# A STUDY OF DOUBLETALK DETECTION PERFORMANCE IN THE PRESENCE OF ACOUSTIC ECHO PATH CHANGES

*Per Åhgren*

Dept. of Systems and Control
Uppsala University, Sweden

*Andreas Jakobsson*

Dept. of Electrical Engineering
Karlstad University, Sweden

**ABSTRACT**

A well-performing double-talk detection (DTD) algorithm is a vital part of an acoustic echo canceller. However, recent algorithms are typically evaluated using a static time-invariant room acoustic impulse response, omitting a proper treatment of the case when the acoustic echo path is changing. In this work, we introduce a common framework to objectively evaluate how path changes affect the DTD performance. Via extensive numerical simulations, we conclude that the main factor in acoustic path changes affecting the DTD performance for some of the more common DTD algorithms is variations in the damping of the echo path.

## 1. INTRODUCTION

The problem of acoustic echo cancellation (AEC) was introduced in [1] and has since been a very active area of research. Briefly stated, an acoustic echo canceller is needed for removing the acoustic echoes resulting from the acoustic coupling between the loudspeaker(s) and the microphone(s) in communication systems. The near-end acoustic signal, $v(t)$, is measured in the possible presence of an echo signal resulting from the far-end signal, $x(t)$, emitted in the near-end room by the loudspeaker. When both such an echo and the near-end signal are present, the so-called *double-talk* (DT) case, the resulting microphone signal, $y(t)$, consists of the near-end signal mixed with the far-end signal filtered by a (typically time-varying) room acoustic filter (which is the impulse response from the loudspeaker to the microphone), $\mathbf{h}_t$. Often, $\mathbf{h}_t$ is modeled as an $n$-tap finite impulse response (FIR) filter, $\mathbf{h}_t = \begin{bmatrix} h_t(0) & h_t(1) & \cdots & h_t(n-1) \end{bmatrix}^T$, yielding the microphone signal

$$y(t) = \mathbf{h}_t^T \mathbf{x}(t) + v(t) + w(t), \qquad (1)$$

where $w(t)$ is additive noise, and

$$\mathbf{x}(t) = \begin{bmatrix} x(t) & x(t-1) & \cdots & x(t-n+1) \end{bmatrix}^T. \qquad (2)$$

THIS WORK WAS PARTLY SUPPORTED BY THE SWEDISH SCIENCE COUNCIL (VR).

To enable the reduction of the undesired echo signal, an adaptive filter, $\hat{\mathbf{h}}_t$, defined similarly to $\mathbf{h}_t$, is used to predict the echo term, $\mathbf{h}_t^T \mathbf{x}(t)$, and subtract this value from the microphone signal, yielding the residual signal $e(t) = y(t) - \hat{\mathbf{h}}_t^T \mathbf{x}(t)$. For simplicity, we will here assume that both $\hat{\mathbf{h}}_t$ and $\mathbf{h}_t$ are of length $n$. When no DT is occurring, the adaptive filter can track the time variations of the room acoustic filter; however, during DT, it is vital to prevent the adaptation of the filter. The residual signal, $e(t)$, will in this case contain the near-end signal which will severely disturb the adaptation, possibly causing the adaptive filter to diverge. As a consequence, the recent literature is abundant with different double-talk detection (DTD) algorithms of varying efficiency [2–8]. The majority of these algorithms are implicitly based on the assumption that the room acoustic filter does not vary over time, with only a minority including simulation studies for the case when the acoustic path is changing [5, 6]. Note that it is very important to differentiate between DT and changes in the acoustic echo path; in the former, adaptation of the adaptive filter should be prevented, while in the latter, the adaptation speed should be increased. It is not clear from most of the above cited references how much changes in the acoustic paths affect the performance of the different DTD algorithms. Furthermore, none of the mentioned papers include information on *how* such time-variations have been simulated (which turns out to be very important). To the best of our knowledge, no detailed study has been made on the performance degradation for different DTD algorithms when the acoustic path is changing, nor has it been investigated what kind of path changes that would significantly affect the DTD algorithms' performance. The purpose of this paper is therefore to formalize an objective approach to examine the performance degradation of DTD algorithms under echo path changes, as well as evaluating which form of changes that affect the performance of the algorithms the most. We note that it is in general difficult to evaluate how DTD algorithms are affected by changes in the acoustic paths; to make the problem feasible, one has to limit the evaluation to specific algorithms, as well as make a series of simplifying assumptions. Here, we will limit our attention to comprise four of the more well-known algorithms, namely the Geigel algorithm [8], the normalized cross correlation (NCR) algorithm [2], the cross correlation (CR) algorithm [6], the short-term normalized correlation-based DTD algorithm (STNC) [4], as well as the recent VIRE algorithm presented in [5] (a brief summary of these methods can also be found in [5]).

