Published in *Proceedings of International Workshop on Acoustic Echo and Noise Control*, Pocono Manor, PA, pp. 40-43, September 1999.

STEREOPHONIC ACOUSTIC ECHO CANCELLATION USING NLMS WITH ORTHOGONAL CORRECTION FACTORS

Sundar G. Sankaran and A. A. (Louis) Beex
Systems Group - DSP Research Laboratory
The Bradley Department of Electrical and Computer Engineering
Virginia Tech, Blacksburg VA 24061-0111

ABSTRACT

A stereophonic echo canceler is proposed based on the Normalized LMS algorithm with orthogonal correction factors (NLMS-OCF). The echo canceler is modeled using a two-input single-output finite-impulse-response (FIR) structure. NLMS-OCF updates the echo canceler coefficients based on multiple input vectors, while NLMS adapts the coefficients based on a single input vector. The proposed algorithm is simulated in MATLAB with two different types of source signals in the far-end room, namely white noise and USASI noise. Simulation results indicate that the NLMS-OCF algorithm produces faster convergence than the well-known stereo projection algorithm, in terms of faster improvement in echo-return loss as well as faster reduction in impulse response misalignment.

1. INTRODUCTION

Echo in stereo-teleconferencing systems is undesirable but inevitable. Echo cancelers are used to mitigate the echo. Stereophonic echo cancelers suffer from two problems that are generally absent in mono echo cancelers. Firstly, the adaptive filter might misconverge due to the correlation between the left and right channel signals [1,4]. Secondly, again due to the correlation between the stereo signals, simple adaptation algorithms such as Least Mean Squares (LMS) and Normalized LMS (NLMS) will converge slowly. Several solutions have been proposed to ameliorate the performance of stereo echo cancelers. Most of these attempt to decorrelate the left and right channel signals by using some pre-processing technique [2,7,8]. These techniques essentially distort the signals, which is undesirable. Furthermore, the achievable performance improvement might be limited by the distortion inaudibility restrictions.

In this paper, we propose an algorithmic solution, which does not introduce any distortion, to surmount the difficulties encountered in typical stereophonic echo cancelers. The proposed strategy involves using the NLMS algorithm with Orthogonal Correction Factors (NLMS-OCF) to adapt the echo canceler, which is modeled as a two-input single-output FIR filter. The NLMS-OCF algorithm [3] updates the echo canceler weight estimates based on multiple input signal vectors, while NLMS updates the estimates based on a single input vector. The well-known stereo projection algorithm [4,6] can be interpreted as a special case of NLMS-OCF. NLMS-OCF provides flexibility in choosing the vectors used for adaptation. This flexibility can be exploited to reduce the misconvergence of echo cancelers and to improve their convergence speed.

The NLMS-OCF adaptation algorithm is presented in Section 2. Sections 3 and 4 describe how NLMS-OCF can be used to mitigate the drawbacks of conventional algorithms such as NLMS in the stereophonic echo cancellation application. In Section 5, we demonstrate the superiority of NLMS-OCF through simulation results. Section 6 provides the conclusion.

2. NLMS WITH OCF

The NLMS-OCF algorithm, as mentioned earlier, adapts the coefficients of the echo canceler based on multiple input vectors. The stereo echo canceler configuration is shown in Figure 1. To avoid clutter, in Figure 1, we show the echo paths corresponding to only one of the two stereo channels. In reality, similar paths exist in the other channel as well.

Let us denote the impulse responses from the speech source in the far-end room to the right and left microphones as $g_{r,n}$ and $g_{l,n}$, respectively. The impulse responses of the echo-paths from the left and right speakers to the left microphone in the near-end room are assumed to be $h_{ll,n}$ and $h_{lr,n}$, respectively. Let $d_{l,n}$ be the total echo received by the left microphone. The adaptive echo canceler (AEC), which is modeled using FIR filters, generates an estimate $\hat{d}_{l,n}$ for the echo, which is subtracted from the true echo to form the error signal $e_{l,n}$. The left-channel echo canceler weight vector $\hat{\mathbf{h}}_{l,n} = \left[\hat{\mathbf{h}}_{ll,n}^T \quad \hat{\mathbf{h}}_{lr,n}^T\right]^T$ is adapted with the objective of minimizing the mean squared left-channel residual echo, $E\left(e_{l,n}^2\right)$, in the absence of near-end speech. The dimensions of $\hat{\mathbf{h}}_{ll,n}$ and $\hat{\mathbf{h}}_{lr,n}$ are assumed to be N_l and N_r, respectively. That is, N_l and N_r are the estimated lengths of the left and right echo paths, respectively.

