

Acoustic Echo Cancellation: Dual Architecture Implementation

¹B. Stark and ²B.D. Barkana

¹Department of Computer Science and Engineering,

²Department of Electrical Engineering,

School of Engineering, University of Bridgeport, Bridgeport CT, 06604 USA

Abstract: Problem statement: With the rise in mobile communication, it is becoming more frequent to use a communication device in an enclosed noisy environment, such as a subway or in a lobby. In this setting however, the received microphone is severely degraded by the echo from the speaker and background noise. The audio processing necessary to clarify the desired speech can be broken down into two parts, removal of the acoustic echo and removal of the background noise. **Approach:** This study proposed an 'external-switched' algorithm of a dual architecture implementation for acoustic echo cancellation. Using the orthogonality property of adaptive algorithms to detect convergence, two complete adaptive filters can be run in parallel to take advantage of each filter's particular configuration. By configuring one filter for fast adaptation and the second for minimizing the steady state error, a system can be designed with the advantages of both without suffering from increased computational cost. **Results:** A slight performance improvement can be demonstrated with this system; however the greatest advantage is in the reduced filter size and calculation cost. **Conclusion:** This parallel approach is suitable for systems in which a single approach to acoustic echo cancellation is insufficient. Disadvantages of one algorithm can be mitigated by being able to switch to a more effective algorithm seamlessly.

Key words: Acoustic echo cancellation, Least-Mean Square (LMS) approximation

INTRODUCTION

With the rise in mobile communication, it is becoming more frequent to use a communication device in an enclosed noisy environment, such as a subway or in a lobby. In this setting however, the received microphone is severely degraded by the echo from the speaker and background noise. The audio processing necessary to clarify the desired speech can be broken down into two parts, removal of the acoustic echo and removal of the background noise. Acoustic Echo Cancellation (AEC) is commonly done with an adaptive filter, frequently done with stochastic-gradient adaptive algorithms that use a Least-Mean Square (LMS) approximation. However, background noise and other non-desired artifacts such as voice reverberation; negatively affect the performance of these filters.

In general, the adaptive algorithm is used to estimate the acoustic echo and subtracts this estimation from the near-end microphone signal. The simplest algorithm uses the previous values to approximate to the gradient vector to solve the steepest-descent problem posed by the Least-Means Square (LMS) approximation. Other algorithms developed to solve the

steepest-descent problem include the Normalized Least Mean Square (NLMS) algorithm, sign-error LMS, Proportionate Normalized Least Mean Squares (PNLMS) algorithm (Gänsler, 2000), robust variable step-size NLMS (RVSS-NLMS) algorithm (Vega, 2008) and momentum NLMS (MNLMS) algorithm (Chhetri *et al.*, 2006). These all have been proven to be effective in removing the acoustic echo to some degree. However, often a residual echo often remains due to several factors, including an insufficient filter length, incorrect echo path estimation and nonlinear signal components (Habbets *et al.*, 2008). A noisy environment can further degrade the effectiveness of the AEC algorithm and the quality of the near-end speech.

Previous study on AEC have focused on minimizing these issues by adding a double-talk detector (Chhetri, 2006), adding a post filter for Noise Suppression (NS) (Habbets, 2008; Gustafsson *et al.*, 2002), improving adaptive algorithms (Chhetri *et al.*, 2008), or by using a nonlinear AEC (Shi, *et al.*, 2008). All of these implementations however, increase the complexity of the system with additional components or more complex algorithms that require more computations.

