Convergence Analyses of Stereo Acoustic Echo Cancelers With Preprocessing

Kazushi Ikeda and Ryohei Sakamoto

Abstract—The convergence properties of stereo acoustic echo cancelers with preprocessing are examined. We first elucidated the conditions of the paths from the talker to the microphones under which the nonuniqueness problem occurs. The degree of nonuniqueness was found to depend on the primeness of the paths. Next, the convergence rates of the acoustic echo cancelers with preprocessing were analyzed based on the independence theory for general cases and on the cyclostationariness of the input signal for the input-sliding method. The derived rates that depended on both of the period of the input signal and the order of the adaptive filter agreed well with the results of computer simulations.

Index Terms—Acoustic echo canceler, adaptive filter, convergence condition, convergence rate.

Fig. 1.   Schematic diagram of a stereo acoustic echo canceler.

I. INTRODUCTION

ACOUSTIC echo cancelers are one of the most attractive applications in adaptive filter technology. Multichannel acoustic echo cancelers are required to remove undesired echoes and to improve reality in teleconferencing; however, they include a fundamental problem that never appears in monaural systems. We consider the two channels' case, i.e., stereo acoustic echo cancelers, in this paper, since we can easily extend it to general cases.

The problem is clearly stated in [1]. In summary, a stereo acoustic echo canceler estimates four unknown paths h_1, h_2, h_3, and h_4 from two sequences of input signals $\{x_a(t)\}$ and $\{x_b(t)\}$ made from one source signal $\{u(t)\}$ and two paths, whose inpulse responses are denoted by a and b, as depicted in Fig. 1. This makes the problem of estimation ill-posed and the nonuniqueness problem occurs since the source signal has two paths from the talker's mouth to a speaker, e.g., $h_1 * a$ and $h_2 * b$ for the upper speaker in Fig. 1, where $*$ denotes the convolution.

When the paths are time-variant enough, the nonuniqueness problem disappears because the estimates $\hat{h}_1$, $\hat{h}_2$, $\hat{h}_3$, and $\hat{h}_4$ have more than two equations to satisfy. Based on this idea, some methods have been proposed (see [2, Introduction]). The input-sliding method originally proposed in [3] is also based on the same idea and is very simple: One of the input signals is identical to the original, and the other is delayed by one sample step. The preprocessing is performed only in the former half of
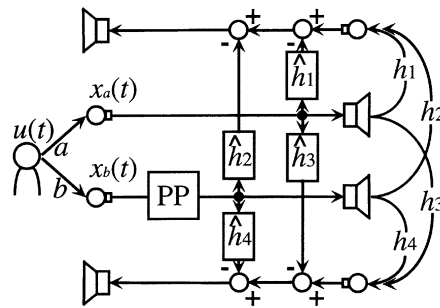
the iteration steps in a constant period $2Q$. In the latter half, the signal is identical to the original. They showed that this simple method vastly improves the convergence property, and the signal distortion is almost inaudible.

The question of why such a simple method works so well motivated us to analyze the convergence properties of stereo acoustic echo cancelers. If we elucidated the mechanism, it might be possible to propose a better method with faster convergence and smaller distortion.

The rest of the paper is organized as follows: Section II contains the mathematical formulation of the stereo acoustic echo canceler problem. The degree of nonuniqueness for a general case is evaluated in Section III. Section IV is devoted to convergence rate analyses based on the independence theory, and the analyses for the input-sliding method are done in Section V. Conclusions are given in Section VI.

II. PROBLEM STATEMENT

The fundamental problem for stereo acoustic echo cancelers has been clearly presented by Sondhi *et al.* in [1], where the input signals are unified, as are the tap-weight vectors. In brief, the problem results from the fact that the input vectors do not span the whole space wherein they belong since both input signals are made from a common source signal as detailed below. We ignore the so-called the double-talk problem, which is another important problem for acoustic echo cancelers, and assume that there are no speech signals except the source signal $\{u(t)\}$.

A. Formulation

Since a stereo acoustic echo canceler has a symmetric structure, we can simply consider the estimation of h_1 and h_2 in Fig. 1 without loss of generality.

Let the source signal be denoted by $u(t)$. Then, two microphones pick it up through two acoustic paths a and b and make the input signals $x_a(t)$ and $x_b(t)$ as

$$x_a(t) = a^* u(t)$$
$$x_b(t) = b^* u(t)$$

respectively. The task of the stereo acoustic echo canceler is to find two FIR filters $\hat{h}_1$ and $\hat{h}_2$ of order N that satisfy

$$\hat{h}_1^T x_a(t) + \hat{h}_2^T x_b(t) = d(t)$$

where the input vectors $x_a(t)$ and $x_b(t)$ and the desired output signal $d(t)$ are, respectively, defined as

$$x_a(t) = (x_a(t), x_a(t-1), \ldots, x_a(t-N+1)) \in R^N$$
$$x_b(t) = (x_b(t), x_b(t-1), \ldots, x_b(t-N+1)) \in R^N$$
$$d(t) = h_1^T x_b(t) + h_2^T x_b(t).$$

For simplicity, we denote

$$x(t) = \left(x_a(t)^T, x_b(t)^T \right)^T$$
$$\hat{h} = \left(\hat{h}_1^T, \hat{h}_2^T \right)^T$$
$$h = \left(h_1^T, h_2^T \right)^T$$

and then, the equation that should be satisfied is rewritten as $d(t) = \hat{h}^T x(t)$. Note that for simplicity, we assume that no noise exists in our analyses.

B. Statistics of Input Signals

We consider the statistical properties of $x(t)$, assuming that the paths from the source to the microphones are FIR filters of order $M+1$ such that

$$a^T = (a_0, a_1, \ldots, a_M) \in R^{M+1}$$
$$b^T = (b_0, b_1, \ldots, b_M) \in R^{M+1}$$

and we can write the input vector as

$$x(t) = Su(t) \tag{1}$$

where

$$D_a = \begin{pmatrix} a_0 & a_1 & \cdots & a_M & 0 & \cdots & 0 \\ 0 & a_0 & \cdots & a_{M-1} & a_M & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & & & a_{M-1} & a_M \end{pmatrix} \tag{2}$$

$$D_b = \begin{pmatrix} b_0 & b_1 & \cdots & b_M & 0 & \cdots & 0 \\ 0 & b_0 & \cdots & b_{M-1} & b_M & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & & & b_{M-1} & b_M \end{pmatrix} \tag{3}$$

$$S = \begin{pmatrix} D_a \\ D_b \end{pmatrix} \in M(2N, N+M) \tag{4}$$

$$u(t) = (u(t), u(t-1), \ldots, u(t-N-M+1)). \tag{5}$$

From (1), $x(t) \in R^{2N}$ exists in the space spanned by the $(N+M)$ column vectors of $S \in M(2N, N+M)$. We assume in the

following that $u(t)$ is fully ranked, that is, rk $E[u(t)u^T(t)] = N + M$. In this case, the dimension spanned by the $(N+M)$ column vectors is determined by the echo paths a and b.

C. Adaptive Algorithm

Standard methods of adaptive filters are applicable to finding an $\hat{h}$ such that $\hat{h}^T x(t) = d(t)$. One of the simplest and most effective methods is the normalized LMS (N-LMS) algorithm [4], [5], which is described as

$$\hat{h}(t+1) = \hat{h}(t) + \mu \frac{x(t)e(t)}{\|x(t)\|^2}$$
$$= \hat{h}(t) + \mu x(t) \frac{d(t) - \hat{h}^T(t)x(t)}{\|x(t)\|^2}$$

where μ is the learning coeffficient. Since $d(t) = h^T x(t)$, the weight-error vector $g(t) = \hat{h}(t) - h$ at time t is updated as

$$g(t+1) = g(t) - \mu \frac{x(t)x^T(t)}{\|x(t)\|^2} g(t)$$
$$= \left[I - \mu \frac{x(t)x^T(t)}{\|x(t)\|^2} \right] g(t). \tag{6}$$

Hence, $\|g(t+1)\|$ is less than $\|g(t)\|$ for $0 < \mu < 2$, and $\|g(t+1)\|$ becomes the minimum for $\mu = 1$.

In the following, we consider the case of $\mu = 1$, where the transition matrix $[I - x(t)x^T(t)/\|x(t)\|^2]$ becomes a projection matrix to a hyperplane orthogonal to $x(t)$. This means that only the component in $g(t)$ parallel with $x(t)$ can change. Therefore, the components orthogonal to the $(N+M)$ column vectors of S never change in adaptation, and $g(t)$ does not converge to the null vector.

The affine projection (AP) algorithm and an LS-type method have been proposed to accelerate the convergence in [6]–[8], respectively, but they do not solve the nonuniqueness problem since the restriction of the input vectors does not change. Moreover, as they are sometimes computationally consuming, we consider only the N-LMS algorithm in this paper.

