PAPER

Frequency domain adaptive algorithm with nonlinear function of error-to-reference ratio for double-talk robust echo cancellation

Suehiro Shimauchi*, Yoichi Haneda and Akitoshi Kataoka

NTT Cyber Space Laboratories, NTT Corporation,
3–9–11, Midori-cho, Musashino, 180–8585 Japan

(*Received 15 January 2004, Accepted for publication 5 July 2004*)

Abstract: Several adaptive algorithms for robust echo cancellation use nonlinear reference and/or error functions. Most of them require time-variant threshold estimators, e.g., noise level estimators or double-talk detectors, since their nonlinearities have to be adjusted in response to changes in near-end noise or speech signal levels. We propose a new frequency domain adaptive algorithm: the gradient-limited fast least-mean-squares (GL-FLMS), in which the coefficients are updated by using a nonlinear function of the error scaled by the reference magnitude, i.e., the error-to-reference ratio (ERR). When the acoustic coupling level between loudspeaker and microphone is bounded, the ERR is also bounded in the case of single-talk, but may increase during double-talk. The GL-FLMS limits unexpected increases in the ERR with fixed thresholds and prevents divergence of the coefficients, while not neglecting updates to adjust when a large reference signal introduces a large error during single-talk.

Keywords: Echo canceller, Adaptive filter, Robust control, Double-talk

PACS number: 43.60.Lq, 43.60.Cg [DOI: 10.1250/ast.26.8]

1. INTRODUCTION

Adaptive filters have been successfully applied to acoustic echo cancellation. In the field, the adaptation must be robust against double-talk, i.e., the situation where the near- and far-end speakers are talking simultaneously. Several ways to fulfill this requirement have been proposed.

In the two echo path model [1] and duo-filter control system [2], an adaptive background filter is used with a fixed foreground filter. This approach works well, but the dual filter structure increases the computational cost by about 50% and necessitates control of the transfer of coefficients. This structure is also incapable of handling echo path changes during double-talk. A three-port echo canceller configuration for the network echo canceller has been proposed, which can make an adaptation during double-talk [3]. It, however, exploits the signals from the two-wire circuit in addition to the four-wire circuit signals. This is not suitable for the acoustic echo cancellation case.

Another solution is to use a double-talk detector (DTD) and freeze updating of the adaptive filter [4–7]. In this case, however, conventional adaptive algorithms, such as the normalized least-mean-squares (NLMS), cannot adapt to echo path changes that occur while double-talk is in progress.

Certain adaptive algorithms in which the coefficients are updated with nonlinear functions of the reference input and/or error signals [8,9] have robust properties. For example, error nonlinearities have been used to achieve robustness against double-talk [10–12]. In [10], the error signal limited by an infinite clipper is used for the adaptation. It is also advised in [10] that the adaptation should be frozen whenever the reference input level is less than a predetermined level above the microphone signal, which indicates double-talk. In the other algorithms of this type [11,12], a variable scale factor is introduced to the error nonlinearity. Although a DTD is still required to control the scale factor, the adaptation is not completely frozen during double-talk. Variable step-size control approaches [13,14] can also be regarded as nonlinear approaches. In [13], as a variant of the NLMS algorithm, the step-size is controlled by using a nonlinear function of the reference power level. A noise level estimator controls the nonlinearity threshold. The fuzzy step-size control in [14] can be understood as nonlinear usage of the reference and error signals. The fuzzy rule is based on a norm estimation of precursor coefficients added to the adaptive filter and a single-talk detection.

In this paper, we focus on frequency domain echo

*e-mail: shimauchi.suehiro@lab.ntt.co.jp cancellers which simply have a single-tap adaptive filter in each frequency bin. Frequency domain adaptive filters have two primary advantages compared to the time domain implementations [15]. One is lower computational complexity due to the efficiency of block processing in connection with the discrete Fourier transform (DFT). The other is fast convergence, especially for colored input signals such as speech, since individual control can be applied to the step-sizes for adaptation in each of the frequency bins according to the input signal spectrum. Our aim in this paper is to propose a new robust frequency domain algorithm that maintains these advantages. The robustness of the proposed algorithm is based on a nonlinear function that provides scaling of the reference and error signal levels. This algorithm achieves good performance without time-variant threshold estimators. Although only one [11] is addressed to the frequency domain among the above conventional algorithms [10–14], some of the others can also be made applicable. However, none of them inherently have signal level scalability, so their robustness is only achieved at the cost of peripheral estimators or detectors. The proposed adaptive filter configuration is based on overlap-save sectioning with the DFT [15]. There are two kinds of basic algorithms for this, i.e., gradient constrained [16] and unconstrained [17] versions. We mainly concentrate on the unconstrained version, since this is less computationally complex and can converge to the Wiener solution when certain conditions are fulfilled: e.g., the length of the unknown system is less than half of the DFT block size [15,17]. The configuration inherently causes a delay equivalent to half of the DFT block size. However, it can still be useful in applications which allow some delay for the response, such as echo cancellation for speech recognition [18].

