Journal of ELECTRICAL ENGINEERING, VOL. 55, NO. 5-6, 2004, 113–121

FAST IMPLEMENTATION OF A SUBBAND ADAPTIVE ALGORITHM FOR ACOUSTIC ECHO CANCELLATION

Khaled Mayyas *

The block subband adaptive algorithm in [1] has illustrated significant improvement in performance over the NLMS and other frequency domain adaptive algorithms. However, it is known that block processing algorithms have lower tracking capabilities than the their sample-by-sample counterparts. The Fast Affine Projection (FAP) algorithm [2] also outperforms the NLMS with a slight increase in complexity, but involves the fast calculation of the inverse of a covariance matrix of the input data that could undermine the performance of the algorithm. In this paper, we present a sample-by-sample version of the algorithm in [1] and develop an exact, low complexity implementation of this algorithm. The new fast algorithm does not require matrix inversion thus alleviating the drawbacks of the FAP algorithm. Moreover, we will show that the new sample-by-sample algorithm approximates the affine projection algorithm and possesses a similar property in reducing the coefficient bias that appears in monophonic and stereophonic teleconferencing when the receiving room impulse responses are undermodelled. The new fast sample-by-sample algorithm is extended for stereo acoustic echo cancellation. Simulations of echo cancellations in actual rooms are presented to verify our findings.

K e y w o r d s: Adaptive algorithms, fast techniques, mono and stereo acoustic echo cancellation.

1 INTRODUCTION

With emerging applications requiring adaptive filter orders of several hundreds or thousands, the implementation complexity of fast versions of the RLS algorithm is still highly costly and beyond the capabilities of current DSP processors. Several algorithms have been proposed outperforming the NLMS algorithm in convergence speed with a reasonable increase in computational complexity [2], or with almost equivalent complexity to the NLMS as in [1,3] as a result of performing block adaptation.

The Affine Projection (AP) algorithm has been used in acoustic echo cancellation (AEC) to speed up convergence [4, 5]. The improvement in convergence speed is known to follow from the fact that AP algorithm decorrelates the input signal thus reducing the eigenvalues disparity of the algorithm autocorrelation matrix. A fast version (FAP) has also been presented in [6] making the complexity of the algorithm acceptable in acoustic echo cancellation. The FAP algorithm involves the computation of the inverse of a $p \times p$ input data covariance matrix, where p is the value of the projection order of the affine algorithm. The inverse is calculated using a sliding-window version of the Fast Transversal Filter (FTF), which is known to have numerical instability problems. In addition, when the input signal is highly correlated, the inversion process causes noise amplification resulting in large misadjustment [7]. This problem appears also in [3] in both the block and sample-by-sample versions of the algorithm. The arithmetic complexity of the sample-by-sample version [3] is in the order of $12N$, where N is the order of the adaptive filter.

The block processing in [1, 3] is advantageous in terms of complexity reduction, however it is known that block adaptation results in lower tracking capabilities compared to its sample-by-sample counterpart.

To solve the above problems, we propose a fast and exact implementation of the sample-by-sample version of the weighted subband algorithm in [1] that was shown to outperform the NLMS and other block transform domain adaptive algorithms. The proposed algorithm combines the low complexity property of the FAP algorithm and the robust performance of the algorithm in [1]. The complexity of the original sample-by-sample algorithm is approximately $O(pN)$. The complexity of the fast implementation is reduced to $O(2N)$ for the same algorithm. We will further illustrate that the original algorithm approximates the AP algorithm and possesses similar decorrelating properties that lead to a decrease in coefficient bias under undermodeling conditions in both monophonic and stereophonic AEC [8, 9]. The problem of coefficient bias is more evident in the stereo case as a result of the high correlation between signals in the two channels. The computationally efficient technique will be extended to the stereophonic case.

2 FAST MONOPHONIC WEIGHTED SUBBAND ADAPTIVE ALGORITHM

We start by deriving a fast sample-by-sample version of the algorithm in [1]. Figure 1 shows the structure of the proposed algorithm that also fits in the monophonic AEC setup in which the input signal $x(n)$ is generated from passing the speech signal through the transmission room impulse response, and the unknown system is the impulse response of the path between the loud speaker

* Department of Electrical Engineering, Jordan University of Science and Technology, Irbid, 22110, Jordan, E-mail: mayyas@just.edu.jo

ISSN 1335-3632

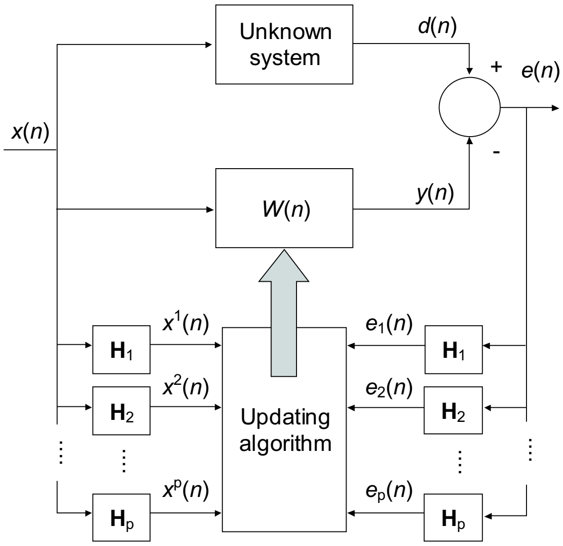

Fig. 1. The structure of the sample-by-sample weighted subband adaptive algorithm.

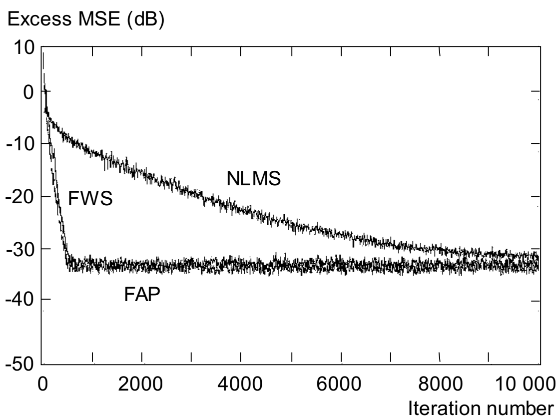

Fig. 2. Comparison of excess MSE between the NLMS, FWS, and FAP algorithm for correlated input signal SNR=30 dB and with $p = 8$.

and the microphone in the receiving room. The structure of Fig. 1 is similar to that in [1] except that in [1] each subband error $e_i(n)$ is downsampled by p before being used in the update scheme. This follows from the block adaptation operation adopted in [1]. The analysis filters $\mathbf{H}_i,\ i = 1, 2, \dots, p$ are assumed to form a perfect reconstruction filter bank. The principal advantage of the structure in Fig. 1 is that it updates a full-band adaptive filter as opposed to the traditional subband adaptive filtering techniques that apply individual adaptive filters in each subband. The delay due to the filter bank is also moved to the adaptation loop out of the input signal path. The sample-by-sample algorithm attempts to minimize the instantaneous sum of the weighted square subbands errors, *ie*, $\sum_{i=1}^{p} \lambda_i e_i^2(n)$, relative to $\mathbf{W}(n)$. Thus, the update recursion of the sample-by-sample algorithm is easily given by [1]