We propose using the NLMS-OCF algorithm to adapt the echo canceler weight-vector. The adaptation equation of NLMS-OCF is as follows [3]:

$$\hat{\mathbf{h}}_{l,n+1} = \hat{\mathbf{h}}_{l,n} + \mu_{l,0}\mathbf{x}_n + \mu_{l,1}\mathbf{x}_n^1 + \ldots + \mu_{l,M}\mathbf{x}_n^M \qquad (1)$$

where $M+1$ is the number of input vectors used for adaptation, $\mathbf{x}_n$ is the input vector at the nth instant given by

$$\mathbf{x}_n = \left[x_{l,n} \;\; x_{l,n-1} \;\ldots\; x_{l,n-N_l+1} \;\vdots\; x_{r,n} \;\; x_{r,n-1} \;\cdots\; x_{r,n-N_r+1}\right]^T \qquad (2)$$

while $\mathbf{x}_n^k$, for $k = 1,2,\ldots,M$, is the component of $\mathbf{x}_{n-kD}$ that is orthogonal to $\mathbf{x}_n, \mathbf{x}_{n-D}, \mathbf{x}_{n-2D}, \ldots, \mathbf{x}_{n-(k-1)D}$, and $\mu_{l,k}$, for $k = 0,1,\ldots,M$ are chosen as in (3).

$$\mu_{l,k} = \begin{cases} \dfrac{\bar{\mu} e_{l,n}}{\mathbf{x}_n^H \mathbf{x}_n} & \text{for } k = 0, \text{ if } \left\|\mathbf{x}_n\right\| \neq 0 \\ \dfrac{\bar{\mu} e_{l,n}^k}{\mathbf{x}_n^{k^T} \mathbf{x}_n^k} & \text{for } k = 1,2,\ldots,M, \text{ if } \left\|\mathbf{x}_n^k\right\| \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Published in *Proceedings of International Workshop on Acoustic Echo and Noise Control*, Pocono Manor, PA, pp. 40-43, September 1999.

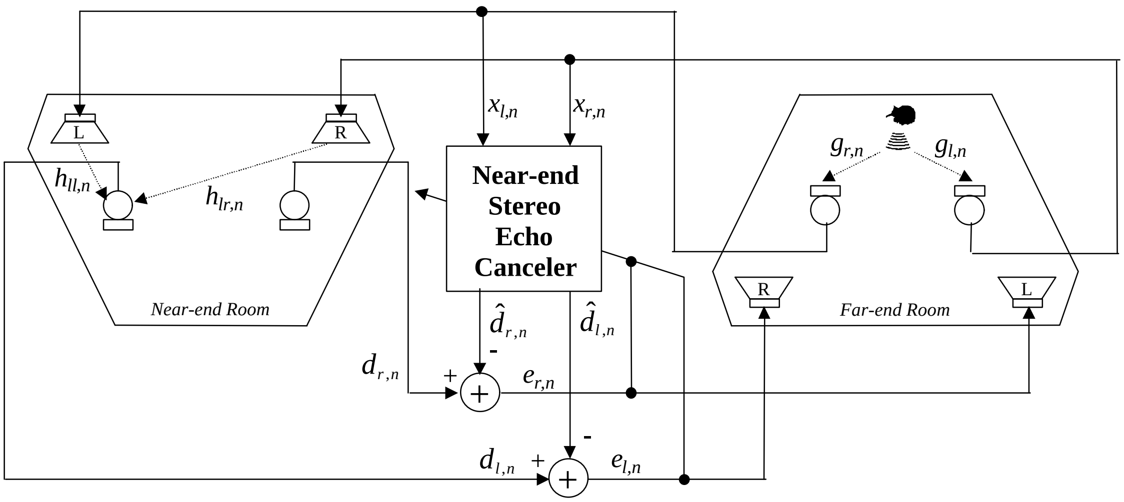


Figure 1. Stereophonic Echo Canceler Configuration.

where

$$e_{l,n} = d_{l,n} - \hat{\mathbf{h}}_{l,n}^{T}\mathbf{x}_n \ ,$$

$$e_{l,n}^{k} = d_{l,n-kD} - \hat{\mathbf{h}}_{l,n}^{kT}\mathbf{x}_{n-kD}, \ \ for \ k = 1,2,...,M, \ and \qquad (4)$$

$$\hat{\mathbf{h}}_{l,n}^{k} = \hat{\mathbf{h}}_{l,n} + \mu_{l,0}\mathbf{x}_n + \mu_{l,1}\mathbf{x}_n^{1} + ... + \mu_{l,k-1}\mathbf{x}_n^{k-1}.$$

The constant $\bar{\mu}$ is usually referred to as the step size.

