A BLOCK LEAST SQUARES APPROACH TO ACOUSTIC ECHO CANCELLATION

Eric Woudenberg†, Frank K. Soong‡ and B.-H. Juang‡

† Advanced Telecommunications Research, Human Information Processing Laboratory, Kyoto, Japan
‡ Bell Laboratories - Lucent Technologies, Murray Hill, New Jersey, USA

ABSTRACT

We propose an efficient block least-squares (BLS) algorithm for acoustic echo cancellation. The high computation and memory requirements associated with a long room echo make the simple, gradient-based LMS filter a more acceptable commercial solution than a full-fledged LS canceler. However, the LMS echo canceler has slower convergence and worse steady-state performance than its LS counterpart. In the proposed BLS approach, the autocorrelation and cross-correlation of the source and echo, required in solving the LS normal equations, are performed once per block using FFT's. With appropriate data windowing the autocorrelation matrix is constrained to be Toeplitz, allowing the corresponding normal equations to be solved efficiently. The positive definiteness of the autocorrelation function eliminates the stability problems of other fast LS algorithms. BLS can reduce the echo residual to the level of background noise, allowing a residual power based, statistical near-end speech detector to be devised. Performance in real environments under various settings of filter length, SNR, near-end speech presence, etc., is investigated.

1. INTRODUCTION

For modern hands-free communications, such as full-duplex teleconferencing in human-to-human communication, or automatic speech recognition in human-machine communication with barge-in capability, the acoustic echo of far-end speech or system messages needs to be cancelled. Inadequate echo cancellation can impede successful full-duplex teleconferencing and significantly degrade automatic speech recognition performance.

In this paper, we propose to use a block least-squares (BLS) based algorithm for cancelling acoustic echoes in a room. Different from typical echo in a telephone network [1], [2] the acoustic echo generally has a much longer time-span. The actual time-span of an echo response can, depending upon the physical size of the room, the wall reflectivity, and the relative positioning between speakers and microphones, easily last up to a second. At a sampling rate of 8 kHz, such a room response requires an adaptive filter of several thousand taps. The high computation and memory requirements associated with such a long filter has made the simple, gradient-based LMS AEC a more desirable solution for commercial applications than its LS counterpart. However, due to its steepest descent nature, the LMS room echo canceler has certain intrinsic performance disadvantages, including: slower convergence and higher echo residuals. Various attempts have been tried to improve the convergence behavior. Most noticeable ones are the NLMS algorithm by Duttweiler [3] and the PNLMS++ by Gay [4] where the far-end signal power and filter coefficient magnitudes are considered in order to improve LMS canceler convergence. A least-squares AEC which is capable of delivering a higher echo cancellation performance, on the other hand, suffers from higher computational complexity and an annoying potential instability.

In this paper we propose a block least-squares (BLS) approach to the room AEC problem. The autocorrelation of the source signal and cross-correlation of the source and the echoes, both required for solving the LS normal equations, are computed efficiently in the frequency domain using FFT's. By appropriate data windowing, the autocorrelation matrix is also constrained to be Toeplitz and the efficient Levinson recursion is used for solving the LS normal equations once per block. Furthermore, the positive definiteness of the autocorrelation function guarantees the stability of the adaptive filter solution and maximum cancellation of the received room echo. The fast convergence and the high cancellation performance of the least-squares algorithm are guaranteed in the BLS canceler. By further exploiting the fact that the echo is cancelled down to the level of background noise, we propose a residual power based, high performance, statistical near-end speech detector. Cancellation performance in a real acoustic environment is evaluated using computer simulation and tested in a live acoustic environment. The performance is investigated by changing the system parameters, including: the misalignment, window size, SNR's etc., with and without near-end speech signals present.