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \sum_{i=1}^{p} \lambda_i e_i(n) \mathbf{X}^{(i)}(n) \qquad (1)$$

where μ is a step size factor, $e(n) = d(n) - \mathbf{X}^\top(n)\mathbf{W}(n)$, $e_i(n) = \mathbf{H}_i^\top \boldsymbol{e}(n)$, $\boldsymbol{e}(n) = [e(n)\ e(n-1)\ \dots\ e(n-K+1)]^\top$, K is the length of the analysis filter $\mathbf{H}_i$, and $\mathbf{X}^{(i)}(n) = \Phi(n)\mathbf{H}_i$. $\Phi(n)$ is an $N \times K$ matrix defined as $\Phi(n) = [\mathbf{X}(n)\ \mathbf{X}(n-1)\ \dots\ \mathbf{X}(n-K+1)]$, where $\mathbf{X}(n) = [x(n)\ x(n-1)\ \dots\ x(n-N+1)]^\top$, and N is the length of the adaptive filter. The weighting factor λ_i is chosen as $\lambda_i(n) = \frac{1}{\|\mathbf{X}^{(i)}(n)\|^2 + \delta}$, $0 < \delta \ll 1$, to normalize the power in each subband to reduce the eigenvalues disparity of the input data autocorrelation matrix. The update algorithm in Eq. (1) requires a total of $N(p+1) + 2p(K+1)$ multiplications, and $Np + 2p(K-1)$ additions, excluding the overhead of calculating $\lambda_i(n),\ i = 1, 2, \dots, p$. It is clear that the complexity of the algorithm grows significantly as p increases. However, the algorithm converges faster than the block one weighted subband [1] since adaptation is performed sample by sample. We will show now how to reduce the complexity of the sample-by-sample update algorithm in Eq. (1) by developing an exact fast implementation of the algorithm.

Define the matrix $H = [\mathbf{H}_1\ \mathbf{H}_2\ \dots\ \mathbf{H}_p]$, and the diagonal matrix $\Lambda^{-1}(n) = \mathrm{diag}\{\mu\lambda_1(n), \mu\lambda_2(n), \dots, \mu\lambda_p(n)\}$, then Eq. (1) can be expressed as

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \Phi(n) H \Lambda^{-1}(n) H^\top \boldsymbol{e}(n). \qquad (2)$$

Define $\mathbf{P}(n) = H\Lambda^{-1}(n)H^\top \boldsymbol{e}(n)$, where $\mathbf{P}(n) = [\rho_1(n)\ \rho_2(n)\ \dots\ \rho_K(n)]^\top$. The second term in Eq. (2) is given by $\Phi(n)\mathbf{P}(n)$ which can be calculated efficiently using a similar approach employed in the FAP algorithm in [1]. Define the intermediate weight vector $\widehat{\mathbf{W}}(n)$ such that

$$\widehat{\mathbf{W}}(n) = \mathbf{W}(n) - \Phi_{K-1}(n)\mathbf{E}(n-1) \qquad (3)$$

where $\Phi(n) = [\mathbf{X}(n)\ \ \Phi_{K-1}(n)]$ and

$$\mathbf{E}(n) = \begin{bmatrix} \rho_1(n) \\ \rho_2(n) + \rho_1(n-1) \\ \vdots \\ \rho_{K-1}(n) + \rho_{K-2}(n-1) + \ldots + \rho_1(n-K+2) \end{bmatrix}. \qquad (4)$$

Using Eq. (2) in (3), $\widehat{\mathbf{W}}(n)$ can be updated as follows

$$\widehat{\mathbf{W}}(n+1) = \widehat{\mathbf{W}}(n) + z(n)\mathbf{X}(n-K+1) \qquad (5)$$

where $z(n) = \rho_K(n) + \rho_{K-1}(n-1) + \cdots + \rho_1(n-K+1)$. Notice that $\mathbf{E}(n)$ and $z(n)$ can be easily computed recursively as

$$\begin{bmatrix} \mathbf{E}(n) \\ z(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{E}(n-1) \end{bmatrix} + \mathbf{P}(n). \qquad (6)$$

Since $\widehat{\mathbf{W}}(n)$ is already available from Eq. 5), Eq. (3) is used to calculate the error $e(n) = d(n) - \mathbf{X}^\top(n)\mathbf{W}(n)$ as

$$e(n) = d(n) - \mathbf{X}^\top(n)\widehat{\mathbf{W}}(n) - \mathbf{B}^\top(n)\mathbf{E}(n-1) \quad (7)$$

where $\mathbf{B}(n) = \Phi_{K-1}^\top(n)\mathbf{X}(n)$. The quantity $\mathbf{B}(n)$ can be calculated efficiently by

$$\mathbf{B}(n) = \mathbf{B}(n-1) + x(n)\overline{\mathbf{X}}(n-1) - x(n-N)\overline{\mathbf{X}}(n-N-1) \quad (8)$$

where $\overline{\mathbf{X}}(n) = [x(n)\ x(n-1)\ \ldots\ x(n-K+2)]^\top$. Table 1 lists the equations for the implementation of the fast weighted subband (FWS) adaptive algorithm along with the number of multiplications and additions needed for each step. Note that the FWS algorithm needs the quantities $\mathbf{X}^{(i)}(n)$, $i = 1, 2, \ldots, p$ to be available to evaluate $\Lambda^{-1}(n)$. The complexity of the FWS algorithm, excluding that of calculating $\Lambda^{-1}(n)$ (the overhead of computing $\mathbf{X}^{(i)}(n)$, $i = 1, 2, \ldots, p$ is included), is $2N + 3K(1+p) + p - 3$ multiplications and $2N + 3K(1+p) - 2p - 2$ additions. For example, for $N = 1024$, $p = 16$, $K = 32$, the complexity of the original sample-by-sample algorithm is 18464 multiplies and 17376 adds, while that of the FWS algorithm is 3693 multiplies and 3646 adds. On the other hand, the FAP algorithm requires 2368 multiplies.

