THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.25, NO.3E 2006. 9 pp. 123~130

A New Least Mean Square Algorithm Using a Running Average Process for Speech Enhancement

Soojeong Lee*, Chansik Ahn*, Jongmu Yun*, Soonhyob Kim*
*Department of Computer Engineering, Kwangwoon University
(Received September 19 2006; accepted September 22 2006)

Abstract

The adaptive echo canceller (AEC) has become an important component in speech communication systems, including mobile station. In these applications, the acoustic echo path has a long impulse response. We propose a running-average least mean square (RALMS) algorithm with a detection method for acoustic echo cancellation. Using colored input models, the result clearly shows that the RALMS detection algorithm has a convergence performance superior to the least mean square (LMS) detection algorithm alone. The computational complexity of the new RALMS algorithm is only slightly greater than that of the standard LMS detection algorithm but confers a major improvement in stability.

Keywords: *Adaptive echo cancellation, LMS, Moving average estimator, Speech enhancement*

I. Introduction

Recently, the use of speech recognition systems, voice-command systems, and especially hand-free cellular phones has increased rapidly. A problem arises when the microphone required in these applications picks up the sound from the loudspeaker. This feedback effect causes an echo, which can be obvious: the longer the delay, the greater the effect the echo has [1].

The most widely used filters for echo cancellation are the least mean square (LMS) adaptive finite impulse response (FIR) filters [2]. The LMS filter is important from a practical standpoint due to its simple implementation and the FIR filter performs effectively to model the unknown path. To improve convergence speed, the length of the filter taps should equal the impulse response of the unknown path. For systems with long impulse responses, however, the performance of adaptive echo canceller (AEC) decreases markedly for two reasons. First, when long delays are involved, which is usually the case in speech communication systems, the filter requires a large number of taps. Second, the unfavorable effect of long taps on convergence speed is emphasized when the input signal to the unknown path and estimator consists of highly correlated speech patterns [3, 4]. In the past few years, considerable research effort has been devoted to systems with long impulse responses having many inactive parts. Examples include room acoustic echo paths [5], mobile radio channels [6], and NLMS detection [4]. In this paper, we propose the running-average least mean square (RALMS) detection algorithm, which is a modification of the LMS detection algorithm [4] and compare it to the standard LMS algorithm and the RLS algorithm [7]. The RALMS filter can be described in terms of a running average of least squares. Particular attention is given to RALMS methods to increase the step size for more robust results using the tap detection algorithm examined by homer [4]. Our simulation result show that the performance of the RALMS method improves the convergence rate.

The paper is organized as follows. A formal description of the echo canceller is given in Section 2. Section3 reviews the fundamentals of adaptive filtering. The running average LMS (RALMS) with tap detection is detailed in Section 4. In Section 5, we compare the outcomes of the simulations and discuss their

Corresponding author: Soojeong Lee (leesoo86@kw.ac.kr)
Department of Computer Engineering Kwangwoon University

implications in detail. Finally, we summarize our conclusions in Section 6.

II. Problem Description

As long impulse responses are inherent in modern speech communication system, some form of echo reduction is required to ensure high quality of service. These echo reduction techniques are described in this section with emphasis on the echo canceller.

Acoustic echo paths are non-stationary. The movement of objects within the room, temperature, pressure, and humidity affect sound and how it is reflected [8]. This is modeled mathematically by using a filter with a time-variable impulse response.

The basic idea behind an echo canceller is illustrated in Fig. 1. An estimate $\hat{y}(n)$ of the echo $y(n)$ is generated and subtracted from the microphone output $v(n)$, and the difference between $v(n)$ and $\hat{y}(n)$, known as the error signal $e(n)$, is transmitted to the far end. When $\hat{y}(n)$ matches $y(n)$ exactly, $e(n)$ contains only the near-end signal. The echo path is generally in the digital domain, modeled as a linear time-varying filter with impulse response $h(n,l)$. Thus, the echo $y(n)$ can be written as

$$y(n)=\sum_{l=-\infty}^{\infty} h(n,l)u(n-l) \tag{1}$$

and the microphone output $v(n)$, consisting of the echo $y(n)$ and the near-end signal $noise(n)$ given by

$$\begin{aligned} v(n) &= y(n)+noise(n) \\ &= \sum_{l=0}^{+\infty} h(n,l)u(n-l)+noise(n) \end{aligned} \tag{2}$$

It is usually assumed that when $l \geq N$, then $h(n,l)=0$. This can be assumed since $h(l)$ is very small for large values of l since it has an exponentially decaying envelope.

