論文 / 著書情報
Article / Book Information

Title	Gradient-Limited Affine Projection Algorithm for Double-Talk-Robust and Fast-Converging Acoustic Echo Cancellation
Authors	Suehiro Shimauchi, Yoichi HANEDA, Akitoshi KATAOKA, Akinori NISHIHARA
Citation	IEICE Trans. Fundamentals., Vol. E90-A, No. 3, pp. 633-641
Pub. date	2007, 3
URL	http://search.ieice.org/
Copyright	(c) 2007 Institute of Electronics, Information and Communication Engineers

PAPER

Gradient-Limited Affine Projection Algorithm for Double-Talk-Robust and Fast-Converging Acoustic Echo Cancellation

Suehiro SHIMAUCHI[†,††a)], **Yoichi HANEDA**[†], **Akitoshi KATAOKA**[†], ***Members***, ***and*** **Akinori NISHIHARA**[††], ***Fellow***

SUMMARY We propose a gradient-limited affine projection algorithm (GL-APA), which can achieve fast and double-talk-robust convergence in acoustic echo cancellation. GL-APA is derived from the M-estimation-based nonlinear cost function extended for evaluating multiple error signals dealt with in the affine projection algorithm (APA). By considering the nonlinearity of the gradient, we carefully formulate an update equation consistent with multiple input-output relationships, which the conventional APA inherently satisfies to achieve fast convergence. We also newly introduce a scaling rule for the nonlinearity, so we can easily implement GL-APA by using a predetermined primary function as a basis of scaling with any projection order. This guarantees a linkage between GL-APA and the gradient-limited normalized least-mean-squares algorithm (GL-NLMS), which is a conventional algorithm that corresponds to the GL-APA of the first order. The performance of GL-APA is demonstrated with simulation results.

key words: *affine projection algorithm, acoustic echo canceller, double-talk, robust control*

1. Introduction

In teleconferencing systems that connect remote sites to communicate with each other, an acoustic echo canceller (AEC) is indispensable at each site. To eliminate acoustic echo from the microphone signal, an adaptive filter with a finite impulse response (FIR) structure is often implemented in the AEC. The adaptive filter adaptively simulates the echo path impulse response between the loudspeaker and microphone, produces an echo replica, and subtracts the echo replica from the microphone signal. However, the adaptive filter often suffers from divergence of its coefficients during double-talk, where the near- and far-end speakers are talking simultaneously. This is because the near-end speech acts as an outlier for the adaptation. Such an outlier causes a large disturbance in the echo path identification of many adaptive algorithms. Thus, robustness against double-talk has been a key issue in AEC implementation.

There are several approaches for achieving robustness. In the dual-filter-structure approach [1], [2], an adaptive filter is used in the background, while a semi-fixed filter achieves echo cancelling in the foreground. The coefficients of the foreground filter are overwritten by the background coefficients only when the background filter is judged to converge sufficiently. Hence, the foreground filter is free from the wrong adaptations that occur in the background. Using a double-talk detector (DTD) [3]–[8] is also a reasonable approach for freezing the adaptation when a double-talk situation is detected. However, when the echo path changes during double-talk, both the dual filter and DTD techniques by themselves can do nothing except stop the (foreground) filter update, even if they can detect echo path changes. This can lead to annoying howling when there is a drastic echo path change. To overcome this problem, the adaptive algorithm itself needs to be improved so that it can track the echo path change even during double-talk.

Direct modification of the least-mean-squares algorithm (LMS) [9] by introducing nonlinearity into the error cost function gives a reliable clue for achieving such a robust adaptation [10]–[12]. We have introduced the gradient-limited normalized LMS (GL-NLMS) [13] based on M-estimation [12], as a robust version of the normalized LMS (NLMS) [14]. As shown in [13], GL-NLMS achieves better echo-path tracking during double-talk than another robust NLMS proposed in [11] does. The difference between their performances is due to their error evaluations. GL-NLMS evaluates the error 'after' normalization by the reference signal vector norm, while the robust NLMS in [11] evaluates the error 'before' the normalization. However, in common with NLMS-based algorithms, GL-NLMS still has a drawback: slow convergence for colored input signals such as speech. This drawback needs to be overcome for AEC applications.

The affine projection algorithm (APA) [15]–[19] is a good candidate to overcome the slow convergence of NLMS-based algorithms. In previous works to achieve fast convergence with low sensitivity to noise, variable step-size control is introduced in the APA [18], [20] or an APA-like signal decorrelation algorithm [21]. Those techniques can be regarded as an extension of variable step-size NLMS, such as that shown in [22]. Parallel subgradient projection techniques [23] are also developed to overcome the sensitivities of APA to noise. However, none of these approaches considers the double-talk situation, where noise corresponds to near-end speech with nonstationarity, and its level is as

Manuscript received May 22, 2006.
Manuscript revised September 29, 2006.
Final manuscript received December 11, 2006.

[†]The authors are with NTT Cyber Space Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[††]The authors are with Tokyo Institute of Technology, Tokyo, 152-8552 Japan.
a) E-mail: shimauchi.suehiro@lab.ntt.co.jp

DOI: 10.1093/ietfec/e90–a.3.633

high as or sometimes higher than that of the echo signal to be eliminated.

In this paper, we focus on deriving a novel adaptive algorithm that can track the echo path change even during double-talk, with faster convergence than that of GL-NLMS and robustness as good as that of GL-NLMS. We extend the gradient-limited (GL-) algorithm based on APA, which we call GL-APA. First, we newly define a cost function for the APA derivation, which does not appear in [15]–[19], but seems promising for our purpose: to apply M-estimation-based nonlinearity to APA. By taking account of the nonlinearity of the gradient, we carefully formulate an update equation consistent with multiple input-output relationships, which the conventional APA inherently satisfies. Furthermore, we also introduce a scaling rule for the nonlinearity, so we can easily implement GL-APA of any projection order by scaling a predetermined primary function. This guarantees a linkage between GL-APA and GL-NLMS by choosing the nonlinear function used in GL-NLMS as the primary function.