We present here an example that examines the performance of the FWS, FAP, and NLMS algorithm. The unknown system to be identified is a 50-coefficient FIR filter, which is a truncation of a 200-tap impulse response of an anechoic room, measured at 8 kHz sampling rate. Perfect modeling of the unknown system is assumed, *ie*, $N = 50$. The input signal is a highly correlated one generated by passing a zero-mean Gaussian signal with unity variance through the filter $H(z) = \frac{1}{1-1.58z^{-1}+0.8z^{-2}}$. White zero-mean Gaussian noise is added to the desired signal such that SNR=30 dB. Results are obtained by averaging over 100 independent runs. Both the FAP and FWS have $p = 8$. The step sizes used are $\mu_{FWS} = 0.07$, $\mu_{FAP} = 0.09$, and $\mu_{NLMS} = 1$, all chosen to achieve the same steady state excess MSE of the NLMS algorithm. The analysis filters are cosine modulated perfect reconstruction filter banks with $K = 32$. It is clear from Fig. 2 that the FWS performs as well as the FAP algorithm with both outperforming the NLMS. Note that though the complexity of the FAP is less than that of the FWS algorithm for the same p, the practical advantages of the FWS, particularly its numerical robustness, make it a safer choice for implementation.

2.1 Monophonic AEC with Undermodeling Conditions

In practical situations, the adaptive filter order is lower than the receiving room impulse response giving rise to a problem of coefficient bias in the adaptive filter coefficients compared to coefficients solution when the adaptive filter order matches that of the receiving room impulse response. In this context, we will show here that the sample-by-sample weighted subband adaptive algorithm has an additional advantage of reducing this bias without extra cost. This property is more important in the stereophonic case where bias is larger due to the strong crosscorrelation between the two-channel signals.

Let $d(n)$ be the desired signal obtained as the output of the unknown system represented by an M-coefficient FIR filter: $[w_0^*, w_1^*, \ldots, w_{M-1}^*]$. The desired signal is corrupted by a zero mean white Gaussian noise $\eta(n)$. The adaptive FIR filter has N coefficients ($N < M$). The desired signal is thus given by

$$d(n) = \mathbf{X}^\top(n)\mathbf{W}^* + \sum_{i=N}^{M-1} w_i^* x(n-i) + \eta(n) \quad (9)$$

where $\mathbf{X}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]$ is the input data vector, and $\mathbf{W}^* = [w_0^*, w_1^*, \ldots, w_{N-1}^*]^\top$. $\mathbf{W}_M^* = [w_0^*, w_1^*, \ldots, w_{M-1}^*]^\top$ is the impulse response of the path between the loudspeaker and the microphone in the receiving room. The second term in Eq. (9) is the part of the desired signal that cannot be modeled by the adaptive filter and will appear in the output error. The input signal $x(n)$ is generated from passing the source signal $s(n)$ through the room impulse response, *ie* $x(n) = \mathbf{S}^\top(n)\mathbf{G}$ where $\mathbf{S}(n) = [s(n), s(n-1), \ldots, s(n-L+1)]^\top$ and $\mathbf{G} = [g_0, g_1, \ldots, g_{L-1}]^\top$ is the impulse response of the path between the source and the microphone in the transmission room.

To study the effect of the unmodeled tail on the adaptation process, we consider the following adaptation equation

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu\mathbf{X}_f(n)e(n) \quad (10)$$

where $\mathbf{X}_f(n)$ is defined as a processed version of the input vector $\mathbf{X}(n)$ [10]. The output error is given by

$$e(n) = d(n) - \mathbf{X}^\top(n)\mathbf{W}(n) \quad (11)$$

The NLMS algorithm has $\mathbf{X}_f(n) = \mathbf{X}(n)$. Using Eq. (9) and (11) in (10), we get

$$\mathbf{V}(n+1) = \big(\mathbf{I} - \mu\mathbf{X}_f(n)\mathbf{X}^\top(n)\big)\mathbf{V}(n) + \mu\eta(n)\mathbf{X}_f(n) + \mu\sum_{j=N}^{M-1} w_j^* x(n-j)\mathbf{X}_f(n) \quad (12)$$

where $\mathbf{V}(n) = \mathbf{W}^* - \mathbf{W}(n)$, and will be referred to as the 'bias in the coefficients'. The mean evolution of Eq. (12) is

$$E\{\mathbf{V}(n+1)\} = (\mathbf{I} - \mu E\{\mathbf{X}_f(n)\mathbf{X}^\top(n)\})\mathbf{V}(n) + \mu\sum_{j=N}^{M-1} w_j^* E\{x(n-j)\mathbf{X}_f(n)\}. \quad (13)$$

Note that if $x(n)$ is a zero-mean white process and the update algorithm uses $x_f(n) = x(n)$, then $E\{x(n-j)\mathbf{X}_f(n)\} = E\{x(n-j)\mathbf{X}(n)\}$ vanishes in Eq. (13) since

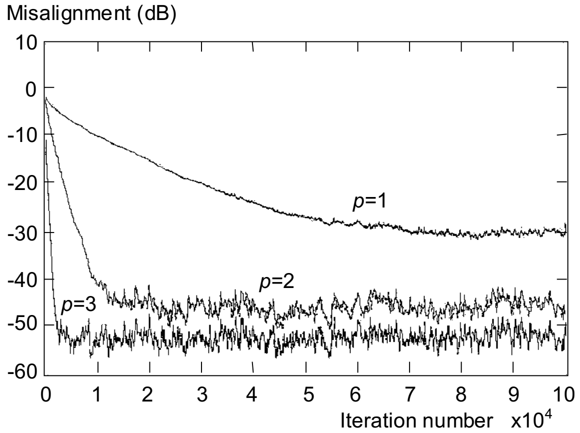

Fig. 3. Misalignment of the FAP algorithm in the monophonic case for $p = 1, 2, 3$ under undermodeling conditions.

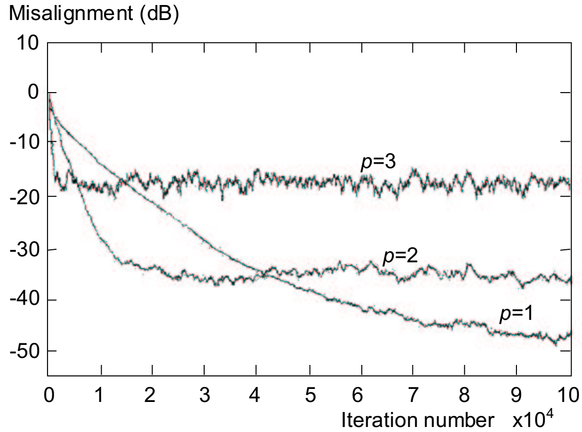

Fig. 4. Misalignment of the FAP algorithm in the monophonic case for $p = 1, 2, 3$ under perfect modeling conditions SNR$\simeq 21$.

$j \geq N$. Accordingly $E\{\mathbf{V}(\infty)\} = 0$. However, under normal circumstances in acoustic environment, $x(n)$ is a correlated sequence and as such $E\{x(n-j)\mathbf{X}(n)\} \neq 0$. This indicates that the undermodelled part of the desired signal in the monophonic case causes a non-zero bias in the filter coefficients, which increases the level of the MSE compared to the case when the adaptive filter order matches completely the unknown system impulse response, *ie*, $N = M$. We assume that the input signal $x(n)$ is of persistent excitation. Note that the smaller the values of $E\{x(n-j)\mathbf{X}_f(n)\}$, $j = N, N+1, \ldots, M-1$ the less the effect of the undermodelled part on the system coefficient bias.

