ELSEVIER

Speech Communication 16 (1995) 255–260

SPEECH COMMUNICATION

Analysis of finite length acoustic echo cancellation system

Sen M. Kuo *, Jier Chen

Department of Electrical Engineering, Northern Illinois University, Dekalb, IL 60115, USA

Received 18 August 1992; revised 15 February 1994 and 31 October 1994

Abstract

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 results in difficulties in convergence and hardware implementation. In this paper, the performance of the finite length adaptive filter is studied. A formula which relates the echo cancellation to the filter size, N, is established. Detailed analysis shows that this finite filter length will have better performance using speech than white noise.

Zusammenfassung

Die Methode zur Aufhebung des akustischen Echos, die ein adaptives, transversales Filter und einen LMS-Algorithmus (least mean-square) verwendet, ist die effizienteste Technik, um akustische Echos in Freihand-Telefonsystemen zu reduzieren. Da aber jedes Mikrofon mit einem Filter höherer Ordnung ausgestattet werden muß, führt dies unweigerlich zu Schwierigkeiten der Konvergenz und der Implementierung des Materials. In diesem Artikel wird die Leistungsfähigkeit eines adaptiven Filters mit finiter Länge untersucht. Eine Formel, die die Echobeseitigung mit der Filtergröße in Beziehung stellt, wird ausgearbeitet. Bei genauerer Analyse zeigt sich dann, daß diese finite Filterlänge bessere Leistungen bei gesprochener Sprache als bei weißem Rauschen bringt.

Résumé

L'annulation d'écho acoustique utilisant un filtre adaptatif transverse et l'algorithme de l'erreur carrée moyenne minimale (LMS) est la plus efficace pour réduire les échos acoustiques d'un téléphone mains libres. Cependant, il est nécessaire d'utiliser un filtre d'ordre très élevé pour chaque microphone, ce qui induit des problèmes de convergence et d'implantation matérielle. Nous étudions dans cet article les performances des filtres adaptatifs de longueur finie. Nous établissons une formule reliant l'annulation d'écho à la taille du filtre. Une analyse détaillée montre que les performances d'un filtre de longueur finie sont meilleures sur de la parole que sur du bruit blanc.

Keywords: Acoustic echo cancellation; Adaptive FIR filter; Performance analysis of finite length AEC

* Corresponding author.

SSDI 0167-6393(94)00057-3

1. Introduction

The basic structure of a conventional acoustic echo canceler (AEC), which consists of a loudspeaker, microphone, and adaptive filter, is illustrated in Fig. 1. This is a classical system identification where the adaptive filter, $W(z)$, adjusts its coefficients via the well-known 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 very high order is needed to cover this type of echo path. As discussed in (Widrow and Stearns, 1985), if a transversal filter with the LMS algorithm is used, we have

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

and

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

where μ is the step size, 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 $\lambda_{\min}$ is the minimum eigenvalue of the autocorrelation matrix of the input signal. Eq. (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). Therefore, the filter is unable to track the transient behavior of $H(z)$.

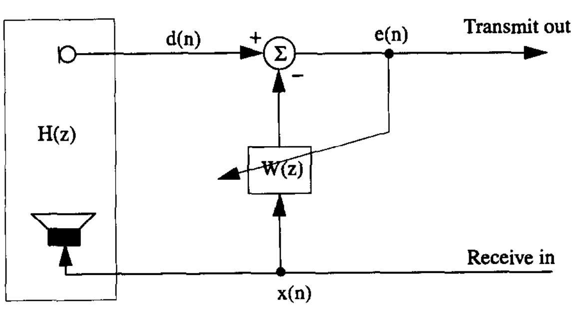


Fig. 1. System diagram of acoustic echo canceller.

Moreover, if fixed point arithmetic is used and the assumption is made that the same word length is used for both data and coefficients and μ is sufficiently small, the total output mean-square error is (Caraiscos and Liu, 1984)

$$\epsilon = \epsilon_{\min} + \frac{\mu \epsilon_{\min} N \sigma_x^2}{2} + \frac{N\sigma_e^2}{2\mu} + \left(\|w^0\|^2 + 1\right)\sigma_e^2, \tag{3}$$

where $\epsilon_{\min}$ is the minimum mean-square error (MSE) of the optimal (Wiener) filter w^0 and σ_e^2 is the variance of the coefficients quantization error. The second term of Eq. (3) shows that an excess MSE is increased when a large N is used, and the third term shows that 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 stalling when a small μ is used (Gitlin et al., 1973). 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 (Oikawa et al., 1988).

