QUALITY EVALUATION OF NOVEL DTD ALGORITHM BASED ON AUDIO WATERMARKING

Andrzej Ciarkowski and Andrzej Czyżewski
Multimedia Systems Department, Gdansk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland

Keywords: Acoustic Echo Cancellation, Doubletalk Detection, Echo Hiding.

Abstract: Echo cancellers typically employ a doubletalk detection (DTD) algorithm in order to keep the adaptive filter from diverging in the presence of near-end speech signal or other disruptive sounds in the microphone signal. A novel doubletalk detection algorithm based on techniques similar to those used for audio signal watermarking was introduced by the authors. The application of the described DTD algorithm within acoustic echo cancellation system is presented. The comparison of the proposed algorithm with very common, but simple Geigel algorithm and representing current state-of-the-art Normalized Cross-Correlation algorithms is performed. Both objective (ROC) and subjective (listening tests) performance evaluation methods are employed to obtain exhaustive evaluation results in simulated real-world conditions. The evaluation results are presented and their relevance is discussed. An issue of algorithms' computational complexity is emphasized and conclusions are drawn.

1 INTRODUCTION

Acoustic echo is one of the most important factors affecting quality and comprehensibility of speech in communications systems. An important reason for increased interest in acoustic echo elimination systems are changing behavioral patterns of telephony users due to the fact that frequently traditional phone handsets are becoming replaced by laptop computers with built-in loudspeaker and microphone acting as a speakerphone terminal. Such a configuration, which is inherently echo-prone due to high coupling between sound source and receiver, is also common in teleconferencing applications and car hands-free adapters making the telecommunications acoustic echo problem even more tangible.

Acoustic echo appears in the conversation when speech signal from far-end speaker, reproduced locally by the loudspeaker is being fed into the receiver (microphone) and returns to the original speaker. High amount of echo mixed with signal from the local (near-end) speaker distorts the communication, making his speech unintelligible and forces the far-end speaker to increase his concentration on understanding the message, which is not only stressful, but can even lead to dangerous accidents in case of car hands-free conversation. To counteract this problem in full-duplex communications setups acoustic echo cancellation (AEC) algorithms are used. Such algorithms typically process the incoming microphone signal in order to remove from it the estimate of echo signal, obtained through the transformation of recently reproduced far-end speaker signal. Most of the AEC algorithms proposed in the literature use adaptive filtering in order to estimate the echo path response. This allows obtaining accurate estimate of echo signal through filter adaptation and effectively eliminate the echo from microphone input by simple signal subtraction, provided that its contents is the sole echo signal (Kuo, Lee and Tian, 2006). Such an assumption however is hardly realistic, as besides the echo signal the microphone signal will typically contain some amount of noise and most importantly, the near-end speech from local speaker. The latter case, which is called doubletalk has to be detected so that the process of filter adaptation could be stopped. This prevents the adaptive filter from diverging from echo path response, which would lead to substantial distortion of microphone signal. The detection of such condition is a task of doubletalk detector (DTD) algorithm, which is considered the most significant and troublesome element of an AEC system.

Ciarkowski A. and Czyżewski A..
QUALITY EVALUATION OF NOVEL DTD ALGORITHM BASED ON AUDIO WATERMARKING.
DOI: 10.5220/0003524701810186
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2011), pages 181-186
ISBN: 978-989-8425-72-0

The subsequent section of this paper introduces a DTD algorithm developed at the Multimedia Systems Department by the authors, based on audio signal watermarking techniques. The detailed description of this algorithm is available in the literature (Szwoch, Czyzewski and Ciarkowski, 2009) (Szwoch and Czyzewski, 2008), therefore only brief description in provided. The actual motivation behind this paper is the objective and subjective quality evaluation of this algorithm, especially against current state-of-the-art NCC algorithm.

In accordance with above-set goal, the next section is devoted to the description of the evaluation procedures applied to the DTD algorithms in order to obtain the results which are consequently presented and discussed.