2. AEC and Conventional Algorithms

2.1 Configuration of AEC

An AEC based on an adaptive filter, as shown in Fig. 1, is used to eliminate acoustic echo caused by acoustic coupling between loudspeaker and microphone in hands-free teleconferencing. The signal $x(n)$ at a discrete time index n is received from the far end. That signal is reproduced by the loudspeaker and picked up by the microphone as echo $d(n)$ after passing through the room echo path, which has an impulse response modeled as $\mathbf{h} = [h_1, \ldots, h_L]^T$, where L is the effective length and T denotes the transpose. For simplicity, here we consider the period where $\mathbf{h}$ is time-invariant, which is assumed in the derivation of most conventional algorithms. The echo $d(n)$ can be denoted as

$$d(n) = \mathbf{x}^T(n)\mathbf{h}, \tag{1}$$

where $\mathbf{x}(n) = [x(n), \ldots, x(n-L+1)]^T$. The microphone signal $y(n)$ may contain a speech signal $s(n)$ and ambient noise signal $a(n)$ at the near end in addition to the echo $d(n)$. On the other hand, the adaptive FIR filter in the AEC has its coefficient vector $\hat{\mathbf{h}}(n)$ to identify the echo path $\mathbf{h}$ and

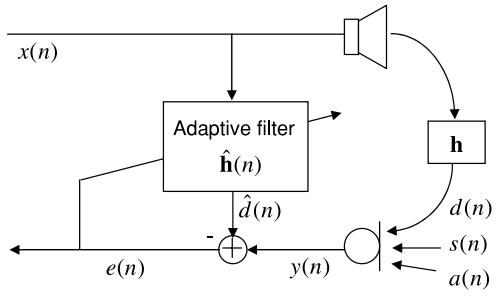

Fig. 1 Configuration of AEC.

generates an echo replica $\hat{d}(n)$ as

$$\hat{d}(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}(n). \tag{2}$$

The residual signal $e(n)$ is fed back to the adaptive filter as the error signal to update $\hat{\mathbf{h}}(n)$ and is sent to the far end as an echo-eliminated signal, where

$$e(n) = y(n) - \hat{d}(n). \tag{3}$$

Here, we suppose that $\hat{\mathbf{h}}(n)$ has the same length as $\mathbf{h}$. A general form of the update equation of $\hat{\mathbf{h}}(n)$ can be expressed as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu\Delta\hat{\mathbf{h}}(n), \tag{4}$$

where the update vector $\Delta\hat{\mathbf{h}}(n)$ depends on each adaptive algorithm, and μ is the step-size. The near-end speech $s(n)$ or ambient noise $a(n)$ contained in the microphone signal $y(n)$ is regarded as an outlier for adaptive algorithms to identify the echo path $\mathbf{h}$. One of the most challenging situations for the adaptation is double-talk, where near- and far-end speakers are talking simultaneously. In such situations, an adaptive algorithm derived from a simple error-minimization criterion, such as NLMS [14], often suffers from misconvergence or divergence of its coefficients because the outlier is dominant in the error signal $e(n)$ to be minimized.

2.2 GL-NLMS

GL-NLMS [13] is a modified version of NLMS, which is more robust against outliers. The modification is based on M-estimation [12] and is described as follows. According to [13], NLMS can be derived from the following cost function

$$J_{\mathrm{NLMS}}(n) = \frac{1}{2}\left(\frac{|e(n)|}{\|\mathbf{x}(n)\|}\right)^2. \tag{5}$$

On the other hand, the cost function for GL-NLMS is given as

$$J_{\mathrm{GLNLMS}}(n) = \rho\left(\frac{|e(n)|}{\|\mathbf{x}(n)\|}\right), \tag{6}$$

where $\rho(v)$ is, in general, an arbitrary function for variable v. Note that, when $\rho(v) = v^2/2$, (6) is equal to (5). The gradient estimate is obtained from (6) as

$$\begin{aligned}\hat{\nabla}_{\mathrm{GLNLMS}}(n) &= \frac{\partial J_{\mathrm{GLNLMS}}(n)}{\partial\hat{\mathbf{h}}(n)} \\ &= -\psi\left(\frac{|e(n)|}{\|\mathbf{x}(n)\|}\right)\cdot\frac{\mathrm{sgn}(e(n))\cdot\mathbf{x}(n)}{\|\mathbf{x}(n)\|},\end{aligned} \tag{7}$$

where the function $\psi(v) = \partial\rho(v)/\partial v$, and $\mathrm{sgn}(v)$ denotes a sign function. The update vector for GL-NLMS is made from the gradient estimate $\hat{\nabla}_{\mathrm{GLNLMS}}(n)$:

$$\Delta\hat{\mathbf{h}}_{\mathrm{GLNLMS}}(n) = -\hat{\nabla}_{\mathrm{GLNLMS}}(n). \tag{8}$$

For good robustness, $\rho(v)$ is chosen so that $\psi(v)$ becomes a bounded function [12]. Examples of functions $\rho(v)$ and $\psi(v)$ for AEC application are shown in [13]:

$$\rho(v) = \begin{cases} \frac{1}{2}v^2, & 0 \le v \le T_1 \\ S_1 v, & T_1 < v \le T_2 \\ S_2 v, & T_2 < v, \end{cases} \tag{9}$$

and

$$\psi(v) = \begin{cases} v, & 0 \le v \le T_1 \\ S_1, & T_1 < v \le T_2 \\ S_2, & T_2 < v, \end{cases} \tag{10}$$

which are specified in the three separated ranges based on the behavior of the value of $v = |e(n)|/\|\mathbf{x}(n)\|$. When $0 \le v \le T_1$, single-talk of the far-end speaker is mostly expected. When $T_2 < v$, single-talk is rarely expected. When $T_1 < v \le T_2$, both single-talk and double-talk, or even an echo path change can be expected. Thresholds T_1 and T_2 correspond to the lower and upper bounds, respectively, of the range of the acoustic coupling level observed after echo cancellation. The lower threshold T_1 can be set by estimating the limit of the adaptive identification because of ambient noise, finite tap length, and so on. The upper threshold T_2 can be set assuming the system setup, such as maximum volume of the loudspeaker. To achieve greater suppression of the update amount for larger v, the values of S_1 and S_2 are chosen as $0 \le S_2 \le S_1$, e.g., $S_1 = 0.5T_1$ and $S_2 = 0.25T_1$.

GL-NLMS achieves significantly more robust adaptation than the ordinary NLMS, as shown in [13]. However, GL-NLMS has the same drawback as the ordinary NLMS, that is, slow convergence for the input of a colored signal such as speech.