This paper is organized as follows. In Chap. 2, some background information on the derivation of the algorithm for frequency domain echo cancellers are given. In Chap. 3, a new robust algorithm is presented and its interpretation and modification are discussed. The simulation results are demonstrated in Chap. 4. The conclusions are given in Chap. 5.

2. FREQUENCY DOMAIN ECHO CANCELLER

2.1. Configuration

The frequency domain echo canceller that we are discussing is shown in Fig. 1. The reference signal $x(n)$ at a discrete time index n is from the far-end and is picked up by the microphone after passing through the room echo path, which has an impulse response modeled as $\boldsymbol{h} = [h_1, \ldots, h_L]^{\mathrm{T}}$, where L is the effective length and T is the transpose. The total microphone signal $y(n)$ is expressed as

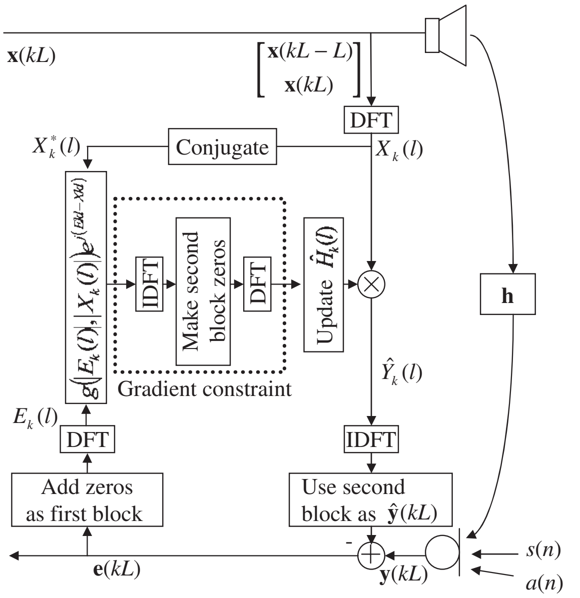


Fig. 1 Frequency domain echo canceller.

$$y(n) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x}(n) + s(n) + a(n), \tag{1}$$

where $\boldsymbol{x}(n) = [x(n-L+1), \ldots, x(n)]^{\mathrm{T}}$, $\boldsymbol{R}$ is the matrix that reverses the order of the elements of $\boldsymbol{x}(n)$, and $s(n)$ and $a(n)$ are, respectively, speech and ambient noise at the near-end. In many actual situations, the impulse response $\boldsymbol{h}$ can be assumed to be fixed or to vary slowly relative to the convergence rate of the adaptive filter. The transformed reference signal in l-th frequency bin at the k-th step, $X_k(l)$, is an element of the DFT of $[\boldsymbol{x}^{\mathrm{T}}(kL-L), \boldsymbol{x}^{\mathrm{T}}(kL)]^{\mathrm{T}}$, where the index k is incremented every time n increases by L, and $l = 0, \ldots, 2L-1$. The filter coefficient for k and l is $\hat{H}_k(l)$. The filter output is $\hat{Y}_k(l) = \hat{H}_k(l)X_k(l)$. The echo replica $\hat{\boldsymbol{y}}(kL)$ corresponds to the last L elements of the inverse DFT (IDFT) of $[\hat{Y}_k(0), \ldots, \hat{Y}_k(2L-1)]^{\mathrm{T}}$. The error is $\boldsymbol{e}(kL) = \boldsymbol{y}(kL) - \hat{\boldsymbol{y}}(kL)$, where $\boldsymbol{y}(kL) = [y(kL-L+1), \ldots, y(kL)]^{\mathrm{T}}$. The transformed error $E_k(l)$ is an element of the DFT of $[\boldsymbol{z}^{\mathrm{T}}, \boldsymbol{e}^{\mathrm{T}}(kL)]^{\mathrm{T}}$, where $\boldsymbol{z}$ is an $L \times 1$ zero vector. We now concentrate on the unconstrained case. We start by omitting the gradient constraint in Fig. 1 and formulate the equation for updating $\hat{H}_k(l)$ as follows:

$$\hat{H}_{k+1}(l) = \hat{H}_k(l) + \mu \cdot g(|E_k(l)|, |X_k(l)|)e^{j(\theta_{Ekl}-\theta_{Xkl})}, \tag{2}$$

where θ_{Ekl} and θ_{Xkl} denote phases of $E_k(l)$ and $X_k(l)$ respectively, $g(|E_k(l)|, |X_k(l)|)$ is an arbitrary function of $|E_k(l)|$ and $|X_k(l)|$, which we abbreviate as $g(\cdot)$ below, and μ is a step-size whose value depends on the full form of $g(\cdot)$.

