An Improved Error Estimation Algorithm for Stereophonic Acoustic Echo Cancellation Systems

T. Nguyen-Ky, J. Leis, W. Xiang
Faculty of Engineering and Surveying
University of Southern Queensland Toowoomba Queensland 4350 ;
Email: nguyentk@usq.edu.au , leis@usq.edu.au , xiangwei@usq.edu.au

Abstract— **In this paper, we propose an error estimate algorithm (EEA) for stereophonic acoustic echo cancellation (SAEC) that is based on an extension of the set-membership normalized least mean-squares (SM-NLMS) algorithm combined with the affine projection (AP) algorithm. In the EEA, with the minimum error signal fixed, we compute the filter lengths so that the error signal may approximate the minimum error signal. When the echo paths change, the adaptive filter automatically adjusts the filter lengths to the optimum values. We also investigate the difference between the adaptive filter lengths. In contrast with the conclusion in [1]–[6], our simulation results have shown that the filter lengths can be different. Our simulation results also confirm that the EEA is better than SM-NLMS algorithm in terms of echo return loss enhancement.**

I. Introduction

An echo is the phenomenon in which a delayed version of an original signal is reflected back to the source. Acoustic echo cancellers (AEC's) are necessary in applications such as mobile phones, hands-free telephony, speakerphones, audio and video conferencing. AEC's rely on an adaptive filter to estimate the echo paths and subsequently use this estimate to reduce the echo in transmitted signals. Typical adaptive algorithms for the filter update procedure in the AEC are the normalized least mean square (NLMS) [7], affine projection (AP) [8], [9], recursive least squares (RLS) [1] and fast recursive least squares (FRLS) [4].

The length of the acoustic echo path is dependent on the environment. Therefore, the computational complexity of the stereophonic acoustic echo cancellation (SAEC) may be very high and critically dependent on the echo cancellation algorithm. Using a long adaptive filter, the adaptive algorithm becomes very slow in terms of convergence speed and is more expensive to implement in terms of memory. In this paper, we present the error estimate algorithm (EEA) to optimize the filter lengths. We also investigate the differing between the adaptive filter lengths. With differing positions of the loudspeakers and the microphones (Fig. 1), the lengths of the acoustic echo paths in the receiving room are different. They will change when the environment is changed. To identify the echo paths in a stereophonic or multichannel system, the lengths of the adaptive filters have different values. Although many papers [1]–[6] assert that the filters lengths of two channel or multichannel systems should be equal, we will show in this paper that they do not necessarily have to be equal.

The organization of this paper is as follows. In Section 2, we introduce a stereophonic acoustic echo canceler. In Section 3, we present the error estimate algorithm for stereophonic acoustic echo canceler to optimize the filter lengths. Section 4 presents simulation results regarding the relationship between the filter lengths and the minimum error signal. The echo return loss enhancement (ERLE) and the convergence of the adaptive filter are also considered in this section. Finally, conclusions are given in Section 5.

II. Stereophonic acoustic echo cancellation system

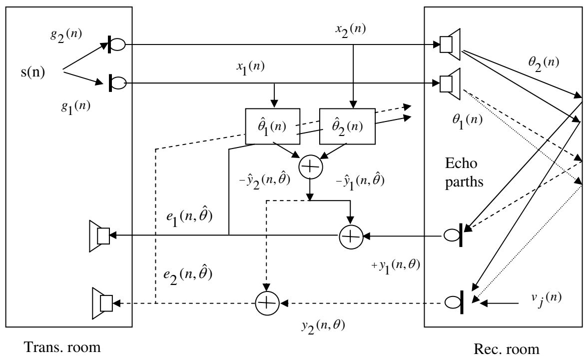


Fig. 1. Stereophonic acoustic echo cancellation system.