Suppose that the update algorithm uses $x_f(n)$ that is un uncorrelated sequence that results from decorrelating $x(n)$, ex., the input sequence $x(n)$ can be modelled as an autoregressive process of order l and is passed through a linear predictor of order l with $x_f(n)$ being the output error of the predictor. Then $E\{x(n-j)\mathbf{X}_f(n)\} = 0 \ \forall\, j = N, N+1, \ldots, M-1$.

The Affine Projection algorithm (AP) [10] is an example of an algorithm that decorrelates the input signal with different decorrelation orders before being used in the adaptive algorithm as in Eq. (10). This was also shown in [11]. It is known that the larger the projection order p the better the decorrelation between the samples of $x_f(n)$, and therefore $E\{x(n-j)\mathbf{X}_f(n)\}$ becomes less for higher p. We present here an example that verifies the above conclusions.

The FAP algorithm is used in an undermodelling monophonic acoustic echo cancellation example. The echo path is that of a room truncated to 300 taps, and measured at 8 kHz sampling rate. The length of the adaptive filter is 200. The input signal is a correlated one generated by passing a zero mean white Gaussian signal with unity variance through the filter $H(z) = \frac{1}{1-1.58z^{-1}+0.81z^{-2}}$. The FAP is used with $\mu = 0.04$, and $p = 1, 2, 3$. Notice that we obtain the NLMS algorithm for p=1. A figure of merit that is used commonly as a measure of the bias in the filter coefficients is the misalignment defined as [8] (which also represents the normalized coefficient error vector norm (in dB))

$$\begin{aligned} \text{Misalignment} &= 10 \log_{10} \frac{\|\mathbf{W}^* - \mathbf{W}(n)\|}{\|\mathbf{W}^*\|} \\ &= 10 \log_{10} \frac{\|\mathbf{V}(n)\|}{\|\mathbf{W}^*\|}. \end{aligned}$$

To isolate the effect of undermodelling on misalignment, no noise is added to the desired signal and the step size is kept the same for different orders, as seen in Eq. (12). Figure 3 shows that misalignment is greatly reduced for $p = 2$, and $p = 3$. No improvement on misalignment is observed for $p > 3$. However, it should be noted that increasing the FAP order has been shown to amplify the noise in the system [2, 12]. This is demonstrated in the next example where perfect modelling is assumed to avoid mixing the impact of undermodelling with that of the noise.

In Fig. 4, the length of the adaptive filter is 300, *ie*, exact modelling of the echo path. White noise of zero mean and 0.1 variance is added to the desired signal (SNR$\simeq$ 21 dB). The FAP is used with same parameters of Fig. 3. Figure 4 shows a typical behavior of the FAP in perfect modeling conditions where the MSE due to coefficient fluctuations increase as the FAP order increases.

2.1.1 Similarities between AP and FWS

Though decorrelating characteristics of AP are very attractive as seen above, in practical situations, its low cost implementation using the FAP has its share of problems. As discussed earlier, the fast calculation of the inverse of $(\phi^\top(n)\phi(n) + \delta\mathbf{I})$, where $\phi(n) = [\mathbf{X}(n)\ \mathbf{X}(n-1)\ \ldots\ \mathbf{X}(n-p+1)]$ is the input data covariance matrix and δ is a small positive regularizing constant, is numerically problematic. Also, the problem of noise amplification occurs. We will show that the sample-by-sample

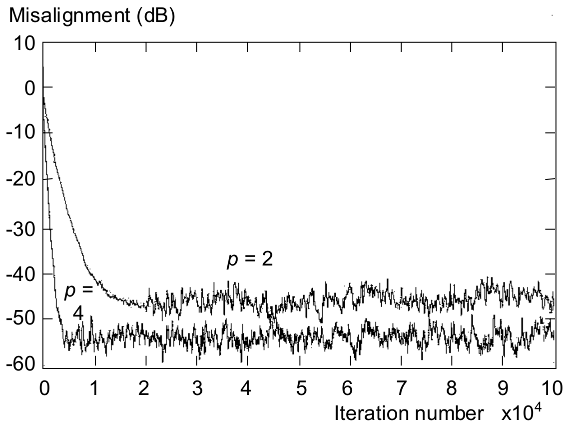

Fig. 5. Misalignment of the FWS algorithm in the monophonic case for $p = 2, 4$ under undermodeling conditions.

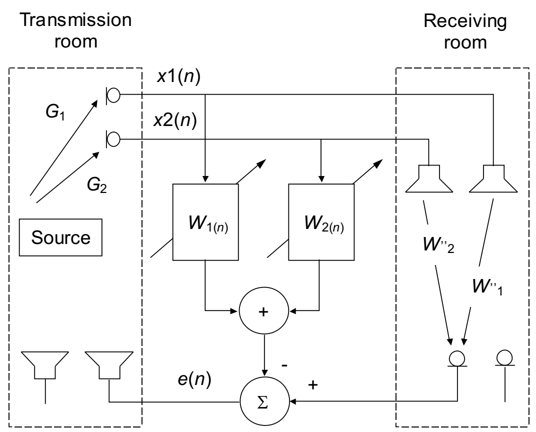

Fig. 6. The setup of a stereophonic acoustic echo cancellation.

weighted subband algorithm has close decorrelating properties to AP. Accordingly, the fast WS algorithm can provide close decorrelating performance to the AP while alleviating the practical pitfalls of the fast AP. For this purpose, we assume that $K = p$ resulting in $\phi(n) = \Phi(n)$.

From Eq. (2), the sample-by-sample algorithm has the coefficient update equation

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \Phi(n)\mathbf{P}(n) \tag{14}$$

where

$$\mathbf{P}(n) = H\Lambda^{-1}(n)H^{\top}\boldsymbol{e}(n). \tag{15}$$

The affine projection algorithm can be put in a relaxed form as [2, 13],

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu\phi(n)\mathbf{P}_{FAP}(n) \tag{16}$$

where

$$\mathbf{P}_{FAP}(n) = [\phi^{\top}(n)\phi(n) + \delta\mathbf{I}]^{-1}\boldsymbol{e}(n)_{FAP} \tag{17}$$

where $\boldsymbol{e}(n)_{FAP} = \boldsymbol{d}(n) - \phi^{\top}(n)\mathbf{W}(n)$, and $\boldsymbol{d}(n) = [d(n)\ d(n-1)\dots d(n-p+1)]^{\top}$. The matrix $\Gamma(n) = (\phi^{\top}(n)\phi(n) + \delta\mathbf{I})$ of the AP algorithm can be written as [14]

$$\Gamma(n) = Q(n)\Lambda(n)Q^{\top}(n) \tag{18}$$

where $Q(n)$ is an orthonormal eigenvector matrix and $\Lambda(n)$ is a diagonal matrix. $\mathbf{P}_{FAP}(n)$ is then given by

$$\mathbf{P}_{FAP}(n) = Q(n)\Lambda^{-1}(n)Q^{\top}(n)\boldsymbol{e}(n)_{FAP}. \tag{19}$$

Note that $Q(n)$ is a linear unitary transform that diagonalizes the matrix $\Gamma(n)$. Note also that H in Eq. (15) is a linear unitary transform that attempts to diagonalize the input data covariance matrix. This fact is clear from the algorithm structure where the input signal is subbanded by the analysis filters and then normalized by their respective output power to generate a white-like input signal to the adaptive algorithm. Therefore, by examining Eq. (15) and Eq. (19) we conclude that the algorithm in Eq. (14) has close approximate decorrelating properties to the AP.

