Exponentially Weighted Stepsize NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response

Shoji Makino, *Member, IEEE,* Yutaka Kaneda, *Member, IEEE,* and Nobuo Koizumi

Abstract—**This paper proposes a new normalized least-mean-squares (NLMS) adaptive algorithm with double the convergence speed, at the same computational load, of the conventional NLMS for an acoustic echo canceller. This algorithm, called the ES (exponentially weighted stepsize) algorithm, uses a different stepsize (feedback constant) for each weight of an adaptive transversal filter. These stepsizes are time-invariant and weighted proportional to the expected variation of a room impulse response. The algorithm is based on the fact that the expected variation of a room impulse response becomes progressively smaller along the series by the same exponential ratio as the impulse response energy decay. As a result, the algorithm adjusts coefficients with large errors in large steps, and coefficients with small errors in small steps. A transition formula is derived for the mean-squared coefficient error of the proposed algorithm. The mean stepsize determines the convergence condition, the convergence speed, and the final excess mean-squared error. The algorithm is modified for a practical multiple DSP structure, so that it requires only the same amount of computation as the conventional NLMS. The algorithm is implemented in a commercial acoustic echo canceller and its fast convergence is demonstrated.**

I. Introduction

An ACOUSTIC echo canceller can overcome the acoustic feedback that interferes with teleconferencing and hands-free telecommunication. It adaptively identifies the transfer function between a loudspeaker and a microphone, and then produces an echo replica which is subtracted from the real echo.

Various adaptive algorithms are applicable to an acoustic echo canceller. The recursive least-squares (RLS) algorithm [1] provides fast convergence at the price of a high computational load. Recently developed fast RLS algorithms still require excessive computation [2], [3]. The least-mean-squares (LMS) algorithm [4], [5], on the other hand, is robust and simple. The normalized LMS (NLMS) algorithm [6], whose convergence speed is independent of input signal power, is widely used in commercial acoustic echo cancellers [7], [8]. However, the major drawback of the LMS and NLMS algorithms is their slow convergence. For example, the mean-squared error in the NLMS takes 2 s to converge for a white noise input signal and 10 s for speech for 8-kHz sampling rate and a filter with an order of 4000. Therefore, there is a strong need to increase the convergence speed of the LMS and NLMS.

A stepsize parameter (feedback constant), used in many gradient-type adaptive algorithms, controls the convergence rate of the filter coefficients but also determines the final excess mean-squared error from the Wiener solution. Therefore, both a time-varying stepsize and a time-varying matrix-form stepsize [9] have been introduced to obtain fast convergence in the transient state and a small excess mean-squared error in the steady state. These time-varying stepsize algorithms, however, require complicated control of the stepsize. On the other hand, the convergence speed of the conventional NLMS with a time-invariant stepsize has a maximum, attained when the stepsize is unity for white noise [10].

Knowledge of the room impulse response is rarely used in conventional algorithms. An adaptive algorithm suited to the variation of an acoustic echo path is expected to improve convergence. In this paper, the room impulse response was measured repeatedly, and the impulse response variation was studied to determine its statistical characteristics. Based on the results, the ES (exponentially weighted stepsize) algorithm is proposed.

This paper is organized as follows. A brief review of acoustic echo cancellers and some conventional adaptive algorithms are given in Section II. The new adaptive algorithm is derived in Section III. Section IV discusses the properties of the proposed algorithm, followed by the modifications for a practical acoustic echo canceller in Section V. The real-time experimental results are shown in Section VI. Section VII summarizes the paper.

II. Acoustic Echo Cancellers and Conventional Adaptive Algorithms

A. Configuration of an Acoustic Echo Canceller

The configuration of an acoustic echo canceller is shown in Fig. 1. The echo canceller identifies the transfer function of the acoustic echo path, *i.e.,* the impulse response $\boldsymbol{h}(k)$ between the loudspeaker and the microphone, where $\boldsymbol{h}(k) = [h_1(k), h_2(k), \cdots]^T$ and $h_1(k), h_2(k), \cdots$ represent coefficients of the impulse response at discrete time k. Since the impulse response $\boldsymbol{h}(k)$ varies as a person moves and varies with the environment, an adaptive filter $\hat{\boldsymbol{h}}(k)$ is used to identify $\boldsymbol{h}(k)$. Usually, $\hat{\boldsymbol{h}}(k)$ is a finite impulse-response (FIR) filter

Manuscript received May 6, 1991; revised May 5, 1992.
The authors are with NTT Human Interface Laboratories, Musashino-shi, Tokyo, 180 Japan.
IEEE Log Number 9203886.

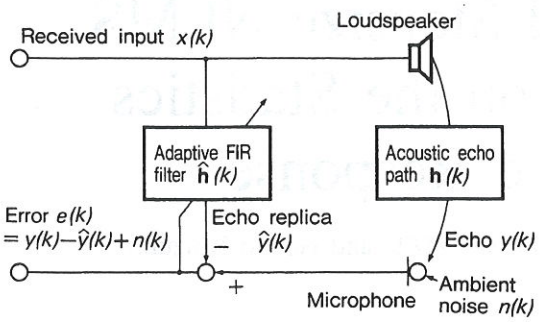

Fig. 1. Configuration of an acoustic echo canceller.

because a wide range of simple adaptive algorithms have been proposed for an FIR filter.