In the unconstrained fast (or frequency domain) LMS (UFLMS) [17] case,

$$g(|E_k(l)|, |X_k(l)|) = |E_k(l)|\frac{|X_k(l)|}{P_k(l)}, \tag{3}$$

where $P_k(l)$ is the smoothed power of $X_k(l)$ as obtained by using a smoothing factor α:

$$P_k(l) = (1-\alpha)P_{k-1}(l) + \alpha|X_k(l)|^2. \tag{4}$$

2.2. Frequency Domain Sign-Sign Algorithm

Our purpose in this paper is to find a desirable $g(\cdot)$ that provides both strong double-talk stability and fast convergence. As an example of a robust algorithm, we review the frequency domain sign-sign algorithm (FSSA) [19], for which $g(\cdot)$ satisfies

$$g(|E_k(l)|, |X_k(l)|) = 1. \tag{5}$$

While the time domain sign-sign algorithm (SSA) is known to be computationally efficient, it converges very slowly [8,9]. The FSSA achieves faster convergence, especially for colored reference signals, since its adaptation is independent of the frequency characteristic of the reference signal's level: $|X_k(l)|$. Since the FSSA only requires estimation of the phase difference $\theta_{Ekl} - \theta_{Xkl}$, it is robust against noise in the error signal as long as the phase difference between the reference signal and the noise is random. However, the FSSA has a similar property to the SSA, in that its region of convergence is a ball around the true solution with a radius proportional to the step-size μ [20]. This limits the accuracy of convergence.

3. NEW ROBUST ALGORITHM

3.1. Scalable Nonlinearities

For robust adaptation, error estimation is important, i.e, to what extent are the near-end signals included in the error signal? The error-to-reference ratio (ERR) is useful as a means for estimating the influence of the near-end signals in the error. In the single-talk case for the far-end talker, the ERR does not exceed the acoustic coupling level (ACL) between the loudspeaker and the microphone, unless the adaptive filter diverges. In the double-talk case, the ERR has a wide distribution around the ratio between the averaged levels of the far- and near-end signals as the mean. The ERR distributions will be different in the single- and double-talk cases. It is thus reasonable to limit the ERR to the level expected during single-talk. Taking this into account, we define $g(\cdot)$ for a new robust version of the UFLMS as

$$g(|E_k(l)|, |X_k(l)|) = \psi_{S_1}\left(\frac{|E_k(l)|}{|X_k(l)|}\right)\frac{|X_k(l)|^2}{P_k(l)}, \tag{6}$$

where $\psi_b(a) = \min\{a, b\}$, which is known as the Huber function [21], and S_1 is the threshold of the limiter. S_1 can be determined from the ACL and the fact that the near- and far-end signals are balanced on average. We can choose a fixed S_1 to suit the specifications of the system in which the algorithm is to be implemented.

In similar previous works [11,12], a nonlinearity is applied to the error scaled by a delicately tuned scaling variable. In contrast to this, our approach in Eq. (6) uses the nonlinearity of the ERR, which provides scalability for the relationship between the error and reference signal levels.

3.2. Gradient-Limited FLMS Algorithm

An alternative to Eq. (6) is given below:

$$g(|E_k(l)|, |X_k(l)|) = \psi_{S_1}\left(|E_k(l)|\frac{|X_k(l)|}{P_k(l)}\right). \tag{7}$$

Equation (7) is a gradient-limited version of Eq. (3). We call the algorithm with Eq. (7) the gradient-limited FLMS (GL-FLMS). If $|E_k(l)| \cdot |X_k(l)|/P_k(l) > S_1$, this corresponds to the FSSA, which uses Eq. (5). The stability of the GL-FLMS is ensured by choosing μ from within the bounds of stability for the UFLMS. There is no essential difference between Eqs. (6) and (7): if $P_k(l) = |X_k(l)|^2 (\alpha = 1)$, Eq. (7) is identical to Eq. (6). So, in the discussion below, we take Eq. (7) as the simpler case.