This simplification yields

$$v(n)=\sum_{l=0}^{N-1} h(n,l)u(n-l)+noise(n) \tag{3}$$

The estimated echo $\hat{y}(n)$ is then obtained by filtering the far-end signal $u(n)$ with a linear time invariant filter $hv(n,l)$ that matches $h(n,l)$.

Thus, the output of the estimated filter is

$$\hat{y}(n)=\sum_{l=0}^{N-1} h(n,l)u(n-l) \tag{4}$$

The difference between the synthesized signal $\hat{y}(n)$ and the desired signal $v(n)$ is sent to the far end:

$$e(n)=v(n)-hv^{*}U \tag{5}$$

The controlling part of an echo canceller is the echo canceling filter $hv(n,l)$. The filter should be chosen such that the difference between $\hat{y}(n)$ and $v(n)$ is as small as possible by a certain means of measurement. A common criterion is

$$hv_{opt}=\arg[\min E|e(n)|^{2}] \tag{6}$$

where hv and hv_{opt} are N by 1 vector, where the l^{th} element is the l^{th} coefficient for the filter coefficients of the echo canceller filter $hv(n,l)$ and the optimal echo canceller filter $hv_{opt}(n,l)$, respectively. The echo canceling filter $hv(n,l)$ is generally made adaptive; that is the coefficients of the filter are updated recursively when new data are available. As the filter is adaptive, it can automatically adjust itself to different environments and track the changes in the environment.

III. Adaptive Filter Review

The main part of an echo canceller is the echo canceling filter. This filter should be such that the difference between the estimated echo and the microphone output is as small as possible in a statistical sense. A commonly used method to measure the size of the mean square error (MSE) involves the well-known Wiener filter. In order to allow the echo canceller to work in different environments and to track environmental changes, the echo canceling filter is exhibits adaptive qualities. This section presents the basics of adaptive filters, starting with the solution of the Wiener filter [7, 9].

3.1. Wiener solution

Referring back to Fig. 1. and assuming $noise(n)=0,$, then $hv(n)$ is ideally equal to $h(n)$ thus, it is desirable for the mean square error (MSE) $|\hat{e}(n)|^{2}$ to be at a minimum. So the cost function can be defined as

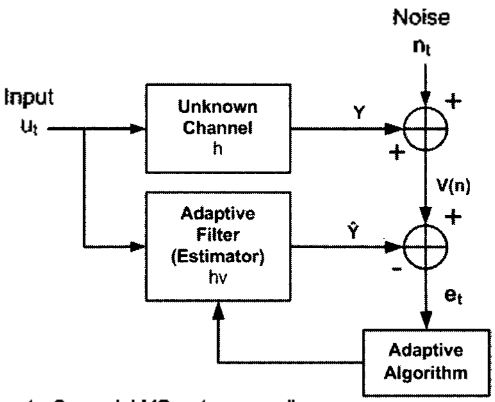

Fig. 1. General LMS echo canceller.

$$J(n) = E[|e(n)|^2] \tag{7}$$

$$e(n) = v(n) - \hat{y}(n) \tag{8}$$

$$\hat{y}(n) = hv^T * U \tag{9}$$

As it is known that $E[v^2(n)]$ is the mean squared power, $v(n)$ of σ_v^2 the expected value, $E[v(n)u(n)]$ is the cross-correlation vector between $v(n)$ and $u(n)$. These two are correlated since $v(n)$ is $u(n)$ after it has been through the filter. The cross-correlation vector is represented by P_N. The autocorrelation matrix is represented by R_{uu}, which is $E[u(n)u(n)^T]$.

The solution is simply

$$R_{uu} hv_{opt} = P_N \tag{10}$$

$$hv_{opt} = R_{uu}^{-1} P_N \tag{11}$$

which gives hv_{opt}, the optimum tap weight vector [9].