In a stereophonic acoustic echo cancellation system shown in Fig. 1, $s(n)$ is the source of $x_1(n)$ and $x_2(n)$ signals in the transmission room. We have $x_i(n) = g_i(n) * s(n)$, where $g_i(n)$ is the impulse response between the source and microphone in the transmission room. We define the echo signals in the receiving room as

$$y_j(n, \theta) = \sum_{i=1}^{2} \boldsymbol{\theta}^T \mathbf{x}_i(n) + v_j(n), \quad j = 1, 2, \tag{1}$$

and the impulse response vector at time n by

$$\boldsymbol{\theta} = [\theta_{ij,0}(n), \theta_{ij,1}(n), ..., \theta_{ij,L-1}(n)]^T \tag{2}$$

is the echo path (length L) of the receiving room between loudspeaker i and microphone j and $(\cdot)^T$ is the transpose.

$$\mathbf{x}_i(n) = [x_i(n), x_i(n-1), ..., x_i(n-L+1)]^T \tag{3}$$

is the far-end speech (loudspeaker) and v_j is the near-end speech added at microphone j in the receiving room. We assume that v_j is uncorrelated with $\mathbf{x}_i(n)$. The estimated response based on the least-squares fit $\hat{\theta}$ can be defined by

$$\hat{y}_j(n, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{2} \hat{\boldsymbol{\theta}}^T \mathbf{x}_i(n), \tag{4}$$

and the adaptive filter vector at time n by

$$\hat{\boldsymbol{\theta}} = [\hat{\theta}_{ij,0}(n), \hat{\theta}_{ij,1}(n), ..., \hat{\theta}_{ij,N_j-1}(n)]^T. \tag{5}$$

N_j is the length of the adaptive filter when we estimate the error $e_j(n)$. The error signal for the estimation is defined by

$$e_j(n) = y_j(n, \boldsymbol{\theta}) - \hat{y}_j(n, \hat{\boldsymbol{\theta}}) + v_j(n). \tag{6}$$

From (6), we can see that once the synthesized echo $\hat{y}_j(n, \hat{\boldsymbol{\theta}})$ is equal to the echo $y_j(n, \boldsymbol{\theta})$, the echo is completely cancelled and the signal transmitted to the transmission room is the near-end speech $v_j(n)$ only. This is the goal of the echo cancellation. An adaptive filter is used to identify the echo paths of the receiving room. The output of the adaptive filter, which is an estimate of the echo signal, can be used to cancel the undesirable echo. The estimated coefficients are chosen through an adaptive filter algorithm such that the cost function $J(e_j(n))$ is minimized. The estimation errors are labelled with two subscripts. The first subscript denotes the filter length N_j and the second subscript indicates the length L of the observed data [10]. We note that when the filter length N_j increases, the error decreases and vice versa. However, if the filter length increases, the adaptive filter algorithm becomes a rather expensive algorithm because its computational complexity grows in proportion to the check of the filter.

III. The error estimate algorithm for SAEC

Conventional filtering schemes estimate the parameter vector $\boldsymbol{\theta}$ so as to minimize a cost function $J(e_j(n))$ of the estimation error $e_j(n)$. The cost function is usually chosen to be a squared error measure $J(e_j(n)) = 1/2E[e_j(n)^2]$. The optimal $\hat{\boldsymbol{\theta}}$ parameters are found by solving $\nabla_{\hat{\boldsymbol{\theta}}} E[e_j(n)^2] = 0$ in [10]. In this section, we propose the new algorithm, achieving a specified bound δ on the magnitude of the estimation error $e_j(n)$ over a model space of interest. Any parameter estimate that results in the error being less than the specified bound δ (Fig. 2) is an acceptable solution. When the bound on the error is properly chosen [11]–[13], we have

$$|e_j(n)| \leq \delta, \qquad (\mathbf{x}_i(n), y_j(n, \boldsymbol{\theta})) \in \boldsymbol{\Omega}. \tag{7}$$