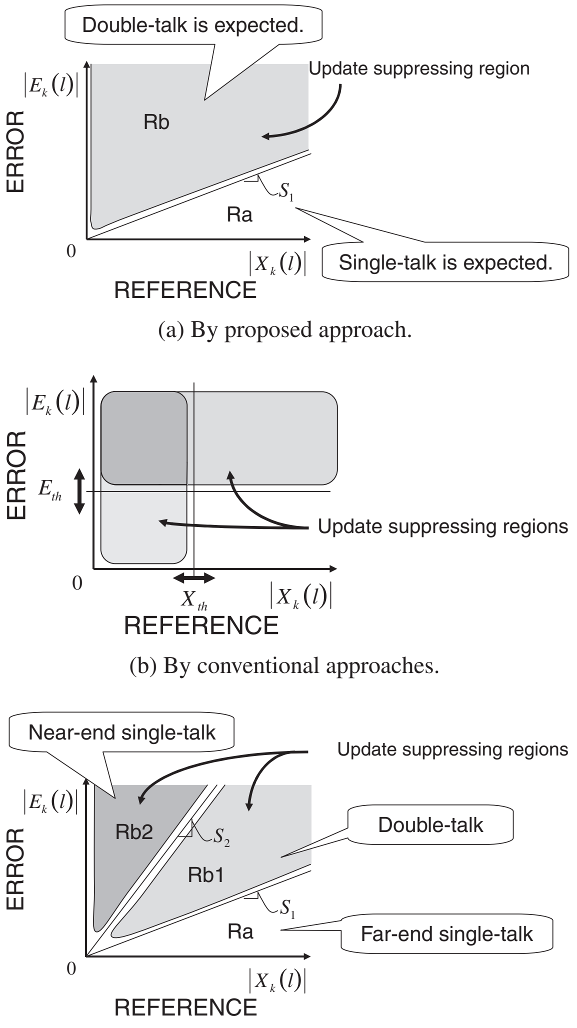
3.3. Interpretation and Improvement

The threshold S_1 of the GL-FLMS corresponds to the slope of the boundary between regions in the reference-error plane, which is shown in Fig. 2(a). The region Ra is for when far-end single-talk is expected and the filter is thus updated by the UFLMS. The region Rb is for when double-talk is expected and the filter is thus updated by the FSSA. Though echo path changes may be covered by the region Rb, the FSSA can still adapt to these without serious loss of convergence rate. For contrast, the form of segmentation used in several conventional approaches [11–13] is shown in Fig. 2(b). The regions are separated by time-variant boundaries that are perpendicular to the reference or error axes. Defining the double-talk region requires adequate and frequent boundary control.

If the maximum of the ACL is known or we can expect it to be bounded by a maximum level S_2, the region Rb in Fig. 2(a) can be separated into Rb1 and Rb2, as in Fig. 2(c). In the region Rb2, near-end single-talk is expected, while double-talk is expected in the region Rb1. Thus, by more strongly limiting the amount of updating in Rb2, the accuracy of convergence can be improved. An improved variant of the GL-FLMS is thus

$$\begin{aligned} & g(|E_k(l)|, |X_k(l)|) \\ & \quad = \psi_{S_1}\left(|E_k(l)|\frac{|X_k(l)|}{P_k(l)}\right)\psi_{\frac{1}{S_2}}\left(\frac{P_k(l)}{|E_k(l)||X_k(l)|}\right)S_2, \end{aligned} \tag{8}$$

$$= \begin{cases} |E_k(l)|\dfrac{|X_k(l)|}{P_k(l)} & \text{(in region Ra)}, \\ S_1 & \text{(in region Rb1)}, \\ S_1 S_2 \dfrac{P_k(n)}{|E_k(l)||X_k(l)|} & \text{(in region Rb2)}. \end{cases} \tag{9}$$



(a) By proposed approach.

(b) By conventional approaches.

(c) By improved proposed approach.

Fig. 2 Region segmentations.

Basically, this approach is similar to that in a much earlier method [10], where the adaptation is frozen when the ERR is above a predetermined level. In fact, however, it is difficult to strictly determine the ACL bound in the acoustic echo cancellation case, although this may be possible in the case of network echo cancellation. Thus, Eq. (9) softly suppresses adaptation in Rb2 by using a term that is inversely proportional to the ERR, so that some adaptation takes place, even in response to unexpected increases in the ACL.

From the above concept, a gradient constrained version of the GL-FLMS, which corresponds to a robust version of the FLMS [16], is also easily derived. In the next section, we give comparative results of simulation on the performance of this version of the GL-FLMS.

4. SIMULATION

In this section, we give results of simulation that demonstrate the effectiveness of the GL-FLMS. In particular, we compare its performance with the performance of conventional frequency domain algorithms, and examine dependence on the difference between the near- and far-end signal levels and on the ACL. We also demonstrate the good performance of the gradient constrained version of the GL-FLMS.

4.1. Comparisons with Conventional Algorithms

We compared the GL-FLMS based on Eq. (8) with the UFLMS based on Eq. (3) and with the two robust frequency domain algorithms described below.

4.1.1. Conventional method I

When we apply one conventional approach [13] to (3), we obtain

$$g(|E_k(l)|, |X_k(l)|) = |E_k(l)| \frac{P_k(l)|X_k(l)|}{P_k^2(l) + P_{\mathrm{th}}^2(l)}, \qquad (10)$$

where $P_{\mathrm{th}}(l)$ is a time-variant threshold based on the estimated noise level, which is updated if the error level is above the filter's output level.