3.2. The least mean square (LMS) algorithm

From the steepest descent, the next updated parameters of the filter is

$$hv(n+1) = hv(n) + \mu'[P_N - R_{uu} hv(n)] \tag{12}$$

Expanding this gives

$$hv(n+1) = hv(n) + \mu' U(n) e(n) \tag{13}$$

The LMS algorithm is often called a stochastic gradient algorithm [7, 9] and is the most commonly used adaptive filtering algorithm because it is both very simple and works well. In particular, the LMS requires relatively little computational resources. Note that the parameter μ plays a very important role in the LMS algorithm. It can also be varied with time, but a constant μ is usually used, which is chosen after experimentation for a given application.

IV. Running-Average LMS (RALMS) with The Detection Algorithm

The FIR filter is a generalization of the concept of a running average filter. The running average method is commonly used whenever data fluctuate and must be smoothed prior to interpretation [10].

The standard detection LMS [4] has a better convergence rate than the LMS, but the detection LMS is quite similar to the LMS in that a problem occurs with fast convergence: the larger the convergence weight factor μ, the smaller the stability, and conversely, the smaller the convergence weight factor μ, the higher the stability. In this section, we will emphasize the convergence rate as well as stability using the new RALMS algorithm. The running average filter is a simple, linear, time-invariant system that is defined by an equation [10]. This system (14) is called an L-point running average of estimation error because the output at time n is computed as the average of $e[n]$ and the $L-1$ previous samples of the input.

$$\begin{aligned} RAVerr[n] &= \frac{1}{L}\sum_{K=0}^{L-1} e[n-k] \\ &= \frac{1}{L}(e[n] + e[n-1] + \cdots + e[n-L+1]) \end{aligned} \tag{14}$$

A new approach to control echoes within a speech communication system is to use echo cancellation, which involves an active detector and running average filter, as shown in Fig. 2.

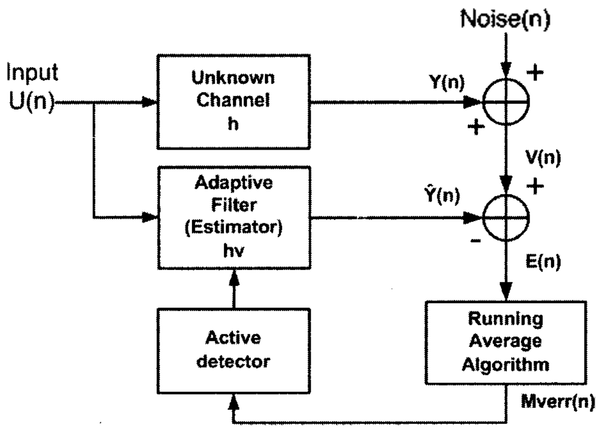

Fig. 2. RALMS echo canceller.

4.1. Detection method

The unknown channel is time-invariant and modeled by an n-delay tap $h=\Theta(z^{-1})$ that has only m < n nonzero taps: where z^{-1} is the unit delay operator.

$$\Theta(z^{-1}) = b_{t_1} z^{-t_1} + b_{t_2} z^{-t_2} + \cdots + b_{t_m} z^{-t_m} \tag{15}$$

where $0 \le t_1 \le t_2 < \cdots < t_m \le n-1$. Consider the unknown channel as described by Eq. (15), which we parameterize by the n-dimensional parameter vector:

$$\Theta = \left[\theta_{j_1-1}, b_{j_1}, \theta_{j_2-j_1-1}, b_{j_2}, \cdots, \theta_{j_m-j_m-1}, b_{j_m}, \theta_{n-1-j_m}\right]^T \tag{16}$$

where $n > j_m > j_{m-1} > \cdots > j_2 > j_1 > 0$, and θ_j is the zero matrix of size $1\times j$ b_{jm} is the non zero matrix. $|b_{ji}| > \sqrt{\mu\sigma^2_{noise} /(\hat{m}\sigma^2_u)}$, where $\hat{m}$ is the estimate of m, σ^2_u and σ^2_{noise} are the variances of $u(k)$ and $noise(k)$. Active parameter has a magnitude greater than the LMS adaptive noise level [4, 13]. Each of the remaining parameters is defined as an inactive parameter [14]. The goal of detection is to determine the positions of the m non zero elements of Θ. We use the following structurally consistent least-squares (SCLS) based cost function [14]:

$$J_{SCLS}(N) = J_{LS}(N) + m\sigma^2_v \log N \tag{17}$$

where $J_{LS}(N) = \sum_{k=1}^{N}[v(k) - hvU(k)^T]^2$; σ^2_v is the variance of $v(k)$; and m is the unknown number of active parameters.

$$\hat{J}_{SCLC} = \sum_{k=1}^{N} v^2(k) - \sum_{i=1}^{m} [X_{ji}(N) - \sigma^2_v \log N] \tag{18}$$

$$X_{ji}(N) = \frac{[\sum_{k=1}^{N} v(k)u_{ji}(k)]^2}{\sum_{k=1}^{N} u^2_{ji}(k)} \tag{19}$$

It is apparent that J_{SCLS} is minimized by those indices $J_i = J$ that satisfy

$$X_j(N) > T(N)$$

$$\text{where } T(N) = \sigma^2_v \log N \approx \frac{\log N}{N}\sum_{k=1}^{N} v^2(k) \tag{20}$$

Where $X_j(N)$ is known as the activity measure, $T(N)$ is the activity threshold.

4.2. RALMS algorithm with detection

This approach is summarized by the following algorithm [14].

1. Choose the LMS, forgetting the factor $\alpha \in [0,1)$. Set $d(k)=0, a_j(0)=0, b_j(0)=b_0>0, c_j(0)=0$, $\hat{\theta}_j(0)=0$, and $j=0,1,\cdots,n-1$.
2. Update $d(k), a_j(j\le k), b_j(j\le k), and\, c_j(j\le k)$ at time k via
   $b_j(k) = b_j(k-1) + v(k)u(k-j)$
   $a_j(k) = a_j(k-1) + u(k-j)^2$
   $c_j(k) = \dfrac{b_j^2(k)}{a_j(k)}$
   $d(k) = d(k-1) + v(k)^2$.
3. Determine the set of indices $\{b_{ji}\}$ that satisfy $x_j(k) > [d(k)\log k]/k$. Construct an $n\times 1$ vector $g(k)$, with ones in the positions corresponding to the set of indices $\{b_{ji}\}$ and zero in the remaining positions.
4. Update the new running average estimator $\hat{\theta}_j(k)$ at time k via where $hv = (\hat{\theta}_0(k), \hat{\theta}_1(k), \cdots, \hat{\theta}_{n-1}(k))^T$
   $n = 48$ $e(n) = v(n) - hv(k)^T * U$
   $$RAVerr[n] = \frac{1}{L}\sum_{K=0}^{L} e[n-k] = \frac{1}{L}(e[n] + e[n-1] + \cdots + e[n-L+1])$$
   $$\hat{\theta}_j(k+1) = \alpha^{1-g_j(k)}\hat{\theta}_j(k) + \mu * RAVerr * g_j(k)u(k-j),$$
   where $g_j(k)$ is the j th element of $g(k)$.
5. Return to step 2.

V. Simulation Result

In this section, we discuss the results of several simulations based on the LMS, RALMS, and RLS algorithms with nonzero tap detection and prewhitening methods. The unknown channel had $m=5$ active taps and a total length of $n=48$. The additional signal $noise$ used was a white zero mean Gaussian signal of unit variance.

5.1. Standard adaptive filter comparison

The filters in the experiments are shown in Table 1:the LMS, NLMS, and RLS. The MSEat 16000 samples is less with the RLS compared to the LMS and NLMS.

Table 1. Comparison of LMS, NLMS, and RLS with mean squared error (MSE).

	LMS	NLMS	RLS
MSE at white input	0.022	0.029–0.065	-
MSE at color input	0.020–0.029	0.043–0.070	0.005–0.006

5.2. Prewhitening methods with the tap detection result

When the input signal is colored, the correlation within the input signal causes coupling, and input signal prewhitening methods should be applied [4].