2.3 APA

APA [15]–[19] is an extended version of NLMS to achieve faster convergence, especially for colored input. The update vector used in APA is given as

$$\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n) = \mathbf{X}(n)\mathbf{R}^{-1}(n)\mathbf{e}(n), \tag{11}$$

where

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \ldots, \mathbf{x}(n-p+1)], \tag{12}$$

$$\mathbf{R}(n) = \mathbf{X}^T(n)\mathbf{X}(n) + \delta_1 \mathbf{I}_{p\times p}, \tag{13}$$

$$\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{X}^T(n)\hat{\mathbf{h}}(n), \text{and} \tag{14}$$

$$\mathbf{y}(n) = [y(n), y(n-1), \ldots, y(n-p+1)]^T. \tag{15}$$

In the above equations, p denotes the projection order, δ_1 is a positive constant for regularization of the inversion stability, and $\mathbf{I}_{p\times p}$ is the identity matrix of $p\times p$. In actual calculations, by regarding δ_1 as sufficiently small, we can approximately obtain the error vector $\mathbf{e}(n)$ in (14) based on a simple past-error correction:

$$\mathbf{e}(n) \simeq \begin{bmatrix} e(n) \\ (1-\mu)\cdot e(n-1) \\ \vdots \\ (1-\mu)^{p-1}\cdot e(n-p+1) \end{bmatrix}. \tag{16}$$

The update vector $\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)$ in (11) is derived as the minimum norm solution among non-unique solutions of p-th order (underdetermined) simultaneous equations

$$\mathbf{e}(n) = \mathbf{X}^T(n)\Delta\mathbf{h}(n), \tag{17}$$

where $\Delta\mathbf{h}(n)$ denotes non-unique unknowns that include the true solution $\mathbf{h} - \hat{\mathbf{h}}(n)$. By satisfying the last p input-output relationships simultaneously, the update vector $\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)$ can be more reliable than that of NLMS, which satisfies only the current relationship. This agrees with the fast convergence of APA.

Our aim in this study is to extend GL-NLMS by incorporating the fast convergence of APA. To achieve this, here we provide another derivation of APA, which does not appear in [15]–[19]. When we give a cost function based on an analogical extension of the cost function in (5) as

$$J_{\mathrm{APA}}(n) = \frac{1}{2}\mathbf{e}^T(n)\mathbf{R}^{-1}(n)\mathbf{e}(n), \tag{18}$$

we obtain

$$\hat{\nabla}_{\mathrm{APA}}(n) = \frac{\partial J_{\mathrm{APA}}(n)}{\partial\hat{\mathbf{h}}(n)} = -\mathbf{X}(n)\mathbf{R}^{-1}(n)\mathbf{e}(n), \tag{19}$$

and thus the update vector of APA

$$\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n) = -\hat{\nabla}_{\mathrm{APA}}(n) \tag{20}$$

is derived.

3. Proposed Robust and Fast-Converging Algorithm

In this section, we propose a robust and fast-converging adaptive algorithm by applying an APA-like extension to GL-NLMS. From the newly defined APA cost function (18), we obtain good prospects for extending the M-estimation so that the proposed algorithm can deal with multiple (the last p) error signals, instead of ordinarily dealing with only the current error $e(n)$. To satisfy the last p input-output relationships simultaneously, as the conventional APA inherently does, we carefully formulate the update equation taking into account the nonlinearity of the gradient estimate. Furthermore, we introduce a scaling rule for the nonlinearity, so we can easily implement GL-APA of any projection order p by using a simple primary function as a basis of scaling with p. We also show that the extra computational complexity of the proposed algorithm is not so large compared with the complexity of the ordinary APA.

3.1 GL-APA

To derive an APA-like robust algorithm based on M-estimation, we modify the cost function in (18) to

$$J_{\mathrm{GLAPA}}(n) = \rho\left(\sqrt{\mathbf{e}^T(n)\mathbf{R}^{-1}(n)\mathbf{e}(n)}\right). \tag{21}$$

By noting that (11) indicates

$$\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\| = \sqrt{\mathbf{e}^T(n)\mathbf{R}^{-1}(n)\mathbf{e}(n)}, \tag{22}$$

we can describe the gradient estimate of (21) as

$$\hat{\nabla}_{\mathrm{GLAPA}}(n) = \frac{\partial J_{\mathrm{GLAPA}}(n)}{\partial \hat{\mathbf{h}}(n)} = -\psi\left(\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|\right) \cdot \frac{\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)}{\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\| + \delta_2}, \tag{23}$$

where a positive constant δ_2 is introduced to avoid division by zero. By expressing the update vector as

$$\Delta\hat{\mathbf{h}}_{\mathrm{GLAPA}}(n) = -\hat{\nabla}_{\mathrm{GLAPA}}(n), \tag{24}$$

we derive a new algorithm: GL-APA. Even though the constants for numerical stability, δ_1 and δ_2, are introduced in (23), the gradient estimate $\hat{\nabla}_{\mathrm{GLAPA}}(n)$ in (23) is essentially equivalent to (7), when $p = 1$. This means that GL-APA includes GL-NLMS.

3.2 Error Vector Approximation

In the case of the ordinary APA, we normally use the error vector $\mathbf{e}(n)$ approximated as (16) to reduce the computational cost of calculating the update vector $\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)$. In the GL-APA case, however, the approximation of (16) is not always accurate because the correction of past errors is affected by not only the step-size μ but also the nonlinearity of $\psi(v)$. Therefore, we apply another approximation to the error vector for GL-APA as follows:

$$\mathbf{e}(n) \simeq \begin{bmatrix} e(n) \\ (1-\gamma(n-1)) \cdot e(n-1) \\ \vdots \\ \left(\prod_{k=1}^{p-1}(1-\gamma(n-k))\right) \cdot e(n-p+1) \end{bmatrix}, \tag{25}$$

where the values of $\gamma(n-1), \ldots, \gamma(n-p+1)$ are calculated in the past steps based on

$$\gamma(n) = \mu \cdot \frac{\psi\left(\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|\right)}{\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\| + \delta_2}. \tag{26}$$

The reasonability of this modification can be easily understood by regarding GL-APA as a kind of variable step-size APA [18], [20]. In other words, the update equation of GL-APA can be written as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \gamma(n)\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n). \tag{27}$$

However, the step-size $\gamma(n)$ in (26) uniquely achieves double-talk robustness by using the bounded nonlinear function $\psi(v)$ derived from the modified cost function (21).