$\boldsymbol{\Omega}$ is the model space comprising input vector-desired output pairs over which we wish to impose the bounded error criterion. From (1-6), We have

$$y_1(n, \theta_1) = x_1(n)\theta_{11}^T + x_2(n)\theta_{12}^T, \tag{8}$$

$$y_2(n, \theta_2) = x_1(n)\theta_{21}^T + x_2(n)\theta_{22}^T, \tag{9}$$

with the transfer functions of the receiving room is

$$\boldsymbol{\theta} = [\theta_1, \theta_2]^T = [\theta_{11,L}, \theta_{12,L}, \theta_{21,L}, \theta_{22,L}]^T. \tag{10}$$

The adaptive filter is used to identify an unknown system (the loudspeaker-to-microphone transfer function in receiving room)

$$\hat{y}_1(n, \hat{\theta}_1) = x_1(n)\hat{\theta}_{11}^T + x_2(n)\hat{\theta}_{12}^T, \tag{11}$$

$$\hat{y}_2(n, \hat{\theta}_2) = x_1(n)\hat{\theta}_{21}^T + x_2(n)\hat{\theta}_{22}^T, \tag{12}$$

with the adaptive filter coefficients is

$$\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2]^T = [\hat{\theta}_{11,N_j}, \hat{\theta}_{12,N_j}, \hat{\theta}_{21,N_j}, \hat{\theta}_{22,N_j}]^T. \tag{13}$$

In Fig. 2, the error will be fedback and compared with δ until

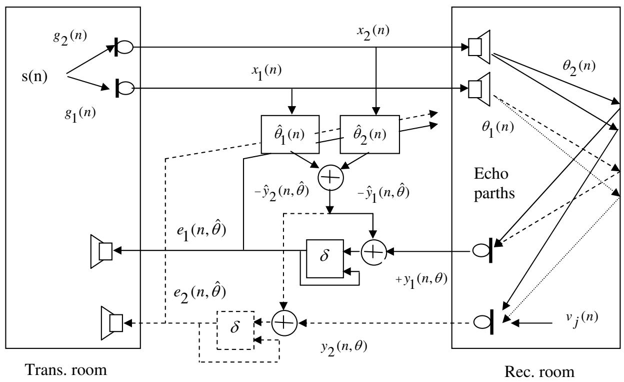


Fig. 2. Stereophonic acoustic echo cancellation control error system .

$|e_j(n)| \leq \delta$. We compare the new algorithm (EEA) with the set-membership normalized least mean-squares (SM-NLMS) algorithm [12].

Algorithm 1: *The set-membership normalized least mean-squares (SM-NLMS)*

Set-membership identification (SMI) theory is a well-established paradigm in the area of system identification that exploits the assumption of a bounded noise process added to a linear-in-parameter model. The set-membership normalized least mean-squares (SM-NLMS) algorithm was presented in [12]. The set-membership filtering criterion is to find $\hat{\theta}$ that satisfies

$$|e_j(n)|^2 \leq \delta^2, \qquad (x_i(n), y_j(n, \boldsymbol{\theta})) \in \boldsymbol{\Omega}. \tag{14}$$

We have

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu \frac{e_j(n)x_i(n)}{x_i^T(n)x_i(n)}, \quad i = 1, 2, \tag{15}$$

$$\mu = \begin{cases} 1 - \frac{\delta}{|e_j(n)|} & \text{if} \qquad |e_j(n)| > \delta, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Algorithm 2: *The error estimate algorithm*

From the SM-NLMS algorithm and affine projection (AP) algorithm, we propose the error estimate algorithm to optimize the filter length, as following

$$e_j(n) = y_j(n, \boldsymbol{\theta}) - \hat{y}_j(n, \hat{\boldsymbol{\theta}}), \tag{17}$$

