# REDUCING INTER-CHANNEL COHERENCE IN STEREOPHONIC ACOUSTIC ECHO CANCELLATION USING PARTIAL UPDATE ADAPTIVE FILTERS

*Andy W. H. Khong and Patrick A. Naylor*

Department of Electrical and Electronic Engineering, Imperial College London, UK
E-mail: {andy.khong, p.naylor} @imperial.ac.uk

**ABSTRACT**

The use of partial update adaptive filters for stereophonic acoustic echo cancellation is investigated. It is proposed to employ subsampling of the tap-input vector, that is intrinsic to partial update schemes, to decorrelate the two tap-input vectors of the stereo adaptive structure thereby improving convergence performance. The problem has been structured as a joint optimization between inter-channel coherence and the $L_1$ norm of the inputs corresponding to the selected filter coefficients. We investigate the trade-off between improvement in convergence due to decorrelation of the two tap-input vectors and any degradation in convergence due to subsampling in the MMax-NLMS partial update scheme. The exclusive MMax approach is proposed which efficiently approximates the joint optimization of these factors. Simulation results show an improvement in the rate of convergence over standard NLMS approaches with a performance close to that of fast-RLS in combination with a non-linear preprocessor.

## 1. INTRODUCTION

A serious problem encountered in stereophonic acoustic echo cancellers (SAEC), shown in Fig. 1, is that the echo canceller coefficients do not necessarily converge to the true impulse responses of the echo path when full modelling of the transmission room exists (nonuniqueness problem) [1] [2]. In a practical case where the lengths of the filters are less than the impulse responses of the transmission room, the problem of nonuniqueness is ameliorated to some degree. However, due to the strong coherence between the two channel input signals, the convergence performance is poor and misalignment is a significant problem.

To overcome the misalignment problem, it is therefore desirable to reduce the coherence between the two input channel signals. This has led to several approaches to the problem that involve techniques to decorrelate the two input signals using, for example, non-linear processing [1][3][4]. Furthermore, the computational complexity of stereophonic echo cancellers is high because the number of taps can be large and also because the use of least-squares-based algorithms is often preferred in order to obtain sufficient levels of cancellation. Therefore, there exists a dual motivation to develop algorithms which have improved convergence performance due to reduction of inter-channel coherence whilst maintaining computational complexity to be as low as possible for practical reasons.

In partial update adaptive filtering, the tap-input vector is subsampled so that only a subset of filter taps is updated at each iteration [5] [6] [7]. The aim of this work is to investigate whether such subsampling can bring a reduction in the inter-channel coherence of the tap-input vectors that results in improved convergence.

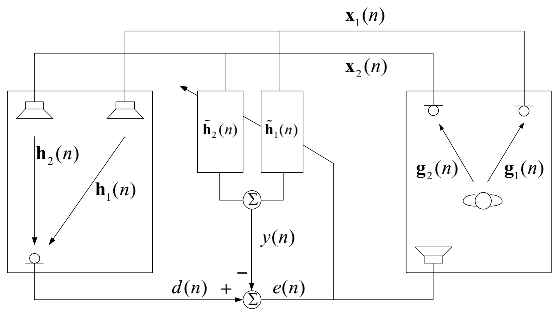


Figure 1: Stereophonic Acoustic Echo Cancellation (after [1])

The problem has been structured as a joint optimization of two scores - one describing the coherence between the tap-input vectors of the two channels and the other describing the 'closeness' of the tap-selection to the MMax criterion of the MMax-NLMS scheme [7]. In this context, the ideal tap-selection is therefore one which selects elements of the tap-input vectors such that the inter-channel coherence is minimized whilst maximizing their $L_1$ norm.

This paper is organized as follows: Section 2 examines the decorrelation effect due to tap-selection. We formulate and present the XMNL-NLMS algorithm in Section 3. Section 4 presents the simulation results while Section 5 concludes our work.

## 2. DECORRELATION USING TAP SELECTION

In the single channel MMax-NLMS algorithm, with filter length $L$, the MMax tap selection criterion selects only those taps corresponding to the $M$ largest magnitude tap-inputs for updating at each iteration. It has been shown that, under specified conditions, the rate of convergence of MMax-NLMS depends on $M$ whilst the same final misadjustment as the NLMS is achieved. This is explained in [8] in terms of the robustness of MMax-NLMS to choices of $0.5L \leq M \leq L$ and graceful degradation in performance for $M < 0.5L$.

