

Selective-Tap Adaptive Algorithms in the Solution of the Nonuniqueness Problem for Stereophonic Acoustic Echo Cancellation

Andy W. H. Khong and Patrick A. Naylor, *Member, IEEE*

Abstract—We investigate stereophonic acoustic echo cancellation in which solutions for the system can be nonunique and propose the use of selective-tap adaptive filters to address this problem. The main concept is to employ tap selection to optimize jointly for minimum interchannel coherence and maximum L_2 -norm of the subselected tap-input vectors. The exclusive maximum (XM) tap-selection approach is proposed and applied to normalized least-mean squares (NLMS) and recursive least-squares (RLS) algorithm. We propose an approach for solving the nonuniqueness problem employing XM tap selection in combination with a nonlinear preprocessor. Simulation results show a significant improvement in convergence rate compared with existing techniques.

Index Terms—Adaptive filtering, selective-tap, stereophonic acoustic echo cancellation.

I. INTRODUCTION

STEREOPHONIC Acoustic Echo Cancellation (SAEC), as shown in Fig. 1, plays a crucial role in applications that include desktop conferencing and hands-free telephony [1], [2]. In SAEC, the solutions for the adaptive filters can be nonunique [1]. Defining L and W as the lengths of the adaptive filters and transmission room's impulse response, respectively, and \mathbf{R}_{xx} as the concatenated input autocorrelation matrix as in [3], two cases have been described for a noiseless system:

- case 1 : $L \geq W \Rightarrow \mathbf{R}_{xx}(n)$ is singular $\forall n$
 case 2 : $L < W \Rightarrow \mathbf{R}_{xx}(n)$ is ill conditioned

giving rise to nonuniqueness and poor convergence due to high interchannel coherence.

The use of nonlinear (NL) distortion has been proposed to solve this problem [1]. Other approaches include the use of spectrally shaped random noise [4] and leaky normalized least-mean squares (NLMS) [5] with the common aim of reducing interchannel correlation without affecting the speech quality and stereophonic perception.

Selective-tap schemes such as [6] update only a subset of taps at each iteration and were originally proposed to reduce computational complexity. Subsequent work using MMax-NLMS [7],

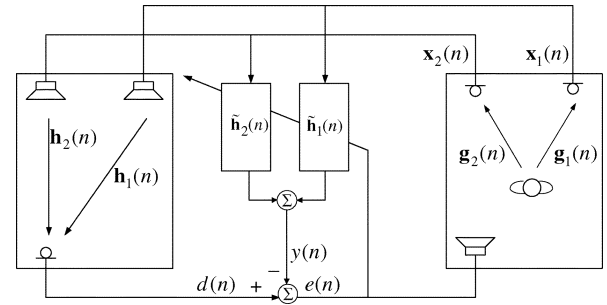


Fig. 1. Stereophonic acoustic echo cancellation system (after [1]).

[8] has shown an improvement in the convergence rate by updating taps corresponding to the M largest magnitude tap inputs. In this letter, we propose the use of an exclusive maximum (XM) tap-selection technique that addresses practical case 2. We further propose an approach for SAEC employing XM tap selection in combination with the NL preprocessor to overcome the nonuniqueness problem with improved convergence speed.

II. MMAX TAP-SELECTION CRITERION

In the single channel MMax-NLMS algorithm [7] for an adaptive filter of length L , only taps corresponding to the M largest magnitude tap inputs are selected for updating at each iteration. Defining the tap-selection matrix $\mathbf{Q}(n) = \text{diag}\{\mathbf{q}(n)\}$, $\mathbf{x}(n)$, and $\tilde{\mathbf{h}}(n)$ as the single channel tap-input vector and filter coefficients, respectively, the MMax-NLMS tap-update equation may be expressed as

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

$$q_i(n) = \begin{cases} 1, & |x_i(n)| \in \{M \text{ maxima of } |\mathbf{x}(n)|\} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $q_i(n)$ are the elements of $\mathbf{q}(n)$ for $i = 1, 2, \dots, L$, and the adaptive step size is μ . The error signal is given by $e(n) = d(n) - \tilde{\mathbf{h}}^T(n) \mathbf{x}(n)$.