An echo replica $\hat{y}(k)$ is created by convolving $\hat{\boldsymbol{h}}(k)$ with the received input vector $\boldsymbol{x}(k)$, where $\boldsymbol{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T$ and L represents the number of taps. The echo replica $\hat{y}(k)$ is then subtracted from the real echo $y(k)$ to give the error $e(k) = y(k) - \hat{y}(k) + n(k)$, where $n(k)$ represents the ambient noise. In the double-talk situation where the near-end speech is added to the microphone, it is common to disable the adaptation in the echo canceller. Therefore, the near-end speech is disregarded in the noise term $n(k)$. The adaptive FIR filter $\hat{\boldsymbol{h}}(k)$ is adjusted to decrease the error power in every sampling interval. The adaptive algorithm should provide real-time operation, fast convergence, and high echo return loss enhancement (ERLE, defined as the ratio of the real echo power to the error power excluding the ambient noise).

B. Conventional Adaptive Algorithms

1) RLS Algorithm: The RLS algorithm [1] updates the filter coefficient vector $\hat{\boldsymbol{h}}(k)$ according to the following equations:

$$\hat{\boldsymbol{h}}(k+1) = \hat{\boldsymbol{h}}(k) + \boldsymbol{k}(k)e(k) \tag{1}$$

$$\boldsymbol{k}(k) = \frac{\nu^{-1}\boldsymbol{P}(k)\boldsymbol{x}(k)}{R + \nu^{-1}\boldsymbol{x}(k)^T\boldsymbol{P}(k)\boldsymbol{x}(k)} \tag{2}$$

$$\boldsymbol{P}(k+1) = \nu^{-1}\boldsymbol{P}(k) - \nu^{-1}\boldsymbol{k}(k)\boldsymbol{x}(k)^T\boldsymbol{P}(k) \tag{3}$$

$$e(k) = y(k) - \hat{\boldsymbol{h}}(k)^T\boldsymbol{x}(k) + n(k) \tag{4}$$

where,

$\hat{\boldsymbol{h}}(k) = [\hat{h}_1(k), \hat{h}_2(k), \cdots, \hat{h}_L(k)]^T$,
$\boldsymbol{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T$: received input vector,
$\hat{h}_i(k)$ where $i = 1, \cdots, L$, coefficients of an FIR filter,
L number of taps,
$\boldsymbol{k}(k)$ Lth-order gain vector,
$\boldsymbol{P}(k)$ $L \times L$ matrix,
ν forgetting factor $(0 < \nu < 1)$,
R variance of $n(k)$.

$\boldsymbol{P}(k)$ is defined as the inverse of the input autocorrelation matrix, and can also be regarded as the coefficient error autocorrelation matrix for constant impulse response $\boldsymbol{h}_0$, *i.e.*,

$$\boldsymbol{P}(k) = E[\{\boldsymbol{h}_0 - \hat{\boldsymbol{h}}(k)\}\{\boldsymbol{h}_0 - \hat{\boldsymbol{h}}(k)\}^T] \tag{5}$$

where $E[\cdot]$ represents the statistical expectation.

The important property of the RLS algorithm is its fast convergence. When $\boldsymbol{h}(k)$ is time-invariant, the mean-squared error in the RLS converges in $2L$ iterations [1]. However, it demands a high computational load, requiring $O(L^2)$ multiply–add operations for each update represented by (1)–(4). Recently developed fast RLS algorithms still require more than $7L$ multiply–add operations [2], [3]. One has been implemented in a prototype of a subband acoustic echo canceller [11], but none have yet been implemented in commercial acoustic echo cancellers.

2) NLMS Algorithm: The NLMS algorithm [6] is expressed by the following equations:

$$\hat{\boldsymbol{h}}(k+1) = \hat{\boldsymbol{h}}(k) + \alpha\frac{e(k)}{\|\boldsymbol{x}(k)\|^2}\boldsymbol{x}(k) \tag{6}$$

$$e(k) = y(k) - \hat{\boldsymbol{h}}(k)^T\boldsymbol{x}(k) + n(k) \tag{7}$$

where α is the scalar stepsize $(0 < \alpha < 2)$ and $\|\cdot\|$ is the Euclidean norm.

The filter coefficient vector $\hat{\boldsymbol{h}}(k)$ is adjusted in proportion to both the error $e(k)$ and the received input $\boldsymbol{x}(k)$. Convergence of the mean-squared coefficient error is guaranteed when $0 < \alpha < 2$, and is fastest at $\alpha = 1$ for white noise. The important property of the NLMS is the relatively small computational load. It requires only $2L$ multiply–add operations for each update in (6) and (7). As a result, the NLMS is used in almost all commercial acoustic echo cancellers. However, it converges very slowly. Assuming $\boldsymbol{h}(k)$ is time-invariant, the mean-squared error in the NLMS converges in about $20L$ iterations for a white noise input signal [1], ten times slower than the RLS.

The NLMS algorithm can be regarded as a simplified version of the RLS. Actually, (6) can be derived from (1)–(3) by setting $R = 0, \nu = 1$, and $\boldsymbol{P}(k) = \boldsymbol{I}$, where $\boldsymbol{I}$ is the unit matrix, and by introducing the stepsize α. Many other algorithms could be classified as being between the RLS and the NLMS. In other words, they can be explained as simplified versions of the RLS, but not as simple as the NLMS. They usually converge faster than the NLMS but slower than the RLS, with a computational load larger than the NLMS but smaller than the RLS.

III. New Adaptive Algorithm

A. Variation of a Room Impulse Response

Variation of a room impulse response, *i.e.*, the amount of change in a room impulse response waveform, was investigated to design an adaptive algorithm suitable for an acoustic echo canceller. The impulse response in a room varies for many reasons. Here, as one example and to simulate teleconferencing, we will discuss two situations.