The previous example of Fig. 3 is repeated using the FWS algorithm with $p = 2$ (two subbands) and $p = 4$ (four subbands). The analysis filters are cosine-modulated perfect reconstruction filter banks with $K = 16$ for $p = 2$, and $K = 32$ for $p = 4$. The step size used is $\mu = 0.04$. Comparing Fig. 5 with Fig. 3 demonstrates the bias reduction advantages of the new proposed algorithm.

3 FAST STEREOPHONIC WEIGHTED SUBBAND ADAPTIVE ALGORITHM

Stereophonic teleconferencing provides more natural acoustic perception than monophonic teleconferencing [8]. With the ongoing prevalence of high speed processors and the introduction of efficient low complexity adaptive algorithms [9, 15, 16], stereophonic acoustic echo cancellation (SAEC) has become an achievable task those days and has been implemented successfully [9].

A stereophonic acoustic echo cancellation setup is shown in Fig. 6. The two microphones in the transmission room pick up the source signal via the acoustic paths characterized by $\mathbf{G}_1$ and $\mathbf{G}_2$, each of length L. We assume that the FIR adaptive filters $\mathbf{W}_1(n)$ and $\mathbf{W}_2(n)$ are each of length N. $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ in the receiving room represent the impulse responses of the paths between the two loudspeakers and the microphone. $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are each of length M. Similar paths couple to the other microphone in the receiving room. Similar analysis applies to the cancellation of echo in that microphone. We assume the practical situation that $N < L$ [8]. Beside the common problem of long adaptive filter order in single channel acoustic echo cancellation, the stereo acoustic

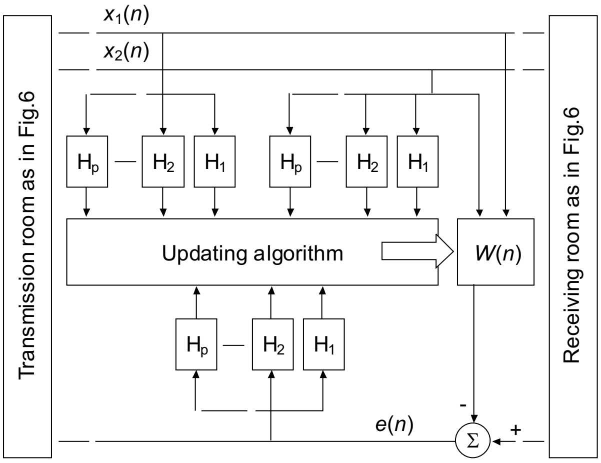

Fig. 7. The structure of the stereophonic weighted subband adaptive algorithm.

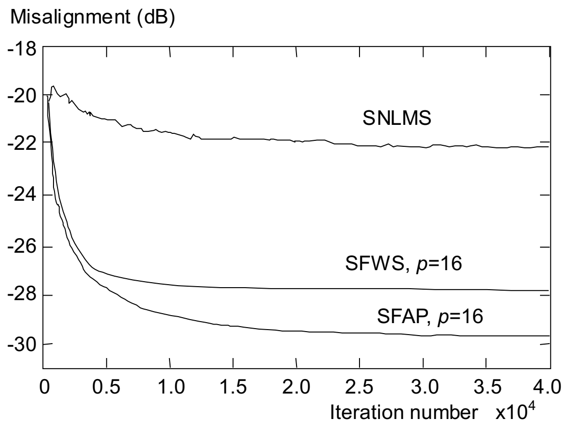

Fig. 8. Comparison of misalignment between the SNLMS, SFAP, and SFWS with $p = 16$ in the stereophonic case.

echo cancellation system has brought with it extra difficulties in implementation. SAEC entails four adaptive filters, two per channel, where each can be of an order of thousands. Additionally, The two transmitted signals ($x_1(n)$ and $x_2(n)$) are highly correlated resulting in an extremely ill-conditioned correlation matrix, and consequently very slow convergence of gradient-based adaptive algorithms. In practical situations, the adaptive filter orders are lower than the receiving room impulse responses giving rise to an additional performance problem of large bias in the adaptive filter coefficients solution [4, 8].

It is clear that the fast sample-by-sample weighted subband algorithm can provide an efficient solution in SAEC due to its low complexity, fast convergence, and bias reduction characteristics.

3.1 The Algorithm

Define $\mathbf{P}(n) = [\mathbf{P}_1^\top(n)\ \mathbf{P}_2^\top(n)]^\top$, $\Phi(n) = \begin{bmatrix} \Phi_1(n) & \mathbf{0} \\ \mathbf{0} & \Phi_2(n) \end{bmatrix}$, where $\Phi_i(n) = [\mathbf{X}_i(n)\ \mathbf{X}_i(n-1)\ \ldots\ \mathbf{X}_i(n-K+1)]$, $\mathbf{X}_i(n) = [x_i(n)\ x_i(n-1)\ldots x_(n-N+1)]^T$, $i = 1, 2$, $\mathbf{P}_1(n) = H\Lambda_1^{-1}(n)H^\top \boldsymbol{e}(n)$, $\mathbf{P}_2(n) = H\Lambda_2^{-1}(n)H^\top \boldsymbol{e}(n)$, and $\mathbf{P}_i(n) = [\rho_{i1}(n)\ \rho_{i2}(n)\ \ldots\ \ \rho_{iK}(n)]^\top$, $i = 1, 2$. Moreover, define $\Lambda_i^{-1}(n) = \mathrm{diag}\{\mu\lambda_{i1}(n),\ \mu\lambda_{i2}(n), \ldots,\ \mu\lambda_{ip}(n)\}$, where $\lambda_{ij}(n) = \frac{1}{\|\mathbf{X}_i^{(j)}(n)\|^2+\delta}$, $0 < \delta \ll 1$, and $\mathbf{X}_i^{(j)}(n) = \Phi_i(n)\mathbf{H}_j$, $i = 1, 2$, $j = 1, 2, \ldots, p$, then the sample-by-sample stereophonic weighted subband adaptive algorithm coefficient equation would have the form

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu\Phi(n)\mathbf{P}(n) \tag{20}$$

where $\mathbf{W}(n) = [\mathbf{W}_1^\top(n)\ \ \mathbf{W}_2^\top(n)]^\top$. The algorithm in Eq. (20) is implemented in Fig. 7 and requires a total of $2N(p+1)+p(3K+4)$ multiplications and $2Np+3p(K-1)$ additions, excluding the overhead of calculating $\lambda_{ij}(n)$, $i = 1, 2$, $j = 1, 2, \ldots, p$. Note that the complexity of the algorithm grows largely as the number of subbands p increases. Following a similar procedure to the single-channel FWS, the complexity of the stereo update version in Eq. (20) can be reduced.