2. Steady state error analysis

Since the impulse response of the room acoustic echo is very long, there are limitations for using a small filter to cancel the acoustic echo. According to Fig. 1, the acoustic echo $d(n)$ picked up by the microphone is

$$d(n) = \sum_{i=0}^{\infty} h_i x(n-i), \tag{4}$$

where h_i is the impulse response of the echo path $H(z)$. If $w_i(n)$ is the ith filter coefficient at time n, the system output is

$$e(n) = d(n) - \sum_{i=0}^{N-1} w_i(n) x(n-i). \tag{5}$$

This equation can be written in vector form as

$$\begin{aligned} e(n) &= d(n) - w_N^{\mathrm{T}}(n) x_N(n) \\ &= d(n) - w^{\mathrm{T}}(n) x(n), \end{aligned} \tag{6}$$

where

$$x_N^T(n) = [x(n)\ x(n-1) \ldots x(n-N+1)],$$

$$w_N^T(n) = [w_0(n)\ w_1(n) \ldots w_{N-1}(n)],$$

$$x^T(n) = [x_N^T(n)\ x(n-N)\ x(n-N-1) \ldots]$$

and

$$w^T(n) = [w_N^T(n)\ 0 \ldots].$$

Since we have finite filter length N, we can express the input vector, $x(n)$, in two portions as

$$x(n) = x_1(n) + x_2(n), \tag{7}$$

where

$$x_1(n) = [x_N^T(n)\ 0 \ldots]^T$$

and

$$x_2(n) = [0 \ldots 0\ x(n-N)\ x(n-N-1) \ldots]^T.$$

Similarly, we can express the impulse response of the acoustic echo path as

$$h = h_1 + h_2, \tag{8}$$

where

$$h_1 = [h_0\ h_1 \ldots h_{N-1}\ 0 \ldots]^T = [h_N^T\ 0 \ldots]^T$$

and

$$h_2 = [0 \ldots 0\ h_N\ h_{N+1} \ldots]^T.$$

Therefore, the echo signal $d(n)$ in Eq. (4) can be written in vector form as

$$d(n) = h^T x(n) = h_1^T x_1(n) + h_2^T x_2(n) = h_N^T x_N(n) + h_2^T x_2(n). \tag{9}$$

Here, we assume that the echo signal given in Eq. (9) is the desired-response signal in the absence of noise and near-end speech. From Eq. (6), the mean-square error can be expressed by

$$\epsilon = E[e^2(n)] = E[d^2(n)] + w_N^T(n)E[x_N(n)x_N^T(n)]w_N(n) - 2w_N^T(n)E[d(n)x_N(n)], \tag{10}$$

where $E[\cdot]$ is the expectation operator. The optimal solution, w^0, can be obtained by setting the derivative of ϵ with respect to $w(n)$ equal to zero. Therefore, we have