Finally, the conclusions regarding the practical implications of obtained results are drawn.

2 WATERMARKING-BASED DTD OVERVIEW

While most DTD algorithms rely on comparison of far-end and microphone signals, the proposed algorithm utilizes a different approach, which is related to the so called “fragile” watermarking techniques, typically used for protection of multimedia contents against tampering. Fragile watermarking has the property that the signature embedded into the protected signal is destroyed and becomes unreadable when the signal is modified. In case of the double-talk detector algorithm such signal protected from “tampering” is the far-end speaker signal, and the tampering is considered an addition of near-end signal to it. Simultaneously, any linear modifications to the signal resulting from the convolution with impulse response of the audio path should not be considered as tampering, so that the embedded signature would be detectible in “sole” echo signal arriving at the microphone and suppressed in combined echo-and-near-end signal. The information contents of the signature in this application is not important, as only the binary decision whether the signature is present or not is required. The applied signature embedding and detection scheme should also be robust against A/D and D/A conversions, which are inevitable in telephony application, being at the same time transparent (i.e. imperceptible) to the listener, not affecting intelligibility of the speech and perceived quality of the signal. Finally, minor addition of noise and non-linear distortions resulting from imperfections of used analogue elements of audio path should not impair the ability of the algorithm to detect presence of signature in echo signal.

The binary decision coming from the signature detection block of above-described arrangement is inverse to the expected output from DTD algorithm. The correct detection of signature in the microphone signal indicates that near-end speech is not present, making it possible to control the adaptation process of adaptive filter. The described concept is presented in Figure 1. Adaptive filter is used to obtain an estimate of audio path impulse response $h_{\mathrm{a}}(n)$ based on original far-end speaker signal $x(n)$ and microphone signal $u(n)$. The far-end speaker signal provides a subject to filtering with estimated impulse response yielding echo estimate $h_{\mathrm{f}}(n)$, which is subtracted from microphone signal $u(n)$ yielding in turn the signal $e(n)$ with cancelled echo. In order to allow DTD operation the far-end speaker signal $x(n)$ passes through signature embedding block prior to reproduction in the loudspeaker, producing the signal $x_{\mathrm{w}}(n)$ with embedded signature. This signature is being detected in the signature detection block yielding detection statistic $f_{\mathrm{d}}(n)$, which is compared to the detection threshold T_{d} bringing in result binary decision $y(n)$ used to control the adaptation process.

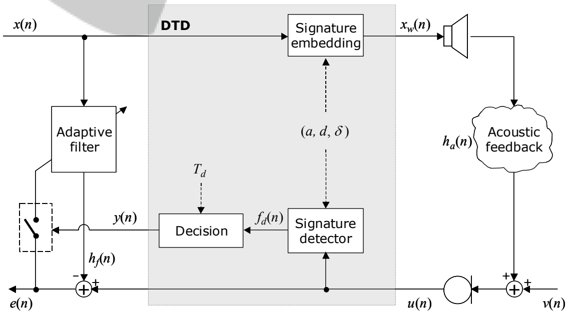

Figure 1: General concept of AEC algorithm with DTD based on audio signal watermarking.

The above-listed requirements regarding the signature embedding and detection process make the choice of a suitable watermarking algorithm problematic. Most commonly used audio watermarking methods are either limited to digital domain only or are too susceptible to noise and reverberation added in the acoustic path. The research on this subject led to the choice of echo hiding method, which adds to the signal single or multiple echoes with short delay (below 30ms), so the effect perceived by the listener is only a slight “coloring” of the sound timbre (Gruhl, Lu and Bender, 1996). In case of watermarking systems the information content of the signature is contained in the modulations of the embedded echo delay, the information being not necessary in this case, therefore constant, predefined echo delay is used during signature embedding, which eases the detection process. It was determined that the use of multiple echoes makes the signature detection more accurate. A detailed description of the design of signature embedding and detection procedure is contained in literature (Szwoch, Czyzewski and Ciarkowski, 2009).