We now examine the dependence of the convergence rate on the subselection by proposing the measure $\mathcal{M}(n)$ as the ratio of the squared L_2 -norm of the M selected tap inputs to that of the full tap-input vector. Thus, $\mathcal{M}(n)$ may be written as

$$\mathcal{M}(n) = \frac{\|\mathbf{Q}(n) \mathbf{x}(n)\|^2}{\|\mathbf{x}(n)\|^2}. \quad (2)$$

Manuscript received May 6, 2004; revised August 18, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yiteng (Arden) Huang.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: andy.khong@imperial.ac.uk; p.naylor@imperial.ac.uk).

Digital Object Identifier 10.1109/LSP.2004.842287

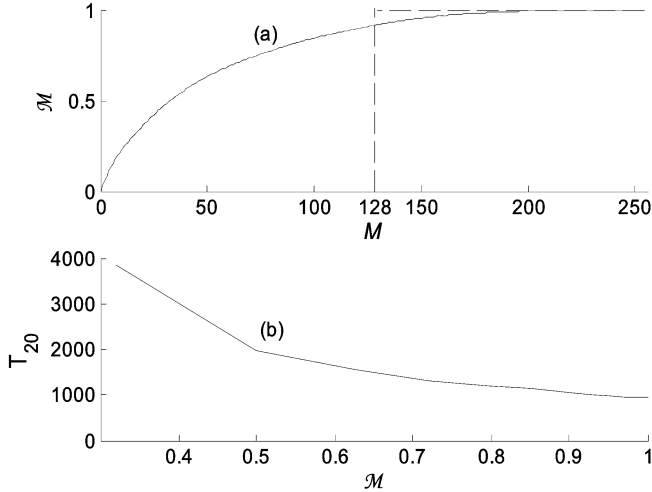


Fig. 2. (a) Variation of \mathcal{M} with subselection parameter M . (b) Dependence of convergence rate on \mathcal{M} .

This measure quantifies the “closeness” of the tap selection to the full tap-input vector in an MMax sense such that when all taps are updated, $\mathcal{M} = 1$. Fig. 2(a) shows how \mathcal{M} varies with the number of selected taps M for zero mean, unit variance white Gaussian noise (WGN) for a particular time iteration n with $L = 256$. We note that \mathcal{M} exhibits only a modest reduction for $0.5L \leq M \leq L$, and hence, a graceful reduction in convergence rate is expected over this range as compared to fully updated NLMS [9]. Fig. 2(b) shows the number of iterations for MMax-NLMS to achieve -20 -dB misalignment for various \mathcal{M} . Since the convergence rate can be seen to increase with \mathcal{M} , we propose that any degradation in convergence performance due to subselection of taps can be minimized by selecting taps to maximize \mathcal{M} .

III. XM TAP-SELECTION CRITERION

Although selective-tap partial update adaptive algorithms [6]–[8] were originally proposed for complexity reduction, this is not our aim. We now propose and develop selective-tap schemes aimed instead to reduce the interchannel coherence of the two tap-input vectors. We note that direct application of MMax tap selection will not serve to decorrelate the two tap-input vectors because since $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are themselves highly correlated, nearly identical tap indices will be selected in both filters. Therefore, we formulate the XM tap-selection criterion that aims jointly to maximize $\mathcal{M}(n)$ and minimize interchannel coherence at each iteration. XM tap selection addresses the minimum coherence condition by constraining tap selections to be exclusive between the two channels so that the same coefficient index may not be selected in both channels [9], [10].

Assuming that $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are highly correlated Gaussian inputs and defining $\mathbf{x}(n) = [\mathbf{x}_1^T(n) \mathbf{x}_2^T(n)]^T$ and $E[\cdot]$ as the mathematical expectation operator, the autocorrelation matrix can be expressed as

$$\begin{aligned} \mathbf{R}_{\mathbf{xx}}(n) &= E[\mathbf{x}(n)\mathbf{x}^T(n)] \\ &= \begin{bmatrix} \mathbf{R}_{11}(n) & \mathbf{R}_{12}(n) \\ \mathbf{R}_{21}(n) & \mathbf{R}_{22}(n) \end{bmatrix}. \end{aligned} \quad (3)$$