$$w_N^0 = R_N^{-1} p_N, \tag{11}$$

where

$$R_N = E[x_N(n)x_N^T(n)]$$

and

$$p_N = E[d(n)x_N(n)].$$

By replacing $d(n)$ using Eq. (9) and rearranging, we have

$$w_N^0 = h_N + R_N^{-1} E[x_N(n)x_2^T(n)]h_2. \tag{12}$$

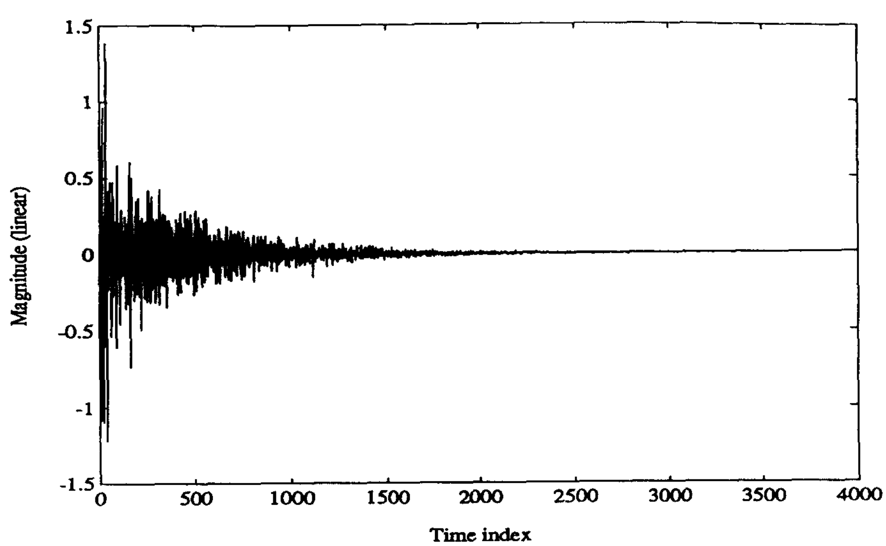


Fig. 2. An impulse response measured in $8 \times 5 \times 3.5$ m^3 room.

Also, by substituting Eq. (12) into Eq. (10), and using the identity in Eq. (9), the minimum MSE for the finite filter length case is

$$\epsilon_{\min} = \boldsymbol{h}_2^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{x}_2(n)\boldsymbol{x}_2^{\mathrm{T}}(n)\right]\boldsymbol{h}_2 - \boldsymbol{h}_2^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{x}_2(n)\boldsymbol{x}_N^{\mathrm{T}}(n)\right]\boldsymbol{R}_N^{-1} \times \left\{\mathrm{E}\left[\boldsymbol{x}_N(n)\boldsymbol{x}_2^{\mathrm{T}}(n)\right]\boldsymbol{h}_2\right\}. \tag{13}$$

This equation shows that, because the filter length is finite, the coefficients of $\boldsymbol{h}_2$, which is the uncovered portion of the impulse response, contributes to both terms of the residual error. If the order of the adaptive filter, N, can be chosen infinitely long, then $\boldsymbol{h}_2$ is the zero vector and it also forces $\epsilon_{\min}$ to zero.

3. Finite filter length effects analysis

The validity of Eqs. (12) and (13) can be verified by considering the case where the input signal is white noise, and, consequently, there is no correlation between $\boldsymbol{x}_N(n)$ and $\boldsymbol{x}_2(n)$. The last terms in both equations have vanished, hence

$$\boldsymbol{w}^0 = \boldsymbol{h}_N \tag{14}$$

and

$$\epsilon_{\min} = \boldsymbol{h}_2^{\mathrm{T}} \mathrm{E}\left[\boldsymbol{x}_2(n)\boldsymbol{x}_2^{\mathrm{T}}(n)\right]\boldsymbol{h}_2 = \sigma_x^2 \sum_{i=N}^{\infty} h_i^2. \tag{15}$$

Therefore, for a white noise input, the minimum MSE is proportional to the sum of the squared terms in $\boldsymbol{h}_2$. Due to the exponential decay char-