On the foundation of described DTD algorithm acoustic echo cancellation system was created in the form presented in Figure 1. The system includes adaptive NLMS filter, whose length (filter order) is determined by the expected echo delay (length of echo path impulse response $h_a(n)$). Each single detection of double-talk condition by the DTD block causes adaptation of the filter to be held for the time period corresponding to the filter length in order to prevent the filter from processing near-end speaker talkspurt “tail” stored within its buffer.

3 OBJECTIVE AND SUBJECTIVE QUALITY ASSESSMENT

The accuracy of DTD algorithms may be assessed objectively using the methodology based on Receiver Operating Characteristic plots proposed by Cho, Morgan and Benesty (1999). This methodology expresses the DTD accuracy in terms of the probability of miss (P_m) that describes the risk of not detecting the doubletalk, and the probability of false alarm (P_f) that describes the risk of declaring a doubletalk that is not present in the signal. The evaluation is based on measuring P_m performance for a given false alarm probability P_f which is measured as the portion of far-end speech in which doubletalk remains declared when there is no near-end speech. The probability of miss P_m is measured as the portion of near-end speech duration that remains undetected at different levels of near-end to far-end speech ratio (NFR).

The evaluation data for all algorithms was prepared as follows. A single 5s-long speech excerpt of male speaker was used during all tests as a far-end sample. For the near-end speakers 4 1s-long speech excerpts were used (2 male, 2 female). Level of all the test samples was normalized. For each NFR value 16 samples were prepared as the combinations of single far-end speech signal and 4 near-end signals introduced at various time instants (0.5s, 1.5s, 2.5s, 3.5s). In order to properly simulate real-life conditions the far-end signal was delayed by 20ms, convolved with pre-recorded room impulse response of length of 340ms. The impulse response signal was normalized prior applying the convolution in order to keep far-end speech level constant. The actual impulse response is presented in Figure 2. Finally, white noise at predefined level was added to the combined near-and-far-end speech signal. The experiments were conducted at 2 noise level presets: -30 and -60dB (relative to far-end speech level).

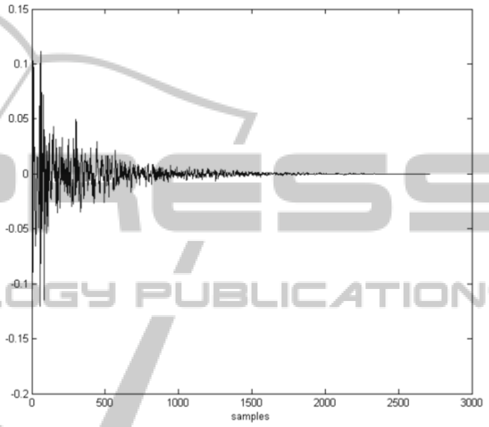

Figure 2: Room impulse response used for simulation of echo signal.

For the purpose of performing performance comparison apart from proposed watermarking-based DTD algorithm, 2 reference DTDs were used, namely the Geigel algorithm introduced by Duttweiler (1978) and more modern Normalized Cross-Correlation (NCC) DTD as described by Benesty, Morgan and Cho (2000). All the DTD algorithms were implemented in MATLAB environment. The choice of the two above-mentioned algorithms was dictated by the fact that the Geigel algorithm, although being very simple, is very commonly found in the literature as a “reference” algorithm (including the papers on the NCC algorithm). On the other hand, the NCC algorithm is relatively recent advancement in the family of time-domain-based DTD methods and demonstrates very good performance, therefore it is considered a state-of-the-art development.

