A1.11

ACOUSTIC ECHO CANCELLER ALGORITHM BASED ON THE VARIATION CHARACTERISTICS OF A ROOM IMPULSE RESPONSE

Shoji Makino Yutaka Kaneda

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180, Japan

ABSTRACT

This paper proposes a new NLMS (normalized LMS) adaptive algorithm for an acoustic echo canceller with double the convergence speed but the same computational load as the conventional NLMS. This algorithm, called the ES (Exponential Step) algorithm, implements a different value of step gain (feedback factor) for each tap coefficient of a canceller. These step gains are weighted exponentially along the series by the same exponential ratio as the expected variation in a room impulse response. This algorithm is implemented in a commercial subband echo canceller and its superiority to the conventional algorithm is demonstrated.

1. INTRODUCTION

In teleconferencing and hands-free telephony, acoustic feedback is a very annoying problem. An acoustic echo canceller is a key technology to overcome this problem. The echo canceller identifies the impulse response between a loudspeaker and a microphone to produce an echo replica which is then subtracted from the real echo.

Recursive LS algorithm and lattice filter are recent topics in adaptive algorithms. These methods converge faster than the LMS algorithm [1] and NLMS (normalized LMS) algorithm [2], but require excessive computation. On the other hand, LMS and NLMS require few computations, and are, therefore, widely applied for acoustic echo cancellers [3]. However, there is a strong need to improve the convergence speed of LMS/NLMS.

In LMS/NLMS, the step gain (feedback factor) in matrix form has been introduced [4,5] to obtain both high convergence speed and high ERLE (Echo Return Loss Enhancement). However, convergence speed is limited to the maximum speed that is attained by setting step gain matrix elements to $1/2\lambda_{max}$ for LMS and to unity for NLMS, where λ_{max} is the maximum eigenvalue of the input autocorrelation matrix.

Knowledge of the room impulse response is rarely used in the conventional LMS/NLMS. An adaptive algorithm suitable for the variation characteristics of an acoustic echo path is expected to improve convergence performance. In this study, the room impulse response was measured repeatedly, and the impulse response variation was studied to determine the statistics of variation in the room impulse response. Based on the results, a new NLMS algorithm called the ES (Exponential Step) algorithm is proposed which has double the convergence speed but the same computational load as the conventional NLMS.

2. CONVENTIONAL ACOUSTIC ECHO CANCELLER ALGORITHM

2.1 Configuration of an acoustic echo canceller

A configuration of an acoustic echo canceller is shown in Fig. 1. The echo canceller identifies the impulse response $\mathbf{h}(k)$ between a loudspeaker and a microphone. Since the impulse response varies as a person moves and varies with the surrounding environment, an adaptive FIR (Finite Impulse Response) filter $\hat{\mathbf{h}}(k)$ is used to identify $\mathbf{h}(k)$. An echo replica $\hat{y}(k)$ is created by convoluting $\hat{\mathbf{h}}(k)$ with the received input $\mathbf{X}(k)$. The echo replica $\hat{y}(k)$ is then subtracted from the real echo $y(k)$ to give the residual echo $e(k)$. $n(k)$ represents the ambient noise. The adaptive FIR filter $\hat{\mathbf{h}}(k)$ is updated to minimize the square of the residual echo in every sampling interval. An echo canceller algorithm should provide real time operation, fast convergence speed and high ERLE.

2.2 Conventional NLMS adaptive algorithm

The conventional NLMS is expressed by the equation

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \alpha \frac{e(k)}{\| \mathbf{X}(k) \|^2} \mathbf{X}(k) \qquad (1)$$

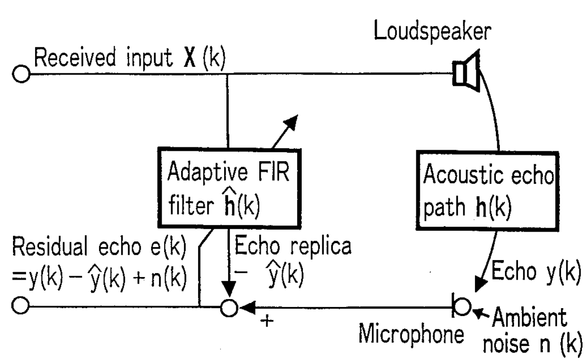


Fig. 1 Configuration of an acoustic echo canceller.