The asymptotic error of 16000 samples is substantially less with the zero tap detection algorithm compared to the LP method LMS without tap detection, which is shown in Figs. 3, as a reduction from 48 active taps detected to only 5 taps detected. The simulation results in Table 2 indicate that AR prewhitening provides a considerable improvement in asymptotic performance of the LMS active tap estimator [11].

Table 2. Comparison of LP method LMS with/without tap detection, SNR=0dB.

	LP method LMS	LP method LMS with tap detection
Active taps detected	48	5
Mean squared error	0.04–0.05	0.0005–0.005

This model corresponds to a first-order LP input. The unknown channel had $m=5$ active taps and a total length of $n=48$. Furthermore, the additional signal $noise(k)$ used was a zero mean white Gaussian signal with unit variance. The results of simulating the first-order LP input model and the same model with zero tap detection and LP prewhitening are shown in Fig. 3

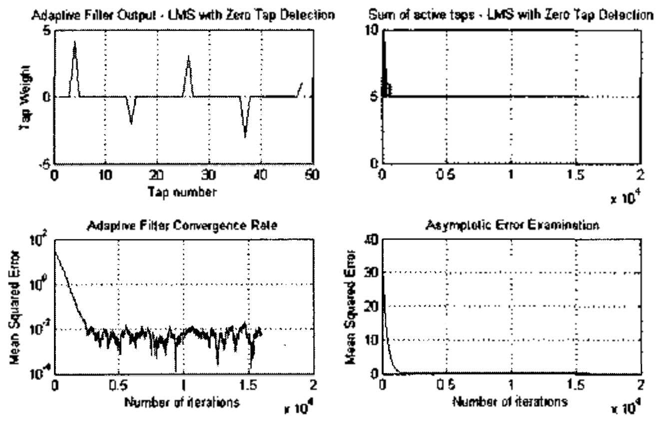

Fig. 3. Mean squared error and number of active taps detected by the LP method LMS, $\mu=0.002$.

5.3. RALMS with the tap detection result

We combined the running average LMS algorithm with the detection method. The LMS [4] with the tap detection is usually unstable, and has a large mean squared error at 16000 samples. However, the RALMS has a considerable ameliorating effect on both the stability and mean squared error, and this performance is better than that of the LMS with the detection method. The performance of the two models, the LMS and RALMS, both with tap detection, are shown in Table 3 below.

The results of simulations for the RALMS model are shown in Fig. 4.

Table 3. Comparison of the new RALMS method and the LP method LMS with tap detection.

	LP method LMS with tap detection				LP method RALMS with tap detection			
Convergence weight factor	μ				μ			
	0.001	0.01	0.05	0.1	0.001	0.01	0.05	0.1
Mean squared error	0.002	0.04	0.4	0.17	20	0.06	0.003	0.001

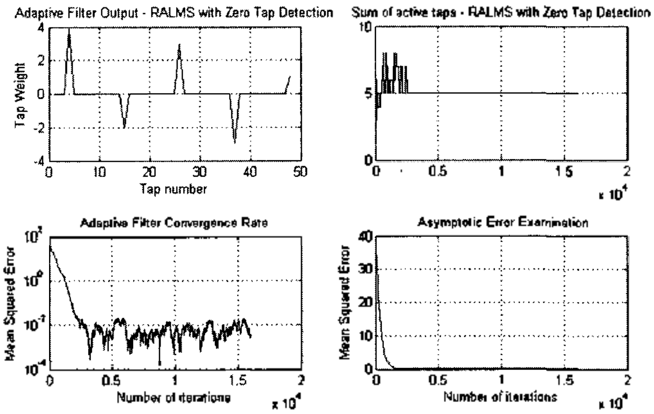

Fig. 4. Mean squared error and number of active taps detected by the RALMS $\mu=0.5$.

Table 4. Comparison of the MSE for RALMS with running average filter size.