2. BLS Acoustic Echo Canceler (AEC)

The following notations are used throughout this paper:
$x(t)$: the sound source signal played through a loudspeaker, $\mathbf{x}(t)$, corresponding vector, and matrix $\mathbf{X}(t)$
$y(t)$: room echo signal of $x(t)$ received at the microphone
$\hat{y}(t)$: estimated echo
$e(t)$: echo residual signal, $\mathbf{e}(t)$, corresponding vector
$v(t)$: background noise sample received at the microphone
$s(t)$: near-end speech signal sample
$z(t)$: $y(t) + v(t) + s(t)$, and $\mathbf{z}(t)$, corresponding vector
$\mathbf{R}_{xx}$: Toeplitz autocorrelation matrix of $x(t)$
$\mathbf{r}_{xz}$: crosscorrelation vector between $x(t)$ and $z(t)$
$h(l)$: l-th tap coefficient, and $\mathbf{h}$ is vector of the AEC filter
N : data block size
L : echo canceler length

We propose a block LS echo canceler as follows:

1. Block N samples of source signal samples, $x(t)$, and observed microphone input samples, $z(t)$.
2. Weight $x(t)$ and $z(t)$ with an appropriate tapering window.
3. Compute the autocorrelation matrix and cross-correlation vector of the current block.
4. Update $\mathbf{R}_{xx}$ and $\mathbf{r}_{xz}$ with auto and cross-correlation of the current block using a leakage integrator.
5. Find the adaptive echo canceler coefficients by solving the normal equations via efficient Levinson recursion, $\mathbf{h} = \mathbf{R}_{xx}^{-1}\mathbf{r}_{xz}$.
6. Compute the estimated echo, $\hat{y}(t)$, and subtract it from the microphone input sample, $z(t)$. Continue at 1 with the next block.

The AEC is formulated as a block LS echo cancellation algorithm. A block, rather than sample, adaptive approach is adopted here for the following two reasons. First, since the positions of speaker(s) and microphone(s) in a room are relatively steady, it is a reasonable assumption that the resultant echo return path is stationary within a block, typically around .5 sec. Even with this stationarity assumption, a slowly time-varying echo return path can still be tracked using our block LS algorithm with memory. To the degree that the echo return path is a linear FIR system, the LS AEC delivers the best achievable cancellation performance.

3. Near-end Speech Detection

It can be shown that the echo residuals of the proposed block LS AEC converge to the uncorrelated background noise level in the mean. This high cancellation performance facilitates a novel, near-end speech detection algorithm proposed as follows. The near-end speech detection algorithm is based upon two estimates: a background noise power estimate and current echo residual power estimate. The near-end speech detection is devised based upon a statistical hypothesis testing procedure. A similar, at least in spirit, near-end speech detection algorithm was proposed by Benesty, Morgan and Cho [6] where a normalized crosscorrelation vector, rather than residual and background noise power estimate, was used as detection parameter.

The background noise power is estimated in the current block and used for near-end speech detection in the next block. This delay is introduced intentionally here to prevent the background power estimate from too closely tracking the current block speech activities. Within a block, from the first frame to δ frames before the first detection of near-end speech and from δ frames after the last detection of near-end speech to the end of the block are used to update the background power estimate. Near-end speech detection is performed from left to right by comparing the smoothed power of every frame with the background power estimate computed in the previous block. If the smoothed power of the current frame is higher than the background estimate by a threshold, η, the frame is declared as near-end, otherwise, non-nearend.