## 2. EVALUATION OF ECHO PATH CHANGES

Changes in the acoustic echo path can occur due to a variety of reasons; the loudspeakers and microphones may be replaced, moved or obscured, and the configuration of the room may vary, e.g., by the motion of people and objects within the near-end room. The nature of these changes differ; some are sudden and occur instantly, others are slowly varying over a period of time. Furthermore, an echo path change may well result in variations of the acoustic damping (e.g., by objects moving into the direct path between the loudspeaker and the microphone), defined as the norm of the impulse response. Here, we consider a combination of acoustic path changes: (i) changes that occur instantly, (ii) changes that takes place continuously over a relatively short period of time, (iii) changes where the damping in the acoustic path is larger before the change than after the change, and (iv) changes where the damping in the acoustic path is larger after the change than before the change. To enable an objective performance evaluation of the examined DTD algorithms, we propose an evaluation scheme reminiscent of the DTD scheme in [3]. The proposed scheme evaluate the effects on the algorithms' probability of false detection, $P_f$, for a given probability of missed detection, $P_m$, in the presence of an acoustic echo path change. Here, $P_m = N_{Dm}/N_D$ and $P_f = N_{Df}/N_{ND}$, where $N_{Dm}$ is the number of samples where DT was not detected but was present, $N_D$ is the total number of samples where DT was present, $N_{Df}$ is the number of samples where DT was detected but where no DT was present, and $N_{ND}$ is the total number of samples where DT was not present. Using $P_m = 0.1$, the evaluation scheme is: (i) Generate seven seconds of data according to (1), with DT present from 5.5 to 6.5 seconds. (ii) Apply the detector to the last 5 seconds of data to allow the adaptive filter time to converge, and choose a detection threshold for each evaluated DTD algorithm such that $P_m = 0.1$. (iii) Create 5 new data sets, with each having different acoustic paths and where there is a change of the acoustic path at time 3.5 seconds. In each data set, the same DT that was present in step (ii) should be present from 5.5 to 6.5 seconds. (iv) Apply the detector to all the 5 data sets and compute the average $P_f$. Each impulse response used before the change in a certain data set should also be used after the change in another data set to ensure that the observed effects of the echo path changes are not caused by the particular impulse responses used. The suggested framework does not consider cases when echo path changes occur when the adaptive filter is still adapting. This is a case most DTD algorithms will have problems with, especially those that rely on a converged estimate of the echo path such as Cheap-NCR. Furthermore, we do not consider changes occurring over a very long time period. This case is usually not a problem since with properly set thresholds the DTD algorithms will generally not detect these kind of changes as doubletalk. Finally, we are not considering cases when echo path changes occur when there is doubletalk, since there is nothing sensible for the AEC or the DTD algorithms to do in such cases, as it is not possible to use the adaptive filter estimate to cancel the echo (as the filter estimate is incorrect), and the adaptive filter estimate can not be improved since there is doubletalk disturbing the adaptation.