3.3 Nonlinearity Scalable to Projection Order

The performance of GL-APA, like that of the ordinary APA, depends on the choice of its projection order p. The requirement for GL-APA is to increase the convergence rate as p is increased while not degrading robustness against outliers. To achieve this, the function $\psi(v)$ should be appropriately chosen depending on the choice of p. Instead of spending time and effort to choose an appropriate $\psi(v)$ for every p, here, we introduce a scaling rule for $\psi(v)$ applicable to any p.

First, as the basis for the scaling, we consider the function $\psi(v)$ used for GL-NLMS and rename it $\psi_1(v)$. The property of the primary function $\psi_1(v)$ can be chosen to fit the statistical behavior of $v = |e(n)|/\|\mathbf{x}(n)\|$, as shown in [13]. In the case of GL-APA, the variable changes to $v = \|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$. Thus, the difference in behavior between $|e(n)|/\|\mathbf{x}(n)\|$ and $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ needs to be investigated. Although their behaviors may also depend on the characteristics of the input signal, here, we simply assume that the input is a white Gaussian signal to obtain a basic clue for estimating the difference. Under the assumption, the non-diagonal elements of the matrix $\mathbf{R}(n)$ take values close to zero, so

$$\mathbf{R}(n) \approx \mathrm{diag}[\|\mathbf{x}(n)\|^2, \|\mathbf{x}(n-1)\|^2, \ldots, \|\mathbf{x}(n-p+1)\|^2]. \tag{28}$$

From (22) and (28), we obtain

$$\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\| \approx \sqrt{\frac{|e_1(n)|^2}{\|\mathbf{x}(n)\|^2} + \frac{|e_2(n)|^2}{\|\mathbf{x}(n-1)\|^2} + \cdots + \frac{|e_p(n)|^2}{\|\mathbf{x}(n-p+1)\|^2}}, \tag{29}$$

where $e_1(n)$, $e_2(n)$, …, and $e_p(n)$ respectively denote the elements of $\mathbf{e}(n)$. Furthermore, by assuming that the reference input and the outlier are stationary at least during p sample periods, and that the adaptation does not progress so much during only p sample periods, we expect

$$E\left[\frac{|e(n)|}{\|\mathbf{x}(n)\|}\right] \approx E\left[\frac{|e(n-1)|}{\|\mathbf{x}(n-1)\|}\right] \approx \cdots \approx E\left[\frac{|e(n-p+1)|}{\|\mathbf{x}(n-p+1)\|}\right], \tag{30}$$

where E denotes expectation. Then, by applying (25) and (30) to (29), we obtain

$$E\left[\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|\right] \approx \kappa_p(n) \cdot E\left[\frac{|e(n)|}{\|\mathbf{x}(n)\|}\right], \tag{31}$$

where

$$\kappa_p(n) = \sqrt{1 + \sum_{m=1}^{p-1}\left(\prod_{k=1}^{m}(1-\gamma(n-k))^2\right)}. \tag{32}$$

The subscript p indicates the projection order, and we define $\kappa_1(n) = 1$. According to the relationship between $|e(n)|/\|\mathbf{x}(n)\|$ and $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ in (31), the statistical behavior of $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ is proportional to that of $|e(n)|/\|\mathbf{x}(n)\|$ scaled with $\kappa_p(n)$. Thus, we obtain a scalable function $\psi_p(v)$ for variable $v = \|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ by scaling the primary function $\psi_1(v)$ as

$$\psi_p(v) = \kappa_p(n) \cdot \psi_1\left(\frac{v}{\kappa_p(n)}\right). \tag{33}$$

We note that the above assumptions that lead to (33) can be applied to the speech input case without causing a significant degradation. Because the matrix $\mathbf{R}(n)$ for the speech input still has diagonal elements larger than its non-diagonal elements. In addition, we can assume that the speech is stationary during a short period (about 30 ms) [24]. Therefore, as long as the projection order p is sufficiently small, we intend to apply the scaling rule in (33) to the gradient estimate (23) even for the speech input.

We also point out that the scale parameter $\kappa_p(n)$ depends not only on the projection order p but also on the past step-sizes $\gamma(n-1), \ldots, \gamma(n-p+1)$, which retain information about past gradient suppressions if $p \geq 2$. Therefore, the scalable function $\psi_p(v)$ is time-variant even after a fixed p is given. On the other hand, even if we could find a manually optimized function instead of a scalable function after an enormous amount of trial and error for each p, it would be only a fixed function that cannot consider the past gradient suppressions. This indicates that the scalable nonlinear function $\psi_p(v)$ in (33) has the potential to deal with the behavior of the last p errors better than the fixed manually optimized function does, as long as the above assumptions that lead to (33) are valid and the primary function $\psi_1(v)$ is selected well.

An example of the scaled function obtained from the primary function $\psi_1(v)$ in (10) is described as

$$\psi_p(v) = \begin{cases} v, & 0 \leq v \leq T_1 \cdot \kappa_p(n), \\ S_1 \cdot \kappa_p(n), & T_1 \cdot \kappa_p(n) < v \leq T_2 \cdot \kappa_p(n), \\ S_2 \cdot \kappa_p(n), & T_2 \cdot \kappa_p(n) < v. \end{cases} \tag{34}$$

3.4 Computation of Variable Step-Size

Here, we discuss the extra computation of GL-APA compared with that of the ordinary APA. GL-APA can be implemented as a variable step-size APA shown in (27). Hence, the extra cost is mostly due to the calculation of the variable step-size $\gamma(n)$. In obtaining the value of $\gamma(n)$ in (26), the calculations of $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ and $\psi(v)$ are dominant.

According to (22), $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ is calculated as

$$\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\| = \sqrt{\mathbf{e}^T(n)\mathbf{g}(n)}, \tag{35}$$

where

$$\mathbf{g}(n) = \mathbf{R}^{-1}(n)\mathbf{e}(n) \tag{36}$$

is the decorrelation vector whose calculation is already included in the ordinary procedure even in the fast versions [18], [19] of APA. Therefore, the calculation for (35) requires p multiplications, $p-1$ additions, and a square root.