	Newly RALMS with tap detection		
Running average point	Mean squared error (μ= 0.05)	Mean squared error (μ= 0.5)	Mean squared error (μ= 0.1)
24 point	0.003	0.04	0.005
48 point	0.001~0.002	0.04	0.001~0.003
64 point	0.002	0.05	0.0005~0.0008

The overall performance of the RALMS with running average filter size is shown in Table 4. The RALMS has a considerable ameliorating effect on the stability and mean squared error (MSE). The simulation results the LMS with detection, RLS, and RALMS schemes is shown in Fig. 5 below.

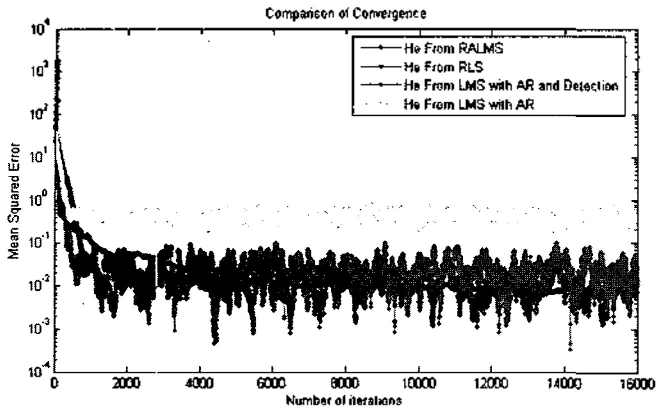

Fig. 5. Comparison of all simulation-mean squared error (MSE).

5.4. Speech signal simulations

Fig. 6 shows the desired signal, adaptive output signal, estimation error, and cost functions for the LMS algorithm with speech input, FIR filter order of 1000, and step size of 0.005. The MSE shows that as the algorithm progresses, the average value of the cost function decreases. This corresponds to the filter's impulse response converging on the actual impulse response, more accurately emulating the desired signal and thus more effectively canceling the echoed signal.

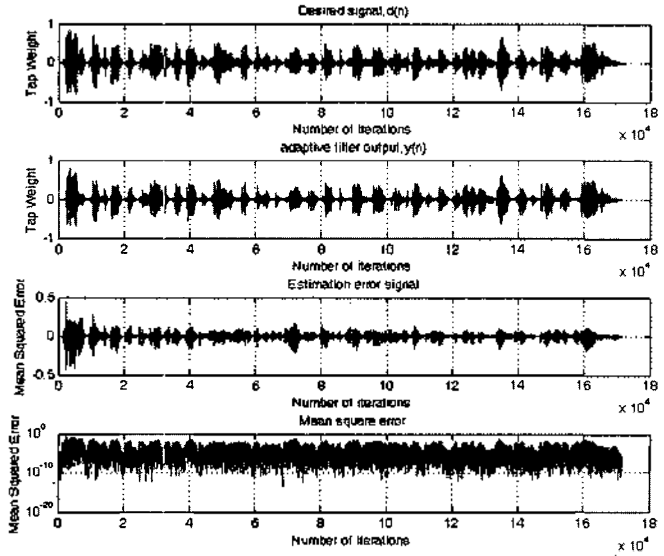

Fig. 6. Standard LMS algorithm for speech input, adaptive filter tap = 1000, step size = 0.005.

The success of the echo cancellation can be determined by the ratio of the desired signal and the error signal. Fig. 7 shows this attenuation as expressed in decibels. The average attenuation for this simulation of the LMS algorithm was − 12.9 dB.

The RALMS algorithm was simulated using Matlab. Fig. 8 shows the results of the RALMS adaptive echo cancellation simulation.

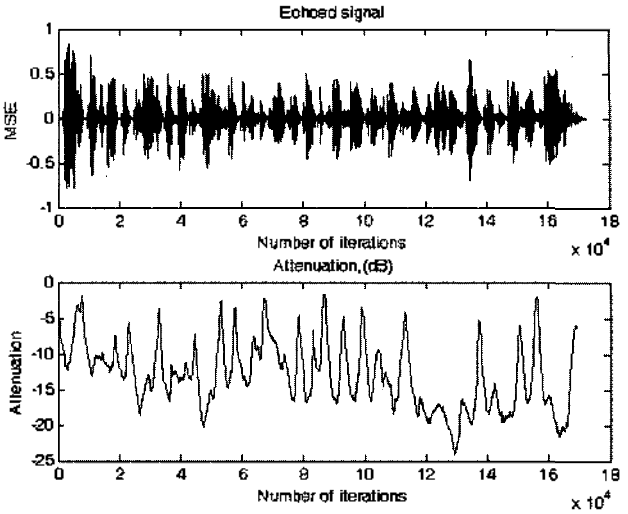

Fig. 7. Attenuation of the echoed signal by the LMS algorithm.