We may examine the effect of exclusive tap selection on $\mathbf{R}_{\mathbf{xx}}(n)$ for which nonselected elements in $\tilde{\mathbf{x}}_1(n) = \mathbf{Q}_1(n)\mathbf{x}_1(n)$ and $\tilde{\mathbf{x}}_2(n) = \mathbf{Q}_2(n)\mathbf{x}_2(n)$ are zero. It can be seen that this exclusive tap selection has the effect of making the diagonals of $\mathbf{R}_{12}(n)$ and $\mathbf{R}_{21}(n)$ zero and adds additional zeros elsewhere in $\mathbf{R}_{12}(n)$ and $\mathbf{R}_{21}(n)$. As a result, the conditioning of $\mathbf{R}_{\mathbf{xx}}(n)$ is improved, and in the limit where $\tilde{\mathbf{x}}_1(n)$ and $\tilde{\mathbf{x}}_2(n)$ are perfectly uncorrelated, the autocorrelation matrix is diagonal $\mathbf{R}_{\mathbf{xx}}(n) = \text{diag}[\sigma_1^2 \dots \sigma_1^2 \sigma_2^2 \dots \sigma_2^2]$, where σ_i^2 is the i th channel input variance, and good convergence is then obtained.

Although an exhaustive search of all exclusive tap selections could be used to find the selection set that maximizes \mathcal{M} , a more efficient method can be found by considering $\mathbf{p}(n) = |\mathbf{x}_1(n)| - |\mathbf{x}_2(n)|$. The tap selection with maximum $\mathcal{M}(n)$ can then be found efficiently by sorting $\mathbf{p}(n)$. Utilizing the robustness of the NLMS algorithm to tap selection for $0.5L \leq M \leq L$ as discussed in the previous section, we select $M = 0.5L$ taps at each iteration.

Consider as a simple example an SAEC system with channels $k = 1, 2$, adaptive filters each of length $L = 4$, and tap-input vectors $\mathbf{x}_k(n) = [x_{k,1} \ x_{k,2} \ x_{k,3} \ x_{k,4}]^T$. Also consider the example case $p_3 > p_2 > p_1 > p_4$, 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 $x_{2,4}$ maximizes $\mathcal{M}(n)$ with the minimum coherence constraint satisfied by the exclusivity. In this two-channel case, $\mathcal{M}(n)$ is then defined as

$$\mathcal{M}(n) = \frac{\|\mathbf{Q}(n)\mathbf{x}(n)\|^2}{\|\mathbf{x}(n)\|^2} \quad (5)$$

where $\mathbf{Q}(n) = \text{diag}\{\mathbf{q}_1^T(n) \ \mathbf{q}_2^T(n)\}$ is the selection matrix such that at each iteration n , element u of $\mathbf{q}_1(n)$ and element v of $\mathbf{q}_2(n)$ are defined for $u, v = 1, 2, \dots, L$ such that

$$\begin{aligned} q_{1,u} &= \begin{cases} 1, & p_u \in \{M \text{ maxima of } \mathbf{p}\} \\ 0, & \text{otherwise} \end{cases} \\ q_{2,v} &= \begin{cases} 1, & p_v \in \{M \text{ minima of } \mathbf{p}\} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

A. XMNL-NLMS

The NLMS algorithm is popular for its robustness and simplicity of implementation, but it is not normally useful for SAEC applications due to its poor convergence. The NL preprocessor [1] has shown to be one of the most effective methods to achieve signal decorrelation by using α as the nonlinearity constant such that

$$\begin{aligned} \mathbf{x}'_1 &= \mathbf{x}_1(n) + 0.5\alpha[\mathbf{x}_1(n) + |\mathbf{x}_1(n)|] \\ \mathbf{x}'_2 &= \mathbf{x}_2(n) + 0.5\alpha[\mathbf{x}_2(n) - |\mathbf{x}_2(n)|]. \end{aligned}$$

We propose a method for SAEC employing the XM tap selection in combination with the NL preprocessor to improve the convergence rate of the NL-NLMS algorithm. We shall show in the next section that the resulting XMNL-NLMS algorithm can

TABLE I
XMNL-NLMS

$\mathbf{p}(n)$	$= \mathbf{x}'_1(n) - \mathbf{x}'_2(n) $
$\mathbf{Q}(n)$	$= \text{diag}\{\mathbf{q}_1^T(n) \mathbf{q}_2^T(n)\}$
$y(n)$	$= \tilde{\mathbf{h}}^T(n) \mathbf{x}'(n)$
$e(n)$	$= d(n) - y(n)$
$q_{1,u}$	$\begin{cases} 1 & p_u \in \{M \text{ maxima of } \mathbf{p}\} \\ 0 & \text{otherwise} \end{cases}$
$q_{2,v}$	$\begin{cases} 1 & p_v \in \{M \text{ minima of } \mathbf{p}\} \\ 0 & \text{otherwise} \end{cases}$

give useful levels of convergence with performance close to that of an existing RLS-based approach [1].