CH2847-2/90/0000-1133 $1.00 © 1990 IEEE

where
$\hat{\mathbf{h}}(k) = (\hat{h}_1(k), \hat{h}_2(k), \cdots, \hat{h}_L(k))^T$,
$\hat{h}_i(k)(i = 1, \cdots, L)$: tap coefficients of a canceller,
L: number of taps,
α: step gain (scalar, $0 < \alpha < 2$),
$e(k) = y(k) - \hat{y}(k) + n(k)$: residual echo,
$\mathbf{X}(k) = (x(k), x(k-1), \cdots, x(k-L+1))^T$: received input,
$\| \mathbf{X} \|$: norm of $\mathbf{X}$.

In the conventional NLMS, convergence speed is fastest at $\alpha = 1$, and it becomes slower and ERLE becomes higher as α becomes smaller.

3. NEW ADAPTIVE ALGORITHM

3.1 Variation in a room impulse response

The impulse response in a room varies for many reasons. Here, as one example and to simulate teleconferencing, we will discuss the case where three seated participants move in front of the microphone. Twenty-one impulse responses IR_i $(i = 1, \cdots, 21)$ were measured in a conference room with a reverberation time of 280 ms. The speaker-microphone distance was 1 m. Two of these waveforms IR_i and IR_j $(i \neq j)$, a variation $(\mathrm{IR}_i - \mathrm{IR}_j)$ and its standard deviation σ are shown in Fig. 2(a). The reverberent energy decay curves [6] of IR_i, IR_j, $\mathrm{IR}_i - \mathrm{IR}_j$ and σ are shown in Fig. 2(b). Figure 2 shows that impulse responses attenuate exponentially, and that the variation in these impulse responses attenuates by the same exponential ratio.

3.2 ES (Exponential Step) algorithm

Because of the variation characteristics of a room impulse response, the expected error in each coefficient becomes progressively smaller along the series by the same exponential ratio as the impulse response. Incorporating this knowledge into the conventional NLMS, we propose to update coefficients with large errors in large steps and coefficients with small errors in small steps. For this purpose, a step gain matrix $\mathbf{A}$ in a diagonal form is introduced,

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & & 0 \\ & \alpha_2 & \\ & & \ddots & \\ 0 & & & \alpha_L \end{pmatrix} \tag{2}$$

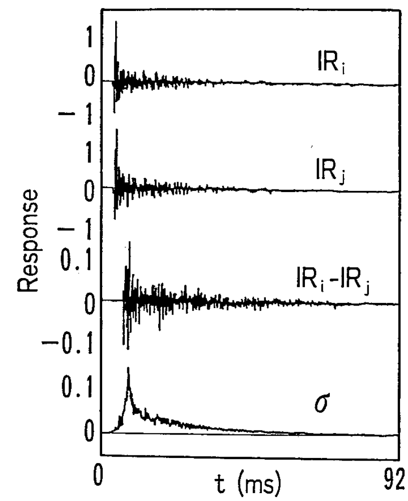
where elements α_i decrease in value exponentially from α_1 to α_L with the same ratio as the impulse response. The new adaptive algorithm, called the ES (Exponential Step) algorithm, is expressed as

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mathbf{A} \frac{e(k)}{\| \mathbf{X}(k) \|^2} \mathbf{X}(k). \tag{3}$$

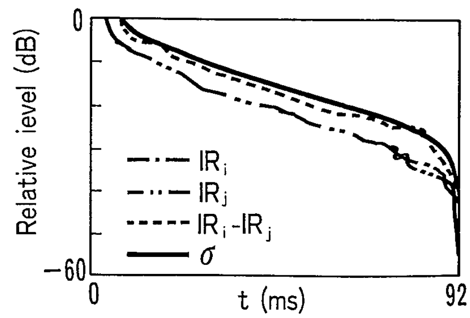
4. DISCUSSION

In this section, convergence characteristics of the proposed algorithm are discussed. A room impulse response is assumed to change at time $k = 0$, and $\hat{\mathbf{h}}(0) \approx \mathbf{h}_0'$

$$\mathbf{h}(k) = \begin{cases} \mathbf{h}_0' & (k < 0) \\ \mathbf{h}_0 & (k \geq 0). \end{cases}$$



(a) Impulse responses and their variation



(b) Reverberent energy decay curves

Fig. 2 Variation characteristics of a room impulse response. IR_i and IR_j are impulse responses. σ is the standard deviation of the variation $(\mathrm{IR}_i - \mathrm{IR}_j)$ $(i, j = 1, \cdots, 21)$ in a impulse response. Reverberation time at 500 Hz is 280 ms.