In the first situation, the distance between the loudspeaker and the microphone changed. Impulse responses were measured for loudspeaker–microphone distances of 1 m (IR_1) and 0.6 m (IR_2) in a conference room. Reverberation time at 500 Hz was 350 ms. These waveforms IR_1 and IR_2, and the variation $(IR_2 - IR_1)$ are shown in Fig 2(a). The reverberent

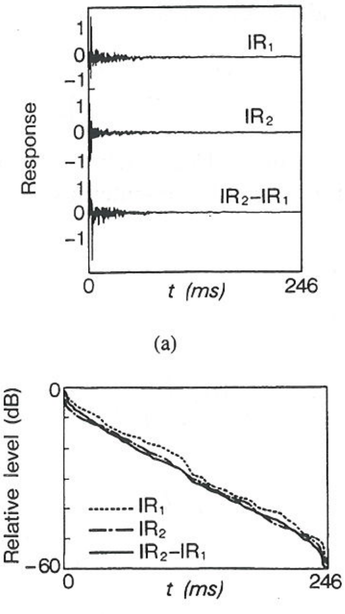

Fig. 2. Variation of a room impulse response when the loudspeaker–microphone distance is changed. IR_1 and IR_2 are impulse responses for loudspeaker–microphone distances of 1 and 0.6 m, respectively. Room reverberation time at 500 Hz is 350 ms. (a) Impulse responses and their variation. (b) Reverberant energy decay curves.

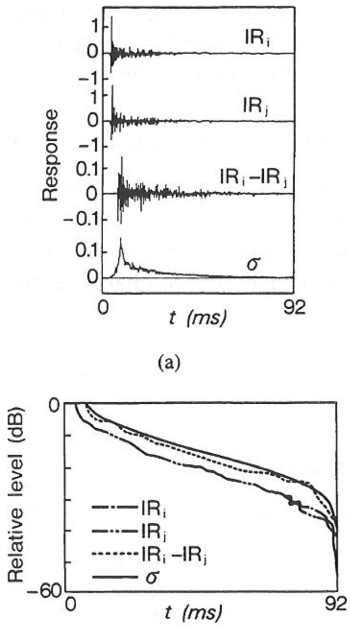

Fig. 3. Variation of a room impulse response when the participants move. IR_i and IR_j are impulse responses. σ is the standard deviation of the variation $(IR_i - IR_j)\,(i,j = 1.\cdots.21)$ of the impulse response. Loudspeaker–microphone distance is 1 m. Room reverberation time at 500 Hz is 280 ms. (a) Impulse responses and their variation. (b) Reverberant energy decay curves.

energy decay curves [12] of IR_1, IR_2, and $IR_2 - IR_1$ are shown in Fig. 2(b).

In the second situation, three seated participants moved around 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 $(i \neq j)$, the variation $(IR_i - IR_j)$, and its standard deviation σ are shown in Fig. 3(a). The reverberent energy decay curves [12] of $IR_i, IR_j, IR_i - IR_j$, and σ are shown in Fig. 3(b).

Figs. 2 and 3 show that impulse responses attenuate exponentially, and that the variation of these impulse responses also attenuates by the same exponential ratio.

The exponential attenuation ratio γ is the same for impulse responses in the same room. It can be derived from the reverberation time which is determined by the acoustical conditions of the room, *i.e.*, size and absorption coefficient. Thus we can estimate the exponential attenuation ratio γ from the room conditions, or determine it by measuring one impulse response.

B. Exponentially Weighted Stepsize (ES) Algorithm

Because of the variation 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 adjust coefficients with large errors in large steps and coefficients with small errors in small steps. For this purpose, a stepsize matrix $\boldsymbol{A}$ with diagonal form is introduced:

$$\boldsymbol{A} = \begin{pmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_L \end{pmatrix} \tag{8}$$

where $\alpha_i = \alpha_0 \gamma^{i-1}\,(i = 1, \cdots L)$ and γ is the exponential attenuation ratio $(0 < \gamma < 1)$.

Elements α_i are time-invariant and decrease exponentially from α_1 to α_L with the same ratio γ as the impulse response $\boldsymbol{h}(k)$. The new adaptive algorithm, called the ES (exponentially weighted stepsize) algorithm [13], is expressed as

$$\hat{\boldsymbol{h}}(k+1) = \hat{\boldsymbol{h}}(k) + \boldsymbol{A}\,\frac{e(k)}{\|\boldsymbol{x}(k)\|^2}\,\boldsymbol{x}(k) \tag{9}$$

$$e(k) = y(k) - \hat{\boldsymbol{h}}(k)^T \boldsymbol{x}(k) + n(k). \tag{10}$$

The scalar stepsize α in (6) is replaced by a stepsize matrix $\boldsymbol{A}$ in (9). This algorithm can also be derived from the RLS by setting the matrix $\boldsymbol{P}(k)$ in the numerator of (2) to the constant $\boldsymbol{A}$ and $\boldsymbol{P}(k)$ in the denominator to $\boldsymbol{I}$, the unit matrix, ν to 1, and by neglecting R. Since each element of the vector $e(k)\boldsymbol{x}(k)/\|\boldsymbol{x}(k)\|^2$ is multiplied by element α_i, this algorithm requires an additional L multiplications. However, this can be avoided with the modifications described in Section V. Consequently, this algorithm requires only $2L$ multiply–add operations (the same as the NLMS).