4. Computational Complexities

The computational requirement of the proposed BLS echo canceler is low when compared with other sample adaptive LS cancelers. This is due to two factors: (1) loading the Toeplitz autocorrelation matrix and the cross-correlation vector is done efficiently via FFT's; and (2) the Levinson recursion used for solving the normal equation, the most computationally intensive module with a complexity αL^2, only needs to be performed once per block. In the following computational breakdown, we list the number of multiply-and-add operations as standard operations for benchmarking. The computation complexities of the BLS AEC are summarized as follows:

data windowing	$2N$
$\mathbf{R}_{xx}$ and $\mathbf{r}_{\mathbf{xz}}$	$2N \log N + N$
Levinson recursion	L^2
leakage integrator	$2L$
echo synthesis	NL
Total	$NL + 2N \log N + 3N + L^2 + 2L$
per sample	$L + 2 \log N + 3 + L^2/N + 2L/N$

For a typical AEC, we choose the block size N to be four times L such that enough data samples are used to load the autocorrelation matrix $\mathbf{R}_{xx}$ and the crosscorrelation vector $\mathbf{r}_{xz}$. The resultant complexity on a per sample basis is then roughly $5/4L$ for a relatively long L. Compared with the popular LMS-based AEC which has a complexity of $2L$, this is favorable. Other optional variations of echo cancelers such as dual path cancellation [5], one for updating the filter coefficients and another for cancelling the echoes, can be implemented. The additional computations for such features increase the computational requirement by the same factor for both the LMS-based and the proposed BLS-based AECs.

5. Delay in BLS AEC

Due to its block processing nature, the BLS-based algorithm has an intrinsic delay of one block in addition to other possible processing delays. A delay of one block may not be acceptable for applications that require low delay, such as full-duplex audio teleconferencing. This delay, despite its intrinsic nature, can be eliminated or circumvented at the price of little or no degradation of cancellation performance.

Our proposed solution is: use the AEC filter coefficients derived from the previous block to cancel the echoes in the current block starting from the first sample. This strategy is quite viable when the following two conditions are met: (1) the echo return path of the previous block is not much different from that of the current block, i.e., the echo return path is quasi-stationary or changes very slowly; and (2) the BLS AEC converges to its optimal solution with a block of data samples. For most applications, the first condition is generally satisfied and the second condition is true for the BLS AEC since the solution of the normal equations guarantees the minimum error squares.

In the experimental results section, we confirm that these two conditions are met in a live acoustic environment. Little or no degradation of cancellation performance is observed.

6. Experimental Results

BLS AEC Performance: WGN Case

The proposed BLS AEC is first evaluated using two 15 sec. of segments white Gaussian noise (WGN) samples, both as an excitation signal for an echo return path (measured from a real room) and as observation noise. Both 15 sec. noise processes are generated independently and twenty such pairs of WGN processes are used in the computer simulation of a fixed echo return path. The cancellation performance of 20 simulation runs are averaged and depicted in Figure 1. The statistical independence of WGN signal, most favorable for LMS convergence, has no effect on the convergence behavior of the proposed BLS AEC. The whitening process of the autocorrelation matrix inverse, $\mathbf{R}_{xx}^{-1}$, decorrelates the excitation signal, and the resultant LS solution therefore converges at the same rate, regardless of the correlation properties of the input excitation. For the case of 20dB SNR and the case of no noise, the BLS reached a decent cancellation performance, i.e., 20.5 dB and 22.5 dB, respectively, at the first block. A 4096 sample block size and 1001 tap BLS filter were adopted in the simulation. Starting with good cancellation at the first block, the BLS improves as more data samples become available.

BLS AEC Performance: Real Speech Case

Using the same measured room impulse response as in the previous experiment but replacing the excitation by 20 seconds of prerecorded female speech, we evaluated the BLS performance with and without additive (at 20 dB SNR) WGN measurement noise. Again, a 4096 sample block and 1001 tap filter were used in this