We emphasize here that we are *not* employing the partial update concept to reduce complexity of the NLMS approach. Instead, we are considering the use of partial updating to reduce the coherence of the two tap-input vectors of the SAEC structure.

Figure 2 shows the effect of tap-selection on the mean coherence (averaged across frequency) between the two channels tap-input vectors which are highly correlated. For zero mean and unit variance white Gaussian noise (WGN) input and filter length $L = 256$, the mean coherence between the tap-input vectors is calculated from the $M$ selected tap-inputs

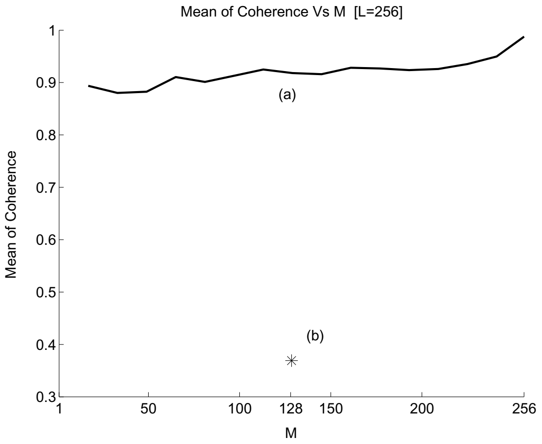


Figure 2: Effect of taps selection on inter-channel coherence for $L=256$ (a)MMax-NLMS (b) XM-NLMS

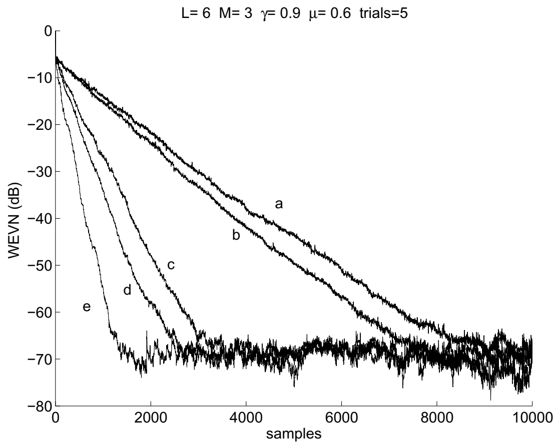


Figure 3: WEVN for (a)$w_m=1$, (b) NLMS, (c) $w_m$= 0.9, (d) 0.7, (e) 0.1. $L$=6, $M$=3, $\mu=0.6$, $\gamma=0.9$.

with the remaining $L-M$ tap-inputs set to zero. The mean coherence is then time averaged and plotted against $M$.

Figure 2(a) shows that the MMax-NLMS provides only a modest decorrelation effect with decreasing $M$. This is due to the high correlation between the two channel input vectors which causes MMax-NLMS to select corresponding weights in both filters for updating. This does not achieve our desired effect of decorrelating the signals. Figure 2(b) shows the effect of decorrelation when exclusive MMax tap-selection is employed. The reduction in mean coherence suggests that tap-selection based on exclusivity may be employed to decorrelate the two channel input vectors.

We now formulate the problem as the joint optimization of the MMax criterion and inter-channel decorrelation (due to tap-selection) by introducing an alternative tap-selection criterion controlled by two variables: magnitude weighting, $w_m$, to describe the 'closeness' of the tap selection to that of the MMax scheme and coherence weighting, $w_c=1-w_m$, to describe inter-channel coherence between the subsampled tap-input vectors respectively. A magnitude weighting of $w_m=1$ corresponds to selecting coefficients based on the MMax criterion only.