The XM tap-selection technique may be incorporated into NL-NLMS by selecting taps corresponding to the M highest value elements of the input magnitude difference vector $\mathbf{p}(n)$ in the first channel. For the second channel, the selected taps will correspond to the M lowest value elements of $\mathbf{p}(n)$. Note that at each iteration, $M = 0.5L$ taps are selected in each channel. Consequently, the XMNL-NLMS has the same complexity as the MMax-NLMS algorithm, which requires, at most, $1.5L + 2\log_2 + 3$ operations (multiplications or comparison) for each filter per sample period, assuming the use of the SORTLINE algorithm. The XMNL-NLMS is expressed in Table I.

B. XMNL-RLS

Direct application of the RLS algorithm to SAEC does not address the nonuniqueness problem, and hence, convergence is poor. We first denote λ as the forgetting factor and $\Psi(n) = \mathbf{X}'(n)\Lambda(n)\mathbf{X}'^T(n)$ as the weighted autocorrelation matrix such that $\mathbf{X}'(n) = [\mathbf{x}'(0) \mathbf{x}'(1) \dots \mathbf{x}'(n)]^T$ and $\Lambda = \text{diag}[\lambda^n \lambda^{n-1} \dots \lambda^0]$. Defining $e_{n-1}(n)$ as the a priori error, the RLS tap update is given by

$$\tilde{\mathbf{h}}(n) = \tilde{\mathbf{h}}(n-1) + \mathbf{k}(n)e_{n-1}(n) \quad (6)$$

where $\mathbf{k}(n) = [\mathbf{k}_1(n) \mathbf{k}_2(n)]^T$ is defined as the concatenated Kalman gain, which is defined at each iteration as

$$\mathbf{k}(n) = \frac{\lambda^{-1}\Psi^{-1}(n)\mathbf{x}'(n)}{1 + \lambda^{-1}\mathbf{x}'^T(n)\Psi^{-1}(n)\mathbf{x}'(n)}. \quad (7)$$

The iterative update of $\Psi^{-1}(n)$ is then given by

$$\Psi^{-1}(n+1) = \lambda^{-1} [\Psi^{-1}(n) - \mathbf{k}(n)\mathbf{x}'^T(n)\Psi^{-1}(n)]. \quad (8)$$

To extend the XM tap-selection approach to NL-RLS, we could first consider the selection of M out of L taps with the largest Kalman gain. However, this is not a good approach because any subselection of $\mathbf{k}(n)$ is used recursively in the updating of $\Psi^{-1}(n+1)$, and this recursion is, therefore, inconsistent from sample to sample since the tap selection changes at each n . Our proposed approach is, therefore, to replace $\mathbf{x}'(n)$ in the Kalman gain calculation (7) and Ricatti update (8) with a subsampled input vector $\tilde{\mathbf{x}}'(n) = \mathbf{Q}(n)\mathbf{x}'(n)$. This ensures that the subsampled input vectors propagate consistently through the memory of the algorithm. Accordingly, we formulate the

TABLE II
XMNL-RLS

$M = 0.5L$	
$\mathbf{p}(n)$	$= \mathbf{x}'_1(n) - \mathbf{x}'_2(n) $
$\mathbf{k}(n)$	$= [\mathbf{k}_1^T(n) \mathbf{k}_2^T(n)]^T$
$\mathbf{Q}(n)$	$= \text{diag}\{\mathbf{q}_1^T(n) \mathbf{q}_2^T(n)\}$
$\tilde{\mathbf{x}}'(n)$	$= \mathbf{Q}(n)\mathbf{x}'(n)$
$\mathbf{k}(n)$	$= \frac{\Psi^{-1}(n)\tilde{\mathbf{x}}'(n)}{\lambda + [\tilde{\mathbf{x}}'(n)]^T \Psi^{-1}(n) \tilde{\mathbf{x}}'(n)}$
$y(n)$	$= \tilde{\mathbf{h}}^T(n) \mathbf{x}'(n)$
$e_{n-1}(n)$	$= d(n) - y(n)$
$\tilde{\mathbf{h}}(n+1)$	$= \tilde{\mathbf{h}}(n) + \mathbf{k}(n)e_{n-1}(n)$
$\Psi^{-1}(n+1)$	$= \frac{1}{\lambda} [\Psi^{-1}(n) - \mathbf{k}(n)[\tilde{\mathbf{x}}'(n)]^T \Psi^{-1}(n)]$
$q_{1,u}$	$\begin{cases} 1 & p_u \in \{M \text{ maxima of } \mathbf{p}\} \\ 0 & \text{otherwise} \end{cases}$
$q_{2,v}$	$\begin{cases} 1 & p_v \in \{M \text{ minima of } \mathbf{p}\} \\ 0 & \text{otherwise} \end{cases}$