In the case of using the nonlinear function $\psi_p(v)$ in (34), which is simply based on a few conditional branch operations, the most expensive cost is the calculation of $\kappa_p(n)$. By taking the square of $\kappa_p(n)$ as

$$K_p(n) = \kappa_p(n)^2, \tag{37}$$

the following recursive equation is obtained:

$$K_p(n) = 1 + \Gamma(n-1)\Big(K_p(n-1) - G_p(n)\Big), \tag{38}$$

where

$$\Gamma(n) = (1-\gamma(n))^2 \text{ and} \tag{39}$$

$$G_p(n) = \prod_{k=2}^{p} \Gamma(n-k). \tag{40}$$

Updating $K_p(n)$ is completed at the cost of p multiplications and three additions at most and a square root operation finally produces

$$\kappa_p(n) = \sqrt{K_p(n)}. \tag{41}$$

As a consequence, the main extra computations required for the variable step-size $\gamma(n)$ are summed up to $2p$ multiplications, $p+2$ additions, and two square root operations, which are spent to calculate $\|\Delta\hat{\mathbf{h}}_{\mathrm{APA}}(n)\|$ and $\kappa_p(n)$. However, this is not a serious obstacle to implement GL-APA instead of ordinary APA because even the fast versions of APA require roughly $2L+20p$ multiplications per sample period [18], [19]. The extra calculations for GL-APA are much smaller than the total computation of APA in many AEC applications, where L is much larger than p, as mentioned in [17].

4. Simulations

We evaluated the performance of the proposed GL-APA, particularly for application to AEC. Although GL-APA may be implemented in practice with a reliable peripheral estimator such as DTD, here we focus on evaluating the performance of the adaptive algorithm alone. The simulation conditions were as follows. The actual echo path $\mathbf{h}$ was formed as an impulse response truncated at 512 samples with an average gain of 0 dB after measurement in an actual conference room with a reverberation time of 200 ms, as shown in Fig. 2. The adaptive filter $\hat{\mathbf{h}}(n)$ also had the same length,

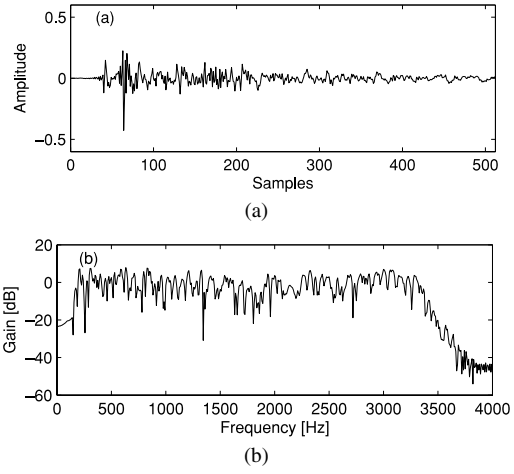

Fig. 2 (a) Impulse response and (b) frequency characteristic of echo path.

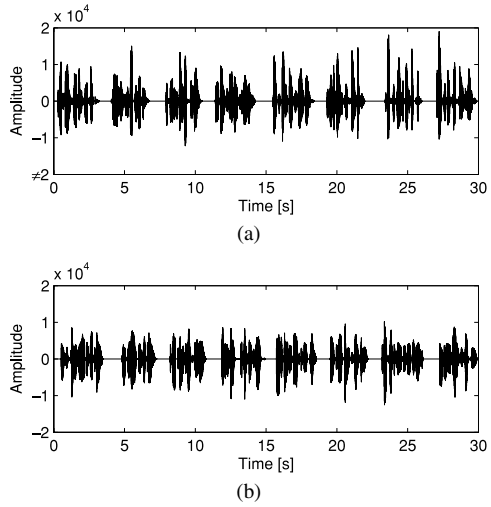

Fig. 3 (a) Far-end signal (male speech) and (b) near-end signal (female speech).

$L = 512$. The sampling frequency was 8 kHz. The far-end speech $x(n)$ for the reference input was male speech, as shown in Fig. 3(a). The near-end speech $s(n)$, which was female speech as shown in Fig. 3(b), was an outlier. The ambient noise $a(n)$ was white Gaussian noise at a level that was 18 dB lower than the echo level. The function $\psi_p(v)$ in (34) was applied to GL-APA, corresponding to each value of p, where the parameters were given as $T_1 = 0.1/\sqrt{L}$, $T_2 = 1/\sqrt{L}$, $S_1 = 0.5T_1$, and $S_2 = 0.25T_1$ after the example shown in [13]. The constant parameters $\delta_1 = 10^8$ and $\delta_2 = 10^{-12}$ were set for all simulations.

For evaluations, the echo return loss enhancement (ERLE) and the coefficient error (CE) defined below were used:

$$\mathrm{ERLE}(n) = 10\log_{10}\frac{\sum_{m=n-M}^{n}[y(m)-s(m)-a(m)]^2}{\sum_{m=n-M}^{n}[e(m)-s(m)-a(m)]^2}\ [\mathrm{dB}]$$

and

$$\mathrm{CE}(n)=20\log_{10}\frac{\|\mathbf{h}-\hat{\mathbf{h}}(n)\|}{\|\mathbf{h}\|}\ [\mathrm{dB}],$$

where ERLE indicates the observed relative echo cancellation level, an appropriate smoothing factor, $M = 8000$, was chosen to show results clearly, and CE indicates a normalized Euclidean distance between the actual echo path and the adaptive filter.

4.1 Improvement in Convergence Rate

Here, we show the improvement in the convergence rate of GL-APA compared with that of GL-NLMS. We compared the cases for different projection orders including $p = 1$ (equivalent to GL-NLMS). The others, $p = 2$, 4, and 8, were chosen from a range where GL-APA can be implemented without significant extra computation compared with APA, as discussed in 3.4. The constant step-size μ used in (26) was set differently for each p as follows. First, according to [13], we set $\mu = 1$ for $p = 1$ as a reference. Then, we found appropriate step-sizes for $p = 2$, 4, and 8, so that each of them including the standard case ($\mu = 1$, $p = 1$) achieved a similar steady-state ERLE level for the stationary signal input, which had an averaged speech spectrum. This was achieved with $\mu = 1$, 0.8, 0.75, and 0.55 for $p = 1$, 2, 4, and 8, respectively, as shown in Fig. 4. The CE performance is also shown in Fig. 5. As shown in Figs. 4 and 5, ERLE and CE gave different characteristics for the order p, e.g., while the ERLEs of any p converged to a similar steady-state level, the CEs did not. This was because the filter coefficient error vector, $\mathbf{h} - \hat{\mathbf{h}}(n)$, for each p might have different frequency characteristics, as discussed in [25]. For the choice of parameters, we mainly concentrated on ERLE characteristics because they directly correspond to the actual echo cancellation level that we will experience in AEC use.