III. NONUNIQUENESS PROBLEM

Since $x(t) \in R^{2N}$ does not span the whole $2N$-dimensional space, the weight vector $\hat{h}$ that satisfies $\hat{h}^T x(t) = d(t)$ is not determined uniquely. In other words, the weight-error vector $g(t) = \hat{h}(t) - h$ can include some components perpendicular to the column vectors of S. This is the origin of the nonuniqueness problem.

Since the residual of the weight-error vector $g(t)$ is composed of vectors orthogonal to the column vectors of S, the nonuniqueness problem vanishes when the vectors increase, and the space they span is spread by means of some tricks.

A. Heuristic Methods

Two kinds of methods have been proposed to overcome the problem. One is to add noise to the input signal (or to assume the existence) [7], [8], whereas the other is to give time-variation to the relation of the input signals. Both increase the rank of S to $2N$ in time-average, and the tap-weight vector $\hat{h}$ is uniquely

determined. However, as the former greatly degrades the quality of input signals, we consider the latter in this paper.

Shimauchi and Makino [6] clearly showed that projection-type algorithms, e.g., the N-LMS and the AP algorithms, can make the weight-error vector converge to null when the learning coefficient μ is set to one that minimizes $\|\boldsymbol{g}(t+1)\|$, as mentioned before.

The method proposed in [2] and [3] is also based on the same idea, that is, giving a time-varying nonlinear transformation to the input signals. One of the input signals, e.g., $\boldsymbol{x}_b(t)$, is preprocessed with a two-tap time-varying filter with a period $2Q$. The filter is 1 for the first Q steps and z^{-1} for the following Q steps. They also proposed making transition areas in switching to avoid "clicks" that result from sudden delays. However, for simplicity, we do not pay attention to this. Their method gives a good performance in computer simulations and listening tests, but the reason why such a simple method works so well is unknown. The purpose of our research is to elucidate these reasons by theoretically analyzing the convergence properties.

Hirano *et al.* [9]–[11] generalized the preprocessing in [2] and [3] and gave some theoretical analyses. They considered the relation of $\boldsymbol{g}(t)$ and $\boldsymbol{g}(t+2Q)$ when $\boldsymbol{x}_a(t)$ is preprocessed with each of F_a and F_a' for Q steps and $\boldsymbol{x}_b(t)$ with F_b and F_b' in the same way. The results showed, however, that time-variation is necessary for convergence and that the convergence is experimentally faster when the change is larger. Since a large amount of preprocessing the input signal greatly distorts or degrades the quality, it is necessary to find a preprocessor with fast convergence and small distortion. The solution is uniquely determined if the preprocessing is time-variant under some conditions [9], [10].

B. Nonuniqueness Theorem

The nonuniqueness problem results from the properties of S, as mentioned above. Therefore, we first calculate the strict rank of S.

Since S is a $2N \times (N+M)$ matrix and the left upper $N \times N$ elements become a triangular matrix, $N \leq \mathrm{rk}\, S \leq (N+M)$ holds true anytime. Considering the structure of S, if $\boldsymbol{b}$ is a combination of $\boldsymbol{a}$ and its shifted ones, $\mathrm{rk}\, S$ may be less than $N+M$. An extreme example is the case of $\boldsymbol{b} = c\boldsymbol{a}$, where c is a constant, which is treated in [6]. Then, $\mathrm{rk}\, S = N$ is easily shown. However, unless $\boldsymbol{b}$ is represented as a combination of $\boldsymbol{a}$ and its shifted ones, S is expected to become fully ranked, that is, $\mathrm{rk}\, S = N+M$. Indeed, the following can be proven.

Theorem 1 (Rank of S): Let $A(z)$ and $B(z)$ be the z-transforms of the paths $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. If $A(z)$ and $B(z)$ are mutually prime, in other words, there exist FIR systems $P_1(z)$ and $P_2(z)$ s.t. $A(z)P_1(z) + B(z)P_2(z) = 1$, then $\mathrm{rk}\, S = N+M$.

The proof is in Appendix A. When the order M_a of $A(z)$ and the order M_b of $B(z)$ are different, $\mathrm{rk}\, S = N + \max(M_a, M_b)$ is shown in the same way.

In the above statement, $M \leq N$ is implicitly assumed. When $M \geq N$, $\mathrm{rk}\, S = 2N$ is easily shown using the first N columns of $D_a^{\,T}$ and the last N columns of $D_b^{\,T}$. Note that the nonuniqueness problem does not exist in this case, which is also mentioned in [9] and [10].

C. Examples

We here show two examples to evaluate the degree of nonuniqueness using the above theorem.

One is the case $\boldsymbol{x}_2(t) = c\boldsymbol{x}_1(t)$ treated in [6]. Since $\boldsymbol{x}_2(t) = c\boldsymbol{x}_1(t)$ and then $\boldsymbol{b} = c\boldsymbol{a}$, $A(z)$ and $B(z)$ are not mutually prime. If we regard $\boldsymbol{a}^T\boldsymbol{u}(t)$ as a source signal, $D_a = I_N$ and $D_b = cI_N$ hold true, where I_N is an $N \times N$ identity matrix, and they are mutually prime with $M = 0$. This means that $\mathrm{rk}\, S = N$ and that the solution of the stereo acoustic echo canceler has N degrees of freedom if $\boldsymbol{a}^T\boldsymbol{u}(t)$ has a fully ranked covariance matrix.

The other is the case of $\boldsymbol{x}_2(t) = c\boldsymbol{x}_1(t-k)$ treated in [12]. In the same way as above, we redefine $u(t)$ as $\boldsymbol{a}^T\boldsymbol{u}(t)$. Then, $D_a = [I_N, O] \in M(N, N+k)$ and $D_b = [O, cI_N] \in M(N, N+k)$, where O is a null matrix of an appropriate size. This means that $\mathrm{rk}\, S = N + k$ and that the solution of the stereo acoustic echo canceler has $N - k$ degrees of freedom. Note that the solution is uniquely determined if $k \geq N$ is satisfied.

D. Input-Sliding Method

As the main purpose of the paper is to analyze the properties of the input-sliding method proposed in [2] and [3], we explicitly show that the input-sliding method can remove the nonuniequeness.

We redefine S and define S' as

$$S = \begin{pmatrix} D_a \mathbf{0}_N \\ D_b \mathbf{0}_N \end{pmatrix} \in M(2N, N+M+1)$$

$$S' = \begin{pmatrix} D_a \mathbf{0}_N \\ \mathbf{0}_N D_b \end{pmatrix} \in M(2N, N+M+1)$$

respectively, where $\mathbf{0}_N$ is an N-dimensional null vector. The fact of $\mathrm{rk}\,(S, S') = 2N$ is shown as follows:

$$\begin{aligned} \mathrm{rk}\,(S, S') &= \mathrm{rk}\,(S', S) \\ &= \mathrm{rk} \begin{pmatrix} D_a \mathbf{0}_N & D_a \\ \mathbf{0}_N D_b & D_b \end{pmatrix} \\ &= \mathrm{rk} \begin{pmatrix} (D_a \mathbf{0}_N) - (\mathbf{0}_N D_a) & D_a \\ \mathbf{0}_N O_N & D_b \end{pmatrix} \\ &= 2N \end{aligned}$$

where O_N is an $N \times N$ null matrix since $\mathrm{rk}\,[(D_a \mathbf{0}) - (\mathbf{0} D_a)] = N$. Note that this result is a special case of the theorem in [9] and [10].

IV. CONVERGENCE RATE BASED ON INDEPENDENCE THEORY

In this paper, we evaluate the convergence rate, which is defined as the amount of how much the weight-error vector $\boldsymbol{g}(t)$ decreases in magnitude, i.e., the average of $\|\boldsymbol{g}(t+1)\|/\|\boldsymbol{g}(t)\|$. The "independence theory" employed in this section is a classical assumption that the input signal vector $\boldsymbol{x}(t)$ is independent of the weight-error vector $\boldsymbol{g}(t)$. Nevertheless, even though the assumption is far from true, the results predicted by the independence theory are usually found to be in agreement with experiments and computer simulations [5].

Based on the independence theory, we derive the convergence rate from the maximum eigenvalue of the expected transition matrix on the input vector $\boldsymbol{x}(t)$ as in [5], [13], and [14]. We evaluate the convergence rate for time-variant input signals in

this section and devote the next section to analyses of the input-sliding method.

A. Convergence Rate of N-LMS Algorithm

First, we show the convergence rate of the adaptive FIR filter of length N with the N-LMS algorithm of $\mu = 1$.

According to the method based on the independence theory in [5], the convergence rate is represented by the maximum eigenvalue of the transition matrix of the weight error vector $\|g(t)\|$ averaged over the input signal $x(t)$, which obeys a normal distribution $N(0, I_N)$. In the case of the N-LMS algorithm, the transition matrix is shown as in (6). Hence, the convergence rate is described as the maximum eigenvalue of $E[(I_N - x(t)x^T(t))/(\|x(t)\|^2)]$. Although this is difficult to calculate directly, we can approximate it by the large number law to $(1 - 1/N)I_N$ using

$$\|x(t)\|^2 = \sum_{i=0}^{N-1} x(t-i)^2 \approx E\left[\|x(t)\|^2\right].$$

Therefore, the approximate value of the convergence rate is $1 - 1/N$ for the adaptive filter of order N.