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_j(n)x_i(n), \quad i = 1, 2, \tag{18}$$

$$\mu = P(n+1) = \left[P(n) - \frac{P(n)x_i(n)x_i^T(n)P(n)}{1 + x_i^T(n)P(n)x_i(n)} \right], \quad (19)$$

where μ is defined from AP algorithm [9].
Let $N_j = 5$ be the start length, the length of adaptive filter is computed by

$$N_j = \begin{cases} N_j + 1 & \text{if} \quad |e_j(n)| > \delta, \\ N_j & \text{if} \quad |e_j(n)| \leq \delta, \end{cases} \quad (20)$$

$$N = max(N_j). \quad (21)$$

When the algorithm is to converge, we have the simulation results in Table. I. In the error estimate algorithm, when the minimum error signal is fixed, we compute the filter lengths so that the error signal may approximate the minimum error signal δ. When the echo paths change, the adaptive filter automatically adjusts the filter lengths to the optimum values. The optimum value of the filter lengths is the N in (21). It is the necessary minimum value for the minimum error signal.

IV. SIMULATION RESULTS

(8) and (9) show the relationship between $x_1(n)$, $x_2(n)$ and $y_1(n)$, $y_2(n)$. Using the transfer function matrix in [14]

$$\mathbf{G}(x) = \begin{bmatrix} \frac{0.1x^{-8} - 0.3x^{-9}}{1+0.2x^{-1}-0.2x^{-2}} & \frac{0.01x^{-6}-0.03x^{-7}}{1+0.02x^{-1}-0.01x^{-2}} \\ \frac{-0.02x^{-7}-0.02x^{-8}}{1+0.01x^{-1}-0.01x^{-2}} & \frac{-0.2x^{-8}-0.3x^{-9}}{1+0.1x^{-1}-0.2x^{-2}} \end{bmatrix}, \quad (22)$$

we have

$$\theta_{11} = \frac{N_{11}}{D_{11}}, \quad \theta_{12} = \frac{N_{12}}{D_{12}},$$

$$\theta_{21} = \frac{N_{21}}{D_{21}}, \quad \theta_{22} = \frac{N_{22}}{D_{22}}, \quad (23)$$

$N_{11} = [0.1, -0.3]$,
$D_{11} = [1, 0.2, -0.2, 0, 0, 0, 0, 0, 0, 0]$,
$N_{12} = [0.01, -0.03]$,
$D_{12} = [1, 0.02, 0.01, 0, 0, 0, 0, 0]$,
$N_{21} = [-0.02, -0.02]$,
$D_{21} = [1, 0.01, -0.01, 0, 0, 0, 0, 0, 0]$,
$N_{22} = [-0.2, -0.3]$,
$D_{22} = [1, 0.1, -0.2, 0, 0, 0, 0, 0, 0, 0]$.

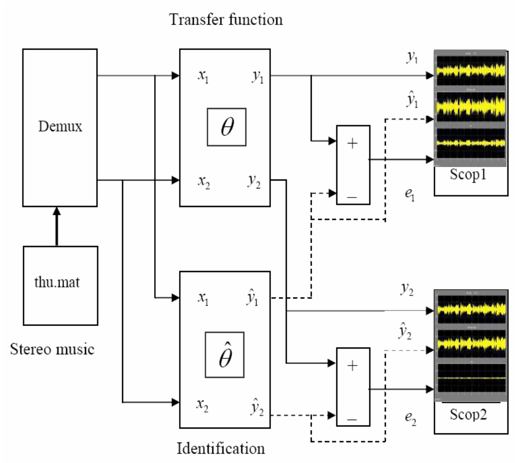


Fig. 3. SAEC control error simulation .