4.1.2. Conventional method II

The other conventional approaches [11,12] correspond to

$$g(|E_k(l)|, |X_k(l)|) = \psi_{k_0(l)}\left(\frac{|E_k(l)|}{s(l)}\right) \frac{s(l)|X_k(l)|}{P_k(l)}, \qquad (11)$$

where $k_0(l)$ is a constant, and the scale factor $s(l)$ is controlled by a DTD.

4.1.3. Results

The conditions were as follows. The desired echo was made by using a 1024-tap FIR filter to form an echo path with the average ACL of 0 dB (Fig. 3). The sampling rate was 16 kHz. A smoothing factor α of 0.8 was used in all cases, after the choice in one pioneering work [17]. For the GL-FLMS, $S_1 = 0.5$ and $S_2 = 2$. For conventional methods I and II, the parameters were basically those used in the original papers, although some, such as those for the estimation of noise level, were rescaled to take the

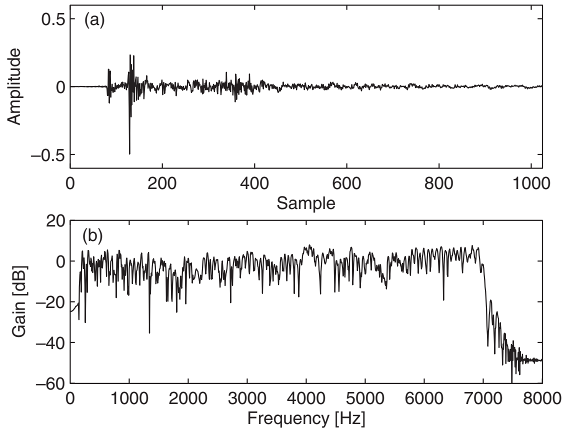


Fig. 3 (a) Impulse response and (b) frequency characteristic of the echo path used in simulation.

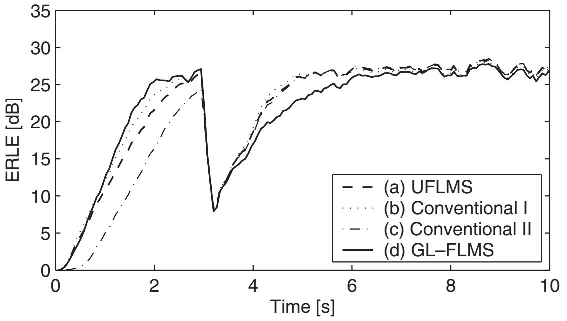


Fig. 4 ERLEs for a stationary signal input with averaged speech (single-talk with ENR = 20 dB). (a) UFLMS, (b) conventional I, (c) conventional II, and (d) GL-FLMS.

sampling rate, block size, or tap length into account. Choosing step-sizes for comparable performance of the algorithms is quite difficult. We chose them that led to the same steady-state echo return loss enhancement (ERLE) for a stationary signal input with averaged speech spectrum, where

$$\mathrm{ERLE}(n) = 10\log_{10}\frac{\sum\limits_{m=n-M}^{n}[y(m)-s(m)-a(m)]^2}{\sum\limits_{m=n-M}^{n}[e(m)-s(m)-a(m)]^2}\ [\mathrm{dB}], \tag{12}$$

and M is a large enough value for smoothing of the data (see Fig. 4). White Gaussian noise was added to the echo as ambient noise with an overall echo-to-noise ratio (ENR) of 20 dB. The echo path was changed at 3 seconds. We thus obtained $\mu = 0.2$ for the UFLMS algorithm, $\mu = 0.23$ for conventional method I, $\mu = 0.2$ for conventional method II, and $\mu = 0.32$ for the GL-FLMS algorithm.

The results for the speech signals were obtained in the following way. The far- and near-end signals (Fig. 5) were male and female speech, respectively. The average ENR was 20 dB. To ensure a minimum stability for all algorithms, the adaptation was frozen when the far-end signal level was below 500 on the 16-bit PCM scale, i.e., at least 12 dB lower than the average signal level. The echo path was changed at 3 seconds. Figure 6 shows the results for ERLE performance in the far-end single-talk case. We can see that conventional method I and GL-FLMS were more robust against ambient noise than UFLMS and conventional method II. Figure 7 shows the results for ERLE performance in the double-talk case. The GL-FLMS was the most robust. For conventional method II, we applied an 'ideal' DTD with advance detection of the near-end speech. The other methods did not require the DTD. Although other parameter choices might have led to better performance for the conventional methods, note that the GL-FLMS provided robust performance even though it was assigned fixed parameters, that is, parameters adjusted by neither the noise level estimator nor the DTD.