The exponentially weighted diagonal stepsize matrix was also proposed in [14] for application to adaptive line enhancers. However, the algorithm in [14] aims to reduce the effect of old input data. It reflects the degree of variation in the input signal statistics, and the exponential attenuation ratio is determined only by how nonstationary the input signal is.

IV. Properties of the Proposed Algorithm

This section discusses various properties of the ES algorithm. The canceller is assumed to have identified the room impulse response well before time $k = 0$, and the room impulse response is assumed to change at time $k = 0$, *i.e.*,

$$\boldsymbol{h}(k) = \begin{cases} \boldsymbol{h}_0', & k < 0 \\ \boldsymbol{h}_0, & k \geq 0 \end{cases} \tag{11}$$

$$\hat{\boldsymbol{h}}(0) \approx \boldsymbol{h}_0'. \tag{12}$$

A. Coefficient Error Transition

The coefficient error vector $\boldsymbol{v}(k)$ is defined as

$$\boldsymbol{v}(k) = \boldsymbol{h}_0 - \hat{\boldsymbol{h}}(k). \tag{13}$$

Then, combining (13) and (9), and using $e(k) = \boldsymbol{v}(k)^T\boldsymbol{x}(k) + n(k)$:

$$\boldsymbol{v}(k+1) = \boldsymbol{v}(k) - \frac{\boldsymbol{v}(k)^T\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}\boldsymbol{A}\boldsymbol{x}(k) - \frac{n(k)}{\|\boldsymbol{x}(k)\|^2}\boldsymbol{A}\boldsymbol{x}(k). \tag{14}$$

The ith component of vector $\boldsymbol{v}(k)$ is given by

$$v_i(k+1) = v_i(k) - \frac{\sum_{j=1}^{L} v_j(k)x(k-j+1)}{\sum_{j=1}^{L} x(k-j+1)^2}\alpha_i x(k-i+1) - \frac{n(k)}{\sum_{j=1}^{L} x(k-j+1)^2}\alpha_i x(k-i+1). \tag{15}$$

Here, $\boldsymbol{x}(k), \boldsymbol{v}(k)$, and $n(k)$ are assumed to have zero mean and to be mutually statistically independent stationary signals. The elements of vector $\boldsymbol{x}(k)$ are assumed to be mutually uncorrelated, *i.e.*,

1) $E[x(k-i+1)x(k-j+1)] = \begin{cases} p_x, & i = j \\ 0, & i \neq j. \end{cases}$

Assuming that the number of taps L is large enough, the following approximations hold:

2) $\|x(k)\|^2 \approx L \cdot p_x, \qquad$ for all k,

3) $\sum_{i=1}^{L} E[v_i(k)^2]E[x(k-i+1)^2 x(k-j+1)^2] \approx p_x^2 \sum_{i=1}^{L} E[v_i(k)^2].$

Based on these assumptions, taking the expected values of the mean-square of (15) for $\boldsymbol{x}(k), \boldsymbol{v}(k)$, and $n(k)$ gives

$$\begin{aligned} E[v_i(k+1)^2] &= b_i(k+1)^2 \\ &= b_i(k)^2 - 2\frac{\alpha_i}{L}b_i(k)^2 \\ &\quad + \frac{\alpha_i^2}{L^2}\sum_{j=1}^{L} b_j(k)^2 + \frac{\alpha_i^2}{L^2}\frac{p_n}{p_x} \end{aligned} \tag{16}$$

where $b_i(k)^2$ and p_n are defined as

$$b_i(k)^2 = E[v_i(k)^2]$$
$$p_n = E[n(k)^2].$$

When the last term in (16), which is due to ambient noise, can be ignored, (16) can be rewritten as

$$\begin{pmatrix} b_1(k+1)^2 \\ b_2(k+1)^2 \\ \vdots \\ b_L(k+1)^2 \end{pmatrix} = \begin{pmatrix} (1-\alpha_1/L)^2 & (\alpha_1/L)^2 & \cdots & (\alpha_1/L)^2 \\ (\alpha_2/L)^2 & (1-\alpha_2/L)^2 & \cdots & (\alpha_2/L)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha_L/L)^2 & (\alpha_L/L)^2 & \cdots & (1-\alpha_L/L)^2 \end{pmatrix} \cdot \begin{pmatrix} b_1(k)^2 \\ b_2(k)^2 \\ \vdots \\ b_L(k)^2 \end{pmatrix} \tag{17}$$

or in matrix form:

$$\boldsymbol{b}(k+1) = \boldsymbol{Q}\boldsymbol{b}(k). \tag{18}$$

Equations (16)–(18) give the transition formula of the mean-squared coefficient error of the ES algorithm.

B. Formula for α_i

The optimum stepsizes α_i which minimize

$$E[\boldsymbol{v}(k+1)^T\boldsymbol{v}(k+1)] = \sum_{j=1}^{L} b_j(k+1)^2 \tag{19}$$

can be derived by setting the derivative of (16) with respect to α_i equal to 0. Again for the noise-free case, assuming the last term in (16) can be disregarded, this gives

$$\alpha_i = \frac{Lb_i(k)^2}{\sum_{j=1}^{L} b_j(k)^2} \qquad i = 1, \cdots, L. \tag{20}$$

This indicates that each stepsize should be proportional to the mean-squared error of the corresponding coefficient.

Although $[b_1(k)^2, \cdots, b_L(k)^2]$ has the exponential decay characteristics of the room impulse response variation at time $k = 0$, it changes as the algorithm converges. This means that $b_i(k)^2$ should be estimated and α_i should be adjusted for each step of k. Some other algorithms, such as the RLS algorithms estimate $b_i(k)^2$ at the price of a high computational load.