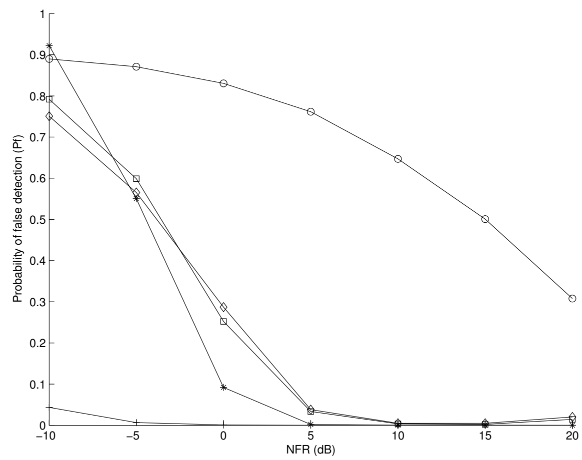


**Fig. 1**. $P_f$ as a function of the NFR for the Geigel (○), VIRE (+), Cheap-NCR (∗), STNC (□) and CR (◇) algorithms for data set (i).

## 3. NUMERICAL SIMULATIONS

In order to evaluate the DTD performance degradation in the presence of acoustic echo path changes, a simulated microphone signal has been constructed according to (1), using an 11 kHz sampling frequency, and a set of real-world impulse responses of lengths $n = 900$, obtained using a teleconferencing setup and normalized to give a certain damping. To update the adaptive filter estimate, the NLMS algorithm, with $\mu = 0.7$ (in order to give reasonably fast convergence while still not making NLMS too sensitive to noise), has been used. Furthermore, both the VIRE and the Cheap-NCR algorithms use a forgetting factor of $\lambda = 0.99$ which gives a good probability of detection, providing a stable performance but still allowing the algorithms to adapt to changes in the echo paths [5]. The following data sets with different echo path changes were created to simulate the time-varying echo path: (i) the impulse responses are exactly the same before and after the change, (ii) the *damping* of the impulse responses are the same before and after the change and the change occurs instantaneously, (iii) the damping of the impulse responses before the change is 10 times smaller than the damping of the impulse responses after the change and the change occurs instantaneously, (iv) the damping of the impulse responses before the change is 10 times larger than the damping of the impulse responses after the change and the change occurs instantaneously, The evaluation scheme was also applied to the data sets (i)-(iv) when the echo path change took place continuously during a time period of 1000 samples. This change was implemented by recomputing the impulse response for each sample during the change by interpolating between the impulse responses before and after the change. Surprisingly, the results from applying the evaluation scheme to the data sets with continuous echo path changes were almost identical to those for the data sets where the change took place instantaneously. We refer the reader to [5] for a further discussion on this aspect. The results [1] obtained by the DTD evalu-

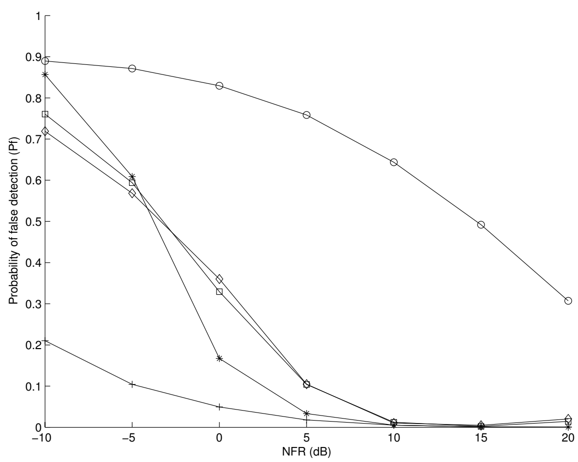


**Fig. 2**. $P_f$ as a function of the NFR for the Geigel ($\circ$), VIRE ($+$), Cheap-NCR ($*$), STNC ($\Box$) and CR ($\diamond$) algorithms for data set (ii).

