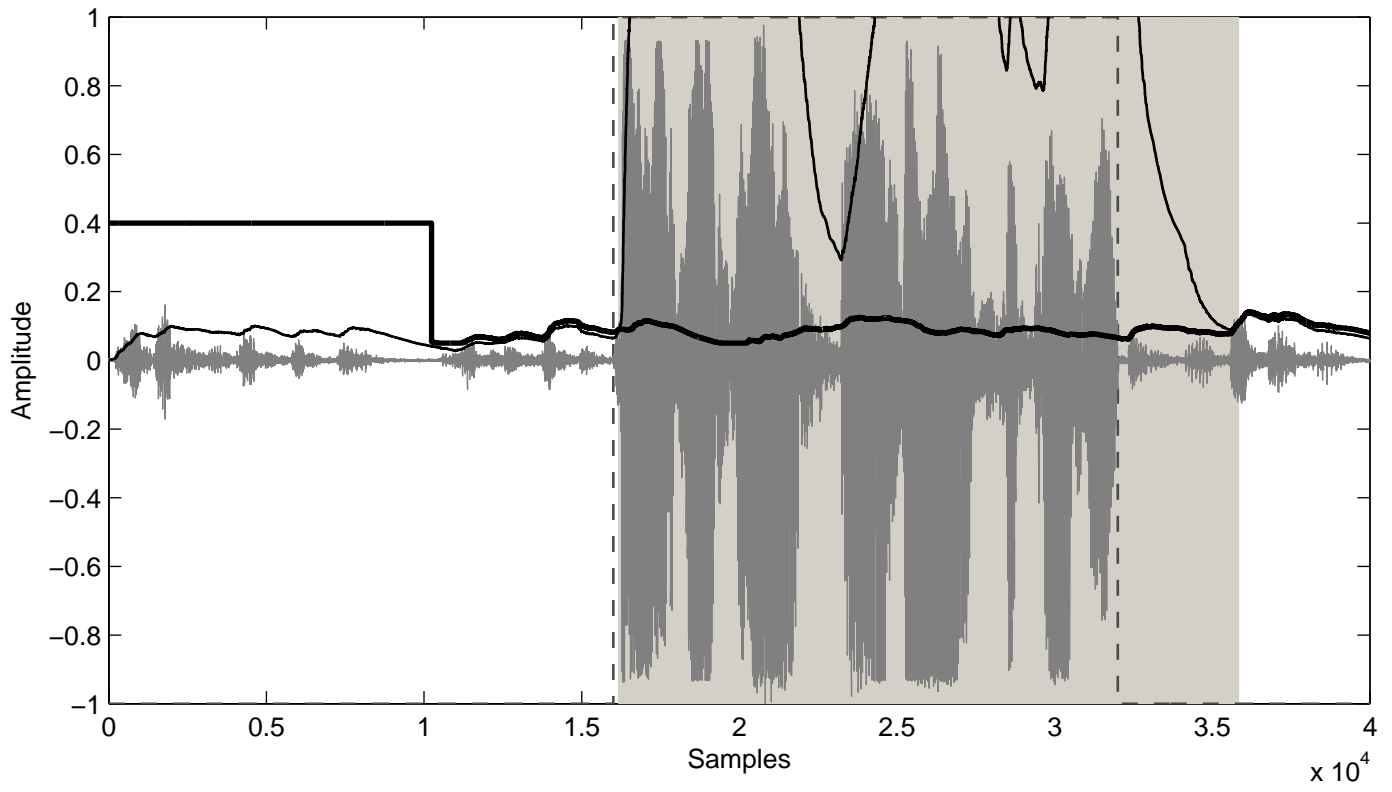


Figure 4

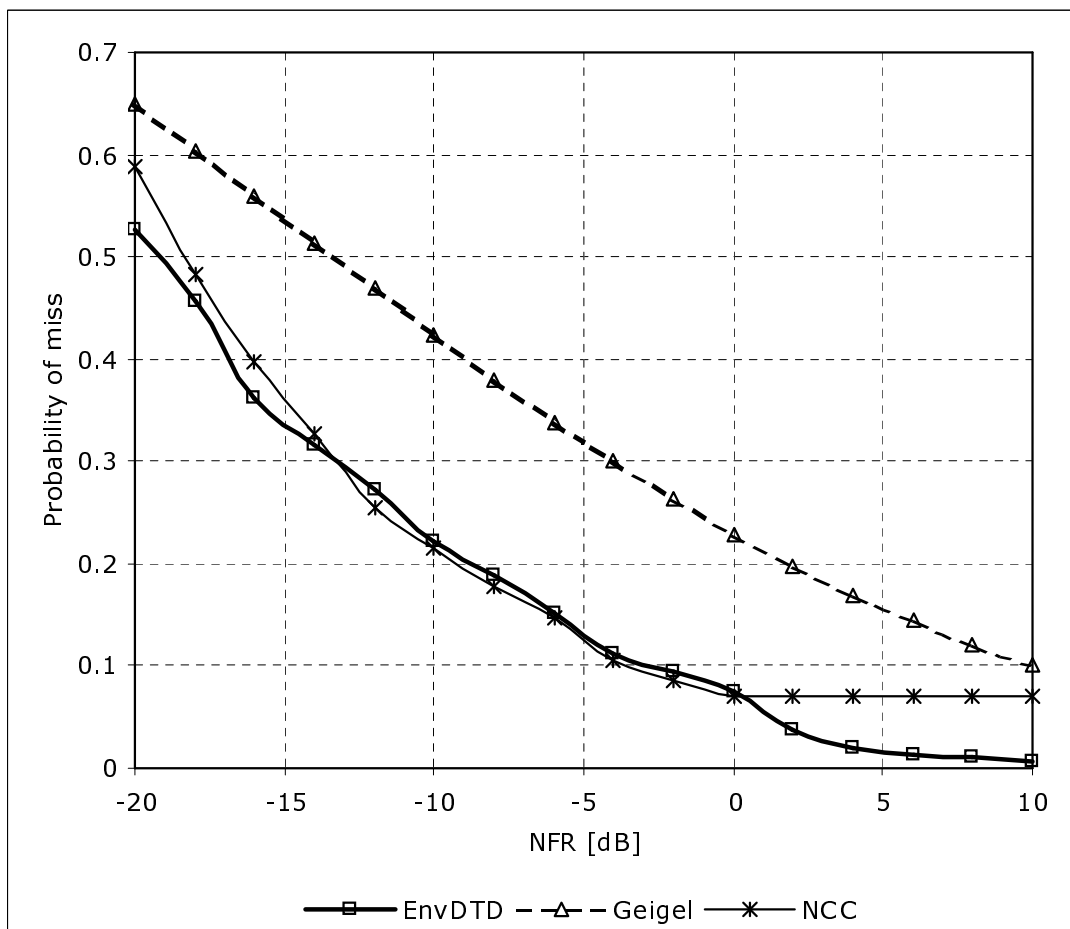


Figure 5