Corresponding Author: B. Stark, Department of Computer Science and Engineering, School of Engineering, University of Bridgeport, Bridgeport CT, 06604 USA

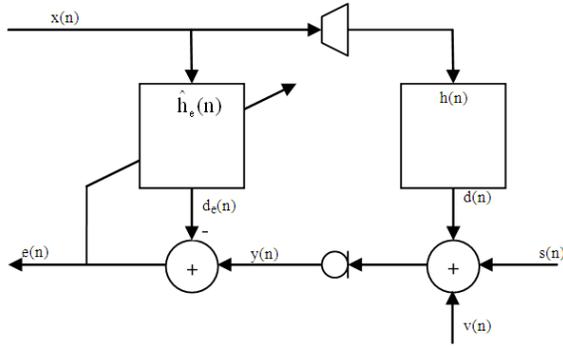


Fig. 1: Single microphone AEC system

This study proposes the use of a type of algorithms described as ‘external-switched’ in which two or more adaptive filters are run in parallel and the final result is determined by which filter is most accurate at the specified time. In this study, a dual architecture implementation of the simple NLMS algorithm is proposed. By configuring one NLMS filter for fast adaptation and one to minimize the steady state error and selecting between the two depending on which one is more accurate at the current time, the system receives the benefit of both configurations, reducing both convergence time and steady state error with results comparable to more complex and costly algorithms.

Acoustic echo cancellation using NLMS: In a typical AEC algorithm, we can model the process with a single microphone system as seen in Fig. 1.

The far-end speech $x(n)$ is played out of the speaker and is picked up on the microphone as an echo $d(n)$. The output of the adaptive filter $d_e(n)$ is intended to cancel out the echo from the microphone signal $y(n)$. The microphone signal is composed of the far-end speech echo $d(n)$, the near-end speech $s(n)$ and background noise $v(n)$. The difference between the microphone signal and the estimated echo forms the near-end speech $e(n)$, which is fed back into the adaptive filter to update the taps.

In this model, the acoustic echo can be assumed to be a linear filter, which takes the form of the following equation:

$$d(n) = \sum_{j=0}^{N_h-1} h_j(n)x(n-j) \quad (1)$$

Where:

- N_h = The length of the true echo filter
- h_j = The filter coefficient
- $x(n)$ = The far-end speech

Using the NLMS algorithm, we can model a time based adaptive filter by the following equation:

$$\hat{h}_e(n+1) = \hat{h}_e(n) + \mu \frac{x(n)e(n)}{\|x(n)\|^2 + \delta_{NLMS}} \quad (2)$$

Where:

- $\hat{h}_e(n)$ = The estimated impulse response vector
- μ = The step-size factor
- δ_{NLMS} = The regularization factor to prevent division by zero
- $x(n)$ = The far-end speech signal

The estimated echo, $\hat{d}(n)$ can then be calculated using:

$$\hat{d}(n) = \sum_{j=0}^{N_e-1} \hat{h}_{e,j}(n)x(n-j) \quad (3)$$

Where:

- N_e = The filter size
- $\hat{h}_e(n)$ = The estimated impulse response vector
- $x(n)$ = The far-end speech

The goal of all acoustic echo cancellation is to minimize the residual echo, which can be defined as the slight difference between the true echo and the estimated echo. This is simply calculated to be:

$$e_r(n) = d(n) - \hat{d}(n) \quad (4)$$

Due to the limitations of the NLMS algorithm, the residual echo is rarely zero. There have been many papers on improving the effectiveness of the AEC by improving the adaptive filter. The simple NLMS algorithm is effective, but other proposed algorithms have been shown to be more accurate. One variant proposed by Vega *et al.* (2008) is the RVSS-NLMS where the step-size solution at each iteration switches between an NLMS $\mu = 1$ or a Normalized Sign Algorithm (NSA) where $\mu = \sqrt{\delta}i-1$. This ‘switched-norm’ algorithm allows for the fast convergence provided by NLMS and the robust performance against noise provided by NSA. The downside of this algorithm and many other complex algorithms is the computation cost. An estimated computation cost can be determined by examining the number of arithmetic operations needed at each iteration. The majority of LMS-based algorithms can be described as being in the order of $O(M)$, where M is the size of the filter (Sayed, 2008).

The simple LMS and NLMS algorithms require 2 and 3 M additions and multiplications respectively, while more complex algorithms such as RVSS-NLMS may require three times as many calculations Vega (2008).