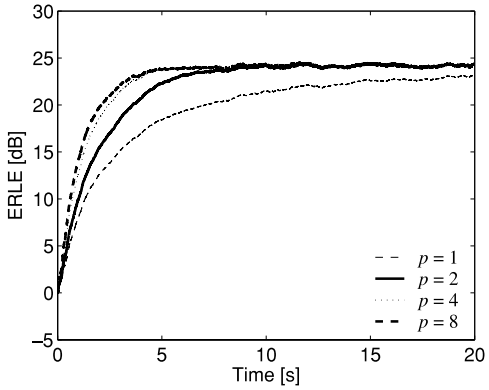

Fig. 4 Comparison of GL-APA performance with different projection orders: ERLEs for colored stationary signal input (single-talk).

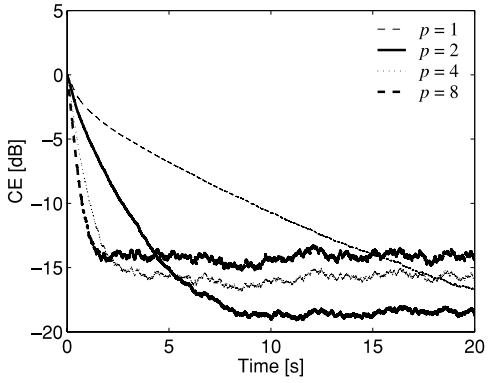

Fig. 5 Comparison of GL-APA performance with different projection orders: CEs for colored stationary signal input (single-talk).

Using those parameters given above, we obtained results for actual speech input. ERLEs and CEs in the single-talk case, where male speech (Fig. 3(a)) was the reference input and the echo path was changed at 12 s, are shown in Figs. 6 and 7, respectively. From these results, we found that convergence rates for larger p were more improved than that

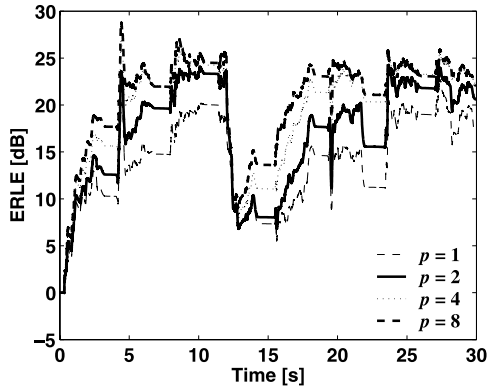

Fig. 6 Comparison of GL-APA performance with different projection orders: ERLEs for male speech input (single-talk).

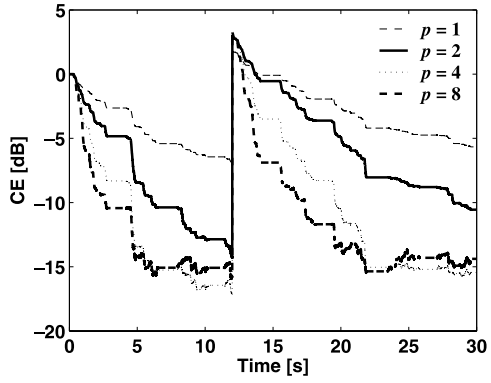

Fig. 7 Comparison of GL-APA performance with different projection orders: CEs for male speech input (single-talk).

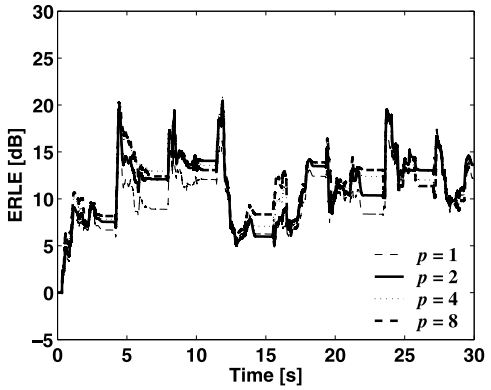

Fig. 8 Comparison of GL-APA performance with different projection orders: ERLEs for male speech input and female speech outlier (double-talk).

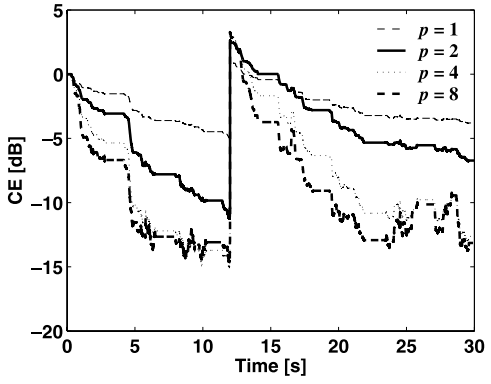

Fig. 9 Comparison of GL-APA performance with different projection orders: CEs for male speech input and female speech outlier (double-talk).

for $p = 1$ (GL-NLMS). ERLEs and CEs in the double-talk case, where the female speech (Fig. 3(b)) was an outlier for the male speech reference and the echo path was changed at 12 s again, are shown in Figs. 8 and 9, respectively. Even in the case of double-talk with echo path change, as p increased, GL-APA achieved faster convergence to the steady state, and ERLE levels in the steady state were as high as the ERLE level when $p = 1$ (GL-NLMS). Although we used the scaling rule for the function $\psi_p(v)$ based on the assumption of a white Gaussian signal input, the results obtained here demonstrate that it is also useful for the speech input, as we expected in 3.3.