We first assume a standard FIR adaptive filter configuration and define $\mathbf{A}$ and $\mathbf{C}$ as square matrices containing elements $a_{ij}$=$|x\{\beta_{ij}\}|$ and $c_{ij}=coh\{\beta_{ij}\}$ respectively. We define $\{\beta_{ij}\}$ as a tap-selection set with $i$ and $j$ representing the different combinations of selecting $M$ out of $L$ coefficients in each of the two filters $(i,j=1,\ldots,{}^{L}C_M)$. The absolute sum of the selected tap-inputs in a particular combination $i$ and $j$ for the two channels is defined as the $L_1$ norm $a_{ij}$=$|x\{\beta_{ij}\}|$ while $coh\{\beta_{ij}\}$ is the coherence (averaged over frequency) of the two tap-input vectors with $L-M$ unselected inputs in each channel set to zero. Elements $a_{ij}$ and $c_{ij}$ are each associated with a cost such that the least cost is allocated to combinations having the maximum magnitude in $\mathbf{A}$ and the minimum coherence in $\mathbf{C}$. A total cost matrix is then derived by summing matrices $\mathbf{A}$ and $\mathbf{C}$ after they are multiplied by $w_m$ and $w_c$ respectively. Defining $u$ and $v$ as the tap-indices of channels 1 and 2 respectively $(u,v=1,\ldots,L)$, the update equation incorporating tap-selection is then given by:

$$\tilde{\mathbf{h}}(n+1)=\tilde{\mathbf{h}}(n)+\mathbf{Q}(n)\mu\frac{\mathbf{x}(n)e(n)}{\|\mathbf{x}(n)\|^2}\quad . \tag{1}$$

Here $\mathbf{Q}(n)=diag\{[\mathbf{q}_1(n)\ \mathbf{q}_2(n)]\}$ such that at each time iteration $n$,

$$\{q_1(u),q_2(v)\}=\begin{cases}1 & \text{if } u,v\in\{\beta_{min}\}\\ 0 & \text{otherwise}\end{cases}$$

where the dependence of $q_1$, $q_2$, $x_1$ and $x_2$ on time $n$ has been temporarily omitted for clarity of notation. The tap-selection set for minimum cost in matrix $w_m\mathbf{A}+w_c\mathbf{C}$ is defined as $\{\beta_{min}\}$ while $\tilde{\mathbf{h}}(n)=[\tilde{\mathbf{h}}_1^T(n)\ \tilde{\mathbf{h}}_2^T(n)]^T$ and $\tilde{\mathbf{x}}(n)=[\tilde{\mathbf{x}}_1^T(n)\ \tilde{\mathbf{x}}_2^T(n)]^T$ represent the concatenated filter coefficient vector and tap-input vector of the filters respectively. Vector transposition is represented by the notation $^T$.

Figure 3 shows simulation results for the weight error vector norm (WEVN) defined as

$$WEVN=\frac{\|\mathbf{h}-\tilde{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2} \tag{2}$$

for different values of magnitude weighting $(w_m=0.1,0.7,0.9,1.0)$. In this simulation, the two channel inputs are zero mean and unit variance WGN sequences. The coherence between the two channel inputs is controlled by $\gamma$ $(0\leq\gamma\leq1)$, where $\gamma$=0 represents independent signals and $\gamma$=1 implies the two channel inputs being identical to one another. In this simulation, $\gamma$=0.9 is used to reflect the high coherence of the two channel tap-input vectors in practical applications. The adaptive filters have 6 taps per channel and for every iteration, 3 taps are updated $(L=6,M=3)$. For clarity, WEVN for only one of the two channels is plotted for each case of $w_m$.

The simulation result shows that $w_m$=1 coincides with MMax-NLMS where performance is close to that of the full update NLMS as expected. The highest convergence rate can be seen when $w_m$=0.1 ($w_c$=0.9) where there is a high emphasis on selecting the exclusive set of weights for updating (i.e. none of the same coefficient indices can be selected in both channels' tap-input vectors). This is similar to finding (out of ${}^{L}C_M$ combinations) the exclusive set of weights for the two filters such that the $L_1$ norm of the active tap-input vector is maximized.

## 3. XMNL-NLMS ALGORITHM

In this section, we develop an efficient algorithm for tap-selection so as to exploit the technique for coherence reduction demonstrated in Section 2.