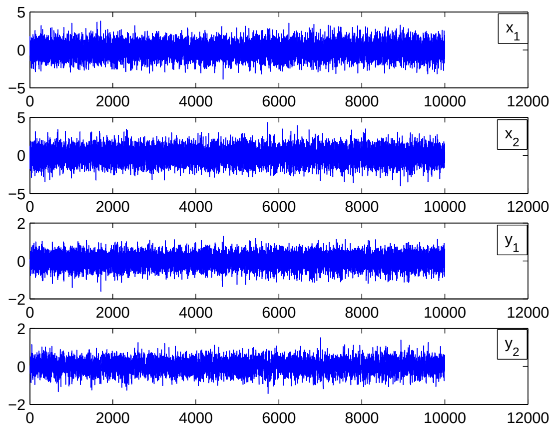


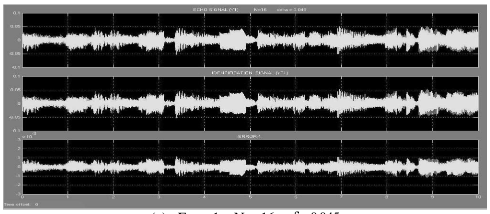
Fig. 4. x_1, x_2 input signals and y_1, y_2 output signals .

The structure shown in Figure 3 is the echo canceller simulation system. Using Matlab and Simulink software, we have the function blocks as:
- *Stereo music block*: read time and output values from the first matrix in the specified MAT file. In this block, `thu.mat` file is made from `thu.wav` file with T_{sam} sample time, $T_{sam} = \frac{1}{F_s} = \frac{1}{8\text{Khz}}$. In this paper, we used the stereo music file because stereo music frequency is higher than talk signal frequency. And the stereo music identification is more difficult than the talk signal identification.
- *Demux block*: split thu.mat file into two vector $x_1(n)$ and $x_2(n)$.
- *Transfer function block*: The top two panels of Figure 4 shown the loudspeaker signals, $x_1(n)$ and $x_2(n)$. The bottom two panels show the echo signals, $y_1(n)$ and $y_2(n)$. Equations

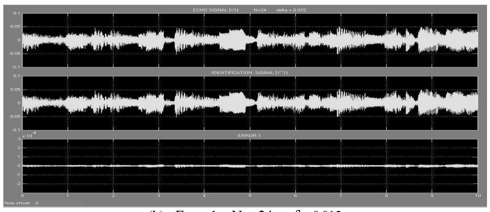
- *Identification block*: simulate the adaptive filter with the $\hat{\boldsymbol{\theta}}$ parameters. Equations (11) and (12) show the $\hat{y}_1, \hat{y}_2$ output signal. The Matlab programs compute the parameters $\hat{\boldsymbol{\theta}}$ and link these results with identification block.
- *Scop block*: The error signals between the echo signals and the identification signals are shown on the *scope*1 and *scope*2 (Fig. 5).
In this simulation, $v_j(n)$ is assumed to be zero when there is the stereo signal activity at the far-end. We consider the following parameters:
- The error signals,
- The relationship between δ and N_j,
- The echo return loss enhancement $(ERLE)$,
- The convergence between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$.

A. The error signals

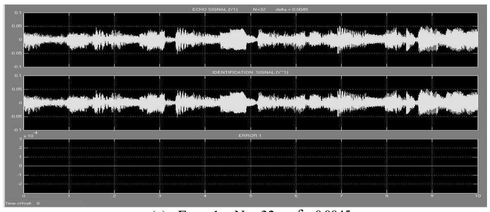
We can observe the y_1, $\hat{y}_1$ and e_1 signals of the scopes in Figure 3. In Figure 5, panel (a) show the error with $\delta = 45 \times 10^{-3}$, when the filter length have the value $N_1 = 16$. With $\delta = 15 \times 10^{-3}$ and $\delta = 4.5 \times 10^{-3}$, we have $N_1 = 24$ and $N_1 = 32$ in panels (b) and (c). When the δ decreases, the filter length increases.



(a) Error 1, N = 16, δ=0.045



(b) Error 1, N = 24, δ=0.015



(c) Error 1, N = 32, δ=0.0045

Fig. 5. y_1, $\hat{y}_1$ and e_1 signals with N=16, 24, 32.