Similarly the right-channel echo canceler weights $\hat{\mathbf{h}}_{r,n} = \left[\hat{\mathbf{h}}_{rl,n}^{T} \quad \hat{\mathbf{h}}_{rr,n}^{T}\right]^{T}$ are estimated so that the corresponding mean squared residual echo, $E\left(e_{r,n}^{2}\right)$, is minimized.

The NLMS-OCF algorithm with $M = 0$ results in the widely used NLMS algorithm. The well-known affine projection algorithm (APA) corresponds to the special case of NLMS-OCF with $D = 1$. Thus, NLMS-OCF is more general than APA.

3. NON-UNIQUENESS PROBLEM

A serious problem encountered in stereophonic echo cancelers is that the echo canceler coefficients do not converge to the true impulse response of the echo path. Due to the cross-correlation between the left and right channel signals, the weight estimate that minimizes the error between echo d_n and estimated echo $\hat{d}_n$ is not unique. In fact there is an affine space of equally good minimizing solutions, which is determined by the cross-correlation between the left and right channel signals, and the weight estimate converges to the point in this affine space that is nearest to the initial guess. Hence, the estimated weights need not necessarily match the true weights. If there is any change in the far-end room, the correlation between the left and right signal changes, which in turn leads to a change in the minimizing-solution space. Consequently, the adaptation algorithm readapts the parameters not only when there is a change not only in the near-end room, but also when there is a change in the far-end room. However, the estimation error between the true and estimated weights reduces with every variation in the cross-correlation between the stereo signals. The stereo projection algorithm, which is equivalent to the affine projection algorithm (APA), is used to emphasize the variations in the cross-correlation by using a block of input vectors for adaptation, thereby accelerating the convergence of the echo canceler weights. When there is a change in the cross-correlation, the input vector block used for adaptation consists of input signals with different cross-correlations. This accelerates the convergence of the estimated weights to the true weights. The variations in cross-correlation can be further emphasized by using the NLMS-OCF algorithm as explained below.

Consider the scenario where one talker in the far-end room stops talking and another starts talking. This results in an abrupt (step) change in the far-end room impulse responses $g_{r,n}$ and $g_{l,n}$, say at time instant k. Suppose NLMS-OCF is used for adaptation, with M orthogonal correction factors based on input vectors that are spaced D apart, then the input vectors used for adaptation from $n = k$ to $n = k + MD$ consist of input signals that are correlated differently. Thus a larger D helps to emphasize the variations. The stereo projection algorithm, which restricts D to unity, does not provide as much flexibility.

Consider another scenario where a time-varying all-pass filter is used to artificially introduce variations in the cross-correlation between the left and right signals [7]. The time variation of the all-pass filter has to be slow so that the changes are inaudible to the listener. However, for the stereo projection algorithm to successfully exploit the variations in cross-correlation between the left and right channels, there should be a significant change in the all-pass filter characteristics within the time window of data used for adaptation. By using NLMS-OCF with a D larger than unity, the time window of data used for adaptation is expanded. Hence, by using NLMS-OCF with a larger D the variations of the all-pass filter can be slowed down without compromising on the convergence speed of the echo canceler.

In practice, the cross-correlation between the stereo signals varies slightly even when the talker does not move the head or body while speaking [8]. Even though such variations are small, they can be effectively emphasized by using NLMS-OCF since by choosing a large D we can adapt using input signals that are farther apart in time (hence with larger variation in cross-correlation).

Published in *Proceedings of International Workshop on Acoustic Echo and Noise Control*, Pocono Manor, PA, pp. 40-43, September 1999.

4. CONVERGENCE RATE

The rate of convergence of the echo canceler estimates depends on the eigenvalue spread of the input vector (also referred to as the information vector) autocorrelation matrix. An important distinction between the mono echo canceler and stereo echo canceler is that the eigenvalue spread depends only on the input signal in the mono case while the spread is significantly larger in the stereo case, due to the correlation between the left and right signals, irrespective of the input signal. This results in slow convergence of the stereo echo canceler coefficients. This problem can be mitigated by using NLMS-OCF. It has been observed that, unless the step size μ is "too small," a larger value of D results in faster convergence [9]. Thus, the flexibility provided by NLMS-OCF in selecting the input vectors used for adaptation can be exploited to accelerate the convergence of the echo canceler.