Here, a time-invariant stepsize is used to avoid increasing the computational cost for practical usage. Convergence of the mean-squared coefficient error in the ES algorithm is calculated using (16) for various exponential stepsizes $\alpha_i = b_i(0)^r$ with a parameter r, as shown in Fig. 4. The number of taps is 3840 and the ambient noise with a fixed SNR of 30 dB is added, where SNR is defined as the ratio of the real echo power to the ambient noise power. Here, $\alpha_i = b_i(0)^0 = 1$ corresponds to the conventional NLMS with scalar

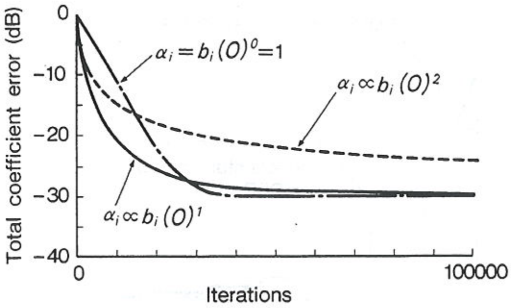

Fig. 4. Convergence of the mean-squared coefficient error for various stepsizes $(\alpha_1, \cdots, \alpha_L)$. Here, $\alpha_i = b_i(0)^0 = 1$ corresponds to the conventional NLMS with $\alpha = 1$ and $b_i(0)^1$ represents the standard deviation of the room impulse response variation. The number of taps is 3840 and the ambient noise with a fixed SNR of 30 dB is added.

stepsize $\alpha = 1$ which results in the fastest convergence in the conventional NLMS.

When α_i are set proportional to $b_i(0)^2\,(r = 2)$, (20) is satisfied only at time $k = 0$. Consequently, convergence is fastest only in the initial state and gets worse later. On the other hand, when α_i are set proportional to $b_i(0)^0 = 1\,(r = 0)$, (20) is satisfied in the steady state, since all the coefficients have the same mean-squared error in the steady state for a white noise input signal. However, (20) does not hold in the initial state. Consequently, convergence is slow in the initial state. Between these two cases, when α_i are set proportional to $b_i(0)^1\,(r = 1)$, (20) is roughly approximated in the transient state. In a practical acoustic echo canceller, convergence to the ERLE of about 20 dB is important. Therefore, we select the time-invariant stepsizes α_i to be proportional to $b_i(0)^1 (r = 1)$, *i.e.*, the expected variation of a room impulse response.

$$\alpha_i \propto b_i(0), \qquad i = 1, \cdots, L. \tag{21}$$

Since $b_i(0)$, *i.e.*, the expected variation of a room impulse response, attenuates by the same exponential ratio as the impulse response, it attenuates to -60 dB in $i = T_R/T_S$ samples, where T_S is the sampling interval and T_R is the reverberation time of the room, defined as the time for the sound energy to attenuate to -60 dB. On the other hand, α_i is expressed by $\alpha_i = \alpha_0\gamma^{i-1}\,(i = 1, \cdots, L)$; therefore, the exponential attenuation ratio γ can be calculated by the following equation:

$$\gamma = \exp\left(-6.9\frac{T_S}{T_R}\right) \tag{22}$$

where -6.9 is $\log_e 10^{-3}$.

When the impulse response has a known fixed flat delay in the acoustic echo path between the loudspeaker and the microphone in a practical situation, the corresponding α_i is set to zero.

C. Steady-State ERLE

In the steady state, the following equation is assumed to hold:

$$b_i(k+1)^2 = b_i(k)^2 = b_i^2. \tag{23}$$

Then, (16) yields

$$b_i^2 = \frac{\alpha_i}{2L}\left(\sum_{j=1}^{L} b_j^2\right) + \frac{\alpha_i}{2L}\frac{p_n}{p_x}. \tag{24}$$

Summing b_i^2 in (24) for all i gives

$$\sum_{i=1}^{L} b_i^2 = \frac{1}{2}\bar{\alpha}\sum_{j=1}^{L} b_j^2 + \frac{1}{2}\bar{\alpha}\frac{p_n}{p_x} \tag{25}$$

hence,

$$\sum_{i=1}^{L} b_i^2 = \frac{\bar{\alpha}}{2-\bar{\alpha}}\frac{p_n}{p_x} \tag{26}$$

where,

$$\bar{\alpha} = \frac{1}{L}\sum_{i=1}^{L}\alpha_i = \frac{\alpha_0}{L}\frac{1-\gamma^L}{1-\gamma}. \tag{27}$$

Now the squared expectation of the residual echo $\tilde{e}(k) = y(k) - \hat{y}(k)$ is

$$\begin{aligned} p_e &= E[\tilde{e}(k)^2] \\ &= E[\{\boldsymbol{v}(k)^T\boldsymbol{x}(k)\}^2] \\ &= \sum_{i=1}^{L} b_i^2 p_x = \frac{\bar{\alpha}}{2-\bar{\alpha}}p_n. \end{aligned} \tag{28}$$

Dividing both sides of (28) by $p_y = \mathrm{E}[y(k)^2]$ and taking logarithms yields the steady-state echo return loss enhancement, ERLE_∞

$$\mathrm{ERLE}_\infty = \mathrm{SNR} + 10\log_{10}\left(\frac{2}{\bar{\alpha}} - 1\right) \quad \text{(dB)} \tag{29}$$

where

$$\begin{aligned} \mathrm{ERLE}_\infty &= 10\log_{10}(p_y/p_e) \\ \mathrm{SNR} &= 10\log_{10}(p_y/p_n). \end{aligned}$$

Equation (29) is the same as the equation of the steady-state ERLE in the conventional NLMS, with α replaced by the mean stepsize $\bar{\alpha}$.