B. Convergence Rate for Input Signals in Subspace

When the input signal vectors $x(t)$ only span an N-dimensional subspace in R^{2N}, the convergence properties are a little different. Let S and $S^\perp$ denote the N-dimensional subspace spanned by the input vectors $x(t)$ and its orthogonal complement, respectively. Then, the components of $g(0)$ in $S^\perp$ are never changed by $x(t)$, whereas the theory in the previous subsection holds true in S. This means that when t input vectors are drawn from the same distribution and given to an adaptive filter of order $2N$, then the convergence rates of the components in S and $S^\perp$ are, respectively, $1 - 1/N$ and 1. Hence, $\|g(t)\|$ can be described as

$$\|g(0)\|\sqrt{\cos^2\theta' + \sin^2\theta'\left(1 - \frac{1}{N}\right)^{2t}} \quad (7)$$

if the components of $g(0)$ in S and $S^\perp$ are $\|g(0)\|\sin\theta'$ and $\|g(0)\|\cos\theta'$ in magnitude, respectively.

C. Convergence Rate for Time-Variant Signals

In the problem of stereo acoustic echo cancelers with algorithms based on the time-variation of signal statistics such as the input-sliding method, the input signal vectors $x(t)$ have some kinds of statistical properties, one of which is employed at each time t. Let us consider the case that the input vectors $x(t)$ are drawn from one probability distribution for Q steps and from the other for the next Q steps. Therefore, the input signal has a period of $2Q$ as in [2], [3], [9], and [10].

Based on the independence theory, Hirano *et al.* [9], [10] considered the transition matrix and gave some intuitive comments that a large amount of change in time may result in a faster convergence. We here derive the convergence rate of an FIR filter
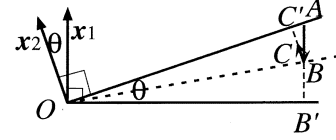

Fig. 2. Convergence properties for two kinds of input signal vectors when $Q \ll N$.

given input signals with two kinds of statistics, each of which spans only an N-dimensional subspace, i.e.,

$$x_1(t) = D_a u \in R^{2N}$$
$$x_2(t) = D_b u \in R^{2N}$$
$$D_a, D_b \in M(2N, N)$$
$$u(t) \sim N(0, I_N), \text{i.i.d.}.$$

The angle θ of $x_1(t)$ and $x_2(t)$ is defined as

$$\cos\theta = \frac{E\left[\|x_1\|^2\right] E\left[\|x_2\|^2\right]}{E\left[\|x_1\|^2\right]^{1/2} E\left[\|x_1\|^2\right]^{1/2}}.$$

Let us consider first the case of $Q \ll N$. From (7), the convergence rate depends on the angle θ' between the current error vector $g(t)$ and the input signal $x(t)$. We show that the angle θ' approximately fluctuates around $\theta/2$.

Fig. 2 shows a schematic view of what happens in the space of weight-error vectors when two statistical kinds of input signals are given. $\overline{OA}$ shows the spaces spanned by the weight-error vectors orthogonal to the input vector x_2 and the vectors in $\overline{OB'}$ that are perpendicular to the input vector x_1. Assume that the weight-error vector is changed by adaptation from A to B and that $\angle AOB$ is almost half of $\angle AOB'$, meaning $(1 - 1/(2N))^Q \approx 1/2$ in (7). Then, C, given by the next step, is located around the point where $\angle BOC$ is half of $\angle BOC'$ since the rate of $\overline{AB}$ to $\overline{AB'}$ is the same as that of $\overline{BC}$ to $\overline{BC'}$ due to (7). By repetition, the points given by the adaptation are located around $\theta' = \theta/2$ with width $\theta/6$ since $2\theta/3*1/2 = \theta - \theta/3$ in this case if θ is small, and $\theta \approx \sin\theta$. In general, when the components in S are shrunk at the rate of $(1 - \eta)$, the points are located around $\theta/2$ with width $\theta/2 - \varphi$, where φ satisfies

$$\eta = \frac{\tan\varphi}{\tan(\theta - \varphi)}.$$

The proof is in Appendix B.

Due to the assumption of $Q \ll N$, $\eta \approx 1$ is satisfied, and then, $\varphi \approx \theta/2$, and $\theta' \approx \theta/2$. Hence, from (7), the weight-error vector $g(t)$ is shrunk at the rate of

$$\sqrt{\cos^2\left(\frac{\theta}{2}\right) + \sin^2\left(\frac{\theta}{2}\right)\left(1 - \frac{1}{N}\right)^{2Q}}$$

$$\approx \sqrt{1 - \frac{2Q\sin^2\left(\frac{\theta}{2}\right)}{N}}$$

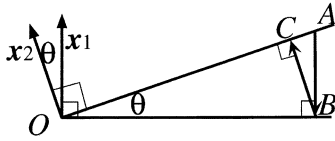$$\approx 1 - \frac{Q\sin^2\left(\frac{\theta}{2}\right)}{N}$$

Fig. 3. Convergence properties for two kinds of input signal vectors when $Q \gg N$.

by Q iteration steps, where $(1 - 1/N)^{2Q} \approx (1 - 2Q/N)$, etc., are applied since $Q \ll N$. Therefore, the convergence rate is written as

$$\left(1 - \frac{Q \sin^2 \left(\frac{\theta}{2}\right)}{N}\right)^{1/Q} \approx 1 - \frac{\sin^2 \left(\frac{\theta}{2}\right)}{N} \qquad (8)$$

which does not depend on the period $2Q$.

Next, we consider the case of $Q \gg N$. Since enough updates are done for one statistical kind of input signal vectors, the components in S become null, and the weight-error vector $\boldsymbol{g}(t)$ is projected onto $S^\perp$ in Q iteration steps. Fig. 3 shows a schematic view of weight-error vectors in such a situation. A moves to B in Q steps, where $\overline{OB}$ are orthogonal to $\boldsymbol{x}_1$ and B to C in the next Q steps, where $\overline{OC}$ are orthogonal to $\boldsymbol{x}_2$. The convergence rate is, therefore, described as

$$(\cos \theta)^{1/Q} = \left(1 - \sin^2 \theta\right)^{1/(2Q)}$$
$$\approx 1 - \frac{\sin^2 \theta}{2Q} \qquad (9)$$

which only depends on the period $2Q$.

Finally, we give the result for general cases: When θ is small

$$\left[\cos^2 \theta' + \sin^2 \theta' \left(1 - \frac{1}{N}\right)^{2Q}\right]^{1/2Q}$$
$$\approx 1 - \frac{\sin^2 \theta'}{2Q} \left[1 - \left(1 - \frac{1}{N}\right)^{2Q}\right] \qquad (10)$$

which depends on both N and Q. We cannot determine the angle θ'. However, $\theta/2$ is a lower bound of θ' since $\theta' = \theta - \varphi$ and $\varphi < \theta/2$, as mentioned in the case of $Q \ll N$. Hence, (10) with $\theta' = \theta/2$ gives a lower bound of the convergence rate.

D. Computer Simulations

To confirm the validity of the theoretical analyses above, some computer simulations were done. In each of the experiments below, the initial value $\boldsymbol{g}(0)$ of the weight-error vector was randomly chosen from $N(0, I_{2N})$, and the experimental convergence rate was calculated as the average of $\|\boldsymbol{g}(t)\| / \|\boldsymbol{g}(t-1)\|$ from $t = 1001$ to $t = 2000$ over eight trials. Since the rate was near unity, the difference from unity was plotted instead in the figures. The two kinds of input signal vectors $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ were made according to

$$\boldsymbol{x}_1(t) = \begin{pmatrix} I_N \\ O_N \end{pmatrix} \boldsymbol{u}(t)$$
$$\boldsymbol{x}_2(t) = \begin{pmatrix} I_N \cos \theta \\ I_N \sin \theta \end{pmatrix} \boldsymbol{u}(t)$$
$$\boldsymbol{u}(t) \sim N(0, I_N), \text{i.i.d.}$$
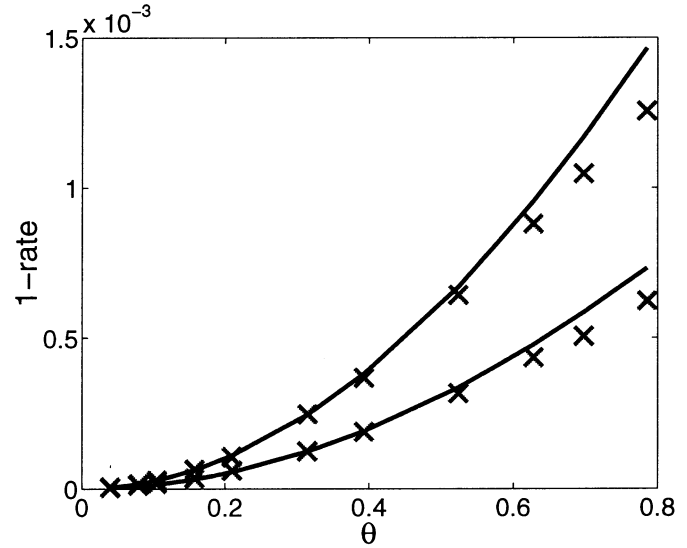
respectively, s.t. their angle becomes θ.