The intermediate weight vector $\widehat{\mathbf{W}}(n)$ is defined as

$$\widehat{\mathbf{W}}(n) = \mathbf{W}(n) - \Phi_{K-1}(n)\mathbf{E}(n-1) \tag{21}$$

where

$$\Phi(n) = \begin{bmatrix} \mathbf{X}_1(n)\ \Phi_{1(K-1)}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2(n)\ \Phi_{2(K-1)}(n) \end{bmatrix} \tag{22}$$

and

$$\Phi_{K-1}(n) = \begin{bmatrix} \Phi_{1(K-1)}(n) & \mathbf{0} \\ \mathbf{0} & \Phi_{2(K-1)} \end{bmatrix}, \tag{23}$$

$$\mathbf{E}(n) = \begin{bmatrix} \mathbf{E}_1(n) \\ \mathbf{E}_2(n) \end{bmatrix} \quad \text{and} \quad \mathbf{E}_i(n) =$$
$$\begin{bmatrix} \rho_{i1}(n) \\ \rho_{i2}(n) + \rho_{i1}(n-1) \\ \vdots \\ \rho_{i(K-1)}(n) + \rho_{i(K-2)}(n-1) + \cdots + \rho_{i1}(n-K+2) \end{bmatrix}, \quad i = 1, 2. \tag{24}$$

Using Eq. (20) in (21), $\widehat{\mathbf{W}}(n)$ is updated as

$$\widehat{\mathbf{W}}(n+1) = \widehat{\mathbf{W}}(n) + \begin{bmatrix} z_1(n)\mathbf{X}_1(n-K+1) \\ z_2(n)\mathbf{X}_2(n-K+1) \end{bmatrix} \tag{25}$$

where $z_i(n) = \rho_{iK}(n)+\rho_{i(K-1)}(n-1)+\cdots+\rho_{i1}(n-K+1)$, $i = 1, 2$. Notice that $\mathbf{E}(n)$ and $z_i(n)$, $i = 1, 2$ can be computed recursively as follows

$$\begin{bmatrix} \mathbf{E}_1(n) \\ z_1(n) \\ \mathbf{E}_2(n) \\ z_2(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{E}_1(n-1) \\ 0 \\ \mathbf{E}_2(n-1) \end{bmatrix} + \mathbf{P}(n). \tag{26}$$

Using Eq. (21) and Eq. (25), the error, $e(n) = d(n) - \mathbf{X}^\top(n)\mathbf{W}(n)$, where $\mathbf{X}(n) = [\mathbf{X}_1^\top(n)\ \mathbf{X}_2^\top(n)]^\top$, is calculated as

$$e(n) = d(n) - \mathbf{X}^\top(n)\widehat{\mathbf{W}}(n) - \mathbf{B}^\top(n)\mathbf{E}(n-1) \tag{27}$$

where $\mathbf{B}(n) = \phi_{K-1}^\top(n)\mathbf{X}(n)$. The quantity $\mathbf{B}(n)$ can be calculated efficiently by

$$\mathbf{B}(n) = \mathbf{B}(n-1) + \begin{bmatrix} x_1(n)\overline{\mathbf{X}}_1(n-1) \\ x_2(n)\overline{\mathbf{X}}_2(n-1) \end{bmatrix} - \begin{bmatrix} x_1(n-N)\overline{\mathbf{X}}_1(n-N-1) \\ x_2(n-N)\overline{\mathbf{X}}_2(n-N-1) \end{bmatrix} \tag{28}$$

where $\overline{\mathbf{X}}_i(n) = [x_i(n)\ x_i(n-1)\ \dots\ x_i(n-K+2)]^\top$, $i = 1,2$. Table 2 lists the equations for the implementation of the stereophonic fast weighted subband (SFWS) adaptive algorithm along with the number of multiplications and additions needed for each step. Note that the SFWS algorithm needs the quantities $\mathbf{X}_i^{(j)}(n)$, $i = 1,2$, $j = 1,2,\dots,p$ to evaluate $\Lambda_i^{-1}(n)$, $i = 1,2$. Also, note that the calculation of $H^\top \boldsymbol{e}(n)$ is common between $\mathbf{P}_1(n)$ and $\mathbf{P}_2(n)$. The complexity of the stereophonic FWS algorithm, excluding that of calculating $\Lambda_i^{-1}(n)$ (the overhead of computing $\mathbf{X}_i^{(j)}(n)$, $i = 1,2$, $j = 1,2,\dots,p$ is included), is $4N + K(6+5p) + 2p - 6$ multiplications and $4N + K(6+5p) - 3p - 8$ additions. For example, for $N = 1024$, $p = 16$, $K = 32$, the complexity of the original stereo sample-by-sample algorithm is 36416 multiplies and 34256 adds, while that of the SFWS algorithm is 6874 multiplies and 6792 adds. On the other hand, the stereophonic FAP (SFAP) algorithm requires 4736 multiplies. The stereo NLMS (SNLMS) requires 4097 multiplies and 4094 adds.

3.2 Stereophonic AEC with Undermodeling Conditions

The stereophonic acoustic echo cancellation setup is shown in Fig. 6. The desired signal can be expressed as

$$d(n) = \mathbf{X}_1^\top(n)\mathbf{W}_1^* + \mathbf{X}_2^\top(n)\mathbf{W}_2^* + \sum_{i=N}^{M-1} w_{1i}^* x_1(n-i) + \sum_{i=N}^{M-1} w_{2i}^* x_2(n-i) + \eta(n). \tag{29}$$

Similar to the monophonic case, we choose the following equation for coefficients adjustment

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu\mathbf{X}_f(n)e(n) \tag{30}$$

and

$$e(n) = d(n) - \mathbf{X}^\top(n)\mathbf{W}(n) \tag{31}$$

where $\mathbf{X}_f(n) = [\mathbf{X}_{1f}^T(n)\ \mathbf{X}_{2f}^T(n)]^T$. Using Eq. (31) and (29) in (30), we get

$$\mathbf{V}(n+1) = \big(\mathbf{I} - \mu\mathbf{X}_f(n)\mathbf{X}^\top(n)\big)\mathbf{V}(n) + \mu\eta(n)\mathbf{X}_f(n) + \mu\sum_{j=N}^{M-1} w_{1j}^* x_1(n-j)\mathbf{X}_f(n) + \mu\sum_{j=N}^{M-1} w_{2j}^* x_2(n-j)\mathbf{X}_f(n). \tag{32}$$