B. Relationship between δ and N_j

Table I shows the simulation results for the relationship between the length N_j of the adaptive filter and the bounded error δ. When the bounded error decreases, the length N_j of the adaptive filter will increase. N_j is the necessary minimum length of the adaptive filter. Let N_1 and N_2 be the result when we use the EEA. When the bounded error criterion is chosen, the values of N_1 and N_2 are different. In [1]–[6], the authors defined the value of N_j to be the same. But in this paper, we show that they are different when we use the conditioning in (7). Therefore, when we design the adaptive filter, we must choose the length of adaptive filter equal to the maximum of N_j.

C. The echo return loss enhancement $(ERLE)$

Let $\sum_{n=1}^{k} y_j^2(n,\theta)$ be the power of the echo signal $y_j(n,\theta)$ at time n, and $\sum_{n=1}^{k} e_j^2(n)$ be the power of the residual-echo signal. The $ERLE$ is defined as

$$ERLE = 10 \log_{10} \frac{\sum_{n=1}^{k} y^2(n)}{\sum_{n=1}^{k} e^2(n)}. \quad (24)$$

There are two important performance measures for echo cancellation: the convergence rate and the steady-state residual

TABLE I
THE BOUNDED ERRORS AND THE LENGTHS OF THE ADAPTIVE FILTER

The bounded error δ	$e_1(n) < \delta$ N_1	$e_2(n) < \delta$ N_2
10×10^{-3}	25	45
7.0×10^{-3}	27	51
5.0×10^{-3}	31	58
3.0×10^{-3}	37	63
1.0×10^{-3}	52	74
0.7×10^{-3}	55	79
0.5×10^{-3}	57	82
0.3×10^{-3}	62	87
0.1×10^{-3}	77	98

echo. The steady-state residual echo equates the true echo subtracted by the synthesized echo after the algorithm is converged. We see that $ERLE$ is a measure of how good an echo canceller is in terms of steady-state residual echo and convergence time. To see the results more clearly, in Figure 6, we have presented plots showing the echo return loss enhancement (ERLE) of the lengths of the adaptive filter ($N = 16$, $N = 24$, $N = 32$). The ERLE is lower for the case $\delta = 10 \times 10^{-3}$ and higher for the case $\delta = 0.1 \times 10^{-3}$ than the ERLE compute with $\delta = 1 \times 10^{-3}$. When the δ decreases then the ERLE increases. This result confirms the high robustness of the EEA. We also compared the ERLE of

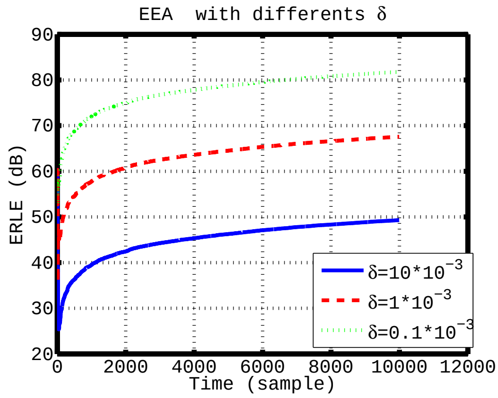


Fig. 6. ERLE with N=45, N=74, N=98.

the EEA with the SM-NLMS algorithm. Figure 7 shows the ERLE of the proposed EEA and the SM-NLMS algorithm. The ERLE simulation results show that the EEA is better than the SM-NLMS algorithm.

D. The convergence between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$

The adaptive filtering algorithm for echo cancellation adapts the adaptive filter ($\hat{\boldsymbol{\theta}}$) by minimizing the error between the echo and the synthesized echo. Once the error is minimized, $\hat{\boldsymbol{\theta}}$ is said to converge with the impulse respones $\boldsymbol{\theta}$. Figure 8 and 9 show the simulation results for the convergence of the adaptive filter with the EEA and the SM-NLMS algorithm. We set $\hat{\theta} \equiv \theta^*$, $\delta = 0.01$ $(N = 25)$. In Figure $9(b,c)$, we observe that $\hat{\boldsymbol{\theta}}_{12}$ is not to converge with $\boldsymbol{\theta}_{12}$. They only converge in Figure 8.

