DIGITAL SIGNAL PROCESSING 3, 54–63 (1993)

Multiple-Microphone Acoustic Echo Cancellation System with the Partial Adaptive Process

Sen M. Kuo* and Jier Chen

Department of Electrical Engineering, Northern Illinois University, DeKalb, Illinois 60115

1. SUMMARY

The acoustic echo canceler using an adaptive transversal filter and the least-mean square (LMS) algorithm is the most effective technique to reduce acoustic echoes in a hands-free telephone system. However, the requirement of a very high order filter for each microphone echo path results in difficulties in convergence speed and hardware implementation. In this paper, a partial adaptive process (PAP) which utilizes the exponential decay characteristics of the acoustic echo path impulse response is developed. Simulations are conducted to show that the performance of the PAP method could be acceptable in practical use. A combination FIR and IIR filter structure and a new time-weighted LMS (TW-LMS) algorithm to complement the PAP are also proposed. This new adaptation algorithm and filter structure further improve the performance of the PAP method in terms of echo reduction and computational savings.

2. INTRODUCTION

The basic structure of a conventional acoustic echo canceller (AEC), which consists of a loudspeaker, microphone, and an adaptive filter, is illustrated in Fig. 1. This is a classical system identification problem where the adaptive filter, $W_n(z)$, adjusts its coefficients via the well-known least-mean square (LMS) algorithm to model the echo path, $H(z)$, between the loudspeaker and the microphone so that the system output, $e(n)$, which contains the residual echo, is minimized. However, since the reverberation of a conference room causes a long acoustic echo tail, an adaptive transversal filter with high order is needed to cover this type of echo path. As discussed in Ref. [1], if a transversal filter with the LMS algorithm is used, we have

$$0 < \mu < \frac{1}{N\sigma_x^2} \tag{1}$$

and

$$\tau_{\text{mse}} \approx \frac{1}{\mu\lambda_{\text{min}}}, \tag{2}$$

where μ is the convergence factor, N is the order of the filter, σ_x^2 is the variance of the input signal to the filter, τ_{mse} is the convergence time constant, and λ_{min} is the minimum eigenvalue of the autocorrelation matrix of the input signal. Equation (1) shows that if a large N has to be incorporated, as in the traditional AEC case, a small μ should be used. Unfortunately, this results in slow convergence, as shown in Eq. (2). As a result, a long adaptation time is generally needed, therefore the filter is unable to track the transient behavior of $H(z)$.

Moreover, if fixed point arithmetic is used and the assumption is made that the same word length B is used for both data and coefficients and μ is sufficiently small, the total output mean square error is [2]

$$\varepsilon = \varepsilon_{\text{min}} + \frac{1}{2}\mu\varepsilon_{\text{min}}N\sigma_x^2 + \frac{N\sigma_e^2}{2\mu} + (|\mathbf{w}^*|^2 + 1)\sigma_e^2, \tag{3}$$

where ε_{min} is the minimum mean square error (MSE) of the optimal (Wiener) filter $\mathbf{w}^*$, and σ_e^2 is the variance of the coefficient quantization error. The second

* To whom correspondence should be addressed.

1051-2004/93 $4.00

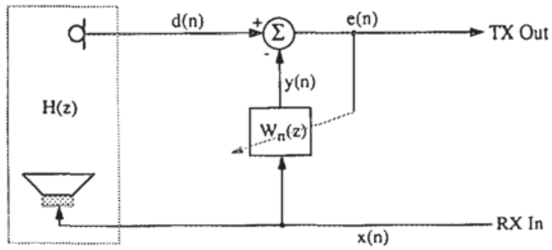

FIG. 1. System diagram of acoustic echo canceller.

term of Eq. (3) shows that excess mean square error is increased when a large N is used, and the third term shows the numerical error (due to coefficient quantization and roundoff) is increased with a large N at a small μ. Furthermore, in a fixed-point processor, the problem of roundoff causes early termination of the adaptation of the coefficients when a small μ is used [3]. In order to alleviate these problems, a higher dynamic range can be achieved by using floating-point arithmetic, with the added cost of a more expensive hardware implementation [4].

A multiple-microphone acoustical echo cancellation system is shown in Fig. 2. In this figure, the near-end conference room consists of a loudspeaker and an array of microphones. Because it is a full duplex system, the transmit signal (sent out) contains the desired speech from the near-end talker, ambient noise, and the acoustic echo. The acoustic echo canceller is employed in the system to remove the unwanted echo. Since more than one microphone is used in the conference room, we encounter a problem of multiple echo paths from the loudspeaker to each microphone.