Consider as an example an SAEC system with channels $k = 1,2$, adaptive filters each of length $L = 4$ and tap-input vectors $\mathbf{x}'_k(n) = [x'_{k,1}, x'_{k,2}, x'_{k,3}, x'_{k,4}]^T$, where the input signals $\{x_k(n)\}$ have been pre-processed by a non-linear processor block [1] to give $\{x'_k(n)\}$. Let $\mathbf{p}(n)$ be the magnitude vector difference such that $\mathbf{p}(n) = |\mathbf{x}'_1(n)| - |\mathbf{x}'_2(n)|$,

$$\begin{pmatrix} p_1(n) \\ p_2(n) \\ p_3(n) \\ p_4(n) \end{pmatrix} = \begin{pmatrix} |x'_{1,1}(n)| \\ |x'_{1,2}(n)| \\ |x'_{1,3}(n)| \\ |x'_{1,4}(n)| \end{pmatrix} - \begin{pmatrix} |x'_{2,1}(n)| \\ |x'_{2,2}(n)| \\ |x'_{2,3}(n)| \\ |x'_{2,4}(n)| \end{pmatrix}. \quad (3)$$

Our objective is to select $M$ out of $L$ taps for updating for which the corresponding (subsampled) tap-input vectors have the maximum absolute sum, so as to approximate MMax-NLMS as closely as possible, but also have the minimum inter-channel coherence. Whereas in principle an exhaustive search of all possible tap selection sets could be made for small $L$, the XMNL-NLMS algorithm finds an efficient approximation to the optimum tap-selection by constraining the search to tap-selections that are exclusive between the two channels so as to minimize the inter-channel coherence. These exclusive sets can be pre-determined off-line for any $L$ and $M = 0.5L$. Within this constrained search space, the tap-selection with maximum absolute sum can then be found efficiently by sorting $\mathbf{p}(n)$. Consider an example of $p_3 > p_2 > p_1 > p_4$ for a particular time instance, since $p_3 + p_2 > p_1 + p_4$, it can be shown that

$$|x'_{1,3}| + |x'_{1,2}| + |x'_{2,1}| + |x'_{2,4}| > |x'_{1,1}| + |x'_{1,4}| + |x'_{2,2}| + |x'_{2,3}|. \quad (4)$$

Thus the tap-selection corresponding to inputs $x'_{1,3}$, $x'_{1,2}$, $x'_{2,1}$ and $x'_{2,4}$ is closest to the MMax-NLMS with the minimum coherence constraint satisfied by the exclusivity of the tap-selection. It is worthwhile to note that it is irrelevant to consider other tap-selection combinations since they have smaller magnitude sum. This approach allows us to eliminate $({}^{L}C_M - 1)$ possible combinations thus allowing efficient implemention of the algorithm. The XMNL-NLMS algorithm is shown in Table 1.

## 4. RESULTS

In the first part of the experiment, we compare the performance of XMNL-NLMS with that of NLMS and NLMS with non-linear pre-processor (NL-NLMS) [1] in a stereophonic system using WGN input sequence. The microphone signals were obtained by convolving the source with two impulse responses $\mathbf{g}_1$ and $\mathbf{g}_2$ of length $T = 256$. In this simulation, two microphones are placed one metre apart in the centre of both the transmission and receiving rooms, each 3x4x5 metres in size. The source is then positioned such that it is one metre away from each of the microphones in the transmission room. The desired response in the receiving room is obtained by summing two convolutions $\mathbf{h}_1^T\mathbf{x}'_1$ and $\mathbf{h}_2^T\mathbf{x}'_2$. The impulse responses $\mathbf{g}_1$, $\mathbf{g}_2$, $\tilde{\mathbf{h}}_1$ and $\tilde{\mathbf{h}}_2$ were generated using method of images [9] in all our experiments.

XMNL-NLMS Algorithm

$M = 0.5L$

| | | |
|---|---|---|
| input vector: | $\mathbf{x}'(n)$ | $= [\mathbf{x}'^T_1(n)\ \ \mathbf{x}'^T_2(n)]^T$ |
| filter coefficients: | $\tilde{\mathbf{h}}(n)$ | $= [\tilde{\mathbf{h}}^T_1(n)\ \ \tilde{\mathbf{h}}^T_2(n)]^T$ |
| filter output: | $y(n)$ | $= \tilde{\mathbf{h}}^T(n)\mathbf{x}'(n)$ |
| error: | $e(n)$ | $= d(n) - y(n)$ |
| difference vector: | $\mathbf{p}(n)$ | $= \lvert\mathbf{x}'^T_1(n)\rvert - \lvert\mathbf{x}'^T_2(n)\rvert$ |
| weight update: | $\tilde{\mathbf{h}}(n)$ | $= \tilde{\mathbf{h}}(n) + \mu\mathbf{Q}(n)\frac{\mathbf{x}'(n)e(n)}{\lvert\lvert\mathbf{x}'(n)\rvert\rvert^2}$ |
| selection matrix: | $\mathbf{Q}(n)$ | $= diag\{[\mathbf{q}_1(n)\ \ \mathbf{q}_2(n)]\}$ |

where $q_1(u) = \begin{cases} 1 & x_1(u) \text{ is one of the } M \text{ maxima of } \mathbf{p}(n) \\ 0 & \text{otherwise} \end{cases}$