4.1 Coefficient error transition

Coefficient error vector $\mathbf{V}(k)$ is defined as

$$\mathbf{V}(k) = \mathbf{h}_0 - \hat{\mathbf{h}}(k). \tag{4}$$

Then, using equations (3), (4) and $e(k) = \mathbf{V}(k)^T \mathbf{X}(k) + n(k)$

$$\begin{aligned} \mathbf{V}(k+1) &= \mathbf{V}(k) - \frac{\mathbf{V}(k)^T \mathbf{X}(k)}{\| \mathbf{X}(k) \|^2} \mathbf{A}\, \mathbf{X}(k) \\ & \quad - \frac{n(k)}{\| \mathbf{X}(k) \|^2} \mathbf{A}\, \mathbf{X}(k). \end{aligned} \tag{5}$$

Here, following conditions are assumed.

1) $E[x(i)x(j)] = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j), \end{cases}$
2) $\mathbf{V}(k)$ and $\mathbf{X}(k)$ are independent,
3) $E[x(k)/ \| \mathbf{X}(k) \|^2] \approx E[x(k)]/ \mathrm{E}\, [\| \mathbf{X}(k) \|^2]$.

Then, the mean-square of the i-th component of $\mathbf{V}(k)$ is derived as

$$E\left[V_i(k+1)^2\right] = b_i(k+1)^2$$
$$= \; b_i(k)^2 - 2\frac{\alpha_i}{L} b_i(k)^2 + \frac{\alpha_i^2}{L^2}\sum_{j=1}^{L} b_j(k)^2 + \frac{\alpha_i^2}{L^2} p_n \qquad (6)$$

where $b_i(k)^2 = E[\; V_i(k)^2]$ and $p_n = E[n(k)^2]$.

Assuming that p_n can be neglected, equation (6) is rewritten in a matrix formula as

$$\mathbf{b}(k+1) = \mathbf{Q}\; \mathbf{b}(k) \qquad (7)$$

where

$$\mathbf{b}(k) = [b_1(k)^2, \cdots, b_L(k)^2]^T$$
$$(\mathbf{Q})_{ij} = \begin{cases} (1-(\alpha_i/L))^2 & (i=j) \\ (\alpha_i/L)^2 & (i \neq j) \end{cases}$$

Equations (6) and (7) give the transition formula of the mean-square coefficient error in the convergence process of the proposed algorithm.

4.2 Formula of α_i

The optimum step gain elements α_i which minimize $E[\mathbf{V}(k+1)^T\mathbf{V}(k+1)] = \sum_{j=1}^{L} b_j(k+1)^2$ can be derived from equation (6). When p_n can be neglected, they are

$$\alpha_i = \frac{L\; b_i(k)^2}{\sum_{j=1}^{L} b_j(k)^2} \qquad (i = 1, \cdots, L). \qquad (8)$$

This indicates that each step gain element should be proportional to the mean-square error of the corresponding coefficient.

Although $\{b_1(k)^2, \cdots, b_L(k)^2\}$ has the exponential decay characteristic of the room impulse response variation at $k = 0$, it changes as the algorithm proceeds to converge. This means that $b_i(k)^2$ should be estimated and α_i should be adjusted for each step of k. Here, time invariant step gain is considered for practical usage.