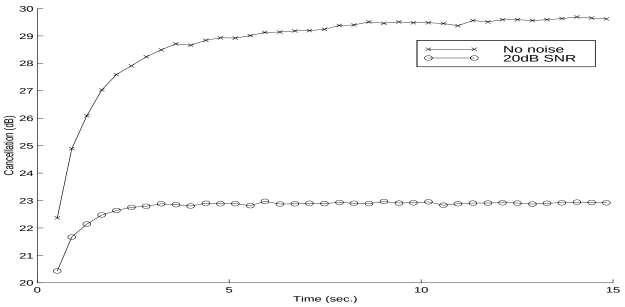

Figure 1: BLS AEC cancellation performance for WGN

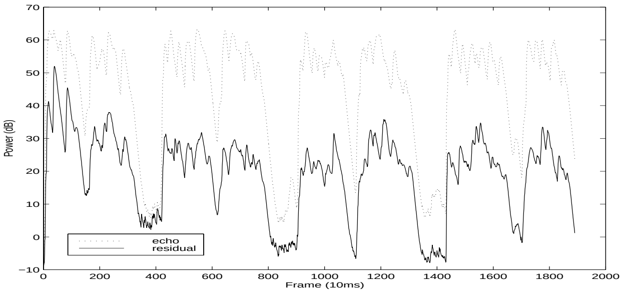

Figure 2: BLS AEC performance for speech input, no noise

experiment. The result is depicted in Figure 2 (no noise) and in Figure 3 (20dB SNR) which plot residual power, estimated at a frame rate of 100 frames/sec, along with measured echo power. Due to the nonstationary nature of speech signals and the leakage integrated auto- and cross-correlation estimates in the BLS, the proposed BLS-based AEC achieved 10-15 dB of cancellation performance in the first block, or first 50 frames, and it continuously improved its cancellation performance until it reached around 30 dB for the no noise case or the background WGN level for the 20 dB SNR case.

Zero Delay BLS AEC Performance

The intrinsic delay of BLS is circumvented here by using echo canceler filter coefficients derived from the previous block to cancel the echoes in the current block. The performance of this technique is compared in the Figure 4. The speech samples used were recorded in a large room with a segment of near-end speech. The echo residuals of the proposed BLS canceler are plotted in the figure, with and without the block delay, together with the microphone input power before cancellation. The fact that the two curves of the echo residual power, with and without delay, are virtually on top of each other demonstrates the validity of our proposed zero-delay solution. The near-end speech signal is remarkably prominent after cancellation, with or without delay.