5. SIMULATION RESULTS

A stereo-echo canceler with its coefficients adapted using the NLMS-OCF algorithm is simulated in MATLAB. The near-end and far-end room left-channel impulse responses are modeled using 1024th order FIR filter estimates based on measurements from two actual rooms. Slight perturbations of the left-channel impulse responses are used to model the right-channel impulse responses. Two different types of signals, namely white noise and USASI noise, are used as source signals in the far-end room in these simulations. USASI noise is a white noise source filtered by a bandpass filter with cut-off frequencies at 100 Hz and 320 Hz [11] and with 6 dB/octave roll-off rate. This simulates the long-term average spectrum of a typical speech signal. We assume a sampling rate of 8 kHz. Note that this does not mean that the left and right channel signals are of white or USASI type, due to the coloration provided by the impulse responses $g_{\bullet,n}$. All simulations use soft initialization; that is, the delay line of the adaptive filter is filled with true data values before the simulation is started. All the adaptive filter coefficients are initialized to zero. Since the adaptive filters are usually adapted under far-end single talk condition only, we assume this condition for our simulations. We further assume that there is only one far-end talker at any time.

The simulated echo canceler uses a 512th order FIR filter to model each of the left and right channels of the near-end room. Since the model order is less than the true order of the channels, there will be an error due to undermodeling. The echo canceler coefficients are adapted using the NLMS-OCF algorithm. The step-size $\bar{\mu}$ is chosen to be unity. The number of vectors used for adaptation is fixed at 32. Different delay values are chosen, viz. D = 1, 16, and 32, to select the input vectors used for adaptation.

Figure 2 shows the learning curves, i.e. the plots of echo return loss (ERL) in dB versus iteration number, corresponding to different delays. The results shown are ensemble average of 25 independent realizations. The echo return loss is defined as the ratio of residual echo energy to original echo energy. Note that the ERL is limited to nearly –70 dB due to the modeling error.

The misalignment curve is shown in Figure 3. Here also we show the ensemble average of 25 independent realizations. The misalignment ξ_n is defined as

$$\xi_n = \frac{\left\| h_n - \hat{h}_n \right\|}{\left\| h_n \right\|} \tag{5}$$

where $\|\cdot\|$ is the standard Euclidean norm.

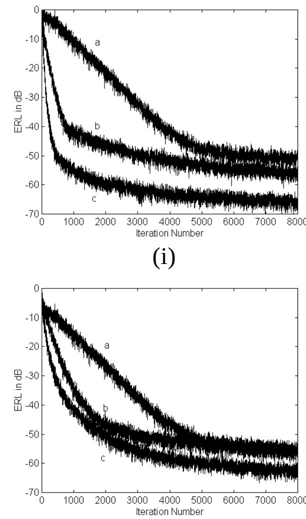
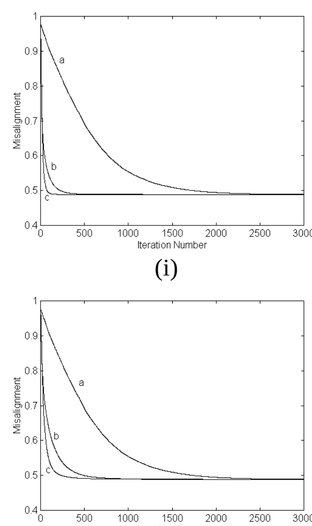


Figure 2. Learning Curves of NLMS-OCF corresponding to (i) white input and (ii) USASI noise input, for different values of delay *D*: (a) *D* = 1, (b) *D* = 16, and (c) *D* = 32.

The steady-state misalignment, shown in Figure 3, is large (nearly 0.5), even though the steady-state ERL is more than -50 dB, due to the non-uniqueness of the minimizing solution. Consequently, if there is any change in the far-end room impulse responses, the residual echo level temporarily increases, even if there is no change in the near-end room.



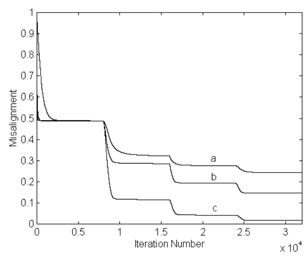
(ii)

Figure 3. Misalignment of echo canceler coefficients with (i) white input and (ii) USASI noise input, for different values of delay *D*: (a) *D* = 1, (b) *D* = 16, and (c) *D* = 32.