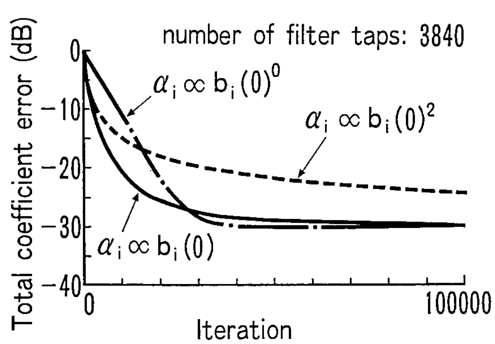


Fig. 3 Comparison of convergence characteristics for various step gain elements $(\alpha_1, \cdots, \alpha_L)$. Here, $\alpha_i = b_i(0)^0$ corresponds to the conventional NLMS and $b_i(0)$ represents the standard deviation of the impulse response variation.

Convergence characteristics are calculated using equation (7) for various exponential step gains $\alpha_i = b_i(0)^r$ with a parameter r. Here, $\alpha_i = b_i(0)^0 = 1$ corresponds to the the conventional NLMS with scalar step gain $\alpha = 1$ which results in the maximum convergence speed in the conventional NLMS. The results shown in Fig. 3 indicate that the best convergence speed is attained when α_i are set proportional to $b_i(0)$ $(r = 1)$.

4.3 Convergence condition

When the maximum squared eigenvalue λ_{max}^2 of $\mathbf{Q}$ in equation (7) is less than unity, the proposed algorithm converges. Many practical calculations of λ_{max}^2 and simulation results indicate that the convergence condition is satisfied when

$$0 < \bar{\alpha} = \frac{1}{L}\sum_{i=1}^{L} \alpha_i < 2. \qquad (9)$$

4.4 Steady-state ERLE

The steady-state ERLE can be calculated by setting $b_i(k+1)^2 = b_i(k)^2$ and summing $b_i(k+1)^2$ in equation (6) for all i. The results are expressed by the following equation.

$$\mathrm{ERLE}_{\infty} = \mathrm{SN} + 10\; log(\frac{2}{\bar{\alpha}} - 1)(dB) \qquad (10)$$

where $\mathrm{ERLE}_{\infty} = 10\; log(p_y/p_e)$, $\mathrm{SN} = 10\; log(p_y/p_n)$, $p_y = E[y(k)^2]$ and $p_e = E[e(k)^2]$.

Equation (10) is the same formula as the equation of the steady-state ERLE in the conventional NLMS with α replaced by $\bar{\alpha}$.

This exponentially attenuating step gain can be applied to other LMS type algorithms.

5. PRACTICAL MODIFICATION

In a practical system constructed with multiple DSP (Digital Signal Processor) chips, step gain α_i is set in discrete steps with one constant value per DSP chip as shown in Fig. 4. This practical modification allows the proposed algorithm to have the same computational load as the conventional NLMS.

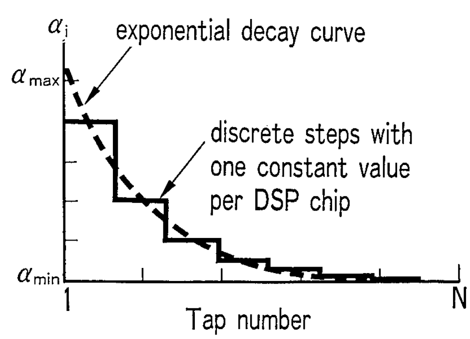


Fig. 4 Step gain element α_i of matrix $\mathbf{A}$ when α_i is set in discrete steps with one constant value per DSP chip.

6. REAL TIME EXPERIMENTS

The proposed algorithm was implemented in a commercial acoustic echo canceller constructed with multiple DSP chips [7], and real-time experiments were performed. The 7-kHz frequency range was separated into two bands by the subband technique. The sampling frequency was 8 kHz in each band. The specifications of each band are shown in Table 1. For real-time operation, α_i was set to α_{1-256}=2.5, $\alpha_{257-512}$=1.5 and $\alpha_{513-last}$=0.3. The speaker-microphone distance was 2.5 m and reverberation time of the conference room was 300 ms.

Figure 5 shows the results of the real-time experiments on residual echo level convergence performance. When using a white noise input signal (Fig. 5(a)), the residual echo level decayed to −20 dB at triple the speed of the maximum speed (α =1.0) for the conventional NLMS. When using a speech input signal (Fig. 5(b)), the residual echo level decayed to −20 dB at double the speed of the maximum speed (α =1.0) for the conventional NLMS. ERLE was over 30 dB in the 7-kHz frequency range.