Fig. 4. Convergence rate versus the angle of two statistics of input signals in the case of $Q \ll N$, where $Q = 2$ and $N = 100$ or 200.
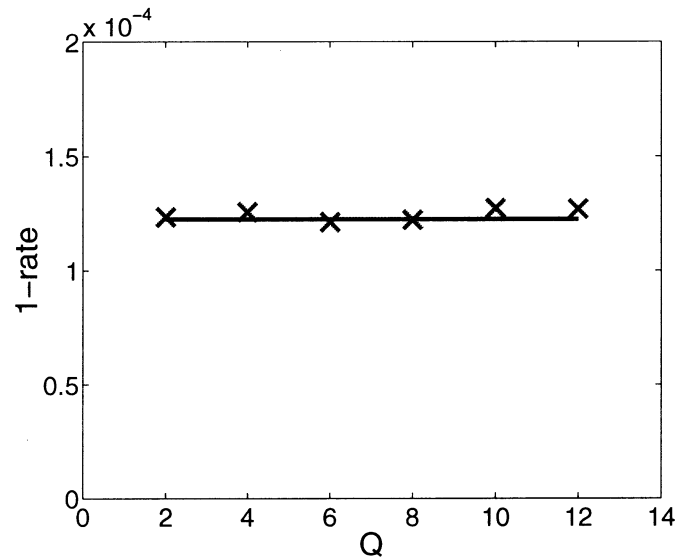

Fig. 5. Convergence rate versus the period $2Q$ of input signals in the case of $Q \ll N$, where $N = 200$ and $\theta = \pi/10$.

Fig. 4 shows the convergence rate versus the angle θ of input signals in the case of $Q \ll N$. The crosses show the experimental convergence rates and the solid lines represent the theoretical convergence rates given by (8). The upper part is the case of $Q = 2$ and $N = 100$ and the lower $Q = 2$ and $N = 200$. The theoretical values agree well with the experimental data. Fig. 5 shows the convergence rate versus the period $2Q$ of input signals in the case of $Q \ll N$. The experimental convergence rate does not depend on Q, as the theory predicts.

Fig. 6 shows the convergence rate versus the angle θ of input signals in the case of $Q \gg N$. The crosses show the experimental convergence rate and the solid line represents the theoretical convergence rate given by (9). The theoretical values agree well with the experimental data. Fig. 7 shows the convergence rate versus the order $2N$ of the adaptive filter in the case
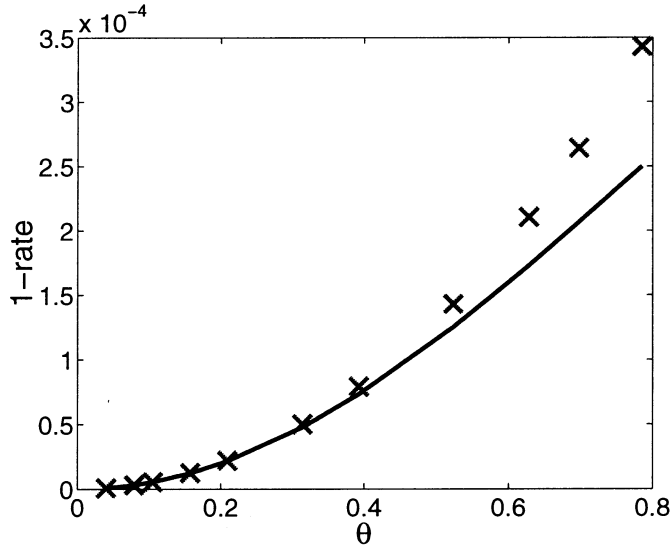
Fig. 6. Convergence rate versus the angle of two statistics of input signals in the case of $Q \gg N$, where $Q = 1000$, and $N = 20$.
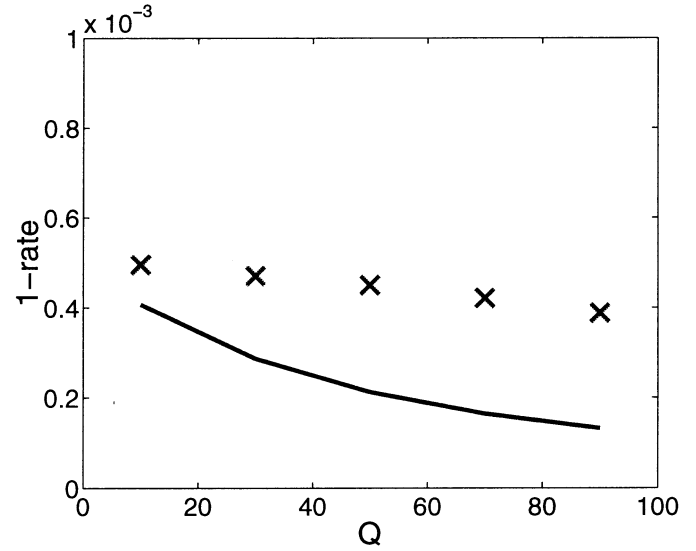

Fig. 8. Convergence rate versus the period $2Q$ of input signals in the case of $Q \approx N$, where $N = 50$ and $\theta = \pi/10$.
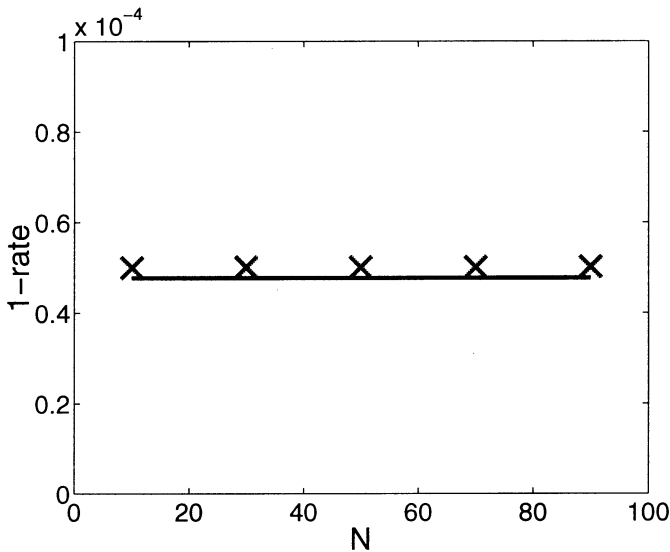

Fig. 7. Convergence rate versus the order $2N$ of the adaptive filter in the case of $Q \ll N$, where $Q = 1000$, and $\theta = \pi/10$.
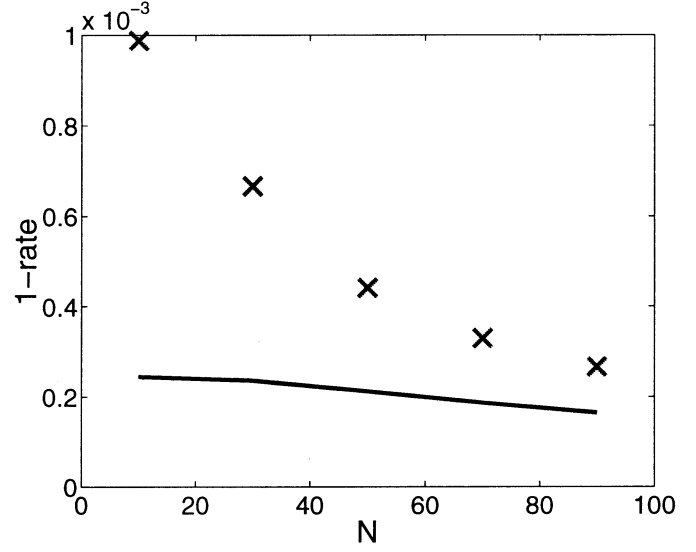

Fig. 9. Convergence rate versus the order $2N$ of the adaptive filter in the case of $Q \approx N$, where $Q = 50$ and $\theta = \pi/10$.

of $Q \gg N$. The experimental convergence rate does not depend on N, as the theory predicts.

Figs. 8 and 9 show the convergence rate versus the period $2Q$ and the order $2N$, respectively, in the case of $Q \approx N$. The crosses show the experimental convergence rate, and the solid line represents the theoretical lower bound of the convergence rate given by (10). The crosses are located above the theoretical line.

V. CONVERGENCE RATE OF THE INPUT-SLIDING METHOD

A. Input-Sliding Method

The input-sliding method proposed in [2] and [3] is one of the methods based on preprocessing that gives the statistics of the input signals a time-variation. It employs a two-tap filter $c(t) + [1 - c(t)]z^{-1}$ as a preprocessor for one of the channels

where $c(t) = 1$ for the first Q iteration steps and $c(t) = 0$ for the following Q iteration steps in the period $2Q$. The other channel has no preprocessor. This method showed a good performance in their computational experiments and listening tests; however, the reason was not clear. We derive the convergence rate of the method analytically in this section. Note that $c(t)$ is modified to vary smoothly in the original method, i.e., transition areas are given, in order to avoid the degradation of signals by clicks due to sudden delays and restoration. However, we neglect this for simplicity. In this section, we assume $Q \ll N$ since it holds true in most applications.