We assume the practical situation such that $L > N$, as defined earlier, that guarantees a unique solution to the MSE minimization algorithm [8]. To get a clear picture of the stereophonic case, we take the expected value of both sides of Eq. (32) and break it into its two components,

$$\begin{aligned} E\{\mathbf{V}_1(n+1)\} &= (\mathbf{I} - \mu E\{\mathbf{X}_{1f}(n)\mathbf{X}_1^\top(n)\})E\{\mathbf{V}_1(n)\} \\ &+ \mu\sum_{j=N}^{M-1} w_{1j}^* E\{x_1(n-j)\mathbf{X}_{1f}(n)\} \\ &- \mu E\{\mathbf{X}_{1f}(n)\mathbf{X}_2^\top(n)\}E\{\mathbf{V}_2(n)\} \\ &+ \mu\sum_{j=N}^{M-1} w_{2j}^* E\{x_2(n-j)\mathbf{X}_{1f}(n)\} \end{aligned} \tag{33}$$

and

$$\begin{aligned} E\{\mathbf{V}_2(n+1)\} &= \big(\mathbf{I} - \mu E\{\mathbf{X}_{2f}(n)\mathbf{X}_2^\top(n)\}\big)E\{\mathbf{V}_2(n)\} \\ &+ \mu\sum_{j=N}^{M-1} w_{2j}^* E\{x_2(n-j)\mathbf{X}_{2f}(n)\} \\ &- \mu E\{\mathbf{X}_{2f}(n)\mathbf{X}_1^\top(n)\}E\{\mathbf{V}_1(n)\} \\ &+ \mu\sum_{j=N}^{M-1} w_{1j}^* E\{x_1(n-j)\mathbf{X}_{2f}(n)\}. \end{aligned} \tag{34}$$

Assuming the standard NLMS adaptation, $\mathbf{X}_f(n) = \mathbf{X}(n)$. To investigate the difference between the stereophonic and monophonic case, we consider Eq. (33) (or Eq. (34)) and Eq. (13). The stereophonic case is characterized by the strong correlation between $x_1(n)$ and $x_2(n)$ resulting in the appearance of two additional terms at the end of Eq. (33) (or Eq. (34)). Note that due to undermodeling conditions $E\{\mathbf{V}_i(n)\}(\infty) \neq 0$, $i = 1,2$. The last two terms contribute highly to the increase in the bias in filter coefficients in the stereophonic case compared to the monophonic one. This results in a much larger mismatch between the room response and the adaptive filter solution and manifests itself as an increase in the output error. Eq. (33) illustrates that the amount of mismatch (and hence the MSE) depends also on the level of undermodeling (influenced by N relative to M, and the values of w_{ij}^*, $j = N, N+1, \dots, M-1$, $i = 1,2$). Notice from Fig. 6 that $x_1(n) = \mathbf{S}^\top(n)\mathbf{G}_1$ and $x_2(n) = \mathbf{S}^\top(n)\mathbf{G}_2$. Consequently, the strength of crosscorrelation between $x_1(n)$ and $x_2(n)$ is not only a function of the source signal $s(n)$ but also essentially dependent on the transmission room characteristics (represented by $\mathbf{G}_1$ and $\mathbf{G}_2$). It is clear that the effect of those factors in the stereophonic case is very interdependent as well as considerable.

Table 1. The proposed fast weighted subbnad (FWS) adaptive algorithm for monophonic AEC

Algorithm steps	multiplications	additions
Computation: For $n = 0, 1, 2, \ldots$		
(1) $\mathbf{B}(n) = \mathbf{B}(n-1) + x(n)\overline{\mathbf{X}}(n-1) - x(n-N)\overline{\mathbf{X}}(n-N-1)$	$2K-2$	$2K$
(2) $e(n) = d(n) - \mathbf{X}^{\top}(n)\widehat{\mathbf{W}}(n) - \mathbf{B}^{\top}(n)\mathbf{E}(n-1)$	$N+K-1$	$K+N-1$
(3) $\mathbf{X}^{(i)}(n) = \Phi(n)\mathbf{H}_i\,,\ i = 1, 2, \ldots, p$	pK	$p(K-1)$
(4) $\mathbf{P}(n) = H\Lambda^{-1}(n)H^{\top}\mathbf{e}(n)$	$2pK+p$	$2pK-p-K$
(5) $\begin{bmatrix} \mathbf{E}(n) \\ z(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{E}(n-1) \end{bmatrix} + \mathbf{P}(n)$		$K-1$
(6) $\widehat{\mathbf{W}}(n+1) = \widehat{\mathbf{W}}(n) + z(n)\mathbf{X}(n-K+1)$	N	N
Total	$2N+3K(1+p)+p-3$	$2N+3K(1+p)-2p-2$

Table 2. The proposed Stereophonic fast weighted subband (SFWS) adaptive algorithm

Algorithm steps	multiplications	additions
Computation: For $n = 0, 1, 2, \ldots$		
(1) $\mathbf{B}(n) = \mathbf{B}(n-1) + \begin{bmatrix} x_1(n)\overline{\mathbf{X}}_1(n-1) \\ x_2(n)\overline{\mathbf{X}}_2(n-1) \end{bmatrix} - \begin{bmatrix} x_1(n-N)\overline{\mathbf{X}}_1(n-N-1) \\ x_2(n-N)\overline{\mathbf{X}}_2(n-N-1) \end{bmatrix}$	$4K-4$	$4K-4$
(2) $e(n) = d(n) - \mathbf{X}^{\top}(n)\widehat{\mathbf{W}}(n) - \mathbf{B}^{\top}(n)\mathbf{E}(n-1)$	$2N+2K-2$	$2N+2K-2$
(3) $\mathbf{X}_i^{(j)}(n) = \Phi_i(n)\mathbf{H}_j\,,\ i = 1, 2,\ j = 1, 2, \ldots, p$	$2pK$	$2p(K-1)$
(4) $\mathbf{P}_i(n) = H\Lambda_i^{-1}(n)H^{\top}\mathbf{e}(n)\,,\ i = 1, 2$	$3pK+2p$	$3Kp-2K-p-K$
(5) $\begin{bmatrix} \mathbf{E}_1(n) \\ z_1(n) \\ \mathbf{E}_2(n) \\ z_2(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{E}_1(n-1) \\ 0 \\ \mathbf{E}_2(n-1) \end{bmatrix} + \mathbf{P}(n)$		$2K-2$
(6) $\widehat{\mathbf{W}}(n+1) = \widehat{\mathbf{W}}(n) + \begin{bmatrix} z_1(n)\mathbf{X}_1(n-K+1) \\ z_2(n)\mathbf{X}_2(n-K+1) \end{bmatrix}$	$2N$	$2N$
Total	$4N+K(6+5p)$ $+2p-6$	$4N+K(6+5p)$ $-3p-8$

As in the monophonic case, if one can generate a decorrelated version of $\mathbf{X}(n)$ then the cross correlation between $x_1(n)$ and $x_2(n)$ is largely reduced and the bias will drop as a result as seen from Eq. (33). Also, convergence speed increases since the effect of coupling, as represented by the third term in Eq. (33) (and Eq. (34)) is reduced. Again, the decorrelation characteristics of the SFAP and SFWS can come very handy in this context to improve the performance of the stereophonic setup. An example is presented here to verify the above conclusions.