D. Convergence Condition

When the eigenvalues λ_i of $\boldsymbol{Q}$ in (18) are such that $|\lambda_i| < 1$, the mean-squared coefficient error in the ES algorithm converges. The necessary and sufficient condition for convergence is

$$0 < \bar{\alpha} = \frac{1}{L}\sum_{i=1}^{L}\alpha_i < 2. \tag{30}$$

This is proved in the Appendix.

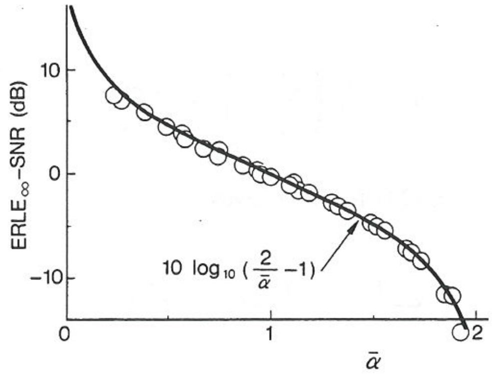

Fig. 5. Steady-state echo return loss enhancement ($ERLE_{\infty}$). $\bar{\alpha}$ is the mean stepsize. The number of taps is 500 and the ambient noise with a fixed SNR of 30 dB is added.

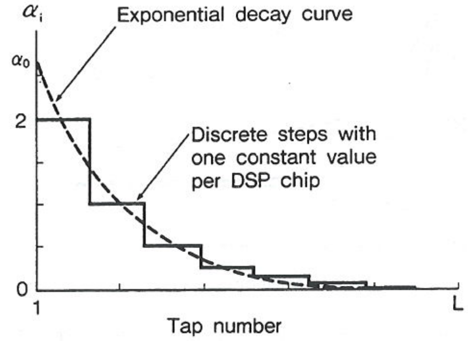

Fig. 6. Stepsize α_i of matrix $\boldsymbol{A}$ when α_i is set in discrete steps with one constant value per DSP chip.

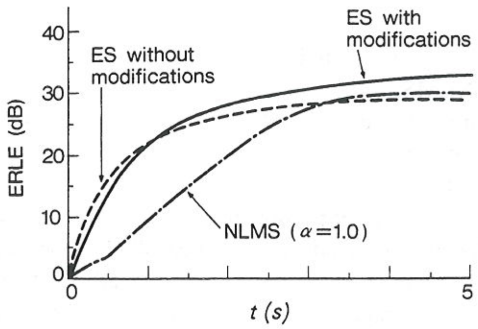

Fig. 7. Convergence of the ERLE in the ES algorithm with and without the practical modifications. The number of taps is 3840 and sampling frequency is 8 kHz. The input signal is white noise. Ambient noise with a fixed SNR of 30 dB is added.

E. Computer Simulation Results

The above results derived in Section IV-A–D show good agreement with computer simulation where a white noise is used as the input signal, and the number of taps L is relatively large. Here, as one example, computer simulation results of the steady-state ERLE for various mean stepsizes $\bar{\alpha}$ are shown in Fig. 5. The number of taps is 500 and the ambient noise with a fixed SNR of 30 dB is added. The theoretical curve given by (29) is also shown in the figure. Fig. 5 shows that the simulation results agree with the theoretical results (29).

In the ES algorithm, the mean stepsize $\bar{\alpha}$ plays a very important role, as it determines the convergence condition and controls the tradeoff between convergence speed and excess mean-squared error.

V. Practical Modifications

A. Practical Modifications for Speech Input and a Multiple DSP Structure

In a practical acoustic echo canceller, which deals with speech input and is constructed with multiple digital signal processor (DSP) chips, the ES algorithm previously described needs slight modifications.

First of all, the maximum stepsize is limited. The convergence condition (30) was derived assuming a white noise input which satisfies assumptions 1)–3) in Section IV-A. However, when these assumptions are not satisfied, *i.e.*, when the input signal is nonstationary such as speech, convergence condition (30) is necessary but not sufficient. Therefore, we propose to use the sufficient condition

$$0 < \alpha_i < 2, \qquad i = 1, \cdots, L \tag{31}$$

to guarantee convergence in a practical acoustic echo canceller.

Next, the $2L$ multiply–add plus L multiply operations of the ES algorithm can be reduced to $2L$ multiply–add operations. In a practical system constructed with multiple DSP chips, the exponential decay curve is approximated stepwise, and stepsize α_i is set in discrete steps with one constant value per DSP chip, as shown in Fig. 6. This practical modification allows the proposed algorithm to have the same computational load, $2L$, as the conventional NLMS.

B. Convergence of the ES Algorithm With and Without the Modifications

Computer simulation results on the convergence of the ERLE in the ES algorithm with and without the practical modifications are shown in Fig. 7. The number of taps is 3840 and the sampling frequency is 8 kHz. The impulse response used in the computer simulation was measured in a room with a reverberation time of 500 ms. The initial condition of the filter coefficients were set to zero. The received input was white noise. Ambient noise with a fixed SNR of 30 dB was added. The curve is an average of 10 independent results. The signal powers of the echo and residual echo used for ERLE were calculated from the squared average of 100 data samples. $\bar{\alpha}$ in the ES algorithm is 0.45 with modifications, and 1.0 without modifications. α in the conventional NLMS algorithm is 1.0.