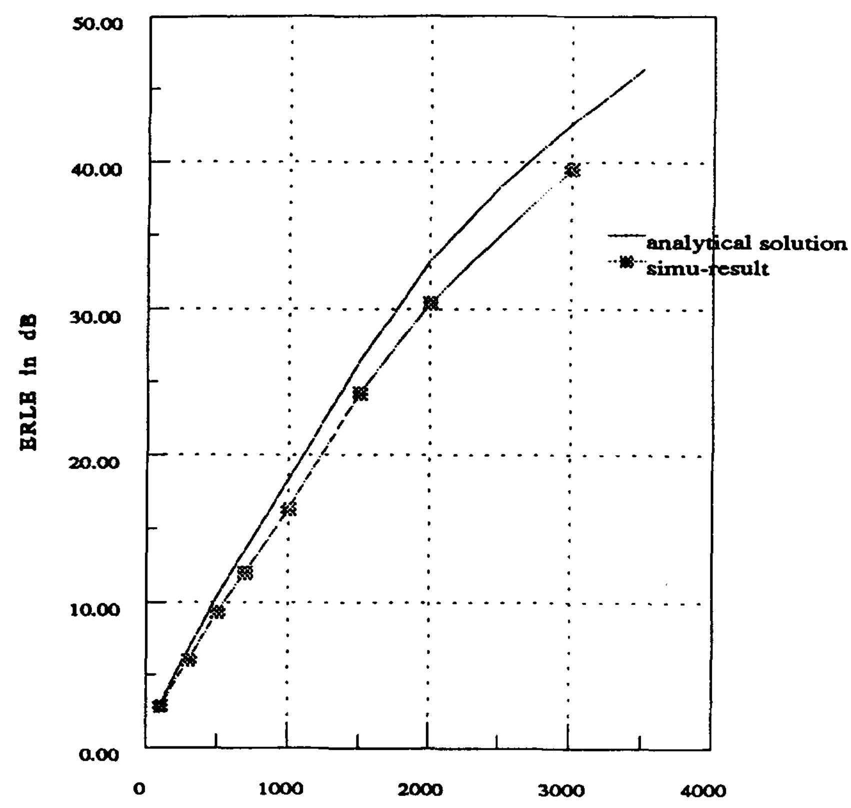
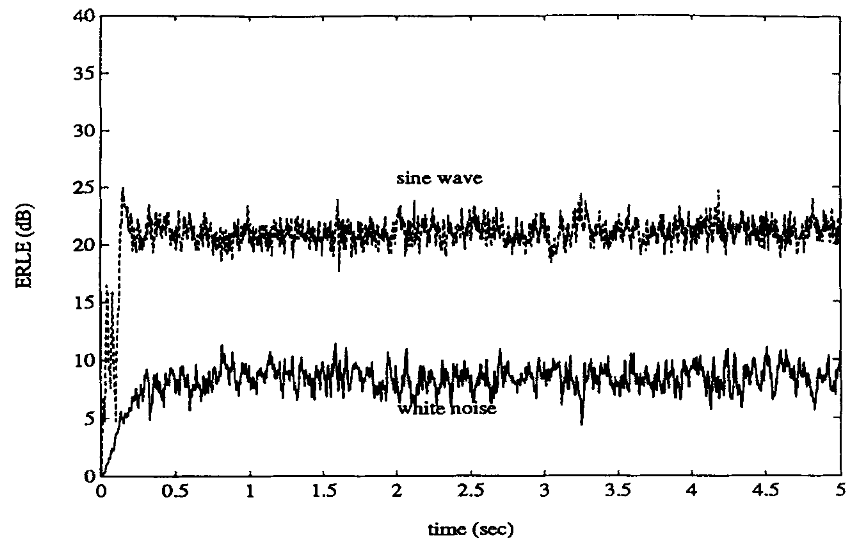


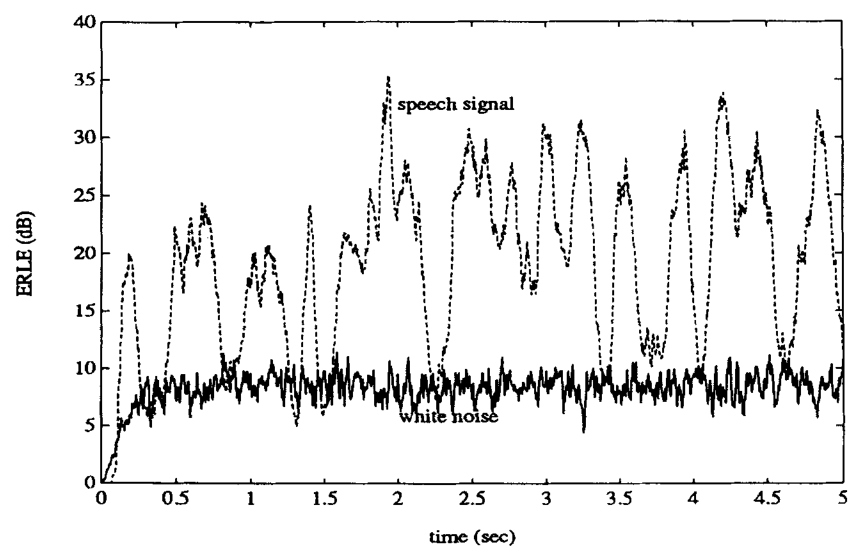
Fig. 3. Simulation result for different N compared with analytical prediction.

acteristics of the impulse response, these terms will contribute little, if N is large.