$q_2(v) = \begin{cases} 1 & x_2(v) \text{ is one of the } M \text{ minima of } \mathbf{p}(n) \\ 0 & \text{otherwise} \end{cases}$

and the dependence of $q_1$, $q_2$, $x_1$ and $x_2$ on time $n$ has been temporarily omitted for clarity of notation.

Table 1: XMNL-NLMS Algorithm

The lengths of the filters have been chosen to be equal to that of $\mathbf{g}_1$ and $\mathbf{g}_2$ ($L = T$=256) as the purpose of this simulation is to show the effect of decorrelation introduced by XMNL-preprocessor and not due to the 'tail' effects of the transmission room impulse responses [1]. For XMNL-NLMS, $M$ is chosen to be 128. The lengths of the unknown system $\mathbf{h}_1$ and $\mathbf{h}_2$ are both $R = 256$. The signal-to-noise ratio of 20dB is obtained by the addition of uncorrelated noise to the desired response. In all our experiments, the non-linearity constant $\alpha = 0.5$ is used [1].

Figure 4 shows the WEVN for NLMS, XMNL-NLMS and NL-NLMS. We have used an adaptive step-size of $\mu = 0.4$. We see that NLMS converges to a poor misalignment because of the non-uniqueness problem. The convergence rate of the NL-NLMS can be seen to increase significantly due to the additional decorrelation property of the XM pre-processor. This means that a lower non-linearity constant can be used to acheive the same rate of convergence hence reducing non-linear distortion.

Figure 5 shows the comparison between the performance of the XMNL-NLMS and fast-recursive least squares algorithm with NL-preprocessor (NL-FRLS) [1]. The room impulse responses are of length $T = R = 1600$, $L$=800, $M$=400 and the input sequence is WGN as before. The adaptive step size of NL-NLMS and XMNL-NLMS are $\mu = 0.8$ while the forgetting factor for NL-FRLS is $\lambda = 1 - \frac{1}{10L}$ [10]. It can be seen that the performance of XMNL-NLMS exceeds that of NL-NLMS and is close to that of NL-FRLS algorithm in this room model example of realistic dimension.

In the last experiment, we have used a speech source as our excitation signal shown in Fig. 6. The impulse responses are of length $T = R = 800$, $L = 256$ and $M = 128$. As before, we notice that the performance of XMNL-NLMS is close to that of the NL-FRLS while exceeding that of NL-NLMS algorithm.

## 5. CONCLUSION

This paper has proposed the use of partial update tap-selection to decorrelate the tap-input vectors in stereophonic AEC. We have illustrated this concept and shown that tap-selection can indeed effectively decorrelate the signals. This

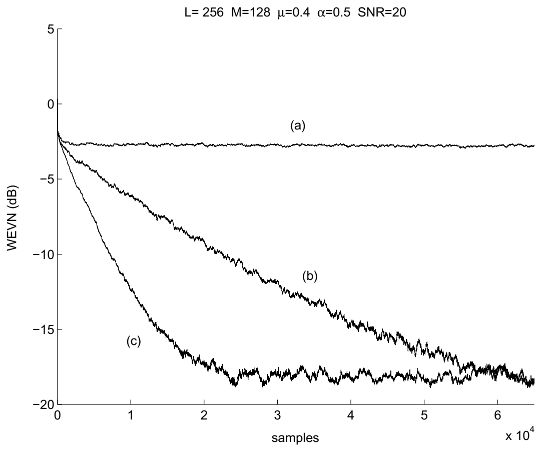


Figure 4: WEVN for WGN sequence (a)NLMS, (b)NL-NLMS and (c)XMNL-NLMS. ($L = T = R = 256$, $M$=128, $\alpha = 0.5$, $\mu = 0.4$)