Fig. 7 shows that the modifications do not have much effect on convergence, and that ES algorithms with and without modifications both have twice the convergence speed of the NLMS. (Detailed comparison indicates that the modifications result in slightly slower convergence in the transient state but a higher ERLE in the steady state. This is due to the tradeoff between convergence speed and excess mean-squared error.)

VI. Real-Time Experiments

Real-time experiments were performed with the ES algorithm implemented in a commercial acoustic echo canceller [8] constructed with multiple DSP chips [15]. The subband

TABLE I
SPECIFICATION OF THE ECHO CANCELLER

Band	Frequency Range (kHz)	No. of Taps
Lower	0.15–4	3840
Higher	4–7	1792

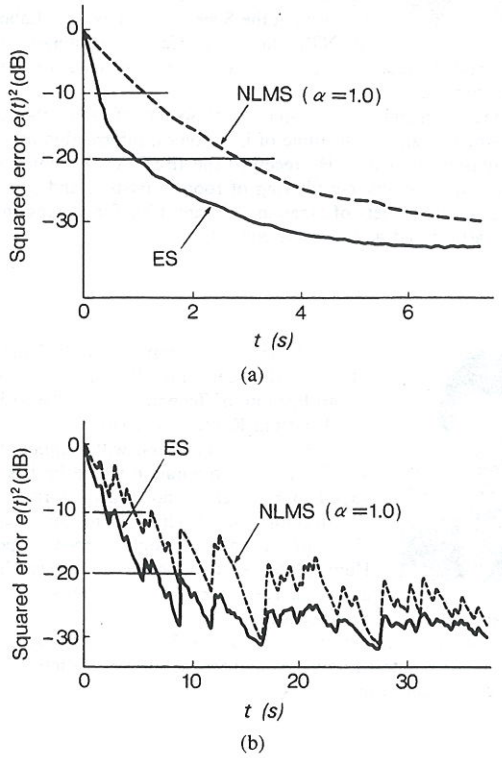

Fig. 8. Real-time experimental results on squared error level convergence using a commercial acoustic echo canceller constructed with multiple DSP chips. The number of taps is 3840 in the lower band and 1792 in the higher band. Sampling frequency is 8 kHz in both bands. Room reverberation time at 500 Hz is 300 ms. (a) Input signal: white noise. (b) Input signal: Speech (male).

technique was used to separate the 7-kHz frequency range into two bands, each sampled at 8 kHz. Table I lists the specifications of the echo canceller.

For real-time operation, α_{1-256} were set to 2.0, $\alpha_{257-512}$ to 1.0, and $\alpha_{513-\text{last}}$ to 0.3. In the lower and higher bands $\bar{\alpha}$ is 0.46 and 0.64, respectively. The speaker–microphone distance was 2.5 m and the reverberation time of the room was 300 ms.

Fig. 8 shows the real-time measurements of squared error level convergence. With a white noise input, Fig. 8(a), the squared error level decayed to -20 dB three times as fast as the conventional NLMS (with $\alpha = 1.0$). With a speech input, Fig. 8(b), the squared error level decayed to -20 dB, twice as fast as the conventional NLMS (with $\alpha = 1.0$). The steady-state ERLE was over 30 dB in the 7-kHz frequency range.

Thus the ES algorithm can easily replace the NLMS algorithm and improve the convergence of practical acoustic echo cancellers.

VII. CONCLUSIONS

A new NLMS adaptive algorithm has been developed for use in acoustic echo cancellers. This algorithm adjusts each filter coefficient by a different stepsize. These time-invariant stepsizes α_i are set proportional to the expected variation of a room impulse response. By modifying α_i in a practical multiple-DSP structure, this algorithm requires only the same computational load as the conventional NLMS. Studying the convergence gives a transition formula for the mean-squared coefficient error of the proposed algorithm. This also indicates that the mean stepsize determines the convergence condition, convergence speed, and excess mean-squared error. The algorithm has been implemented in a commercial acoustic echo canceller constructed with multiple DSP chips. Real-time experiments in a room showed that this algorithm converges faster than the conventional NLMS: three times as fast for a white noise input signal, and twice as fast for speech.

APPENDIX
PROOF OF THE CONVERGENCE CONDITION

When (30) holds, the eigenvalues λ_i of $\boldsymbol{Q}$ in (18) obey $|\lambda_i| < 1$, as proven below.

$\boldsymbol{Q}$ is written as

$$\boldsymbol{Q} = \begin{pmatrix} (1-a_1)^2 & a_1^2 & \cdots & a_1^2 \\ a_2^2 & (1-a_2)^2 & \cdots & a_2^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_L^2 & a_L^2 & \cdots & (1-a_L)^2 \end{pmatrix} \tag{32}$$

where

$$a_i = \alpha_i / L > 0.$$

Consider

$$|\boldsymbol{I} - \boldsymbol{Q}| = \begin{vmatrix} 2a_1 - a_1^2 & -a_1^2 & \cdots & -a_1^2 \\ -a_2^2 & 2a_2 - a_2^2 & \cdots & -a_2^2 \\ \vdots & \vdots & \ddots & \vdots \\ -a_L^2 & -a_L^2 & \cdots & 2a_L - a_L^2 \end{vmatrix}$$
$$= 2^{L-1} a_1 a_2 \cdots a_L \left(2 - \sum_{i=1}^{L} a_i \right). \tag{33}$$