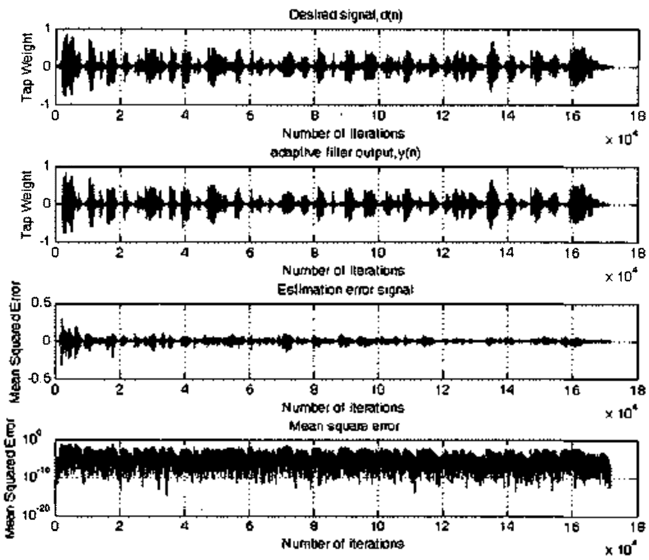

Fig. 8. proposed RALMS algorithm for speech input, adaptive filter tap =1000, step size=0.5.

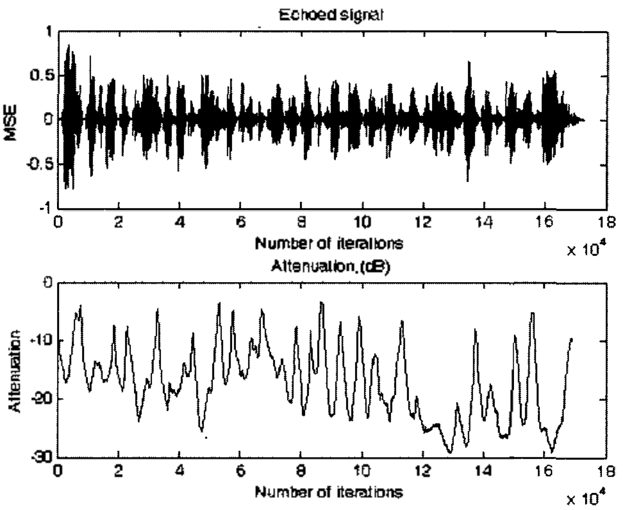

Fig. 9. RALMS attenuation of echoed signal (dB).

Fig. 9 shows the dB attenuation of the echoed signal by the RALMS adaptive filter, which had an average attenuation of -19.5 dB, an approximate improvement of 6 dB over the standard LMS algorithm.

5.5. Computational cost

The LMS active tap estimator requires $4n+2$ multiplications per sample interval, which is essentially twice that of the LMS estimator [4]. The computational cost of RALMS is as follows: 5n+2.

1) RALMS update of $hv(k)$ in Step 4 3n;
2) Active detection in Steps 2 and 3 2n+2;
   Item 2 can broken in the following manner:
   2.1) Update of $b_j(k), j=0:n-1$
   requires 2n multiplications.
   2.2) Update of $a_j(k), j=0:n-1$
   requires a total of one multiplication.
   2.3) $(\log k)/k$ can be obtained from a lookup table.
   There, $[d(k)\log k]/k$ requires one multiplication per sample interval.