We examine the SFWS, the stereophonic FAP (SFAP) that is a straightforward extension of the single-channel one [2], and the SNLMS. The source signal $s(n)$ is zero mean white Gaussian with unity variance. The transmission room impulse responses $\mathbf{G}_1$ and $\mathbf{G}_2$ are each of length $L = 4096$, measured in a teleconferencing room [8]. The receiving room impulse responses $\mathbf{W}^{*}{}_1$ and $\mathbf{W}^{*}{}_2$ are also of length $M = 4096$, measured in a teleconferencing room, and all at 16kHz sampling rate. The length of the adaptive filter is $N = 1200$. Both the SFAP and SFWS have $p = 16$. The analysis filters are cosine-modulated perfect reconstruction filter banks with $K = 64$. Figure 8 demonstrates the decorrelating properties of the SFAP and SFWS under undermodeling conditions as they provide less misalignment than the SNLMS. While the SFAP has less misalignment than the SFWS, the SFWS has the practical advantages of numerical robustness discussed earlier. Moreover, Fig. 8 shows the fast convergence characteristic of the SFWS that rivals that of SFAP.

4 CONCLUSIONS

In this paper, we have presented a fast weighted subbands adaptive algorithm that leads to considerable improvement over the NLMS algorithm with a reasonable level of complexity. The new algorithm also performs as well as the FAP algorithm for the same p. The algorithm was extended for stereophonic acoustic echo cancellation. The arithmetic complexity of the fast monophonic algorithm amounts to $2N + 3K(1 + p) + p$ multiplications and $2N + 3K(1 + p) - 2p$ additions, and that of the fast stereophonic algorithm is $4N + K(6 + 5p) + 2p - 6$ multiplications and $4N + K(6 + 5p) - 3p - 8$ additions. It was demonstrated that the sample-by-sample weighted subband algorithm has similar approximate decorrelating property to the Affine Projection algorithm. This feature helps in lowering the coefficient bias of the adaptive filter that results from undermodeling the receiving room echo paths in the stereophonic as well as the monophonic case.

References

[1] COURVILLE, M.—DUHAMEL, P. : Adaptive Filtering in Subbands Using a Weighted Criterion, IEEE. Trans. Signal Processing **46** No. 9 (1998), 2359–2371.

[2] GAY, S.—TAVATHIA, S. : The Fast Affine Projection Algorithm, Proc. IEEE Int. Conf. Acoustic, Speech and Signal Processing, pp. 3023–3026, 1995.

[3] MONTAZERI, M.—DUHAMEL, P. : A Set of Algorithms Linking NLMS and Block RLS Algorithms, IEEE Trans. on Signal Processing **43** No. 2 (1995), 444–453.

[4] SHIMAUCHI, S.—MAKINO, S. : Stereo Projection Echo Canceller with True Echo Path Estimation, Proc. ICASSP'95, pp. 3059–3062, 1995.

[5] BENESTY, J.—DUHAMEL, P.—GRENIER, Y. : A Multi-Channel Affine Projection Algorithm with Applications to Multichannel Acoustic Echo Cancellation, IEEE Signal Processing Lett. **3** No. 2 (1996), 35–37.

[6] GAY, S. L.—TAVATHIA, S. : The Fast Affine Projection Algorithm, Proc. ICASSP'95, pp. 3023–3026, 1995.

[7] MAOUCHE, K.—SLOCK, D. T. : A Fast Instrumental Variable Affine Projection Algorithm, Proc. IEEE Int. Conf. Acoustic, Speech and Signal Processing, pp. 1481–1484, 1998.

[8] BENESTY, J.—MORGAN, D. R.—SONDHI, M. M. : A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation, IEEE Trans. Speech Audio Processing **6** (1998), 156–165.

[9] SHIMAUCHI, S.—MAKINO, S.—HANEDA, Y.—NAGAWA, A.—SAKAUCHI, S. : A Stereo Echo Canceller Implemented Using a Stereo Shaker and a Duo-Filter Control System, Proc. ICASSP'99, vol. 2, pp. 857–860, 1999.

[10] RUPP, M. : A Family of Adaptive Filter Algorithms with Decorrelating Properties, IEEE. Trans. Signal Processing **46** No. 3 (1998), 771–775.

[11] JUNG, Y. W.—LEE, J. H.—PARK, Y. C.—YOUN, D. H. : A New Adaptive Algorithm for Stereophonic Acoustic Echo Cancellation, Proc. ICASSP'2000, pp. 801–804, 2000.

[12] TANAKA, M.—KANEDA, Y.—MAKINO, S.—KOJIMA, J. : Fast Projection Algorithm and its Step Size Control, Proc. ICASSP'95, pp. 945–948, 1995.

[13] MORGAN, D. R.—KRATZER, S. G. : On a Class of Computationally Efficient, Rapidly Converging, Generalized NLMS Algorithms, IEEE Signal Processing Lett. **3** No. 8 (1996), 245–247.

[14] DOUGLAS, S. C. : Efficient Approximate Implementations of the Fast Affine Projection Algorithm Using Orthogonal Transforms, Proc. IEEE ICASSP'96, pp. 1656–1659, 1996.

[15] MAYYAS, K. : Stereophonic Acoustic Echo Cancellation Using Lattice Orthogonalization, IEEE Transactions on Speech and Audio Processing **10** No. 7 (2002), 517–525.

[16] MAYYAS, K. : MSE Analysis of A Reduced Complexity LMS-Type Algorithm for Stereophonic Acoustic Echo Cancellation, International Conference on Signal Processing, Pattern Recognition and Applications SPPRA 2003, pp. 709–712, 2003.

Received 21 August 2003

Khaled Mayyas received the BSE and MSc degrees from the Jordan University of Science and Technology, Irbid, Jordan, in 1990 and 1992 and the PhD degree from the University of Ottawa, Ottawa, Ont., Canada, in 1995, all in electrical engineering. He is currently an Associate Professor with the Department of Electrical Engineering, Jordan University of Science and Technology. He was a visiting Research Associate at the School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada, in summers of 96, 97, 98, 99, and 2002. His research interests are in various aspects of adaptive digital signal processing and their applications, including the development of efficient algorithms.