During the experiments the DTD algorithms were coupled with NLMS adaptive filter in order to create functional AEC system. This allowed to obtain not only the DTD output pattern, but also the resulting signal with cancelled echo, useful for the second part of the experiments involving the listening tests. Moreover, a practical implementation of NCC DTD relies on the reuse of room impulse response estimate obtained from the adaptive filter. The length of the NLMS filter used was L=512 and the NCC algorithm used the window of length W=500 to obtain estimates of correlation vectors. These values were chosen identical to those used by Cho et al. (1999), so that the direct comparison of results was possible.

The objective evaluation was performed as follows. A decision threshold of each algorithm was adjusted so as to achieve requested probability of false alarm P_f in an iterative process. The values of P_f were chosen identical to those used by Cho et al. (1999) – 0.1 and 0.3. Then, with fixed P_f, for each NFR value (in the range -20 to 20dB with step 5dB) a series of 16 aforementioned simulated echo excerpts was prepared and probability of miss P_m for each excerpt was calculated. Therefore, the actual P_m values plotted in the results section are obtained as an average over the whole series.

Apart from objective evaluation, also subjective one was performed in the form of listening tests. The purpose of such arrangement was to check the perception of quality of speech subjected to AEC system based on various DTD algorithms. The group of 12 experts (PhD students and staff members of Multimedia Systems Department at GUT) was asked to assign their scores to some of the excerpts obtained during objective evaluation. The excerpts in a single comparison group consisted of the pure near-end speech signal, unprocessed simulated echo signal and output signals from AEC system with each DTD algorithm. All the excerpts within the group were obtained at the same NFR value, noise level and consisted of the same near-end speech signal introduced at fixed moment. Overall 4 groups were prepared with NFR values of 0 and -15dB and noise level of -30 and -60dB, therefore each expert was subjected to the rehearsal of the total of 20 excerpts. The scores used Mean Opinion Score (MOS) scale (1-5).

4 EVALUATION RESULTS

The results of objective evaluation are presented in Figure 3 and Figure 4 for noise levels -30dB and -60dB, respectively. At the higher noise level and decision threshold set for false alarm probability P_f=0.1 all DTD algorithms demonstrate similar performance. While Geigel algorithm in all conditions performs the worst, the difference between NCC and watermarking-based DTD for P_f=0.3 is almost negligible.

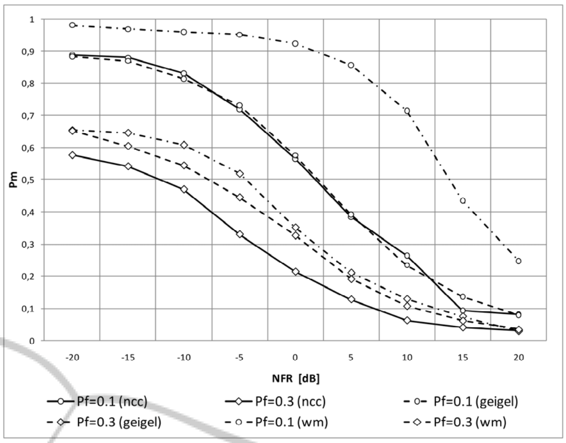

Figure 3: Performance of tested DTD algorithms, noise level at -30dB.

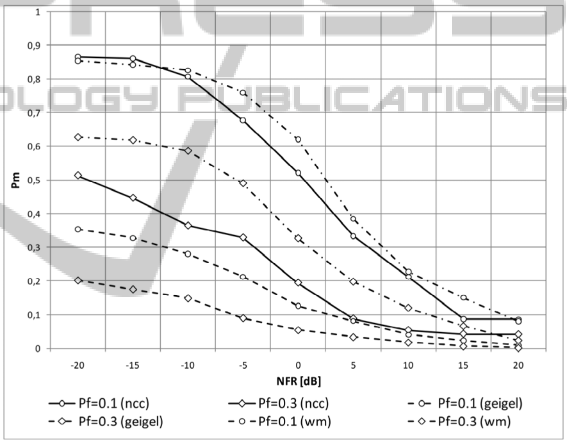

Figure 4: Performance of tested DTD algorithms, noise level at -60dB.