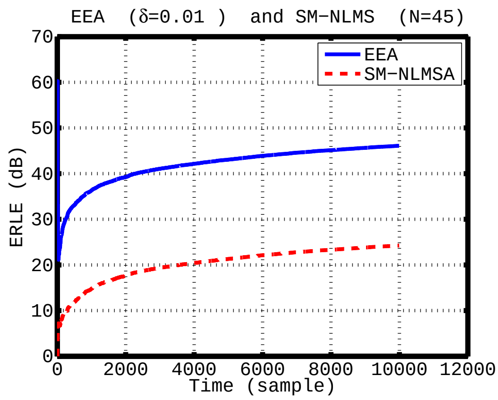


Fig. 7. ERLE of the EEA and SM-NLMS algorithm.

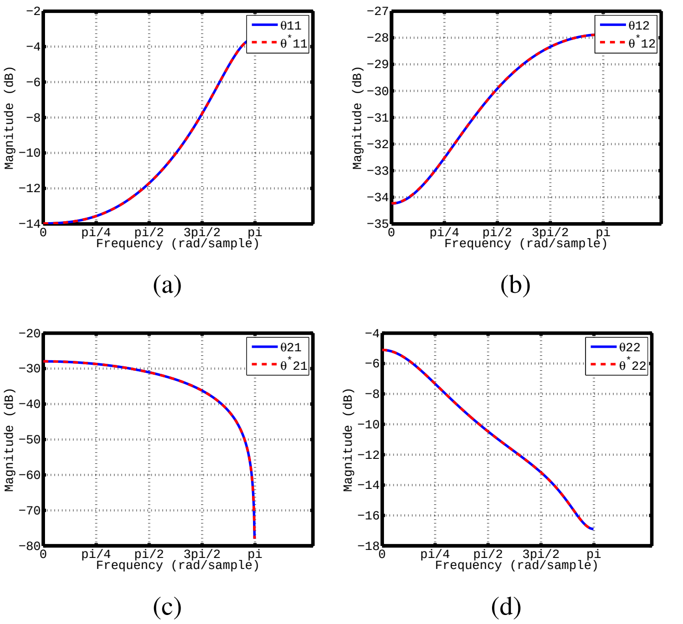


Fig. 8. Error estimate algorithm, magnitude response plots for: (a) $\theta_{11}, \hat{\theta}_{11}$ (b) $\theta_{12}, \hat{\theta}_{12}$ (c) $\theta_{12}, \hat{\theta}_{12}$ (d) $\theta_{22}, \hat{\theta}_{22}$.