Beyond the adaptive algorithm, there are several external features that can be added to improve the effectiveness of an AEC system. A post filter, appended to the system, has been demonstrated to be an effective addition (Habbets, 2008; Gustafsson *et al.*, 2002). Habets *et al.* (2008) provides an excellent overview of post filters designed to mitigate the limitations of a deficient adaptive filter. The addition of a robust post filter has also been demonstrated to help alleviate adaptive algorithm computation complexity by allowing the filter to use a smaller filter order. A smaller filter order has several advantages, including a faster convergence time, less sensitive to noise and reduced computational complexity at the cost of a higher steady state error. On the other hand, post filters have been demonstrated to introduce distortion and other artifacts during the processing. Nonlinear processes such as center clipping have a notable distortion effect (Chhetri *et al.*, 2006). As such, it has been well documented that there is a tradeoff between not only between adaptation time and steady state error, but between balancing the computational complexity of the adaptive filter and the post filter (Chhetri *et al.*, 2006).

Double Talk Detectors (DTD) have also been frequently added to AEC systems. An occurrence of speech by both the far end speaker and the near end speaker into a system often disrupts the acoustic echo cancellation process. The simplest double talk detectors simply prevent the filter coefficients of the adaptive algorithm from changing during the double talk which is determined by comparing the magnitude of the far end and near end signals. Several other DTDs have been proposed, however, of note, a novel DTD proposed by (Ye *et al.*, 1991) uses the orthogonality property of adaptive algorithms, wherein when the echo canceller has converged, the AEC output signal is orthogonal to the speaker signal. The cross correlation thus can be used to determine whether or not the adaptive algorithm has converged. This was further explored by (Chhetri *et al.*, 2006) to create a convergence detector. This property is explored in greater detail as the convergence detector for the ‘external-switched’ algorithm in the dual architecture implementation.

Dual architecture implementation: The ‘external-switched’ adaptive algorithm is the backbone of the dual architecture implementation. In all of the

previously discussed AEC systems, each strive to maintain a balance between fast convergence, a low steady state error, computation cost and hardware complexity. With the large number of possibilities, it is difficult to create an optimized configuration for all cases. In this implementation, the goal is to maximize fast convergence time, a reduce steady state error and computation cost at the expense of hardware complexity and size. With the ever decreasing size of electrical components, hardware size is less significant.

The ‘external-switched’ adaptive filter portion of the dual architecture implementation, as seen in Fig. 2, consists of two NLMS adaptive algorithms (NLMS₁, NLMS₂) running in parallel, one configured for fast convergence, NLMS₁ and the second configured to minimize the steady state error, NLMS₂. In general, for all stochastic gradient adaptive algorithms, the approximation for the steepest descent is based off two major variables; the size of the filter and the step-size for adjustment. A larger filter size provides the greatest accuracy in terms of steady state error; however it is both costly computation-wise and reacts poorly to sudden changes (Sayed, 2008). In regards to step-size, in the NLMS algorithm, the step-size is normalized to be in proportion to the squared-norm of the input signal. This is particularly useful in speech signals, where the input signal fluctuates frequently due to pauses in speech. This way the filter taps are not overly adjusted when there is a pause.

With the effectiveness of the NLMS algorithms in these configurations well known, the critical addition to this ‘external-switched’ algorithm is the convergence detector. At each sample, the output signal from NLMS₂, $e_2(n)$ is processed by the convergence detector. If NLMS₂ has converged, $e_2(n)$ is used as the final AEC output; otherwise the output from NLMS₁, $e_1(n)$ is used.