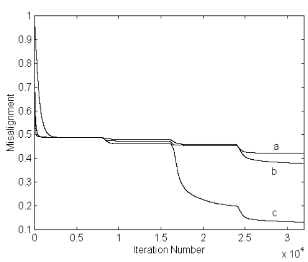
Published in *Proceedings of International Workshop on Acoustic Echo and Noise Control*, Pocono Manor, PA, pp. 40-43, September 1999.

We see that both residual echo and misalignment decrease faster for larger values of *D*. NLMS-OCF with *D* = 1 corresponds to the stereo projection algorithm. Hence, the flexibility of choosing input vector delays in the NLMS-OCF adaptation is exploited here to achieve faster convergence than that obtained with the stereo projection algorithm (the *D* = 1 case).

Now, we present results that show that the NLMS-OCF algorithm can emphasize variations in cross-correlation better than the stereo projection algorithm, thereby converging faster to the true solution for the echo channel weights. All the parameters are chosen as before. We simulate the condition where one far-end talker stops talking and another far-end talker starts talking. This results in an abrupt step change in the channel impulse responses in the far-end room, thereby introducing variations in the near-end received signal cross-correlation. This change is assumed to occur at *n* = 8000, 16000, and 24000. The simulation results (ensemble average of 10 independent realizations) are shown in Figure 4. We observe that NLMS-OCF with larger delays converges faster than the stereo projection algorithm, which uses unit delay by default.



(i)



(ii)

Figure 4. Misalignment of echo-canceler with step changes in cross-correlation of stereo signals with (i) white noise input and (ii) USASI noise input, for different values of delay *D*: (a) *D* = 1, (b) *D* = 16, and (c) *D* = 32.

Remarks: While increasing *D* results in performance improvement, there are some drawbacks in using a larger *D*. Firstly, a larger D necessitates storing more of the past data and hence requires more memory. Secondly, using a very large *D* can affect the near-end tracking performance of the echo canceler. If the time window of data used for adaptation is so long that the near-end room characteristics change significantly within that time, then the weight estimates can never track the weights correctly. This problem can be mitigated by using an exponential weighting on the orthogonal correction factors.

The NLMS-OCF algorithm as presented in Section 2 has a complexity of $O(NM^2)$. A fast version of NLMS-OCF with complexity of $O(NM)$ was derived [9]. The fast NLMS-OCF algorithm uses a lattice structure to compute orthogonal correction factors and avoids explicit computation of any matrix inverses. On the other hand, the fast stereo projection algorithm [5] computes a matrix inverse using a sliding-window version of the fast transversal filter, which is known to have numerical problems [10].

6. SUMMARY

We presented a stereo echo canceler adapted using the NLMS-OCF algorithm. The NLMS-OCF algorithm is shown to produce faster convergence than the well-known stereo projection algorithm, in terms of faster improvement in echo-return loss as well as faster reduction in impulse response misalignment.

REFERENCES

1. M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation--An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, Vol. 2, No. 5, pp. 148-151, August 1995.
2. Y. Joncour and A. Sugiyama, "A Stereo Echo Canceler with Pre-processing for Correct Echo-Path Identification," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Seattle, WA, pp. 3677-3680, May 1998.
3. S. G. Sankaran and A. A. (Louis) Beex, "Normalized LMS Algorithm with Orthogonal Correction Factors," *Proceedings of the Thirty-First Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, pp. 1670-1673, November 1997.
4. S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, pp. 3059-3062, May 1995.
5. F. Amand, J, Benesty, A. Gilloire, and Y. Grenier, "A Fast Two-Channel Projection Algorithm for Stereophonic Acoustic Echo Cancellation," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, pp. 949-952, May 1996.
6. J. Benesty, F. Amana, A. Gilloire, and Y. Grenier, "Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, pp. 3099-3102, May 1995.
7. M. Ali, “Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering for Signal Decorrelation,” *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Seattle, WA, pp. 3689-3692, May 1998.
8. S. Shimauchi, S. Makino, Y. Haneda, A. Nakagawa, S. Sakauchi, “A Stereo Echo Canceller Implemented Using a Stereo Shaker and a Duo-Filter Control System,” *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, AZ, pp. 857-860, March 1999.
9. S. G. Sankaran and A. A. (Louis) Beex, “Fast Generalized Affine Projection Algorithm,” Submitted to *International Journal of Adaptive Control and Signal Processing*, October 1998.
10. S. Haykin, *Adaptive Filter Theory*, Second Edition, Prentice Hall, Englewood Cliffs, NJ, 1991.
11. EIA Standard, *NRSC AM Preemphasis, Deemphasis, and Broadcast Audio Transmission Bandwidth Specifications*, EIA-549, Electronics Industries Association, July 1988.