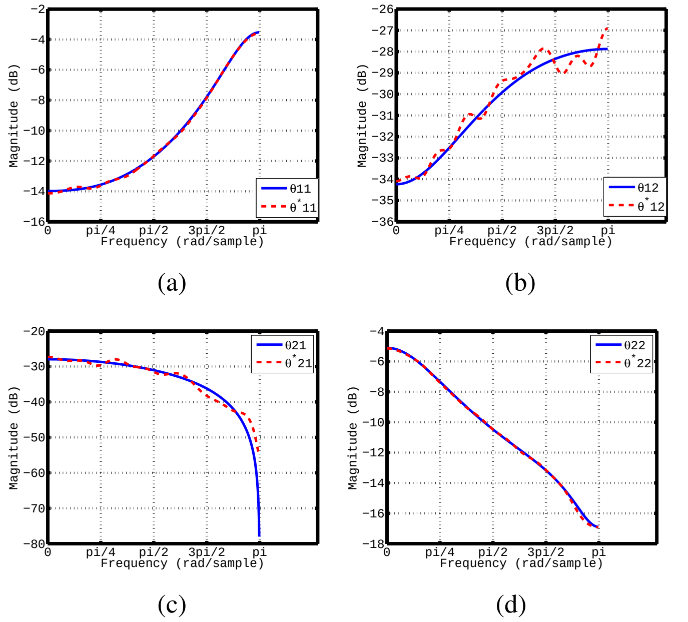


Fig. 9. Set-membership algorithm, magnitude response plots for: (a) $\theta_{11}, \hat{\theta}_{11}$ (b) $\theta_{12}, \hat{\theta}_{12}$ (c) $\theta_{12}, \hat{\theta}_{12}$ (d) $\theta_{22}, \hat{\theta}_{22}$.

V. CONCLUDING REMARKS

Stereophonic acoustic echo cancellation with two acoustic paths becomes problematic since the two excitation signals are highly correlated. In this paper, we propose the EEA for SAEC and compare the ERLE of the EEA with the SM-NLMS algorithm. Simulation results show that the ERLE of EEA is higher than the ERLE of SM-NLMS algorithm. Our results show the convergence of the EEA is better than that of the SM-NLMS algorithm. The results in Table. I also show that the filter lengths are different. This result will help us to choose the optimal filter length. Although we only discussed the two-channel case here, the approach can be extended to the multi-channel case.

REFERENCES

[1] T. Gänsler and J. Benesty. Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview. *International journal of adaptive control and Signal Processing*, 14:565–586, 2000.

[2] J. Benesty and T. Gänsler. A multichannel acoustic echo canceler double-talk detector based on a normalized cross-correlation matrix. *Acoustic echo and noise control*, 13(2):95–101, Mar. 2002.

[3] T. Gänsler and J. Benesty. A frequency-domain double-talk detector based on a normalized cross-correlation vector. *Signal Processing*, 81:1783–1787, 2001.

[4] P. Eneroth, S.L. Gay, T. Gänsler, and J. Benesty. A real-time implementation of a stereophonic acoustic echo canceler. *IEEE Trans. on speech and audio processing*, 9(5):513–523, July 2001.

[5] T. Gänsler and J. Benesty. New insights into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution. *IEEE Trans. on Speech and Audio Processing*, 10(5):257–267, July 2002.

[6] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay. *Advances in Network and Acoustic echo cancellation*. Springer-Verlag, Berlin, 2001.

[7] F. Lindstrom, C. Schüldt, and I. Claesson. Efficient multichannel NLMS implementation for acoustic echo cancellation. *EURASIP Journal on audio, speech and music processing*, 2007(ID 78439).

[8] H. Choi and H-D Bae. Subband affine projection algorithm for acoustic echo cancellation system. *EURASIP Journal on Advances in Signal Processing*, 2007(Article ID 75621).

[9] A. Carini and G. L. Sicuranza. Filtered-x affine projection algorithms for active noise control using volterra filters. *EURASIP Journal on Applied Signal Processing*, 2004(12):1841–1848.

[10] B. Farhang-Boroujeny. *Adaptive filters Theory and applications*. John Wiley and Sons, 1998.

[11] I. Yamada, K. Slavakis, and K. Yamada. An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques. *IEEE Trans. Signal Processing*, 50(5):1091–1101, May 2002.

[12] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y-F Huang. Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size. *IEEE Signal Processing Letters*, 5(5):111–114, May 1998.

[13] T. V. Waterschoot, G. Rombouts, P. Verhoeve, and M. Mooonen. Double-talk-robust prediction error identification algorithms for acoustic echo cancellation. *IEEE Trans. Signal Processing*, 55(3):846–858, Mar. 2007.

[14] Thomas Meurers, Sándor M. Veres, and Stephen J. Elliott. Frequency selective feedback for active noise control. *IEEE control systems magazine*, pages 32–41, Aug. 2002.