B. Convergence Rate Analysis

Fig. 10 shows a schematic view of how the input-sliding method works in a parameter space. Different from the situations in the previous section, the statistical characteristics
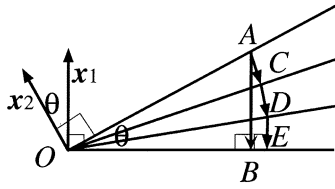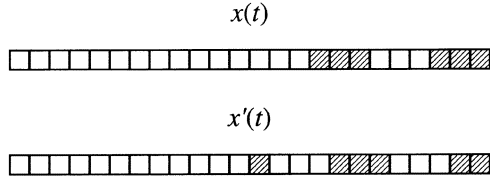
Fig. 10.　Effects of a gradual change of input signal.

Fig. 11.　Structure of input signal vector with $N = 12$ and $Q = 3$.

gradually change, as does the hyperplane onto which the weight-error vector should be projected. Let us consider, concretely, the case where the input signal vectors are orthogonal to $\overline{OA}$ and where the error vector is located at A at time t (see Fig. 10). If the input signal vector suddenly changes in statistics and becomes orthogonal to $\overline{OB}$, the error vector moves in the direction of B. However, the preprocessor can affect only the current input signals $x_1(t)$ and $x_2(t)$, which are the first and the $(N + 1)$th elements of the input vector $\boldsymbol{x}(t)$.

Fig. 11 shows the structures of the input signals $\boldsymbol{x}(t)$ and $\boldsymbol{x}'(t)$ before and after the statistics change, respectively, in the case of $Q \ll N$. Each square shows an input signal, and the elements in the shadowed squares are delayed. From the relation between $\boldsymbol{x}(t)$ and $\boldsymbol{x}'(t)$ seen in Fig. 11, the angle θ of $\overline{OA}$ and $\overline{OC}$ in Fig. 10 satisfies

$$\cos\theta = \frac{E\left[\boldsymbol{x}^T(t)\boldsymbol{x}'(t)\right]}{E\left[\|\boldsymbol{x}(t)\|^2\right]E\left[\|\boldsymbol{x}'(t)\|^2\right]}$$
$$= \frac{\left(2N - \frac{N}{Q}\right)}{2N} = 1 - \frac{1}{(2Q)}. \tag{11}$$

Note that $\boldsymbol{x}'(t)$ is a vector made by shifting the shadowed components in $\boldsymbol{x}(t) = (x(t), x(t-1), \ldots, x(t-N+1))$. Hence, as seen in the previous section, the convergence rate for $Q \ll N$ is approximately written as (8)

$$1 - \frac{\sin^2\left(\frac{\theta}{2}\right)}{N} = 1 - \frac{1 - \cos\theta}{2N}$$
$$= 1 - \frac{1}{(4QN)}. \tag{12}$$

C. Modification of Input-Sliding Method

The convergence rate $1 - 1/(4QN)$ of the input-sliding method is much larger than $1 - \sin^2(\theta/2)/N$ or $1 - \sin^2\theta/(2Q)$, which are given when the input vectors are mutually independent. To cope with this, we have the concept of the use of the input signal vector where all components are delayed, instead

of the vector delayed componentwise, in the adaptation of $\hat{\boldsymbol{h}}$. If the independence theory holds true, since

$$\boldsymbol{x}_1(t) = D_a\boldsymbol{u}(t) = \begin{pmatrix} I_N \\ I_N \end{pmatrix}\boldsymbol{u}(t)$$
$$\boldsymbol{x}_2(t) = D_b\boldsymbol{u}(t) = \begin{pmatrix} I_N \\ I_N z^{-1} \end{pmatrix}\boldsymbol{u}(t)$$
$$\boldsymbol{u}(t) \sim N(0, I_N)$$

the angle θ between $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ satisfies

$$\cos\theta = \frac{N}{(2N)} = \frac{1}{2}$$

and $\theta = \pi/3$ due to the whiteness of $\boldsymbol{u}(t)$. This implies that the modified version may converge much faster than the original input-sliding method. In fact, the modified version has much slower convergence than the original one. An explanation follows.

The input vectors $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ are rewritten more precisely as

$$\boldsymbol{x}_1(t) = D_a\boldsymbol{u}_{N+1}(t) = \begin{pmatrix} I_N & \boldsymbol{0}_N \\ I_N & \boldsymbol{0}_N \end{pmatrix}\boldsymbol{u}_{N+1}(t)$$
$$\boldsymbol{x}_2(t) = D_b\boldsymbol{u}_{N+1}(t) = \begin{pmatrix} I_N & \boldsymbol{0}_N \\ \boldsymbol{0}_N & I_N \end{pmatrix}\boldsymbol{u}_{N+1}(t)$$

where $\boldsymbol{u}_{N+1}(t) = (u(t), u(t-1), \ldots, u(t-N)) \in R^{N+1}$ and assume the whiteness of $\boldsymbol{u}_{N+1}(t)$, that is, $\boldsymbol{u}_{N+1}(t) \sim N(0, I_{N+1})$ i.i.d.

Since the input signal is cyclostationary, we consider the expectation of the transition matrix over the period $2Q$. This means that we assume the independence of every $2Q$ steps instead of each iteration step, as in the independence theory.

The expectation of the transition matrix over $2Q$ steps becomes

$$E\left[\prod_{i=0}^{2Q-1}\left(I_{2N} - \frac{\boldsymbol{x}(t+i)\boldsymbol{x}^T(t+i)}{\|\boldsymbol{x}(t+i)\|^2}\right)\right]$$
$$\approx E\left[I_{2N} - \sum_{i=0}^{2Q-1}\frac{\boldsymbol{x}(t+i)\boldsymbol{x}^T(t+i)}{\|\boldsymbol{x}(t+i)\|^2}\right]$$
$$= I_{2N} - Q\left(E\left[\frac{\boldsymbol{x}_1(t)\boldsymbol{x}_1^T(t)}{\|\boldsymbol{x}_1(t)\|^2}\right] + E\left[\frac{\boldsymbol{x}_2(t)\boldsymbol{x}_2^T(t)}{\|\boldsymbol{x}_2(t)\|^2}\right]\right)$$

since $\boldsymbol{x}(t)\boldsymbol{x}^T(t)/\|\boldsymbol{x}(t)\|^2$ is as small as of the order of $1/N$. Hence, its average in one step is written as

$$\left[I_{2N} - Q\left(E\left[\frac{\boldsymbol{x}_1(t)\boldsymbol{x}_1^T(t)}{\|\boldsymbol{x}_1(t)\|^2}\right] + E\left[\frac{\boldsymbol{x}_2(t)\boldsymbol{x}_2^T(t)}{\|\boldsymbol{x}_2(t)\|^2}\right]\right)\right]^{1/2Q}$$
$$\approx I_{2N} - \frac{1}{2}\left[\frac{1}{2N}\begin{pmatrix} I_N & I_N \\ I_N & I_N \end{pmatrix} + \frac{1}{2N}\begin{pmatrix} I_N & J_N^T \\ J_N & I_N \end{pmatrix}\right]$$
$$= I_{2N} - T$$

using the same method in Section IV, where

$$T = \frac{1}{2N}\begin{pmatrix} I_N & \frac{1}{2}\left(I_N + J_N^T\right) \\ \frac{1}{2}\left(I_N + J_N^T\right) & I_N \end{pmatrix}$$
$$J_N = \begin{pmatrix} \boldsymbol{0}_{N-1} & I_{N-1} \\ 0 & \boldsymbol{0}_{N-1}^T \end{pmatrix}.$$

Suppose that

$$J_N' = \begin{pmatrix} \mathbf{0}_{N-1} & I_{N-1} \\ 1 & \mathbf{0}_{N-1}^T \end{pmatrix}$$

is substituted for J_N, and consider $I_{2N} - T'$ instead of $I_{2N} - T$, where

$$T' = \frac{1}{2N} \begin{pmatrix} I_N & \frac{1}{2}\left(I_N + J_N'^T\right) \\ \frac{1}{2}\left(I_N + J_N'^T\right) & I_N \end{pmatrix}.$$

Then, $I_{2N} - T'$ has a null eigenvalue with an eigenvector $\mathbf{e} = (\mathbf{1}_N^T, -\mathbf{1}_N^T)^T$. Since T is similar to T', $I_{2N} - T$ also has a small magnitude eigenvalue, and the convergence is much slower.