7. CONCLUSION

A new NLMS (normalized LMS) adaptive algorithm, called the ES (Exponential Step) algorithm, for an acoustic echo canceller has been developed. In this algorithm, each tap coefficient of a canceller is adjusted by a different value of step gain. These step gain values α_i are determined proportionally to the expected variation in a room impulse response. This algorithm only requires the same computational load as the conventional NLMS by modifying α_i in a practical multiple DSP structure. Convergence study indicates that the mean value of step gain elements decides the convergence condition and the steady-state ERLE and that some individual step gain elements can exceed the value 2. This algorithm has been implemented in a commercial subband echo canceller constructed with multiple DSP chips. Experiments in a conference room show that this algorithm converges faster than the conventional NLMS: triple the speed for a white noise signal and double the speed for a speech signal.

REFERENCES

[1] B. Widrow, et al.: "Adaptive noise cancelling: Principles and applications", Proc. IEEE, 63, 12, pp.1692-1716 (1975).

[2] J. Nagumo, A. Noda: "A learning method for system identification", IEEE Trans., AC-12, pp.282-287 (1967).

[3] H. Oikawa, S. Minami, T. Saeki: "A new echo canceller realized by high performance digital signal processor", ISCAS88, pp.1329-1332 (1988).

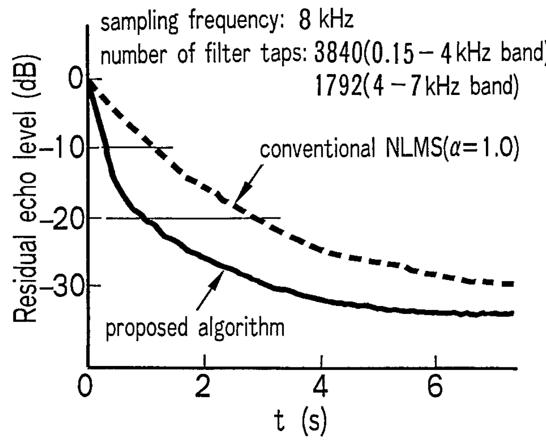
[4] R. W. Harris, D. M. Chabries, F. A. Bishop: "A variable step (VS) adaptive filter algorithm", IEEE Trans., ASSP-34, 2, pp.309-316 (1986).

[5] J. Chen, J. Vandewalle: "A μ-vector LMS adaptive system for enhancing nonstationary narrow-band signal", ISCAS88, pp.771-774 (1988).

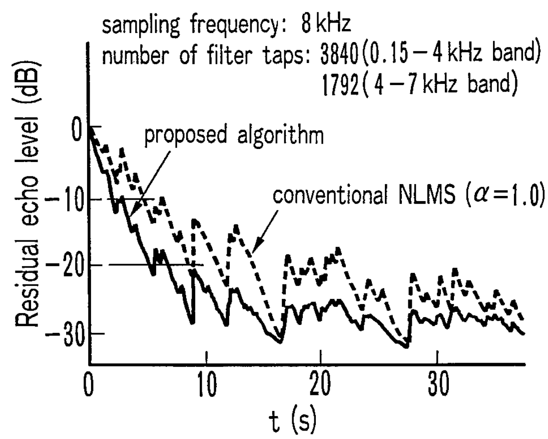
[6] M. R. Schroeder: "New method of measuring reverberation time", J. Acoust. Soc. Am., 37, pp.409-412 (1965).

Table 1 The specifications of experimental setup.

band	frequency range	number of taps	delay processing time
lower	0.15-4 kHz	3840 taps	480 ms
higher	4-7 kHz	1792 taps	224 ms



(a) Input signal: white noise



(b) Input signal: speech (male)

Fig. 5 Real time experimental results on convergence performance using a commercial subband echo canceller constructed with multiple DSP chips. Room reverberation time at 500 Hz is 300 ms.

[7] T. Kaneko, H. Yamauchi, A. Iwata: "A 50 ns floating-point signal processor VLSI", ICASSP86, pp.401-404 (1986).