Three techniques are developed to accommodate multiple microphones into the acoustic echo cancellation system. The first technique simply mixes the output of all microphones [5]. This method has its major drawback of reducing signal-to-noise ratio (SNR) by adding up the ambient room noise, and the speech signal is also distorted by the addition of reverberation. Reduction of the SNR in dB is approximated as $10 \log_{10} L$ [6], where L is the number of microphones used.

The second method uses a channel switching technique, as shown in Fig. 2. Assuming that each near-end person is assigned a microphone and only one person talks at a time, the channel switching unit actuates one of the L microphones corresponding to this near-end talker. In the receiving mode, this microphone is still actuated (since it is more likely for this person to talk again) and the AEC is updated to track this echo path and reduce the acoustic echo. Channel switching occurs only when the near-end talker changes. Because only one channel is updated for this multiple channel system, performance suffers when the microphone switches and the recovery is delayed until the convergence of the filter is achieved.

The third method uses a memory bank to assist the channel switching method as shown in Fig. 3 [7,8]. The transfer functions of the acoustic echo paths are estimated for each microphone off-line, and the optimum echo canceller filter coefficients are stored in memory. In receive mode, only one of the L microphones is actuated (called the ON-MIC). The corresponding filter coefficients are loaded into acoustic echo canceller and AEC is updated to perform echo cancellation task. When channel switching occurs in the subsequent receiving mode, these updated filter coefficients are restored into the memory and the filter coefficients corresponding to the current ON-MIC will be loaded into AEC and updated.

However, this method still has difficulties in real-time implementation due to the following two reasons. First, it requires a large memory to store all the filter coefficients. Since the impulse response of acoustic echo path has very long tails, the order of the echo canceller has to be large for a desired amount of echo cancellation. For example, for a 40-dB echo can-

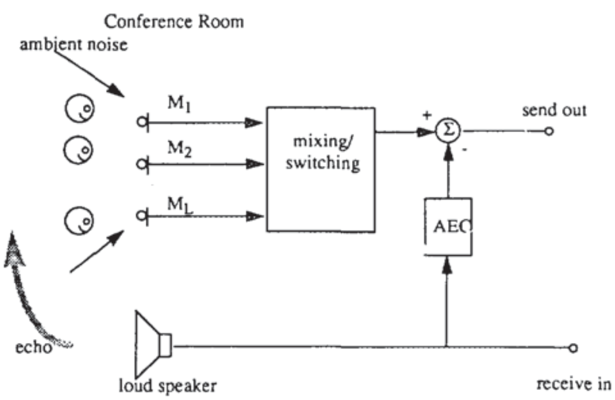

FIG. 2. Block diagram of multimicrophone AEC system.

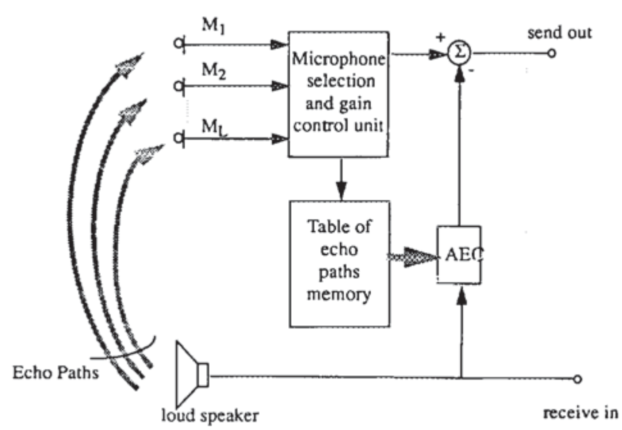

FIG. 3. Switching technique.

cellation in a medium sized room with 0.5 sec. reverberation time, an FIR filter of order over 3000 is required. Assuming that eight microphones are used in a conference room, the total memory required to store all the filter coefficients is 24,000 words (each word may be 16-bit or 32-bit if floating-point arithmetic is used). Second, this method can only work in the static environment where the transfer functions from the loudspeaker to every microphones are assumed to be time-invariant. This assumption is not true in general, since the room environment changes from time to time. Hence the stored filter coefficients are not always optimal for echo cancellation. As a result, this technique still suffers from the abrupt decreasing in performance at the instance of channel switching as well as large memory requirement.