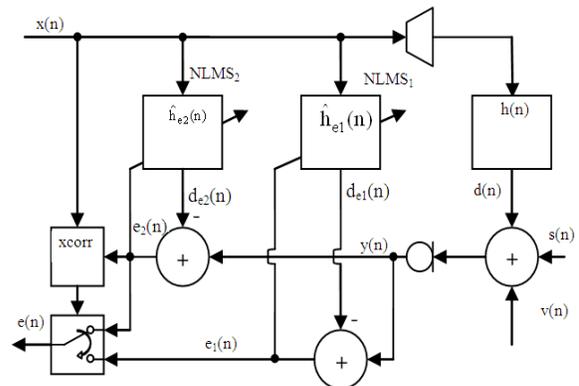


Fig. 2: Dual architecture AEC system

The convergence detector is based on the orthogonality property of adaptive filters, where in a converged adaptive filter; the output signal is orthogonal to the input signal (Sayed, 2008). This property has been used by (Ye *et al.*, 1991) as the basis for a double-talk detector. It was expanded to its current implementation as a convergence detector by (Chhetri *et al.*, 2006). As described in these works, the cross correlation function is large while the filter is adapting and very small once the filter has converged.

With this property, the Average Cross Correlation Coefficients (ACCC) of $e_2(n)$ and $x(n)$ can be used to determine whether NLMS₂ has converged. At every 50 ms frame, the ACCC is compared to a convergence threshold. The convergence threshold is best obtained experimentally; though an approximation for the threshold is the average unwanted noise which can be described as:

$$ACCC_{th} \approx \frac{\sum_{i=0}^{N-1} |v_i(n)|}{N} \quad (5)$$

Where:

$v_i(n)$ = The background noise at sample i

N = The total number of samples

If the inequality $ACCC(n) < ACCC_{th}$ is true, it can be said that \hat{h}_{e2} has converged. Otherwise, NLMS₂ is still adapting which indicates either the filter has not converged or the echo path has changed.

MATERIALS AND METHODS

The ‘external-switched’ algorithm was implemented in MATLAB Simulink using the Signal Processing Blockset, following the block diagram in Fig. 2. NLMS₁ was designed with a filter size of 512 taps, and NLMS₂ had a filter size of 2048 taps. The convergence detector was made with a custom function to calculate the ACCC during a 50 ms frame. A switch compares the result of the ACCC to the threshold value and selects which output should be the system output. The sample signal used was an 8 kHz sample whose Signal-to-Noise Ratio (SNR) was adjusted at each simulation.

The performance of this system was evaluated through two sets of simulations. The first set evaluates the MSE and convergence time of the ‘external-switched’ algorithm using a noisy input signal. The ‘external-switched’ algorithm is compared against a similar NLMS algorithm, with an experimentally optimized filter to achieve the best balance between

convergence time and Mean-Squared-Error (MSE). Convergence time in the context of analysis is defined as when the MSE has reached an asymptote.

The second set of simulations examines the Echo Return Loss Enhancements (ERLE) which is described as:

$$ERLE(n) * \log_{10} \left[\frac{y(n)^2}{e(n)^2} \right] \quad (6)$$

Where:

$y(n)$ = The microphone signal

$e(n)$ = The AEC output

The ERLE is a measure of the reduction in echo from the microphone signal; the larger the dB value, the greater the effectiveness of the AEC system.

For this set of simulations, the proposed algorithm is compared against a Frequency Domain Adaptive Filter (FDAF). Adaptive filters in the frequency domain use a fast convolution technique to compute the output. In the frequency domain, the computational cost is no longer proportional to the filter size, as a result, convergence time is often shorter. The drawback to this class of adaptive filters is the extra hardware necessary to convert into the frequency domain and back to the time domain, and only updating the weights once per frame (Sayed, 2008). The frequency domain NLMS thus provides an excellent comparison to the proposed ‘external-switched’ algorithm because both emphasize speed and accuracy over hardware size.

RESULTS

The ‘external-switched’ algorithm was first tested as a noise cancellation system to demonstrate its proper function. For noise reduction, the convergence time and the MSE were used to analyze the effectiveness of the algorithms. The SNR ranged from 70.1-10.4 dB. The results seen in Fig. 3 and 4 are from a simulation set using a noisy signal with an SNR of 10.4 dB. These results were compared to an experimentally optimized NLMS algorithm with a filter size of 4096.