ation scheme for the first data set are plotted in Fig. 1 in terms of achieved $P_f$ as a function of the Near-end to Far-end speech Ratio (NFR) for a Signal to Noise Ratios (SNR) of 30 dB, where NFR $= 10\log_{10}[Ev^2(t)/E[y(t)-v(t)]^2]$ and SNR $= 10\log_{10}[E[y(t)-w(t)]^2/Ew^2(t)]$. It is clear from the figure that for the first data set, when no changes are occurring, all the DTD algorithms perform well. The results obtained by the DTD evaluation scheme for the second data set are shown in Fig. 2. From the figure, we see that the actual performance degradation is quite small; for most algorithms, the decrease in DTD performance is at most 8 percentage units. The performance decrease for the VIRE algorithm is larger, although it is still outperforming the other algorithms. An interesting feature of the plot is that some of the algorithms even achieve an increase in DTD performance when there is a change in the acoustic path. However, this is only occurring for very low NFR, when it is difficult to set a proper threshold yielding $P_m = 0.1$. This occasional increase can also be seen in the remaining figures. Fig. 3 shows the corresponding figure for the third data set. It is clear that the performance of the detectors is significantly worse when there is a change in the damping of the impulse response. The difference is especially large for the

[1] Further numerical results can be found in [5].

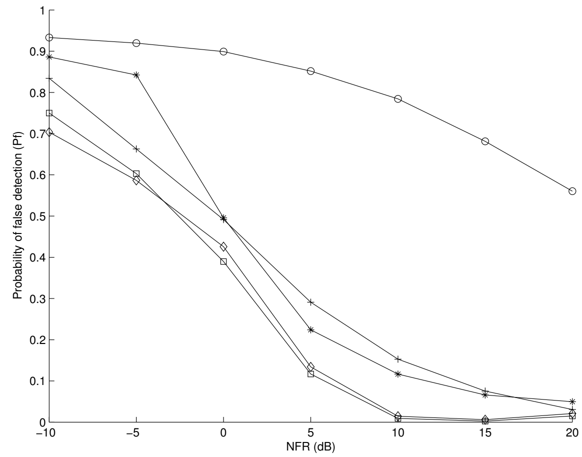


**Fig. 3**. $P_f$ as a function of the NFR for the Geigel ($\circ$), VIRE ($+$), Cheap-NCR ($*$), STNC ($\Box$) and CR ($\diamond$) algorithms for data set (iii).

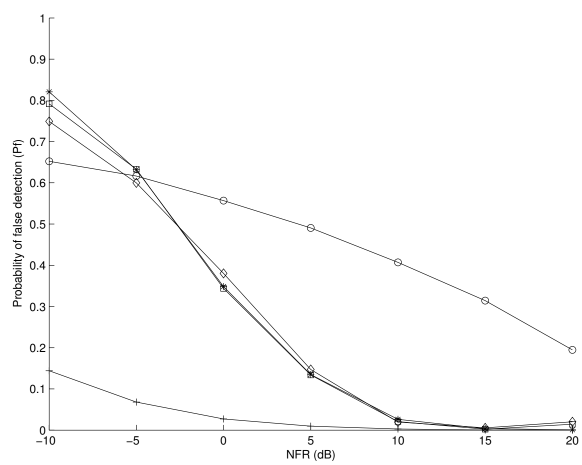


**Fig. 4**. $P_f$ as a function of the NFR for the Geigel ($\circ$), VIRE ($+$), Cheap-NCR ($*$), STNC ($\Box$) and CR ($\diamond$) algorithms for data set (iv).