VI. Conclusions

This paper considered the LMS estimation of channels that have impulse responses consisting of extended regions of active taps. The standard detection LMS algorithm has a better convergence rate than the LMS family algorithm, but since the standard detection LMS is quite similar to the LMS family algorithm, fast convergence is a problem: the larger the convergence weight factor μ, the lower the system stability. Our proposed RALMS detection scheme has a markedly improved convergence weight factor, as well as greater stability. Our future work is to test the robustness of the proposed algorithm against various experiments, such as real time environment. Furthermore, experiment based on the precisely analysis of proposed algorithm.

Acknowledgment

This study was supported by research fund Kwangwoon University, 2005

References

1. WebProForums.Echocancellationtutorial. http://webproforum.com/echo_cancel/index.htm, 2001.
2. J.Homer, *Adaptive echo cancellation in telecommunications*, (Ph.D. dissertation, Australian National University, Canberra, ACT, Australia, 1994)
3. M.M.Sondhi, "Silencing echoes on the telephone network" Proceedings of the IEEE, Berkley, D.A. , 68, pp. 948-963, 1980.
4. J.Homer, "Detection guided NLMS estimation of sparsely parameterized channels". IEEE Transactions on Circuits and Systems U: Analog and Signal Processing, 47, No.12, pp. 1437-1442, 2000.
5. J.R.Casar-Corredera and J.A.Alcazar-Fernandez, "An acoustic echo canceller for teleconference systems",In Proceedings of the IEEE 1986 International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Tokyo, Japan, 1317-1320, 1986.
6. "Digital land mobile radio communications," Office for Official Publication of the European Communities, Luxembourgh, Final Rep. COST 207, 1987.
7. S.Haykin, *Adaptive Filter Theory* (PrenticeHall, Upper Saddle River, New Jersey, 2002), Chap.5, 231-235.
8. S.L.Gay, *Acoustic Signal Processing for Telecommunication* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000)
9. B.Widrow and S.D.Stearns, *Adaptive Signal Processing*, (PrenticeHall, Upper Saddle River, New Jersey, 1985),Chap.6, 99-113.
10. J.H.McClellan, R.W.Schaferand M.A.Yoder, *Signal Processing First*, (PrenticeHall, Upper Saddle River, New Jersey, 2003), Chap.5,102-105.
11. V. Edward, *Signal prewhitening schemes for detection-guided LMS estimates*, (Department of Electrical and Computer Engineering) University of Queensland, Brisbane, Queensland, Australia, 1999.
12. H.Michael, *Acoustic echo cancellation digtal signal processing, Bachelor of Engineering thesis*, (The school of electrical engineering, The university of Queensland,2003)
13. J.Homer and I.Mareels, "LS detection guided NLMS estimation of sparse systems", Proceedings of the IEEE 2004 International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Montreal, Quebec, Canada, pp. 861-864, 2004.
14. J.Homer, I.Mareels, R.R. Bitmead, B.Wahlberg and F. Gustafsson, "Improved LMS estimation via structural detection", IEEE Transactions on Signal Processing, 46, No. 10, 26512663, 1998.

[Profile]

•Soojeong Lee

1997: B.S. degree in Computer science, Korea National Open University
2000: M.S. degree in Computer Engineering, Kwangwoon University
2004~present : Ph.D. course, Department of Computer Engineering, Kwangwoon University

•Chansik Ahn

2002: M.S.degree in Computer Engineering, Kwangwoon University
2002~present: Ph.D. course, Department of Computer Engineering, Kwangwoon University

•Jongmu Yun

2004: B.S.degree in Computer Engineering, Kwangwoon University
2004: M.S.cource in Computer Engineering, Kwangwoon University

•Soonhyob Kim

B.S. degree in electronics engineering from the Ulsan University, Korea, 1970.3~1974.2
M.S. degree in electronics engineering from the Yonsei University, Korea, 1974.3~1976.2
Ph.D. degree in electronics engineering from the Yonsei University, Korea, 1976.3~1983.2
1979. 3~present Professor, Dept. of Computer Engineering, Kwangwoon University
1986. 8~1987. 8 Visiting professor, Dept. of Electrical & Computer Eng. Univ. of Texas at Austin
1998. 1~2000. 12 President, the Acoustical Society of Korea
2000. 10~2004.10 Chairman of the invitation committee ICSLP 2004
2001. 1~present President Emeritus, the Acoustical Society of Korea