When noise level is set at -60dB the proposed algorithm performs substantially better than the other two. For the probability of false alarm P_f=0.1 NCC algorithm is only slightly better than Geigel DTD. This is in clear contradiction to the results obtained by Cho et al. (1999), however the conditions of this experiment are quite different. In the original work, the NCC algorithm used the real room impulse response, while here it uses the estimate obtained from adaptive filter. Therefore, any DTD miss leading to NLMS filter divergence will in turn have great impact on its performance for subsequent audio samples. However, the conditions of this experiment better mimic real-life, whereas room impulse response is not known *a priori*. Another difference from the experiment setup discussed by Cho et al. (1999) is the length of the impulse response (340 compared to 256ms), therefore the effect of unmodeled impulse response “tail” is here far more significant.

The results of objective evaluation are presented in tables 1-4. The NCC algorithm demonstrates almost consistent behaviour regardless of applied noise level. The watermarking-based algorithm is the most susceptible to noise, however when the noise level is low its performance is superior to the other two. The Geigel algorithm consequently receives lowest scores.

Table 1: MOS values for DTD algorithm comparison; noise level at -30dB, NFR=0, P_f=0.1.

Test signal	MOS
Reference (sole near-end speech)	4.83
Reference (unaltered microphone signal)	1.25
AEC output with NCC DTD	3.92
AEC output with Geigel DTD	2.58
AEC output with proposed DTD	3.75

Table 2: MOS values for DTD algorithm comparison; noise level at -60dB, NFR=0, P_f=0.1.

Test signal	MOS
Reference (sole near-end speech)	4.92
Reference (unaltered microphone signal)	1.25
AEC output with NCC DTD	3.75
AEC output with Geigel DTD	3.08
AEC output with proposed DTD	4.33

Table 3: MOS values for DTD algorithm comparison; noise level at -30dB, NFR=-15dB, P_f=0.1.

Test signal	MOS
Reference (sole near-end speech)	4.75
Reference (unaltered microphone signal)	1.16
AEC output with NCC DTD	2.0
AEC output with Geigel DTD	1.25
AEC output with proposed DTD	1.84

Table 4: MOS values for DTD algorithm comparison; noise level at -60dB, NFR=-15dB, P_f=0.1.

Test signal	MOS
Reference (sole near-end speech)	4.83
Reference (unaltered microphone signal)	1.16
AEC output with NCC DTD	1.84
AEC output with Geigel DTD	1.84
AEC output with proposed DTD	3.75

During the preparation of test signals the authors also paid attention to the execution time of DTD simulation. These time spans are presented in Table 5. It is notable that the full iteration over all NFR levels with 16 test signals at each level took over 17 hours with NCC DTD, so it was running on average 88.6 times slower than the real-time. The same data set was processed by the other DTD algorithms in just over 4 minutes. Although unoptimized MATLAB implementations are not credible target for complexity benchmarking, the disparity in achieved execution times is outstanding.

Table 5: Execution time of full DTD simulation for tested algorithms.

DTD algorithm	Execution time [s]
NCC	63793
Geigel	259
Watermarking-based	232

5 CONCLUSIONS

The presented DTD algorithm based on audio watermarking techniques has been exhaustively evaluated against both very simple-yet-common Geigel algorithm and more recent and sophisticated Normalized Cross-Correlation DTD. The evaluation technique proposed by Cho et al. (1999) was used to obtain Receiver Operating Characteristic plots of the aforementioned algorithms, which are the objective means of DTD comparison, based on the foundation of the detection theory. The results show that the proposed novel DTD algorithm performs comparably (or only slightly worse) to the NCC algorithm in high-noise conditions, significantly outperforming it when the near-end background noise level is low, while both “modern” algorithms perform better than Geigel algorithm.