Figure 5 shows the results of the ‘external-switched’ algorithm in comparison to a Frequency Domain Adaptive Filter (FDAF) NLMS algorithm with a frame size of 50 ms. The ‘external-switched’ algorithm starts converging faster, due to NLMS₁, which is configured for fast convergence. Until the slower NLMS₂ converges, the FDAF has a higher ERLE. However, once both AEC’s stabilize, it is apparent that they are comparable.

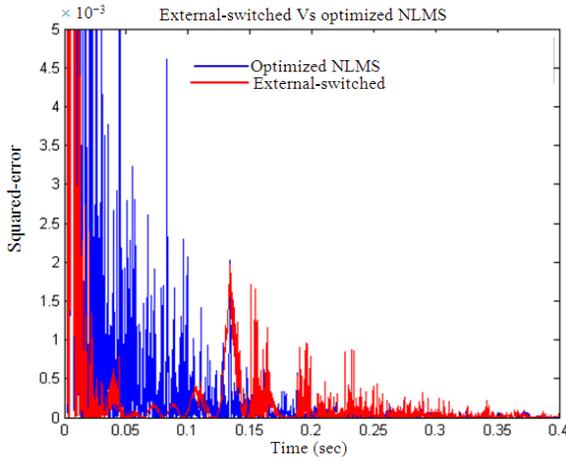


Fig. 3: Convergence time comparison

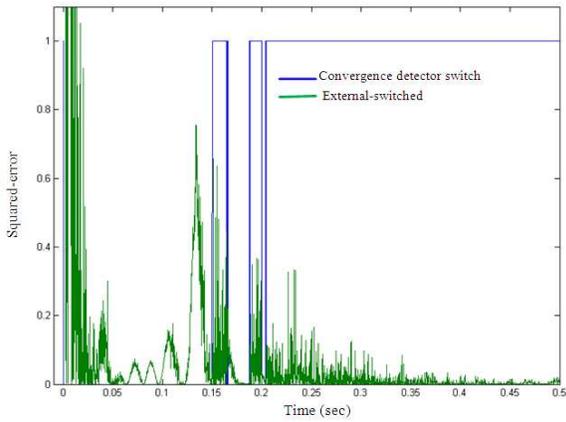


Fig. 4: Convergence detector operation

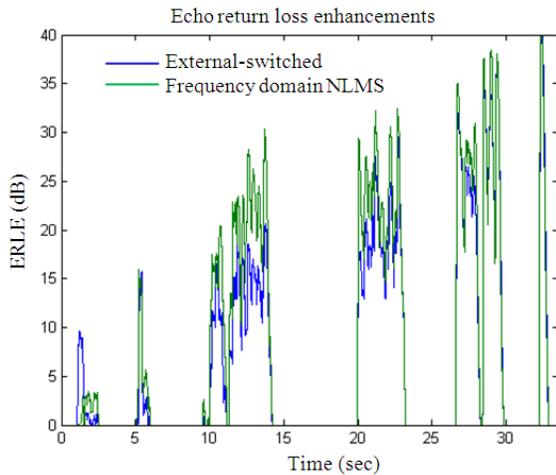


Fig. 5: Echo returns loss enhancements

DISCUSSION

In Fig. 3, the advantages of the ‘external-switched’ algorithm are readily apparent. Although the convergence time for both filters is similar, the instantaneous squared error of the ‘external-switched’ drops rapidly due to the fast convergence of NLMS₁. While the instantaneous squared-error increases due to the change from NLMS₁ to NLMS₂, this is due to a value for the threshold, $ACCC_{th}$, that is not optimal. In practice, an optimized value for $ACCC_{th}$, would be impossible to determine, so for these simulations the approximate value is used which could be calculated from an input signal.