4.2 Comparison with Conventional APA

We compared the proposed GL-APA and the conventional APA when $p = 8$. The step-size of GL-APA was set to $\mu = 0.55$, which is the same value used in 4.1. The echo path was fixed in this simulation. For the comparison, we chose two step-sizes for APA, $\mu = 0.05$ and 0.08 to obtain Fig. 10, in which ERLEs in the single-talk case are shown. As shown in Fig. 10, when we chose $\mu = 0.05$, APA achieved a similar convergence rate to that of GL-APA in the initial periods, while reaching a higher steady-state level. On the other hand, when we chose $\mu = 0.08$, APA reached a steady-state

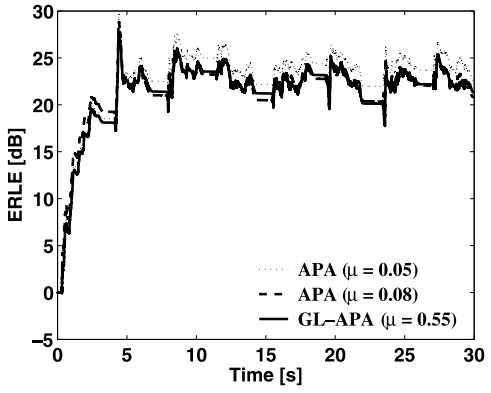

Fig. 10 Comparison between GL-APA and APA for $p = 8$: ERLEs for male speech input (single-talk).

level similar to that of GL-APA, while achieving a faster initial convergence rate. We also note that filter updates of the conventional APA were frozen when $\|\mathbf{x}(n)\| \leq 500\sqrt{L}$, though GL-APA was updated without such a control. CEs in the same situation are shown in Fig. 11.

Then, we evaluated robustness against double-talk under the above-mentioned step-size conditions. ERLEs and CEs in the double-talk case are shown in Figs. 12 and 13, respectively. In the steady state, GL-APA achieved much

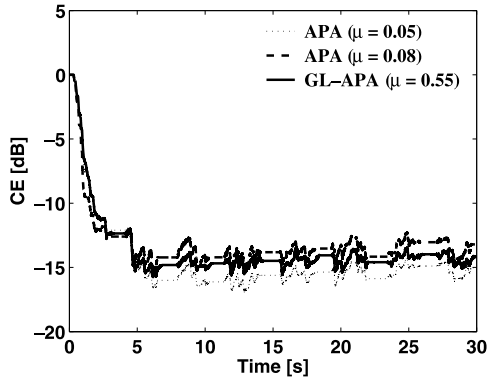

Fig. 11 Comparison between GL-APA and APA for $p = 8$: CEs for male speech input (single-talk).

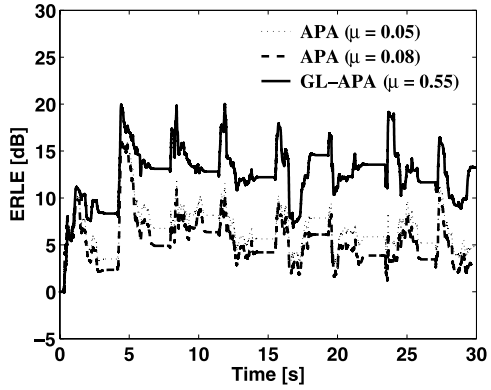

Fig. 12 Comparison between GL-APA and APA for $p = 8$: ERLEs for male speech input and female speech outlier (double-talk).

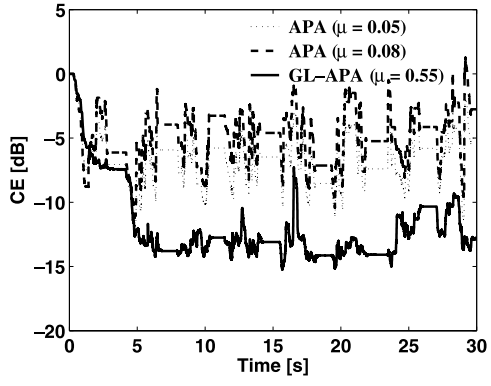

Fig. 13 Comparison between GL-APA and APA for $p = 8$: CEs for male speech input and female speech outlier (double-talk).

better ERLE and CE convergence than APA did. Thus, the GL-APA is more robust against double-talk than the conventional APA at any constant step-size.

According to Figs. 10 and 12, while the ERLE of the APA during double-talk was degraded by about 20 dB from that during single-talk, the ERLE of the GL-APA during double-talk was degraded by only about 10 dB from that during single-talk and was still about 15 dB. In the actual implementation, the degradations caused by both methods can be compensated for by applying post-filtering approaches based on short-time spectral amplitude (STSA) estimation, as shown in [26]. However, the smaller residual echo, which is due to GL-APA, provides better quality of the near-end speech post-filtered during double-talk, because the performance of the STSA estimation depends on the ratio of the near-end speech and residual echo levels.

5. Conclusion

We have presented GL-APA, which can achieve fast and double-talk-robust convergence in AEC applications. GL-APA is derived from the M-estimation-based cost function extended for evaluating multiple (the last p) error signals dealt with in the ordinary APA. Because of nonlinearity of the gradient for GL-APA, we carefully formulated the update equation consistent with multiple (the last p) input-output relationships, which APA inherently satisfies to achieve fast convergence. We also newly introduced a scaling rule for the nonlinearity, so that we can easily implement GL-APA of any p by using a simple primary function as a basis for scaling with the order p. This guarantees a linkage between GL-APA and GL-NLMS. Through the simulations, we confirmed that GL-APA can quickly and robustly track an echo path change even during double-talk and is applicable to AEC. In actual implementations, the combination of GL-APA and a double-talk or echo-path-change detector is also a reasonable solution because such a detector can provide helpful information for controlling GL-APA more precisely.

Acknowledgements

We would like to thank Mr. A. Imamura, a project manager in NTT Cyber Space Laboratories, for fruitful discussions and helpful suggestions.

References

[1] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceller with two echo path models," IEEE Trans. Commun., vol.COM-25, no.6, pp.589–595, June 1977.
[2] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and evaluation of an acoustic echo canceller using duo-filter control system," Proc. EUSIPCO96, vol.2, pp.1115–1118, Sept. 1996.
[3] D.L. Duttweiler, "A twelve-channel digital echo canceller," IEEE Trans. Commun., vol.COM-26, no.5, pp.647–653, May 1978.
[4] H. Ye and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," IEEE Trans. Commun., vol.39, no.11, pp.1542–1545, Nov. 1991.
[5] J. Benesty, D.R. Morgan, and J.H. Cho, "A new class of doubletalk detectors based on cross-correlation," IEEE Trans. Speech and Audio, vol.8, no.2, pp.168–172, Nov. 2000.
[6] T. Gänsler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence," IEEE Trans. Commun., vol.44, no.11, pp.1421–1427, Nov. 1996.
[7] K. Fujii and J. Ohga, "Double-talk detection method with detecting echo path fluctuation," IEICE Trans. Fundamentals (Japanese Edition), vol.J78-A, no.3, pp.314–322, March 1995.