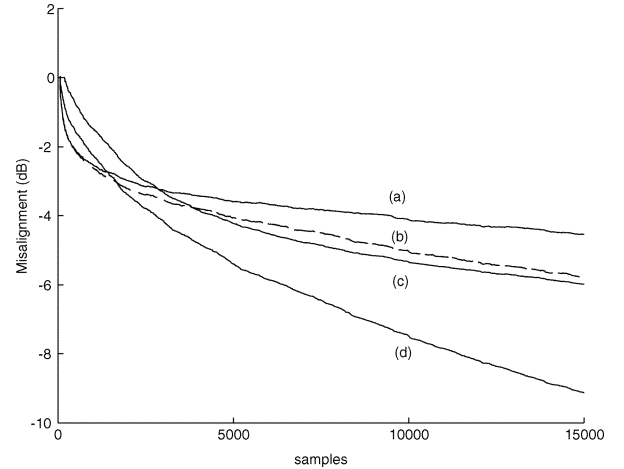


Fig. 3. Misalignment for (a) NLMS, (b) NL-NLMS, (c) XM-NLMS, and (d) XMNL-NLMS [$L = 256$, $M = 128$, $\mu = 0.1$, $\alpha = 0.5$].

XMNL-RLS algorithm as shown in Table II. As before, we have used u and v to denote the elemental indices of $\mathbf{q}_1(n)$, $\mathbf{q}_2(n)$, and $\mathbf{p}(n)$. Using the SORTLINE algorithm, the complexity of XMNL-RLS is, at most, $2.5L(LH) + 3 + 2\log_2 L$ operations per adaptive filter compared to $4L^2 + 3L + 2$ for RLS. Although complexity reduction is not our main aim, it can be seen that XM tap selection reduces computational complexity.

IV. SIMULATION RESULTS

A. Performance of XMNL-NLMS Algorithm

In these tests, the unknown room impulse responses were generated using the method of images [12], with the microphones placed one meter apart and the source positioned one meter away from each of the microphones in the transmission room. In this first experiment, the lengths of the adaptive filters are $L = 256$, while the lengths of the transmission and receiving

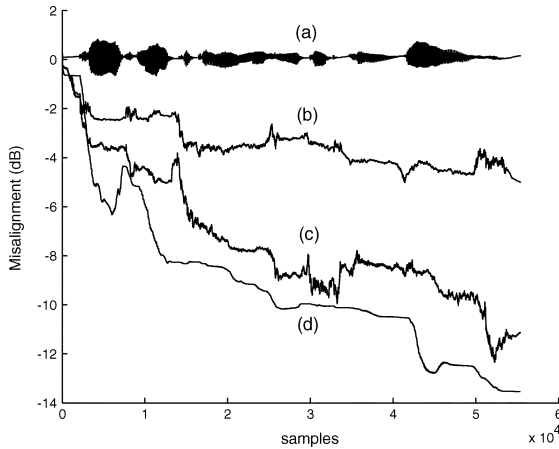


Fig. 4. (a) Speech and misalignment plot for (b) NL-NLMS, (c) XMNL-NLMS, and (d) NL-FRLS [$L = 256$, $M = 128$, $\mu = 0.9$, $\alpha = 0.5$, $\lambda = 0.9996$].

rooms are $W = 1600$ and $N = 256$, respectively. Fig. 3 shows the misalignment $\eta(n)$ for NLMS-based algorithms, where

$$\eta(n) = \frac{\|\mathbf{h} - \tilde{\mathbf{h}}(n)\|^2}{\|\mathbf{h}\|^2}. \quad (9)$$

A WGN input was used with $M = 128$ and a nonlinear distortion factor of $\alpha = 0.5$. We see that NLMS converges to a poor solution. The convergence rate of XM-NLMS and NL-NLMS increases significantly, due to the XM and NL preprocessor, respectively. XMNL-NLMS shows even further improvement compared to NL-NLMS, due to the additional decorrelation property of the XM tap selection. Alternatively, XMNL-NLMS could achieve the same rate of convergence as NL-NLMS but with a lower value of α [1].