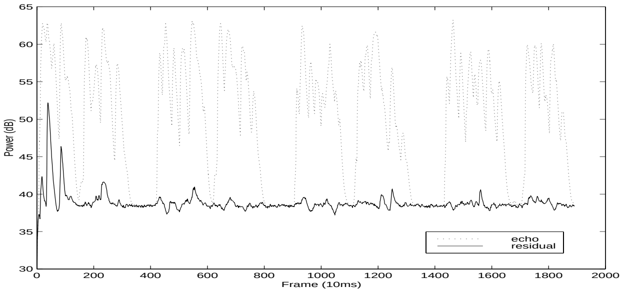

Figure 3: BLS AEC performance for speech input, 20dB SNR

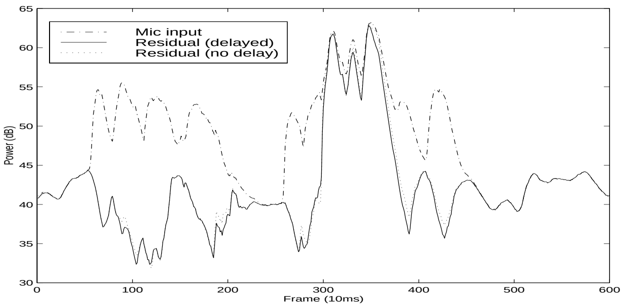

Figure 4: BLS AEC performance with/without delay

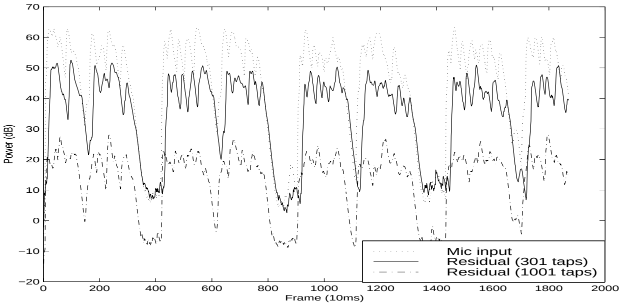

Figure 5: Filter length effects on AEC performance, no noise

The Effects of Filter Length

If the filter length of an AEC is much shorter than the actual room response, a “misalignment” [1], or insufficient coverage of the full time-span of the echo path, can result in inferior cancellation performance. To study misalignment in the presence of measurement noise we created two sets of synthetic echo signals (real speech signals filtered with a 1001 tap pre-measured room response), with and without additive WGN at 20dB SNR. Experiments were conducted with BLS filters of two different lengths: 301 taps and 1001 taps.

The steady-state cancellation performance is depicted in Figures 5 (without noise) and 6 (WGN measurement noise at 20 dB SNR). It is observed from Figure 5 that even in a steady state a BLS AEC with underspecified taps (301) can only deliver limited cancellation which is much inferior to the performance of a BLS AEC with more taps (1001). It is also interesting to observe that while the echo residual output of the BLS AEC of 1001 taps converges to the background noise levels as shown in Figure 6 at a SNR of 20dB, the large modeling errors caused by the insufficient time-span of a short BLS filter with only 301 taps are too high to settle at the background noise levels.

Near-end Speech Detection Performance

The proposed echo-residual power based near-end speech detector is tested both in a hands-free, real-time, human-machine dialogue system and off-line using a recorded speech database where near-

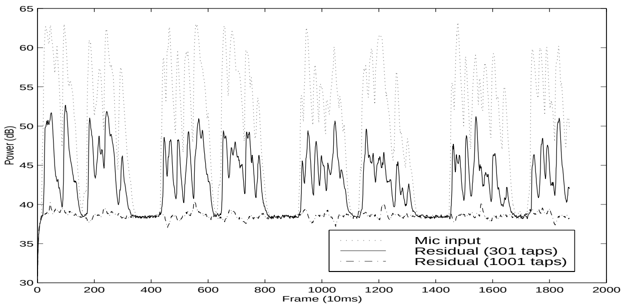

Figure 6: Filter length effects on AEC performance, 20dB SNR

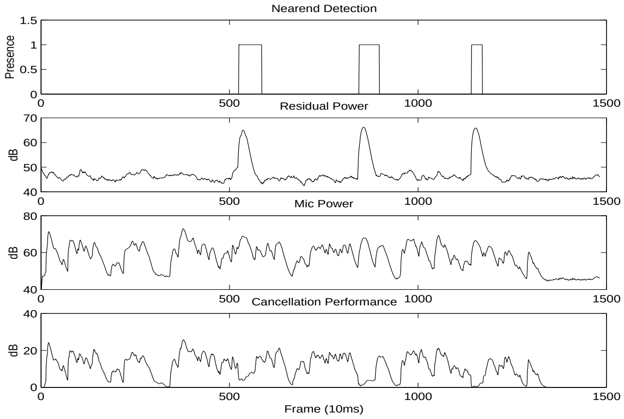

Figure 7: Near-end speech detection performance

end speech is present with far-end echoes. In both cases, near-end speech signals have been detected reliably. Systematic evaluation of the near-end speech detection performance is under current investigation and the results will be reported elsewhere. Figure 7 illustrates robust performance of the new near-end speech detector, displaying 4 traces of critical detector information from a live experiment. The 4 panels, from top to bottom in the figure, are the near-end detection signal, smoothed power of echo residuals, smoothed power of microphone input (echo and near-end speech), and echo cancellation performance. The plot makes obvious the difficulty of using only the microphone input to detect near-end speech. However the smoothed power of the residual signals, being essentially free from any far-end echo, show a clear advantage for near-end speech detection over other signals.