To demonstrate the analysis result in Eqs. (14) and (15), simulations were performed by using a typical room acoustic echo impulse response shown in Fig. 2. Data was taken in a medium size room using an impact testing method. An air balloon was burst, and the noise was sensed by the microphone and recorded. The Areil DSP-16 system that has two 16-bit A/D converters was used for data acquisition. A sampling rate, $f_s = 10$ kHz, was used for the experiment. Fig. 2 indi-



(a)



(b)

Fig. 4. Finite filter length effect on ERLE for various signals. (a) Comparison of white noise to pure sinewave; (b) white noise and speech signal.

cated the exponential decaying nature of the room impulse response.

In computer simulation, the filter length N is varied from 100 to 3000 and the $\epsilon_{\min}$ as a function of N is obtained. These residual errors are then represented by echo return loss enhancement (ERLE) since this is commonly used in evaluating the performance of echo cancellation. It is defined as the logarithmic ratio between the expected values of the squared echo $d(n)$ and the residue error $e(n)$. That is,

$$\mathrm{ERLE} = 10\log_{10}\left(\frac{\mathrm{E}[d^2(n)]}{\mathrm{E}[e^2(n)]}\right). \tag{16}$$

Computer simulation results are presented in Fig. 3, where each mark on the graph indicates the ERLE at a different N and the line is the analytical prediction using Eqs. (15) and (16). The simulation results are in good agreement with the analytical prediction though they deviate slightly as N becomes large. This can be explained as the misadjustment noise introduced in large adaptive transversal filters (Widrow and Stearns, 1985). In these simulations, the input signal is white noise.

In real echo cancellation application, a speech signal is used and it is well known that a speech signal is highly correlated. Therefore, the last terms in Eqs. (12) and (13) will not be zero. This will affect the optimal solution w^0, and the minimum MSE accordingly. As indicated in Eq. (12), the optimal filter coefficients will not be the same as that of the room echo impulse response. Rather they are biased by the second term in this equation. And it is also true from Eq. (13) that the minimum residual error is modified by a second term which involves the cross correlation between different portions of the input signal, $x(n)$.

The second term in Eq. (13) is always a positive number when the input signal is correlated (e.g., speech) and will be zero for a totally uncorrelated signal (white noise). For a finite length adaptive filter, the residue error obtained for correlated signals should be smaller than that obtained for uncorrelated white noise.

Simulations were performed for input signals consisting of white noise, pure sine wave, and a speech signal. The order of the adaptive filter is $N = 256$. The ERLE for each case is shown in Fig. 4. In Fig. 4, a comparison of sine wave and white noise is made, and it is observed that there is a 15 dB improvement by using the sine wave over the white noise since the sine wave is highly correlated. In Fig. 4(b), the ERLE for a speech signal is compared to that of white noise. This figure shows that because of the stationary property of white noise, the ERLE for white noise is nearly flat. But the ERLE obtained for the speech signal varies significantly due to the dynamic nature of speech signals. In general, the ERLE for speech signals is greater than the ERLE for white noise ($\sim$ 10 dB).

4. Conclusion

This paper analyzed the performance of the finite length adaptive transversal filter for acoustic echo cancellation. In the case of undermodeling a long impulse response with an FIR filter of insufficient length, the MSE decreases as the input signal correlation increases.

References

B. Widrow and S.D. Stearns (1985), *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

C. Caraiscos and B. Liu (1984), "A roundoff error analysis of the LMS adaptive algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-32, No. 1, pp. 34–41.

R.D. Gitlin, J.E. Mazo and M.G. Taylor (1973), "On the design of gradient algorithm for digital implemented adaptive filter", *IEEE Trans. Circuit Theory*, Vol. CT-20, No. 2, pp. 125–136.

H. Oikawa, S. Minami and T. Saehi (1988), "A new echo canceller realized by high performance digital signal processor", *Proc. IEEE Internat. Symp. Circuits and Systems*, pp. 1329–1332.