3. THE PARTIAL ADAPTIVE PROCESS

In this section, a technique called the partial adaptive process is developed. This process is proposed to solve the difficulties (memory requirement and time-varying environment criteria) encountered in the switching technique. The partial adaptive process is shown in Fig. 4. In this figure, the far-end speech $x(n)$ from the loudspeaker is picked up by an array of L microphones. Each microphone has its own echo path transfer function, and the impulse response of the lth microphone is $h_l(n)$. The acoustic echo picked up by the lth microphone, $d_l(n)$, is the linear convolution of $x(n)$ and $h_l(n)$.

Suppose an Nth order adaptive filter with coefficients $w_{li}(n)$ and output $y_l(n)$ is used to cancel the lth channel echo $d_l(n)$; the error at the lth output is given by

$$e_l(n) = d_l(n) - y_l(n) = d_l(n) - \sum_{i=0}^{N-1} w_{li}(n)x(n-i), \quad (4a)$$

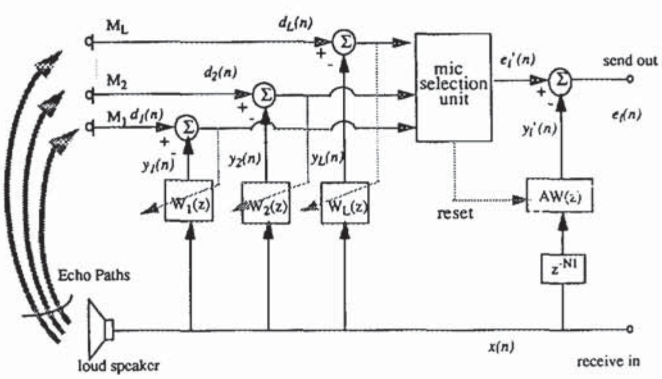

FIG. 4. Block diagram of partial adaptive process.

where l is from 1 to L. In order to reduce the complexity of this echo canceller system, we can break each filter into two sections. Therefore Eq. (4a) becomes

$$e_l(n) = d_l(n) - \sum_{i=0}^{N1-1} w_{li}(n)x(n-i) - \sum_{i=N1}^{N-1} w_{li}(n)x(n-i). \quad (4b)$$

Equation (4b) is identical to (4a) except that the adaptive filter $w_{li}(n)$ now is considered as two filters each with order $N1$ ($<N$) and $N2$ ($=N-N1$), respectively. The first section is a transversal filter that spans the first few hundred taps of the echo impulse response, and the second section approximates the remainder or tail. Based on Fig. 4, Eq. (4b) can be modified as

$$e_l(n) = e_l'(n) - y_l'(n), \quad (5a)$$

where

$$e_l'(n) = d_l(n) - \sum_{i=0}^{N1-1} w_{li}(n)x(n-i), \quad (5b)$$

and

$$y_l'(n) = \sum_{i=N1}^{N-1} w_{li}(n)x(n-i), \quad (5c)$$

where $y_l'(n)$ is the output of the tail filter.

Since the microphone selection unit actuates one microphone at a time, and most of the echo energy is concentrated in the early portion of the impulse response, much of the computational power is wasted in calculating those insignificant coefficients of the tail filters. A different approach which uses a single tail filter or called the auxiliary filter ($AW(z)$) with output $y_l'(n)$ to concatenate with the first stage filters $W_l(z)$, $l = 1, 2, \ldots, L$. In other words, this auxiliary adaptive filter is appended to one of the L small filters corresponding to the actuated microphone, thus forming a single large adaptive filter, as shown in Eq. (4b). This large adaptive filter is fully updated to perform the echo cancellation, while the other $L-1$ small adaptive filters are also updated to keep track of the significant changes in the acoustic echo path of these $L-1$ inactive channels. When channel switching occurs, the microphone selection unit resets all the coefficients of $AW(z)$ to zero and starts to update $AW(z)$ again.

This technique will have certain advantages over the switching method proposed in Ref. [7]. Assuming eight microphones are used, and the order of each

$W_l(z)$ is 500 and the order of the auxiliary filter, $AW(z)$, is 2500. This technique uses less than 7000 words of memory compared with 24,000 words required in [7]. Furthermore, since the filter coefficients for those inactive channels are updated, a dynamic system which has a good performance in real environment can be achieved.

Simulation was performed in a four-microphone system and the result is shown in Fig. 5. In this figure, the symbol "♦" indicates the occurrence of channel switching. In this simulation, the echo path transfer function is assumed to be time-invariant and white noise is used as the input signal, $x(n)$, instead of real speech. During each receiving period, one channel was active, hence, the auxiliary filter $AW(z)$ is combined with the corresponding $w_l(n)$ to produce a large adaptive FIR filter with order $N = 3000$. This large filter is used for echo cancellation, while the other $L - 1$ adaptive filters with orders $N1 = 500$ are updated for tracking purposes. Figure 5 shows that there is about 10 dB of echo return loss enhancement (ERLE) at the incident of channel switching, because the 500 taps of the L filters were already in optimal condition. A much higher ERLE improvement can be achieved if the echo paths have faster changes.