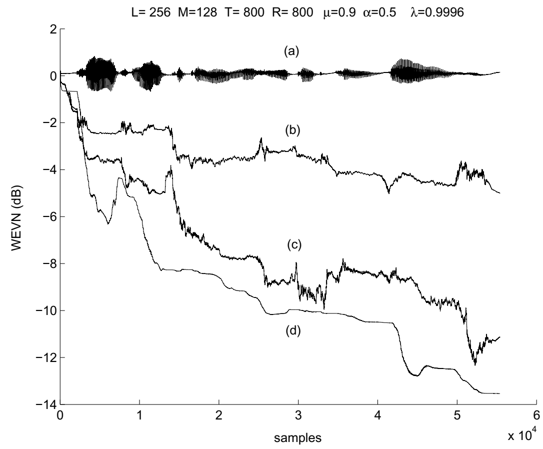


Figure 6: WEVN for (a)Speech Signal, (b)NL-NLMS, (c)XMNL-NLMS and (d)NL-FRLS. ($L$=256, $T = R = 800$, $M$=128, $\alpha = 0.5$, $\mu = 0.9, \lambda = 0.9996$)

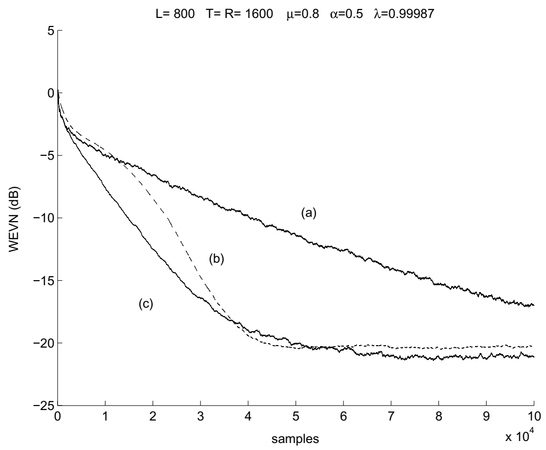


Figure 5: WEVN for (a)NL-NLMS, (b)NL-FRLS and (c)XMNL-NLMS. ($L$=800, $T = R = 1600$, $M$=400, $\alpha = 0.5$, $\mu = 0.8, \lambda = 0.99987$)

approach therefore provides a new technique for the solution of the non-uniqueness problem. A tap-selection method has been proposed employing joint optimization of the MMax criterion and the level of decorrelation of the two channels tap-input vectors. An efficient XMNL-NLMS algorithm has been formulated which approximates the optimum tap-selection using the exclusive MMax criterion. Simulation results have shown (for WGN and speech signals) significant improvement in performance compared to direct application of NLMS and NL-NLMS. The performance of XMNL-NLMS has been found to come close to one of the best existing approaches involving fast-RLS and a non-linear preprocessor. The XMNL-NLMS has the benefits of low complexity and the robustness inherent in NLMS-based algorithms compared to least squares approaches.

## REFERENCES


[1] J. Benesty, D. R. Morgan, and M. M. Sondhi, “A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation,” *IEEE Trans. Speech Audio Processing*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[2] S. Shimauchi and S. Makino, “Stereo projection echo canceller with true echo path estimation,” in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 5, 1995, pp. 3059–3060.

[3] S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, 2001.

[4] S. Emura, Y. Haneda, and S. Makino, “Enhanced frequency-domain adaptive algorithm for stereo echo cancellation,” in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 2, 2002, pp. 1901–1904.

[5] K. Dogancay and O. Tanrikulu, “Adaptive filtering algorithms with selective partial updates,” *IEEE Trans. Circuits Syst. II*, pp. 762–769, Aug. 2001.

[6] P. A. Naylor and W. Sherliker, “A short-sort M-max NLMS partial update adaptive filter with applications to echo cancellation,” in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 5, 2003, pp. 373–376.

[7] T. Aboulnasr and K. Mayyas, “Complexity reduction of the NLMS algorithm via selective coefficient update,” *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1421–1424, 1999.

[8] A. W. H. Khong and P. A. Naylor, “The use of partial update scheme to reduce inter-channel coherence in adaptive stereophonic acoustic echo cancellation,” in *International Workshop on Acoustic Echo and Noise Control*, 2003, pp. 59–62.

[9] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[10] C. Breining, P. Dreiseitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, “Acoustic echo control: An application of very-high-order adaptive filter,” *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, Jul. 1999.