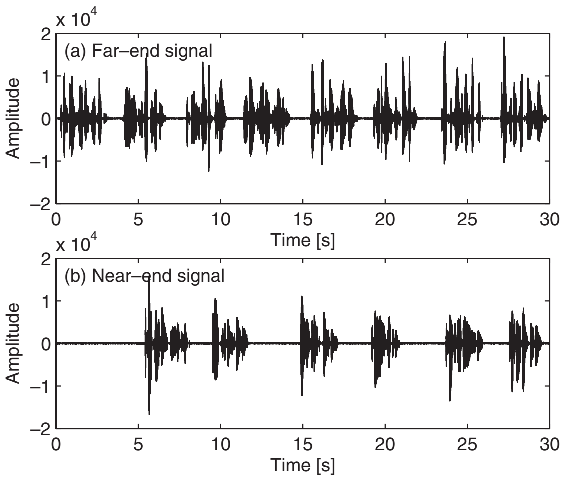


Fig. 5 Speech signals: (a) far-end and (b) near-end.

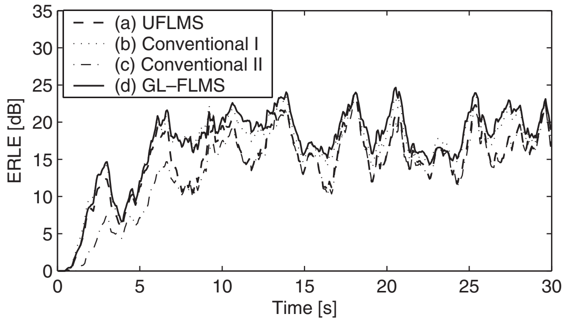


Fig. 6 ERLEs for the single-talk case with the four algorithms. (a) UFLMS, (b) conventional I, (c) conventional II, and (d) GL-FLMS.

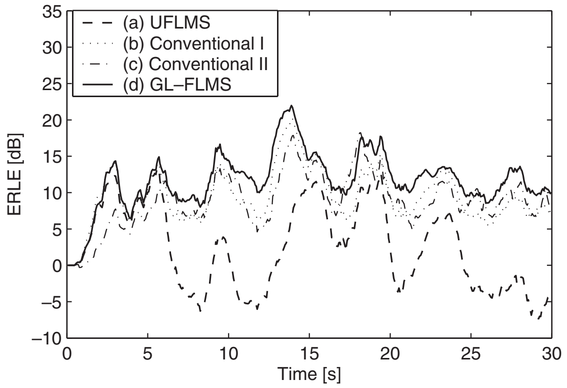


Fig. 7 ERLEs for the double-talk case with the four algorithms. (a) UFLMS, (b) conventional I, (c) conventional II, and (d) GL-FLMS.

The computational complexity of the GL-FLMS is similar to or rather less than that of the conventional method II, since the GL-FLMS does not require peripheral threshold estimators. In either case, there is no serious increase in complexity from the UFLMS.

4.2. Dependence of GL-FLMS Performance on Environmental Conditions

The dependence of GL-FLMS performance on the near- and far-end level balance and on the ACL were examined. The conditions were the same as in Sect. 4.1, except as specified below.

To evaluate the dependence on level balance, we tested three combinations of levels, obtained by scaling the signals in Fig. 5 as follows: (a) the far-end signal by +6 dB and the near-end signal by −6 dB, (b) the far-end signal by −6 dB and the near-end signal by +6 dB, and (c) the original far- and near-end signals. Since the ambient noise level was unchanged, the ERLE characteristics differed with the ENR. So the mean squared error (MSE), as calculated with the near-end speech excluded, was used to compare the absolute residual echo levels.

$$\mathrm{MSE}(n) = 10\log_{10}\sum_{m=n-M}^{n}[e(m) - s(m)]^2 \quad [\mathrm{dB}]. \tag{13}$$

Figure 8 shows results obtained in the absence of the near-end signal. This reflects the dependence on the reference input level during single-talk. The residual echoes converged on a similar steady-state level in all cases. Figure 9 shows results obtained in the double-talk case. Despite the 12-dB difference between levels at the two ends in cases (a) and (b), the results did not deviate greatly from those for case (c).

The dependence on ACL was evaluated by comparing the performance for (a) average ACL = +10 dB, (b) average ACL = −10 dB, and (c) average ACL = 0 dB. The far- and near-end signals of Fig. 5 were used again. Figure 10 shows the MSEs obtained for the single-talk case. The differences between Figs. 8 and 10, particularly those seen in the steady-state, were due to the fixed reference-level threshold for the freezing of adaptation for all of the simulations in which we used real speech signals.

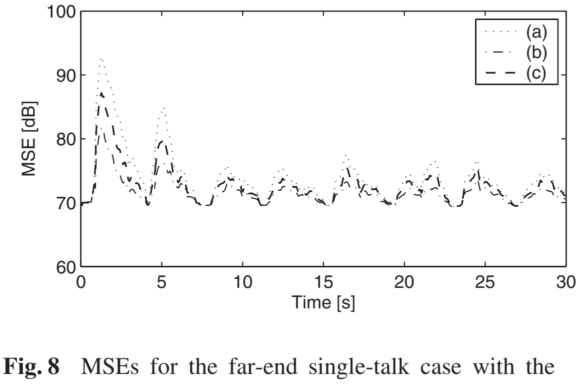


Fig. 8 MSEs for the far-end single-talk case with the signal level scaled by (a) +6 dB, (b) −6 dB, and (c) 0 dB.