4. TIME-WEIGHTED LMS ALGORITHM

As shown in Section 3, there is at least 10 dB of ERLE achievement with the filter order at 500. A filter order of 1000 is required to achieve 18 dB of ERLE. Thus, the extra ERLE will cost more computation. Besides this heavy computational burden, the slow convergence of a high order adaptive filter is also problematic. Several algorithms were developed to improve the convergence speed by using the tap selection method [9,10] and by using an exponential step

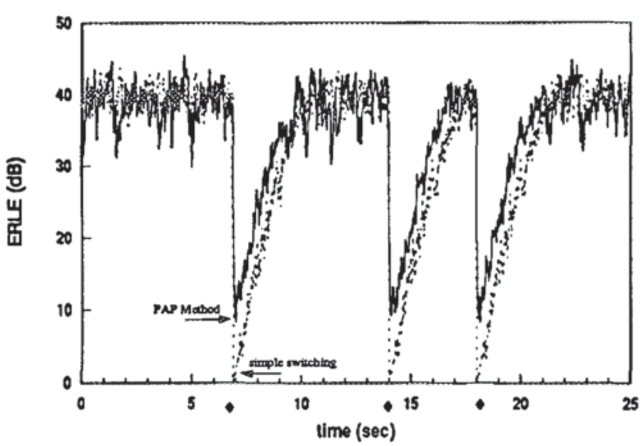

FIG. 5. Echo return loss enhancement for MMAEC.

TABLE 1

Comparison of Mathematical Operation for Adaptive Algorithms

		Convolution	Adaptation	Total
Standard LMS (order $N/2$)	Multiplication	$N/2$	$(N/2) + 1$	$N + 1$
	Addition	$(N/2) - 1$	$N/2$	$N - 1$
TW-LMS (order N)	Multiplication	N	$(N/4) + 1$	$(5N/4) + 1$
	Addition	$N - 1$	$N/4$	$(5N/4) - 1$

size algorithm [11]. In this paper, based on the exponential decay characteristics of acoustic echo impulse response, an alternative approach called the time-weighted LMS algorithm is developed. This algorithm is effective both in convergence speed and computational savings.

Considering an adaptive filter of order 1000 (corresponding to ERLE of about 18 dB), we can split these 1000 filter coefficients into M sections such that each section has length $N1, N2, N3, \ldots, NM$. For each sampling period, only one section's coefficients are updated. The simplest updating strategy is to update sections in their sequential order. This sequence will be repeated after updating the NMth section and can be expressed as

$$N1, N2, \ldots, NM; N1, N2, \ldots, NM; \ldots . \quad (6a)$$

There are alternative updating strategies. For example, one can update the $N1$ section more often than others by using the following updating sequence, for example:

$$N1, N1, N2, N3, \ldots, NM; \\ N1, N1, N2, N3, \ldots, NM; \ldots . \quad (6b)$$

This is a time-weighted sequence and there are variety of sequences we can choose depending on applications.

To illustrate the advantage of the TW-LMS algorithm, two simulations are carried out. The first simulation compares the TW-LMS with the standard LMS. A filter of order $N = 1000$ is used for the TW-LMS algorithm. This N is evenly divided into four sections and each has 250 coefficients. Considering 10 consecutive sampling periods as a cycle, the number of times used for updating $N1$, $N2$, $N3$, and $N4$ is in the ratio of 5:3:1:1. For the standard LMS algorithm, a filter of order $N/2 = 500$ is used and all the filter coefficients are updated at a time. The number of multiplications and additions in each case is listed in Table 1. From this table, it is concluded that the TW-

LMS algorithm uses almost the same computation as the standard LMS algorithm when N is large.

Figure 6 shows the computer simulation result. In this figure, the upper curve is the ERLE obtained by the TW-LMS with $N = 1000$, while the lower curve is for the standard LMS with $N/2 = 500$. From this figure and Table 1, we show that by using the TW-LMS algorithm, we can implement a filter of twice the length to achieve 8 dB of improvement over the standard LMS algorithm, but introduces only one quarter more computation. This TW-LMS algorithm can be used in the first section of PAP filters, $W_l(z)$, as shown in Fig. 4 to improve the transient performance (Fig. 5) from 10 to 18 dB. Also, this algorithm can be employed in the tail filter $AW(z)$ to further reduce the computational requirement.