[8] S. Ohta, Y. Kajikawa, and Y. Nomura, "An acoustic echo cancellation using sub-adaptive filter—Combination of echo path change and double-talk detector," The 20th Signal Processing Symposium, C3-2, Nov. 2005.
[9] B. Widrow and S.D. Stearns, Adaptive Signal Processing, Prentice Hall, 1985.
[10] M.M. Sondhi, "An adaptive echo canceller," Bell Syst. Tech. J., vol.XLVI, no.3, pp.497–511, March 1967.
[11] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," IEEE Trans. Speech and Audio, vol.8, no.6, pp.656–663, Nov. 2000.
[12] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[13] S. Shimauchi, Y. Haneda, and A. Kataoka, "A robust NLMS algorithm for acoustic echo cancellation," IEICE Trans. Fundamentals (Japanese Edition), vol.J88-A, no.8, pp.926–934, Aug. 2005.
[14] J. Nagumo and A. Noda, "A learning method for system identification," IEEE Trans. Autom. Control, vol.AC-12, no.3, pp.282–297, June 1967.
[15] T. Hinamoto and S. Maekawa, "Extended theory of learning identification," IEEJ Trans. EIS, vol.95, no.10, pp.227–234, Oct. 1975.
[16] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," IEICE Trans. Fundamentals (Japanese Edition), vol.J67-A, no.2, pp.126–132, Feb. 1984.
[17] M. Tanaka, S. Makino, and Y. Kaneda, "On the order and performance of the projection algorithm with speech input," Proc. Autumn Meet. Acoust. Soc. Jpn., 1-4-14, pp.489–490, Oct. 1992.
[18] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," Proc. 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.2, pp.945–948, May 1995.
[19] S.L. Gay and S. Tavathia, "The fast affine projection algorithm," Proc. 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp.3023–3026, May 1995.
[20] H.-C. Shin, A.H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," IEEE Signal Process. Lett., vol.11, no.2, pp.132–135, Feb. 2004.
[21] H. Yasukawa and S. Shimada, "An acoustic echo canceller using subband sampling and decorrelation methods," IEEE Trans. Signal Process., vol.41, no.2, pp.926–930, Feb. 1993.
[22] A. Hirano and A. Sugiyama, "A noise-robust stochastic gradient algorithm with an adaptive step size suitable for mobile hands-free telephones," Proc. 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.2, pp.1392–1395, May 1995.
[23] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," IEEE Trans. Signal Process., vol.50, no.5, pp.1091–1101, May 2002.
[24] B. Zhu and E. Micheli-Tzanakou, "Nonstationary speech analysis using neural prediction," IEEE Eng. Med. Biol. Mag., vol.19, no.1, pp.102–105, Jan.-Feb. 2000.
[25] S. Shimauchi, Y. Haneda, and A. Kataoka, "An error-memoryless affine projection algorithm for acoustic echo cancellation," IEICE Trans. Fundamentals (Japanese Edition), vol.J88-A, no.11, pp.1235–1245, Nov. 2005.
[26] S. Sakauchi, Y. Haneda, and A. Kataoka, "Gain emphasis method for echo reduction based on a short-time spectral amplitude estimation," IEICE Trans. Fundamentals (Japanese Edition), vol.J88-A, no.6, pp.695–703, June 2005.

Suehiro Shimauchi received the B.E. and M.E. degrees from Tokyo Institute of Technology in 1991 and 1993, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1993, he has been engaged in acoustic signal processing for acoustic echo cancellers. He is now a Senior Research Engineer at NTT Cyber Space Laboratories and is also pursuing the Ph.D. degree in the Department of Communications and Integrated Systems at Tokyo Institute of Technology. He is a member of ASJ and IEEE.

Yoichi Haneda received the B.S., M.S., and Ph.D. degrees from Tohoku University in Sendai, in 1987, 1989, and 1999, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1989, he has been investigating acoustic signal processing and acoustic echo cancellers. He is now a Senior Research Engineer at NTT Cyber Space Laboratories. He received the paper awards from the Acoustical Society of Japan, and from the Institute of Electronics, Information, and Communication Engineers of Japan in 2001. Dr. Haneda is a member of ASJ and IEEE.

Akitoshi Kataoka received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Doshisha University of Kyoto in 1984, 1986, and 1999, respectively. Since joining NTT Laboratories in 1986, he has been engaged in research on noise reduction, acoustic arrays, and medium bit-rate speech and wideband coding algorithms for the ITU-T standard. He contributed to establishing the ITU-T G.729 standard. He is currently the Manager of Acoustic Information Group at the Media Processing Laboratory, NTT Cyber Space Laboratories, Tokyo, Japan. He received Technology Development Award from ASJ in 1996 and the Prize of the Commissioner of the Japan Patent Office from the Japan Institute of Invention and Innovation in 2003. Dr. Kataoka is a member of ASJ and IEEE.

Akinori Nishihara received the B.E., M.E., and Dr. Eng. degrees in electronics from Tokyo Institute of Technology in 1973, 1975, and 1978, respectively. Since 1978 he has been with Tokyo Institute of Technology, where he is Professor of the Center for Research and Development of Educational Technology. His main research interests are in signal processing, and its application to educational technology. He served as an Associate Editor of the IEICE Trans. Fundamentals from 1990 to 1994, an Associate Editor of the IEEE Transactions on Circuits and Systems II from 1995 to 1997, and Editor-in-Chief of Trans. IEICE Part A from 1998 to 2000. He served as an ExCom member of IEEE Region 10 (Asia Pacific Region), as Student Activities Committee Chair, Treasurer and Edicational Activities Committee Chair. He now serves as a member of IEEE EAB Committee of Global Accreditation Activities, Chair of IEEE Circuits and Systems Society Japan Chapter, and adviser of IEEE Japan Council Women in Engineering Affinity Group. He received the IEICE Best Paper Award in 1999, and the IEEE Third Millennium Medal in 2000. Dr. Nishihara is a Fellow of IEEE, and a member of EURASIP, ECS, JSET, and IEEK.