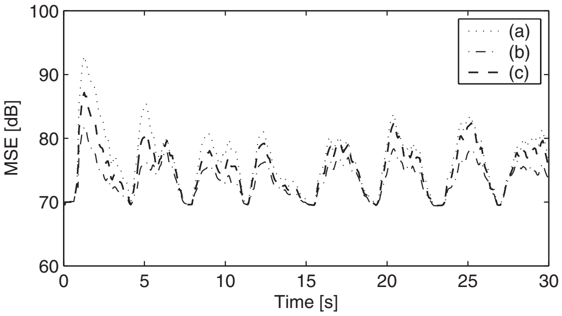


Fig. 9 MSEs for the double-talk case with different balances between the levels of the far- and near-end signals. (a) far-end: +6 dB, near-end: −6 dB, (b) far-end: −6 dB, near-end: +6 dB, and (c) far-end: 0 dB, near-end: 0 dB.

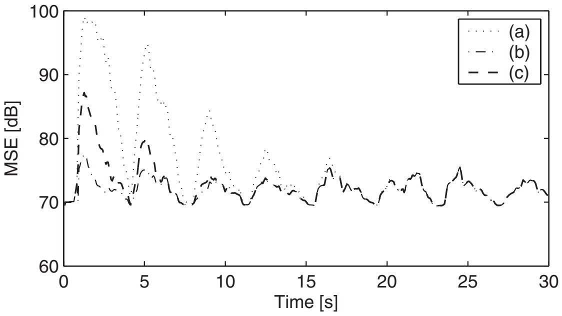


Fig. 10 MSEs for the single-talk case with different ACLs. (a) ACL: +10 dB, (b) ACL: −10 dB, and (c) ACL: 0 dB.

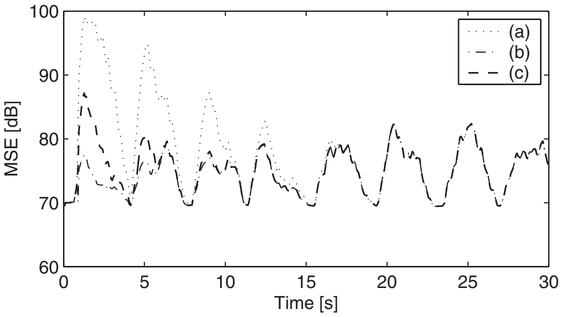


Fig. 11 MSEs for the double-talk case with different ACLs. (a) ACL: +10 dB, (b) ACL: −10 dB, and (c) ACL: 0 dB.

Figure 11 shows MSEs obtained in the double-talk case. Once the residual echo had converged on the steady-state level, the differences in ACL had no effect on the robustness against double-talk.

4.3. Performance of the Gradient Constrained Version

We also evaluated the gradient constrained version of the GL-FLMS. We have not given a detailed description of

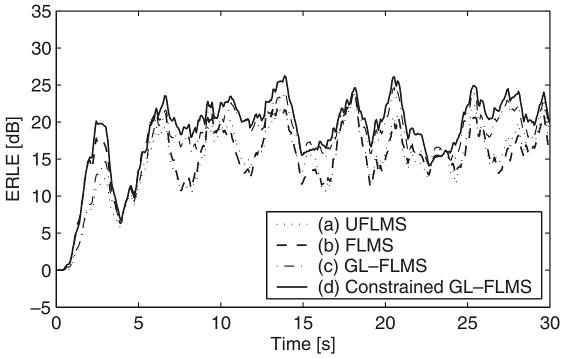


Fig. 12 ERLEs for the single-talk case with unconstrained and constrained algorithms. (a) UFLMS, (b) constrained FLMS, (c) unconstrained GL-FLMS, and (d) constrained GL-FLMS.