The subjective evaluation of the algorithms was carried out in order to assess the perceived quality of echo cancellation performed with them. This allows not only to statistically verify how often the DTD algorithm result matches the reference pattern, but also takes into account how the exact conditions of DTD miss impact the adaptive filter behavior. This is especially important for the NCC algorithm, as its “fast” version relies heavily on the reuse of room impulse response estimated by the adaptive filter. That factor is also clearly visible in the obtained ROC plots, which differ from the reference plots published by Cho et al. (1999), obtained with the fixed, real impulse response, whenever tests conducted by the authors used estimated one to better simulate real-life usage.

The MOS values obtained through the listening tests confirm the observation that the proposed, watermarking-based DTD algorithm is more susceptible to high noise levels that NCC algorithm. The Geigel algorithm in all test cases was rated the lowest, which is consistent with the expectations.

An interesting aspect of the experiment was the preparation of test data, which allowed to observe the practical effects of computational complexity of the algorithms. As a direct consequence of its frame-based operation, the proposed algorithm achieved lower execution times than Geigel DTD, which operates sample-by-sample. On the other hand, the NCC algorithm even in the “fast” version, reusing adaptive filter’s room impulse estimate instead of performing cross-correlation matrix inversion achieved execution times on the order of tens to hundreds times slower than the watermarking-based algorithm. This has direct practical implications, as it restricts the use of NCC algorithm to specialized hardware implementations, while the watermarking-based algorithm with execution time on modern PC machine several times faster than the real-time is perfectly suited for software implementations. This makes it a viable choice as a component of AEC system embedded in e.g. software VoIP terminal.

The hitherto performed evaluation of proposed watermarking-based algorithm included only comparison with the representatives of time-domain-based operation DTD algorithms, therefore the authors would like to enhance this study in future research by including also other recent developments, e.g. the algorithm based on a soft decision scheme in the frequency domain proposed by Park and Chang (2010) or frequency-domain Gaussian-Mixture Model-based DTD algorithm proposed by Song et al. (2010).

ACKNOWLEDGEMENTS

The research was funded by the Polish Ministry of Science and Higher Education within the grant No. PBZ-MNiSW-02/II/2007.

REFERENCES

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

Cho, J. H., Morgan, D. R., Benesty, J., (1999). An Objective Technique for Evaluating Doubletalk Detectors in Acoustic Echo Cancelers. *IEEE Trans. Speech Audio Processing,* 7(6), 718-724.

Duttweiler, D. L., (1978). A Twelve-Channel Digital Echo Canceller. *IEEE Trans. Commun*., 26(5), 647–653.

Gruhl, D., Lu, A., Bender, W., (1996). Echo Hiding. In *Proc. Information Hiding Workshop* (pp. 295–315). Cambridge, UK.

Kuo, S. M., Lee, B. H., Tian, W., (2006). Adaptive Echo Cancellation. In Real-Time Digital Signal Processing: Implementations and Applications. Wiley: NewYork.

Szwoch, G., Czyzewski, A., Ciarkowski, A., (2009). A Double-talk detector using audio watermarking. *J. Audio Eng. Soc.*, 57(11), 916-926.

Czyzewski, A., Szwoch, G., (2008). Method and Apparatus for Acoustic Echo Cancellation in VoIP Terminal. International Patent Application No. PCT/PL2008/000048.

Vu, T., Ding, H., Bouchard, M., (2004). A Survey of Double-Talk Detection Schemes for Echo Cancellation Applications. *Can. Acoust*., 32, 144–145.

Park, Y-S., Chang, J-H., (2010). Double-talk detection based on soft decision for acoustic echo suppression. *Signal Processing*, 90(5), 1737-1741.

Song, J-H., Lee, K-H., Park, Y-S., Kang, S-I., Chang, J-H., (2010). On Using Gaussian Mixture Model for Double-Talk Detection in Acoustic Echo Suppression. In *INTERSPEECH 2010,* 2778-2781.