The slowness can be quantitatively evaluated using the Rayleigh quotient theorem [15]:

Theorem 2 (Rayleigh Quotient): Let $M \in M(n)$ be a normal matrix ($MM^* = M^*M$), and denote its eigenvalues by $\lambda_1, \ldots, \lambda_n$. If the Raileigh quotient $\nu = \mathbf{e}^* M \mathbf{e}$ with $\|\mathbf{e}\| = 1$ satisfies

$$\|M\mathbf{e} - \nu\mathbf{e}\| = \varepsilon$$
$$|\lambda_j - \nu| \geq \delta > \varepsilon \quad (j \geq 2)$$

then

$$|\lambda_1 - \nu| \leq \frac{\varepsilon^2}{\left[\delta\left(1 - \frac{\varepsilon^2}{\delta^2}\right)\right]}.$$

In order to apply Theorem 2 to the matrix T, we set the kth element e_k of the vector $\mathbf{e}$ to

$$e_k = \frac{1}{\sqrt{N}} \sin \frac{\pi}{N}\left(k - \frac{3}{4}\right) \quad (1 \leq k \leq N) \qquad (13)$$
$$e_k = -e_{2N-k+1} \quad (N+1 \leq k \leq 2N) \qquad (14)$$

where $\|\mathbf{e}\| = 1$ is satisfied since

$$\|\mathbf{e}\| = \sum_{k=1}^{2N} e_k^2$$
$$= \frac{2}{N} \sum_{k=1}^{N} \sin^2 \frac{\pi}{N}\left(k - \frac{3}{4}\right)$$
$$= 1 - \frac{1}{N} \sum_{k=1}^{N} \cos \frac{2\pi}{N}\left(k - \frac{3}{4}\right)$$
$$= 1.$$

Denote the kth element of $T\mathbf{e}$ by $(T\mathbf{e})_k$. Then

$$(T\mathbf{e})_k = \frac{1}{2N}\left[e_k + \frac{e_{N+k-1}}{2} + \frac{e_{N+k}}{2}\right]$$
$$= \frac{1}{2N}\left[e_k - \frac{e_{N-k+2}}{2} + \frac{e_{N-k+1}}{2}\right]$$
$$= \frac{1}{2N^{3/2}}\left[\sin \frac{\pi}{N}\left(k - \frac{3}{4}\right)\right.$$
$$\left. - \frac{1}{2}\left\{\sin \frac{\pi}{N}\left(k - \frac{5}{4}\right) \sin \frac{\pi}{N}\left(k - \frac{1}{4}\right)\right\}\right]$$
$$= \frac{1}{2N^{3/2}}\left[\sin \frac{\pi}{N}\left(k - \frac{3}{4}\right) - \sin \frac{\pi}{N}\left(k - \frac{3}{4}\right)\cos \frac{\pi}{2N}\right]$$
$$= \frac{1}{N^{3/2}}\left[\sin \frac{\pi}{N}\left(k - \frac{3}{4}\right)\sin^2 \frac{\pi}{4N}\right]$$
$$= \frac{e_k}{N} \sin^2 \frac{\pi}{4N}$$

for $2 \leq k \leq N$

$$(T\mathbf{e})_1 = \frac{1}{2N}\left[e_1 + \frac{e_{N+1}}{2}\right]$$
$$= \frac{1}{2N^{3/2}}\left[\sin \frac{\pi}{4N} - \frac{1}{2}\sin \frac{3\pi}{4N}\right]$$
$$= \frac{1}{N^{3/2}} \sin \frac{\pi}{4N} \sin^2 \frac{\pi}{4N} - \frac{1}{4N^{3/2}}\sin \frac{\pi}{4N}$$
$$= \frac{e_1}{N} \sin^2 \frac{\pi}{4N} - \frac{e_1}{4N}$$

and $(T\mathbf{e})_k = -(T\mathbf{e})_{2N-k+1}$ for $N+1 \leq k \leq 2N$ from the symmetry of T and $\mathbf{e}$. Hence, the Rayleigh quotient ν becomes

$$\nu = \mathbf{e}^T T \mathbf{e}$$
$$= 2\sum_{k=1}^{N} \frac{e_k^2}{N} \sin^2 \frac{\pi}{4N} - \frac{e_1^2}{2N}$$
$$= \frac{1}{N} \sin^2 \frac{\pi}{4N} - \frac{e_1^2}{2N}$$

and then, the difference between $T\mathbf{e}$ and $\nu\mathbf{e}$ is

$$\|T\mathbf{e} - \nu\mathbf{e}\| = \left\| \frac{1}{N} \sin^2 \frac{\pi}{4N}\mathbf{e} - \left(\frac{e_1}{4N}, \mathbf{0}_{2N-2}^T, -\frac{e_1}{4N}\right)^T \right.$$
$$\left. + \left(\frac{e_1^3}{4N}, \mathbf{0}_{2N-2}^T, -\frac{e_1^3}{4N}\right)^T \right\|$$
$$= O\left(\frac{1}{N^{5/2}}\right).$$

Since ν can be approximated to

$$\nu = \frac{1}{N} \sin^2 \frac{\pi}{4N} - \frac{e_1^2}{2N} \approx \frac{\pi^2}{16N^3}$$

the minimum eigenvalue of T in total is $O(1/N^{\max(5-k,k)})$ when the minimum of the others is $O(1/N^k)$ since ν has a precision with $\varepsilon^2/\delta = O(1/N^{5-k})$ from Theorem 2. Therefore, even in the fastest case, the convergence rate becomes $1 - O(1/N^{5/2})$, which is much slower than that of the original method $1 - 1/(2QN)$.

Although the above explanation only shows that the minimum eigenvalue of T is smaller than or equal to $O(1/N^{\max(5-k,k)})$, the actual minimum eigenvalue is well approximated by the Rayleigh quotient $\nu = \pi^2/(16N^3)$, as shown in Fig. 12, where the crosses represent the minimum eigenvalues, and the solid and the dashed lines show $\pi^2/(16N^3)$ and $1/N^{5/2}$, respectively.

D. Computer Simulations With White Signal

The figures in this section show the results of computer simulations on the convergence properties of the input-sliding method and its modified version.

In each of the experiments by the original input-sliding method, the initial value $\mathbf{g}(0)$ of the weight-error vector is randomly chosen from $N(0, I_{2N})$, and the experimental convergence rate is calculated as the average of $\|\mathbf{g}(t)\|/\|\mathbf{g}(t-1)\|$ from $t = 1001$ to $t = 2000$ over eight trials. The input signal $x(t)$ at time t is drawn from a normal distribution $N(0, 1)$.

Fig. 13 shows the convergence rate versus the order $2N$ of the adaptive filter. The crosses show the experimental convergence rates, and the solid lines represent the theoretical convergence
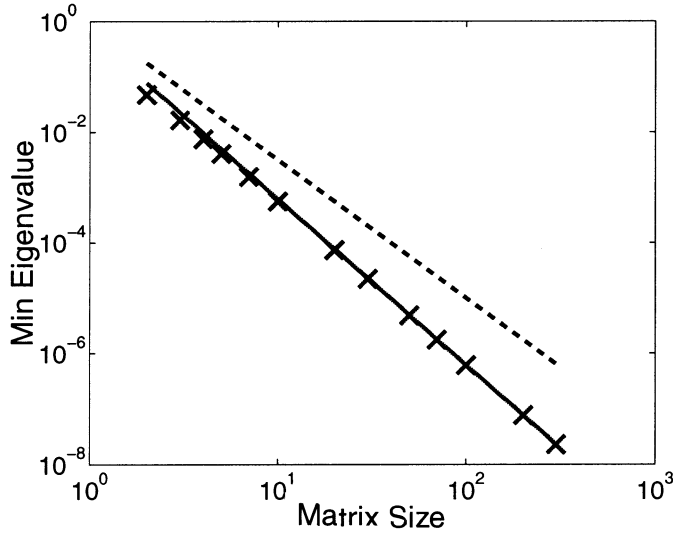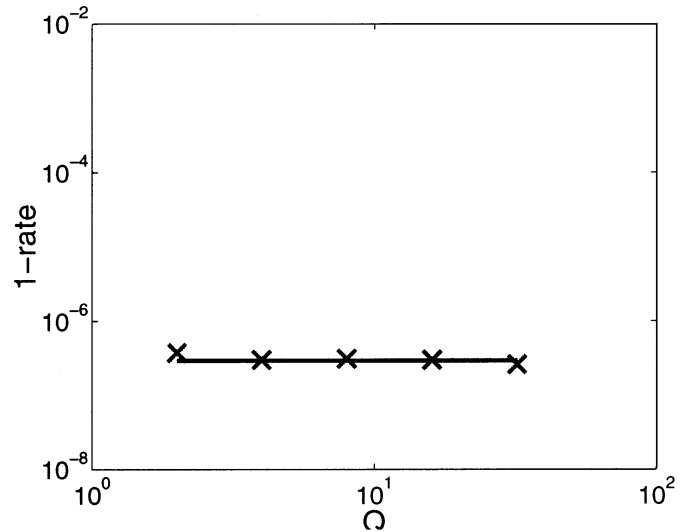
Fig. 12.   Minimum eigenvalue of T versus the size of the matrix.
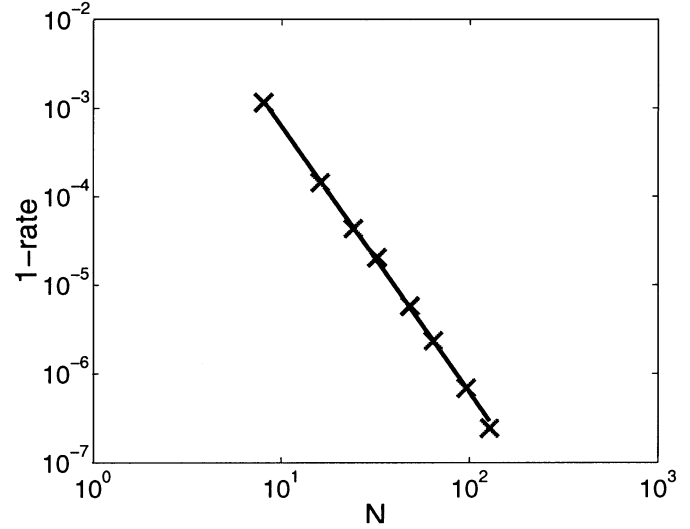
Fig. 14.   Convergence rate versus the order $2N$ of the adaptive filter where $Q = 4$.