If

$$2 - \sum_{i=1}^{L} a_i > 0 \tag{34}$$

then all principal minors are positive. Therefore, $(\boldsymbol{I} - \boldsymbol{Q})$ is positive-definite, *i.e.*, all eigenvalues are positive. Therefore, the eigenvalues λ_i of $\boldsymbol{Q}$ are

$$\lambda_i < 1, \qquad i = 1, \cdots, L. \tag{35}$$

On the other hand, because $\boldsymbol{Q}$ is a positive matrix, *i.e.*, all the elements of $\boldsymbol{Q}$ are positive, the Perron–Frobenius theorem [16] gives

$$-1 < \lambda_i < 1. \tag{36}$$

Acknowledgment

The authors would like to thank Dr. M. Tohyama and Dr. S. Furui for their technical advice and continuous encouragement. They also wish to express their gratitude to Dr. K. Sugiyama, I. Kawashima, and H. Oikawa for their thoughtful comments and interest in this work. The reviewers are thanked for valuable comments and suggestions.

References

[1] S. Haykin, *Introduction to Adaptive Filters.* New York: Macmillan, 1984.

[2] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification.* Cambridge, MA: M.I.T. Press, 1985.

[3] D. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Processing,* vol. 39, pp. 92–114, Jan. 1991.

[4] B. Widrow *et al.,* "Adaptive noise cancelling: Principles and applications," *Proc. IEEE,* vol. 63, pp. 1692–1716, Dec. 1975.

[5] B. Widrow and S. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[6] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.,* vol. AC-12, pp. 282–287, June 1967.

[7] Y. Itoh, Y. Maruyama, N. Furuya, and T. Araseki, "An acoustic echo canceller for teleconference," in *Proc. ICC85,* pp. 1498–1502, June 1985.

[8] H. Oikawa, S. Minami, and T. Saeki, "A new echo canceller realized by high performance digital signal processor," in *Proc. ISCAS88,* pp. 1329–1332, June 1988.

[9] R. Harris, D. Chabries, and F. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 309–316, Apr. 1986.

[10] S. Haykin, *Adaptive Filter Theory.*2nd. edition. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[11] B. Hatty, "Recursive least squares algorithms using multirate systems for cancellation of acoustic echoes," in *Proc. ICASSP90,* pp. 1145–1148, Apr. 1990.

[12] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Amer.,* vol. 37, pp. 409–412, 1965.

[13] S. Makino and N. Koizumi, "Improvement on adaptation of an echo canceller in a room," in *Proc. Spring Meeting of Acoust. Soc. Jap.,* pp. 355–356, Mar. 1988.

[14] J. Chen and J. Vandewalle, "A μ-vector LMS adaptive system for enhancing nonstationary narrow-band signal," in *Proc. ISCAS88,* pp. 771–774, June 1988.

[15] T. Kaneko, H. Yamauchi, and A. Iwata, "A 50 ns floating-point signal processor VLSI," in *Proc. ICASSP86,* pp. 401–404, Apr. 1986.

[16] F. R. Gantmacher, *The Theory of Matrices.* New York: Chelsea, 1959.

Shoji Makino (A'89–M'90) was born in Nikko, Japan, on June 4, 1956. He received the B. E. and M. E. degrees in mechanical engineering from Tohoku University, Sendai, Japan, in 1979 and 1981, respectively.

He joined the Electrical Communication Laboratory of Nippon Telegraph and Telephone Corporation (NTT) in 1981. Since then, he has been engaged in research on electroacoustic transducers and acoustic echo cancellers. He is now a Senior Research Engineer at the Speech and Acoustics Laboratory of the NTT Human Interface Laboratories. His research interests include acoustic signal processing, and adaptive filtering and its applications.

Mr. Makino is a member of the Acoustical Society of Japan and the Institute of Electronics, Information, and Communication Engineers of Japan.

Yutaka Kaneda (M'80) was born in Osaka, Japan, on Feb. 20, 1951. He received the B.E., M.E., and Doctor of Engineering degrees from Nagoya University, Nagoya, Japan, in 1975, 1977 and 1990.

In 1977, he joined the Electrical Communication Laboratory of Nippon Telegraph and Telephone Corporation (NTT), Musashino, Tokyo, Japan. He has since been engaged in research on acoustic signal processing. He is now a Senior Research Engineer at the Speech and Acoustics Laboratory of the NTT Human Interface Laboratories. His recent research interests include microphone array processing, adaptive filtering and sound field control.

Dr. Kaneda is a member of the Acoustical Society of Japan, the Acoustical Society of America and the Institute of Electronics, Information and Communication Engineers of Japan. He received the IEEE ASSP Senior Award in 1990 for an article on inverse filtering of room acoustics, and paper awards from the Acoustical Society of Japan in 1990 and 1992 for articles on adaptive microphone arrays and active noise control.

Nobuo Koizumi was born in Kyoto, Japan, on Jan. 21, 1951. He received the B.E. degree in 1973 from Kyoto Institute of Technology and the M.E. degree in 1975 from Kyoto University.

Since 1975 he has been with Nippon Telegraph and Telephone Corporation where he has been engaged in research of noise and vibration control, and development of telecommunication equipment. He is now a senior research engineer, supervisor at Human Interface Laboratories. His research interests are on active control of sound, telecommunication acoustics, and audio signal processing.

Mr. Koizumi is a member of the Acoustical Society of Japan, the Society of Instrument and Control Engineers, and the Acoustical Society of America, and the Audio Engineering Society.