VIRE algorithm when the NFR is low. Only the STNC and CR algorithms are able to obtain a really low $P_f$ for high NFR. The corresponding results for data set (iv) are shown in Fig. 4. As seen from the figure, for this data set the performance decrease is smaller that for data set (iii). In the data sets where the dampings of the impulse responses differed, the relatively large difference of 10 dB was used. In order to study the effect of the amount of damping on the DTD performance we have performed simulations using a number of data sets with instantaneously changing impulse responses where the amount of damping is varied. The results when the NFR is 10 dB is shown in Figure 5. It is clear that the DTD performance variation of most of the algorithms for different damping ratios is minor, with the Geigel algorithm being the exception. It is also clear that the VIRE and Cheap-NCR algorithms perform worse when the damping after the change is larger than before the change.

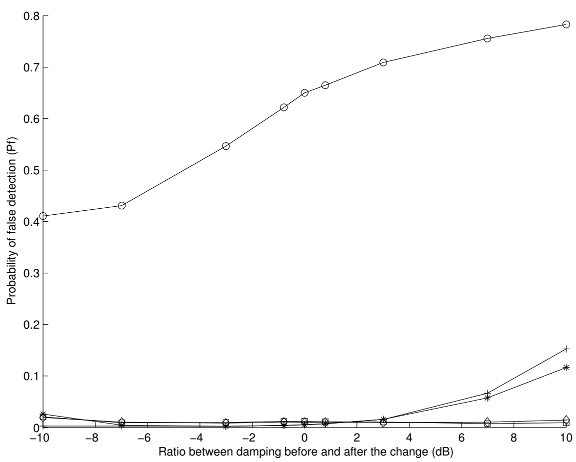


**Fig. 5**. $P_f$ as a function of the the ratio between the dampings before and after the change (a negative value in dB means that the damping for the impulse response after the change is smaller than for the impulse response before the change) Geigel ($\circ$), VIRE ($+$), Cheap-NCR ($*$), STNC ($\Box$) and CR ($\diamond$) algorithms for an NFR of 10 dB.

## 4. ANALYSIS OF THE SIMULATION RESULTS

From the results presented in the previous section (see also [5]), one can conclude that: (i) the duration of the change has almost no impact on the performance of the algorithms under study, and (ii) there is only a minor performance decrease for echo path changes that leave the damping of the impulse responses unchanged after the change. Furthermore, most of the algorithms under study work poorly for low NFR; in fact, the performance of several of the algorithms is so poor that it is not possible to reduce their performance further. Typically, it is hard to find the appropriate threshold for these methods yielding the desired $P_m$ for low NFRs. A general conclusion of the above simulations is that the DTD algorithms primarily experience performance degradation for echo path changes with variations in the damping of the acoustic paths. However, it should be noted that parts of this performance decrease is likely due to the difficulty in making a fair comparison. There are two main reasons for this: as the SNR is computed as an average over the entire data set, it will vary when the damping in the acoustic paths varies. Furthermore, the threshold value is set so that $P_m = 0.1$ in the DT region where the data is always generated using the post-change impulse response. If there is a change in the path damping, and thus in the SNR and the NFR, the selected threshold does not necessarily appropriate for the pre-change data. To take such effects into account would require an adaptive threshold value; few DTD algorithms include such an obviously desirable adaptation.

## 5. REFERENCES


[1] M. M. Sondhi, "An adaptive echo canceler," *Bell System Techn. J.*, vol. XLVI, no. 3, pp. 497–510, 1967.

[2] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 168–172, Mar 2000.

[3] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 718–724, Nov 1999.

[4] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control - An application of very-high-order adaptive filters," *IEEE Signal Processing Mag.*, vol. 16, pp. 42–69, July 1999.

[5] P. Åhgren, *On system identification and acoustic echo cancellation*. PhD thesis, Uppsala University, 2003.

[6] H. Ye and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Commun.*, vol. 39, pp. 1542–1545, Nov 1991.

[7] H. K. Jung, N. S. Kim, and T. Kim, "A new double-talk detector using echo path estimation," in *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. ICASSP-2002.*, vol. 2, pp. 1897–1900, 2002.

[8] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, pp. 647–653, May 1978.