Fig. 13.   Convergence rate versus the order $2N$ of the adaptive filter.
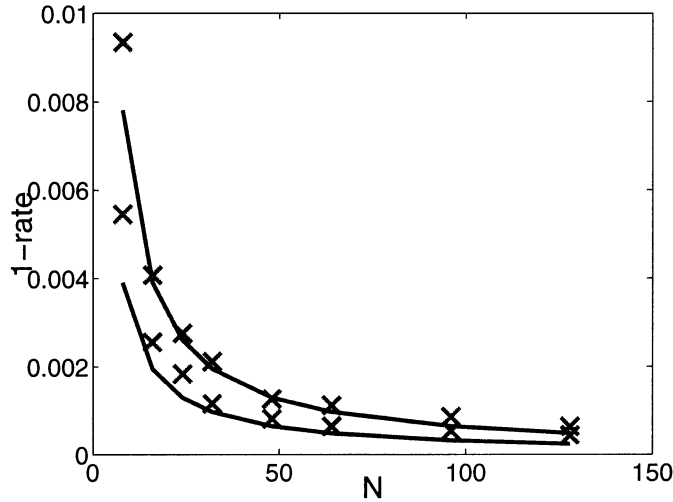
Fig. 15.   Convergence rate versus the period $2Q$ of input signals where $N = 128$.

rates given by (12). The upper part is the case of $Q = 4$ and the lower $Q = 8$. The theoretical values agree well with the experimental data.

In each of the experiments by the modified input-sliding method, the initial value $\boldsymbol{g}(0)$ of the weight-error vector is set to $\boldsymbol{e}$ in (13) and (14), instead of being randomly chosen, in order to reduce the computational complexity by removing the faster modes in advance. The experimental convergence rate is calculated as the average of $\|\boldsymbol{g}(t)\|/\|\boldsymbol{g}(t-1)\|$ from $t = 1001$ to $t = 2000$ over eight trials. The input signal $x(t)$ at time t is drawn from a normal distribution $N(0,1)$.

Fig. 14 shows the convergence rate versus the order $2N$ of the adaptive filter in the case of $Q = 4$. The crosses show the experimental convergence rate, and the solid line represents the theoretical convergence rates given by the Rayleigh quotient $\nu = \pi^2/(16N^3)$. The theoretical values agree well with the experimental data. Fig. 15 shows the convergence rate versus the period $2Q$ of input signals with $N = 128$. The experimental convergence rate does not depend on Q as the theory predicts.

E. Computer Simulations With Speech Signal

The theoretical analyses are based on the assumption that the input signal is white and stationary. In order to elucidate the limits of the analyses for real sound signals, we show the results of computer simulations with a speech signal.

In each of the experiments, the initial value $\boldsymbol{g}(0)$ of the weight-error vector is randomly chosen from $N(0, I_{2N})$, and the experimental convergence rate is calculated as the average of $\|\boldsymbol{g}(t)\|/\|\boldsymbol{g}(t-1)\|$ from $t = 1001$ to $t = 2000$ over eight trials. The input signal $x(t)$ is a real sound shown in Fig. 16, which is correlated, i.e.,

$$\frac{\sum\limits_{t=1001}^{2000} x(t-1)x(t)}{\sqrt{\sum\limits_{t=1001}^{2000} x(t-1)x(t-1)}\sqrt{\sum\limits_{t=1001}^{2000} x(t)x(t)}} = 0.8936.$$
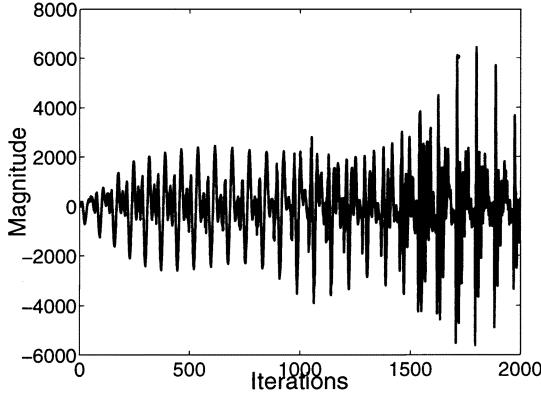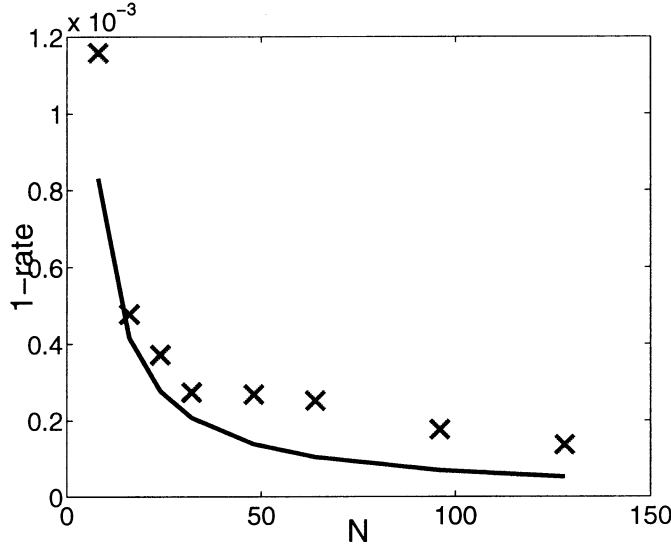
Fig. 16. Waveform of the input speech signal.

Fig. 17. Convergence rate versus the order $2N$ of the adaptive filter.

Hence, $\cos\theta$ in (11) should be substituted with

$$\cos\theta = \frac{1 - 0.1064}{(2Q)}$$

and then, the convergence rate is written as $1 - 0.1064/(4QN)$ instead of (12).

Fig. 17 shows the convergence rate versus the order $2N$ of the adaptive filter. The crosses show the experimental convergence rates, and the solid line represents the theoretical convergence rates given above in the case of $Q = 4$. The difference between the theoretical values and the experimental data is larger than that in Fig. 13. This may result from the time-variation of the input signal since it accelerates the convergence as the preprocessor does.

VI. CONCLUSIONS

The convergence properties of stereo acoustic echo cancelers with preprocessing have been examined. First, the condition of the paths from the talker to the microphones was elucidated under which the nonuniqueness problem exists. The degree of nonuniqueness evaluated depends on the primeness of the paths. Second, the convergence rates of the acoustic echo cancelers with the N-LMS algorithm were shown based on the independence theory. When the input signal vectors are not fully ranked

and time-variant, the convergence rate depends on both the period $2Q$ of the input signal and the order $2N$ of the adaptive filter, but if one of them is sufficiently larger than the other, the rate depends on either of them. Last, the input-sliding method was precisely analyzed. The method has a slow convergence rate since the preprocessor affects only the current input signal. A simple modification, which was expected to accelerate the convergence under the independence theory, actually makes the convergence much slower. The validity of these analyses was confirmed by computer simulations.

APPENDIX A
PROOF OF THEOREM 1

Consider the independency of the column vectors of $S^T = (D_a{}^T, D_b{}^T)$. Let $\mathbf{0}_k^T = (0, \ldots, 0) \in R^k$. Then, N column vectors of $D_a{}^T$

$$D_a^T = \begin{pmatrix} \boldsymbol{a} & 0 & \cdots & 0 \\ 0 & \boldsymbol{a} & & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \boldsymbol{a} \end{pmatrix} \in M(N+M, N)$$

are linearly independent since the upper-right components of the diagonal are null.