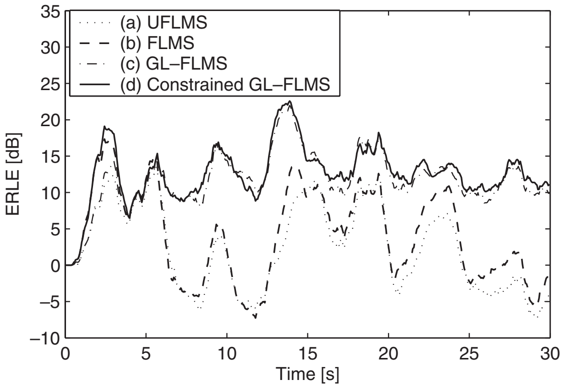


Fig. 13 ERLEs for the double-talk case with unconstrained and constrained algorithms. (a) UFLMS, (b) constrained FLMS, (c) unconstrained GL-FLMS, and (d) constrained GL-FLMS.

the constrained GL-FLMS, because it can be easily derived by following a gradient constraining scheme in the literature [15,16]. The following algorithms were compared: (a) the UFLMS, (b) the FLMS, which is really a constrained version of the UFLMS, (c) the unconstrained GL-FLMS, and (d) the constrained GL-FLMS. The step-sizes were adjusted, in the same manner as above: $\mu = 0.35$ for the FLMS and $\mu = 0.5$ for the constrained GL-FLMS. The other conditions were the same as those used in Sect. 4.1.

Figures 12 and 13 show the ERLEs in the single- and double-talk cases, respectively. Some improvement was obtained by using the constrained GL-FLMS, and the constraint had no negative effect.

5. CONCLUSION

We have proposed the gradient-limited FLMS (GL-FLMS) algorithm as a frequency domain algorithm that is robust against double-talk. In the GL-FLMS, the sizes of updates are nonlinearly controlled according to the ERR. Unlike some conventional robust algorithms, the GL-FLMS achieves its robustness with fixed thresholds predetermined on the basis of the fact that the far- and near-end signals will be balanced on average and of the bounds on the ACL that would be expected in actual situations. The algorithm's effectiveness and the reasonableness of the underlying concepts were confirmed through simulation. The GL-FLMS is not essentially incompatible with approaches in which variable threshold control is applied. Rather, it can be used as the basic algorithm in conjunction with peripheral threshold estimators. In other words, the GL-FLMS has potential for further performance improvement through the use of reasonable time-variant thresholds.

ACKNOWLEDGEMENT

We would like to thank Dr. H. Ohara 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.*, **COM-25**, 589–595 (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. EUSIPCO 96*, Vol. 2, pp. 1115–1118 (1996).

[3] J. Chao and S. Tsujii, "A new configuration for echo canceller adaptable during double talk periods," *IEEE Trans. Commun.*, **37**, 969–974 (1989).

[4] D. L. Duttweiler, "A twelve-channel digital echo canceller," *IEEE Trans. Commun.*, **COM-26**, 647–653 (1978).

[5] H. Ye and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Commun.*, **39**, 1542–1545 (1991).

[6] T. Gänsler, M. Hansson, C.-J. Ivarsson and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, **44**, 1421–1427 (1996).

[7] J. Benesty, D. R. Morgan and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio*, **8**, 168–172 (2000).

[8] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust. Speech Signal Process.*, **30**, 578–586 (1982).

[9] W. A. Sethares, "Adaptive algorithm with nonlinear data and error functions," *IEEE Trans. Signal Process.*, **40**, 2199–2206 (1992).

[10] M. M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, **XLVI**, 497–511 (1967).

[11] T. Gänsler, "A robust frequency-domain echo canceller," *Proc. ICASSP 97*, Vol. 3, pp. 2317–2320 (1997).

[12] T. Gänsler, S. L. Gay, M. M. Sondhi and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio*, **8**, 656–663 (2000).

[13] A. Hirano and A. Sugiyama, "A noise-robust stochastic gradient algorithm with an adaptive step size suitable for mobile hands-free telephones," *Proc. ICASSP 95*, Vol. 2, pp. 1392–1395 (1995).

[14] C. Breining, "A robust fuzzy logic-based step-gain control for adaptive filters in acoustic echo cancellation," *IEEE Trans.*

Speech Audio, **9**, 162–167 (2001).

[15] J. J. Shynk, "Frequency-domain and multi-rate adaptive filtering," *IEEE Signal Process. Mag.*, **9**, 14–37 (1992).

[16] E. R. Ferrara, "Fast implementation of LMS adaptive filters," *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-28**, 474–475 (1980).

[17] D. Mansour and A. H. Gray Jr., "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-30**, 726–734 (1982).

[18] H. Buchner and W. Kellermann, "An acoustic human-machine interface with multi-channel sound reproduction," *2001 IEEE Fourth Workshop on Multimedia Signal Processing*, pp. 359–364 (2001).

[19] S. Shimauchi, Y. Haneda and A. Kataoka, "Study on frequency domain echo canceller based on higher order statistics," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 615–616 (2003).

[20] A. Dasgupta, C. R. Johnson, Jr. and A. M. Baksho, "Sign-sign LMS convergence with independent stochastic inputs," *IEEE Trans. Inf. Theory*, **36**, 197–201 (1990).

[21] P. J. Huber, *Robust Statistics* (Wiley, New York, 1981).

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. He is a member of the ASJ, IEICE, 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 on 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 the ASJ, IEICE, 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 form 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 the ASJ, IEICE, and IEEE.