In Fig. 4, the convergence detector switch is overlaid on the instantaneous square-error graph of the ‘external-switched’ algorithm. In this simulation, the convergence detector switched to the slower adaptation at 0.2 sec. While not optimal, it is still effective enough to be comparable to a NLMS algorithm that requires a filter size nearly twice the size of entire ‘external-switched’ algorithm. The MSE for the ‘external-switched’ algorithm hovered around 0.32×10^{-3} , whereas the MSE for the optimized NLMS algorithm settled at 0.33×10^{-3} .

In subsequent simulations, the ‘external-switched’ algorithm performed similarly. While there was no significant advantage of the algorithm performance-wise, it was easily comparable to an NLMS algorithm that was optimized for each simulation.

The results of the AEC system using the ‘external-switched’ algorithm depict it as comparable to the frequency domain NLMS algorithm in regards to performance. This is not wholly unexpected as FDAF normally perform significantly better than their time based adaptive filter counterparts. However, it should be noted that applying an ‘external-switched’ algorithm to the traditional NLMS algorithm improves its performance to the level of a better performing algorithm, at a reduced computational cost. An even better performance may be gained by combining the ‘external-switched’ algorithm with properly optimized algorithms in the frequency domain.

CONCLUSION

This study proposes an ‘external-switched’ algorithm of a dual architecture implementation for an AEC system. The proposed system was designed as an attempt to maximize convergence speed and to minimize the steady state error, at the expense of extra hardware. While this implementation is effective and comparable to other more refined algorithms, it does

not show a marked improvement in AEC design. The convergence detector developed by (Ye *et al.*, 1991) and expanded upon by (Chhetri *et al.*, 2006) is effective and warrants further exploration. A dual architecture of a more complex algorithm than NLMS may prove to be more effective, albeit at the cost of increased computation requirements.

REFERENCES

- Chhetri, A., J. Stokes and D. Florencio, 2006. Acoustic echo cancellation for high noise environments. Proceedings of IEEE International Conference on Multimedia and Expo (ICME), July 9-12, IEEE Xplore Press, Toronto, Ontario, Canada, pp: 905-908. DOI: 10.1109/ICME.2006.262666
- Fu, J. and W.P. Zhu, 2008. A nonlinear acoustic echo canceller using sigmoid transform in conjunction with RLS algorithm. IEEE Trans. Circ. Syst., 55: 1056-1060. DOI: 10.1109/TCSII.2008.926798
- Gänsler, T., S.L. Gay, M.M. Sondhi and J. Benesty, 2000. Double-talk robust fast converging algorithms for network echo cancellation. IEEE Trans. Speech Audio Process., 8: 656-663. DOI: 10.1109/89.876299
- Gustafsson, S., R. Martin, P. Jax and P. Vary, 2002. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. IEEE Speech and Audio Process., 10: 245-256. DOI: 10.1109/TSA.2002.800553
- Habets, E., S. Gannot, I. Cohen and P. Sommen, 2008. Joint dereverberation and residual echo suppression of speech signals in noisy environments. IEEE Audio, Speech Language Process., 16: 1433-1451. DOI: 10.1109/TASL.2008.2002071
- Sayed, A., 2008. Adaptive Filters. 1st Edn., IEEE Press, Hoboken, New Jersey, ISBN: 9780470253885, pp: 786.
- Shi, K., X. Ma and G.T Zhou, 2008. Adaptive acoustic echo cancellation in the presence of multiple nonlinearities. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Mar. 30-Apr. 4, IEEE Xplore Press, Las Vegas, NV., USA., pp: 3601-3604. DOI: 10.1109/ICASSP.2008.4518431
- Vega, V., H. Rey and J. Benesty, 2008. A new robust variable step-size NLMS algorithm. IEEE Trans. Sign. Process., 56: 1878-1893. DOI: 10.1109/TSP.2007.913142
- Ye, H. and B. Wu, 1991. A new double talk detection algorithm based on the orthogonality theorem. IEEE Trans. Commun., 39: 1542-1545. DOI: 10.1109/26.111430