To further illustrate the high performance of the new echo canceler and near-end speech detector, both the microphone input and the echo cancelled output around a near-end speech segment are displayed in Figure 8 in both waveforms and narrowband spectrograms. Before BLS cancellation, the far-end echo, due to a long echo response, smears the corresponding spectrogram. When this echo mixes with the near-end speech, near-end speech detection becomes very difficult. After cancellation, while almost all far-end echoes disappear, the near-end speech remains intact and becomes distinctively prominent against the low background noise level. The cancelled output contains almost no audible far-end speech and the near-end phrase “sports results” becomes audibly crisp and clear, including the weak fricatives at both ends.

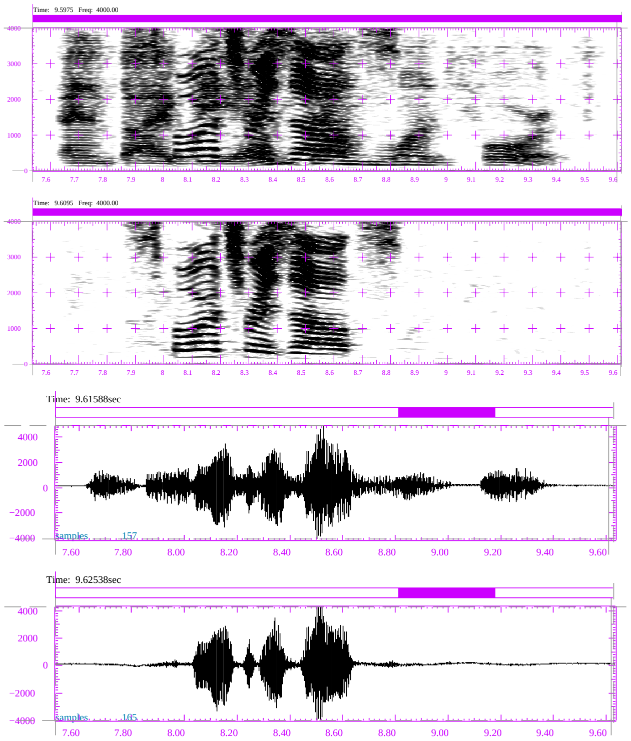

Figure 8: Waveforms and narrowband spectrograms before and after echo cancellation

7. Conclusion

We have proposed a new approach to echo cancellation in a room, and have demonstrated its practical advantages in terms of computational requirements and cancellation performance. The new BLS echo canceler is currently in use in a hands-free Human-Machine Dialog demo. This demo employs automatic speech recognition and allows users to barge in naturally and begin speaking during system prompts. The barge-in nature of the interaction requires a significant attenuation of room echo so that the recognizer will not falsely trigger on system output. The current BLS AEC implementation (using a 1001 tap cancellation filter) takes 50% of the CPU cycles on an SGI O2, and the entire dialog demo, including ASR, runs comfortably on the same machine.

8. REFERENCES

1. M. M. Sondhi, “An Adaptive Echo Canceler,” *The Bell System Tech. Journal*, Vol. 46, pp. 497-510, Mar. 1967.
2. D. Duttweiler, “A Twelve-channel digital echo canceler,” *IEEE Trans. Commun.*, Vol. 26, pp. 647-653, May. 1978.
3. D. L. Duttweiler, “Proportional Normalized Least Mean Squares Adaptation in Echo Canceler,” Bell Labs Technical Memorandum, October, 1996.
4. S. L. Gay,“PNLMS++ for Network Echo Cancellation,” Bell Labs Technical Memorandum, May, 1997.
5. K. Ochiai, T. Araseki and T. Ogihara, “Echo Canceler with Two Echo Path Models,” *IEEE Trans. on Commun.*, Vol. 25, pp. 589-95, June 1977.
6. J. Benesty, D. R. Morgan and J. H. Cho, “A New Class of Doubletalk Detectors Based on Crosscorrelation,” submitted for publication.