B. Comparison of XMNL-NLMS and NL-FRLS

Fig. 4 compares the convergence of NL-NLMS, XMNL-NLMS and fast-RLS with NL preprocessor (NL-FRLS) [1] using a speech signal from a male talker. Both the transmission and receiving room's responses were of length $W = N = 800$. Adaptive filters of length $L = 256$ and $M = 128$ were used with $\lambda = 1 - (1/10L)$ for the NL-FRLS algorithm. It can be seen that the performance of XMNL-NLMS exceeds that of NL-NLMS by around 5 to 8 dB and is close to that of NL-FRLS in this example.

C. Comparison of XMNL-RLS and NL-FRLS

We now compare the rate of convergence between the XMNL-RLS algorithm and the NL-FRLS algorithm using the same experimental setup as the previous experiment but using a different speech signal for generality. We see from Fig. 5 that there is a significant improvement in misalignment of 3 to 6 dB for the XMNL-RLS compared to that of the NL-FRLS.

V. CONCLUSION

We have formulated the XM tap-selection technique and employed it in the proposed XMNL-NLMS and the XMNL-RLS algorithms. These algorithms achieve the required decorrelation of the tap-input vectors in SAEC using this novel selective-tap

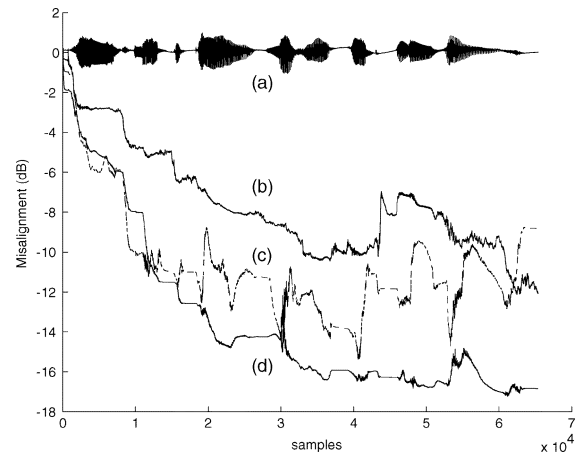


Fig. 5. (a) Speech and misalignment plot for (b) XMNL-NLMS, (c) NL-FRLS, and (d) XMNL-RLS [$L = 256$, $M = 128$, $\mu = 0.9$, $\alpha = 0.5$, $\lambda = 0.9996$].

scheme and give a significant improvement in performance over and above the use of the NL preprocessor alone. Although NLMS is not normally employed for SAEC because of its poor convergence, relatively good performance—close to that of RLS-based schemes—can be obtained, nevertheless, through the use of the proposed XM tap-selection approach. XMNL-NLMS has the benefits of low complexity and robustness compared to least-squares approaches. Additionally, a significant increase in the convergence rate can be seen for XMNL-RLS as compared to that obtained from an RLS-based approach using the NL preprocessor alone.

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 Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.
- [2] S. Emura, Y. Haneda, and S. Makino, "Enhanced frequency-domain adaptive algorithm for stereo echo cancellation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 2, 2002, pp. 1901–1904.
- [3] P. Eneroth, S. L. Gay, T. Gansler, and J. Benesty, "A real-time implementation of a stereophonic acoustic echo canceller," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 513–523, Jul. 2001.
- [4] A. Gilloire and V. Turbin, "Using auditory properties to improve the behavior of stereophonic acoustic echo cancellers," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 6, 1998, pp. 3681–3684.
- [5] T. Hoya, Y. Loke, J. Chambers, and P. A. Naylor, "Application of the leaky extended LMS algorithm in stereophonic acoustic echo cancellation," *Signal Process.*, vol. 64, pp. 87–91, 1998.
- [6] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Trans. Circuits Syst. II*, vol. 44, no. 3, pp. 209–216, Mar. 1997.
- [7] T. Aboulmar and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1421–1424, May 1999.
- [8] 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. Acoust. Speech Signal Process.*, vol. 5, 2003, pp. 373–376.
- [9] 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 *Proc. Int. Workshop Acoust. Echo Noise Control*, 2003, pp. 59–62.
- [10] —, "Reducing inter-channel coherence in stereophonic acoustic echo cancellation using partial update adaptive filters," in *Proc. Eur. Signal Process. Conf.*, 2004, pp. 405–408.
- [11] I. Pitas, "Fast algorithms for running ordering and max/min calculation," *IEEE Trans. Circuits Syst.*, vol. 36, no. 6, pp. 795–804, Jun. 1989.
- [12] 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.