In the second simulation, the ERLE of the TW-LMS algorithm is compared with the traditional LMS algorithm where the adaptations of sections $N1$, $N2$, $N3$, and $N4$ have equal weight. Figure 7 shows the computer simulation result in both cases. In this simulation, the TW-LMS again uses the adaptation ratio 5:3:1:1 for these four sections, and the traditional LMS uses the ratio 1:1:1:1. From this figure, it is obvious that the TW-LMS algorithm has a faster convergence rate compared with the traditional LMS algorithm. Actually, the slopes are 12 dB/10,000 samples and 8 dB/10,000 samples, respectively. In the TW-LMS case, 80% of the adaptation time is used on the first two sections $N1$ and $N2$. These 500 filter coefficients cover the first 62.5 msecs of the impulse response which is contributed by direct coupling and earlier reflections. The faster convergence of the TW-LMS algorithm is important in real-time acoustic echo cancellation applications, since the room environment changes as the result of the movements of the people.

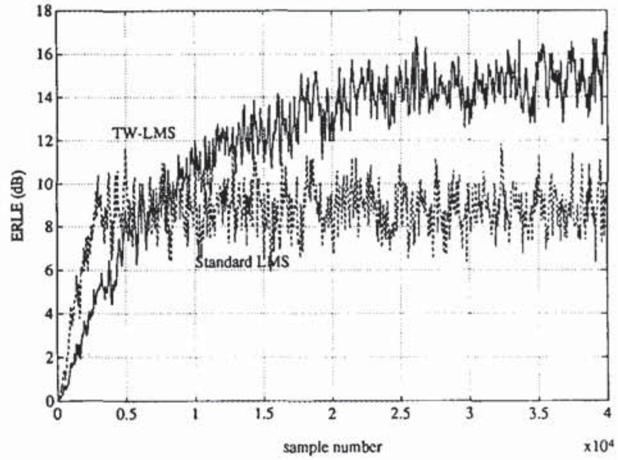

FIG. 6. Comparison of the TW-LMS and the standard LMS algorithms.

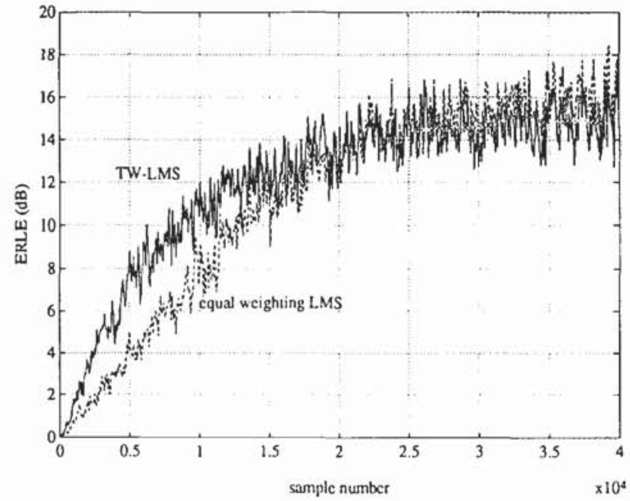

FIG. 7. Comparison of the TW-LMS and the traditional LMS case.

The dynamic (or time-varying) behavior of impulse response due to the movement of people in the room is a complicated problem. In acoustics analysis, the impulse response of acoustic path can be classified into three categories as direct sound, earlier reflection, and long-term reverberation [12,13]. In principle, the earlier reflection involves the first and second reflections from the object existing in the room. Hence for a conference room of size about 8 m long, the earlier echo takes about 47 msec (16/340 secs). Since the direct sound and the earlier reflection are affected more significantly by the environmental change in the room than the long-term reverberations, it is desirable to weight heavier on the first portion of the adaptive filter coefficients so that it can track the change of the echo path. This is the advantage of using the TW-LMS algorithm to replace the traditional LMS algorithm.

5. AEC USING COMBINATIONS OF ADAPTIVE FIR AND IIR FILTERS

In this section, an alternative structure which combines adaptive FIR and IIR filters is proposed. In the multiple microphone echo canceller system described in Fig. 4, an auxiliary filter, $AW(z)$ is assigned to one of the L first-stage filters according to the microphone selection unit. This $AW(z)$ is a high order FIR filter since it is used to model the tail part of the echo. Now considering the situation that $AW(z)$ is an IIR filter with much lower order to perform the same echo cancellation as an FIR filter. The resultant echo canceller for each echo path is shown in Fig. 8.

In Fig. 8, the FIR filter with length $N1$ is used to

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.