Next, we consider whether or not each of the column vectors of $D_b{}^T$ can be represented as a linear combination of the vectors located on its left side in $(D_a{}^T, D_b{}^T)$. If we assume that $(\boldsymbol{b}^T, \mathbf{0}_{N-1}^T)^T$ is a linear combination of the column vectors of $D_a{}^T$, that is, there exists a $\boldsymbol{\gamma}^T = (\gamma_0, \ldots, \gamma_{N-1})^T \in R^N$ s.t.

$$\begin{pmatrix} \boldsymbol{b} \\ \mathbf{0}_{N-1} \end{pmatrix} = D_a^T \boldsymbol{\gamma}$$

then

$$B(z) = A(z)\gamma(z)$$
$$\gamma(z) = \sum_{i=0}^{N-1} \gamma_i z^{-i}$$

holds true. This means that

$$A(z)P_1(z) + B(z)P_2(z)$$
$$= A(z)P_1(z) + A(z)\gamma(z)P_2(z)$$
$$= A(z)\left(P_1(z) + \gamma(z)P_2(z)\right).$$

Therefore, $A(z)P_1(z) + B(z)P_2(z)$ is a multiple of $A(z)$, and it never becomes 1. Therefore, if $A(z)P_1(z) + B(z)P_2(z) = 1$, $(\boldsymbol{b}^T, \mathbf{0}_{N-1}^T)^T$ is independent of the column vectors of $D_a{}^T$. With regard to $(0, \boldsymbol{b}^T, \mathbf{0}_{N-2}^T)^T$, if we assume $A(z)P_1(z) + B(z)P_2(z) = 1$ and that there exist δ_0 and $\boldsymbol{\gamma}^T = (\gamma_0, \ldots, \gamma_{N-1})^T$ s.t.

$$\begin{pmatrix} 0 \\ \boldsymbol{b} \\ \mathbf{0}_{N-2} \end{pmatrix} = D_a^T \boldsymbol{\gamma} - \delta_0 \begin{pmatrix} \boldsymbol{b} \\ \mathbf{0}_{N-1} \end{pmatrix}$$

it leads to

$$z^{-1}B(z) = A(z)\gamma(z) - \delta_0 B(z)$$
$$\delta_0 + z^{-1} = (\delta_0 + z^{-1})\left[A(z)P_1(z) + B(z)P_2(z)\right]$$
$$= A(z)\left(\delta_0 + z^{-1}\right)P_1(z) + B(z)\left(\delta_0 + z^{-1}\right)P_2(z)$$
$$= A(z)\left(\delta_0 + z^{-1}\right)P_1(z) + A(z)\gamma(z)P_2(z)$$
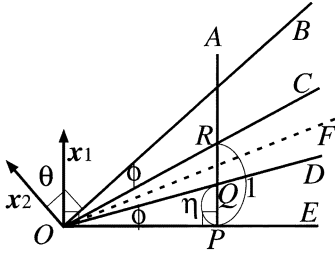$$= A(z)\left(\left(\delta_0 + z^{-1}\right)P_1(z) + \gamma(z)P_2(z)\right).$$

Fig. 18.   Angle of the weight-error vector and the input signal vectors.

Since the right-hand side is a multiple of $A(z)$ with the order of M, it never equals $\delta_0 + z^{-1}$ if $M > 1$, and hence, $(0, \boldsymbol{b}^T, \boldsymbol{0}_{N-2}^T)^T$ is independent of its leftward vectors in $(D_a{}^T, D_b{}^T)$. In the same way, each of $(\boldsymbol{0}_k^T, \boldsymbol{b}^T, \boldsymbol{0}_{N-1-k}^T)^T$ for $0 \le k \le M - 1$ is shown to be independent. This means that rk $S \ge N + M$. On the other hand, rk $S \le N + M$ necessarily holds true since $S \in M(2N, N+M)$, and then, rk $S = N+M$ has been proven if $A(z)P_1(z) + B(z)P_2(z) = 1$.

APPENDIX B
PROOF OF CONVERGENCE TO $\theta/2$

Fig. 18 represents the space of weight-error vectors. We claim that the weight-error vector oscillates on $\overline{OC}$ and $\overline{OD}$. Suppose that the current point A is not on $\overline{OC}$ and that the difference is $\delta\varphi$ in angle, that is, $\angle AOC = \delta\varphi$. Since A is moved to a point on $\overline{AP}$, the angle converges to $\theta/2 \pm (\theta/2 - \varphi)$ if the point has a smaller difference with $\overline{OD}$ in angle. In other words

$$\tan(\varphi + \delta\varphi) \gtrless \tan(\theta - \varphi + \delta\varphi)^* \eta$$

should be satisfied for $\delta\varphi \gtrless 0$.

Consider a function

$$f(\delta\varphi; \theta, \varphi) = \frac{\tan(\varphi + \delta\varphi)}{\tan(\theta - \varphi + \delta\varphi)}$$

of $\delta\varphi \in (\varphi - \theta/2, \pi/2 - \theta + \varphi)$ with parameters θ and φ. Since $f(0; \theta, \varphi) = \eta$, if $f(\delta\varphi; \theta, \varphi)$ is monotonically increasing around $\delta\varphi = 0$, the oscillation around $\theta/2$ with width $\theta/2 - \varphi$ is stable. Because

$$f'(\delta\varphi; \theta, \varphi) = \frac{1}{\tan^2(\theta - \varphi + \delta\varphi)} \left[\frac{\tan(\theta - \varphi + \delta\varphi)}{\cos^2(\varphi + \delta\varphi)} \right.$$
$$\left. - \frac{\tan(\varphi + \delta\varphi)}{\cos^2(\theta - \varphi + \delta\varphi)} \right]$$
$$= \frac{1}{2} \frac{\sin(2\theta - 2\varphi + 2\delta\varphi) - \sin(2\varphi + 2\delta\varphi)}{\cos^2(\varphi + \delta\varphi)\sin^2(\theta - \varphi + \delta\varphi)}$$
$$= \frac{\cos(\theta + 2\delta\varphi)\sin(\theta - \varphi)}{\cos^2(\varphi + \delta\varphi)\sin^2(\theta - \varphi + \delta\varphi)}$$

$f'(\delta\varphi; \theta, \varphi) > 0$ indeed holds true for

$$\varphi - \theta < \delta\varphi < \frac{\pi}{4} - \frac{\theta}{2}. \tag{15}$$

When A is located outside $\angle COD$ (higher than R), the point after the adaptation is located on $\overline{AQ}$, the point monotonically approaches the intersection with $\overline{OE}$, and A satisfies the condition; otherwise, A satisfies the condition of (15).

REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation—An overview of the fundamental problem," *IEEE Signal Processing Lett.*, vol. 2, pp. 148–151, Aug. 1995.

[2] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," *IEEE Trans. Signal Processing*, vol. 49, pp. 2577–2587, Nov. 2001.

[3] Y. Joncour and A. Sugiyama, "A stereo echo canceler with pre-processing for correct echo-path identification," in *Proc. ICASSP*, 1998, pp. 3677–3680.

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

[5] S. Haykin, *Adaptive Filter Theory*, 3rd ed.   Englewood Cliffs, NJ: Prentice-Hall, 1996.

[6] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," in *Proc. ICASSP*, 1995, pp. 3059–3062.

[7] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive filtering algorithms for stereophonic echo cancellation," in *Proc. ICASSP*, 1995, pp. 3099–3102.

[8] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 156–165, Mar. 1998.

[9] A. Hirano, K. Nakayama, and K. Watanabe, "Convergence analysis of stereophonic echo canceller with pre-processing—Relation between pre-processing and convergence," in *Proc. ICASSP*, 1999, pp. 861–864.

[10] ——, "Convergence analysis of a stereophonic acoustic echo canceller with pre-processing" (in Japanese), *Trans. IEICE*, vol. J84-A, pp. 565–575, 2001.

[11] M. Tanaka, A. Hirano, and K. Nakayama, "On uniqueness of filter coefficients in stereophonic acoustic echo canceller" (in Japanese), in *Proc. DSP Symp. IEICE*, 1999, pp. 31–36.

[12] A. Hirano and S. Koike, "Convergence analysis of a stereophonic acoustic echo canceller Part I: Convergence characteristics of tap weights," in *Proc. DSP Symp. IEICE*, 1996, pp. 569–574.

[13] K. Ikeda and H. Sakai, "Convergence properties of the block orthogonal projection algorithm," in *Proc. ICASSP*, 1999, pp. 1665–1668.

[14] K. Ikeda, "Convergence analysis of block orthogonal projection and affine projection algorithms," *Signal Process.*, vol. 82, pp. 491–496, 1997.

[15] H. Shin, *Numerical Analyzes*   Tokyo, Japan, 1982, vol. 13.

Kazushi Ikeda was born in Shizuoka, Japan, in 1966. He received B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1989, 1991, and 1994, respectively.

From 1994 to 1998, he was with the Department of Electrical and Computer Engineering, Kanazawa University, Kanazawa, Japan. Since 1998, he has been with the Department of Systems Science, Kyoto University, Kyoto, Japan. His research interests are focused on the fields of adaptive systems, including adaptive filters, and neural networks.

Ryohei Sakamoto was born in Yamaguchi, Japan, in 1977. He received the B.E. and M.E. degrees from Kyoto University, Kyoto, Japan, in 2000 and 2002, respectively.

Since 2002, he has been with Accenture Corporation, Tokyo, Japan.