Acoustic Echo Canceller Algorithm based on Room Acoustic Characteristics

Shoji Makino, Nobuo Koizumi
NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180, Japan

Summary

A new adaptive algorithm called the exponential step or ES algorithm is proposed for an acoustic echo canceller. Based on room acoustic characteristics, this algorithm adjusts the magnitude and direction of the adjustment vector in the gradient algorithm[1],[2]. Real time experiments show that the ES algorithm converges faster than previous algorithms : triple the speed for white noise signal and double the speed for speech signal without increasing the number of calculations per iteration.

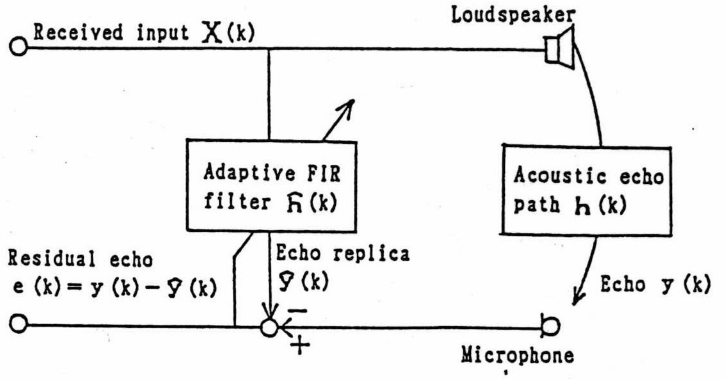


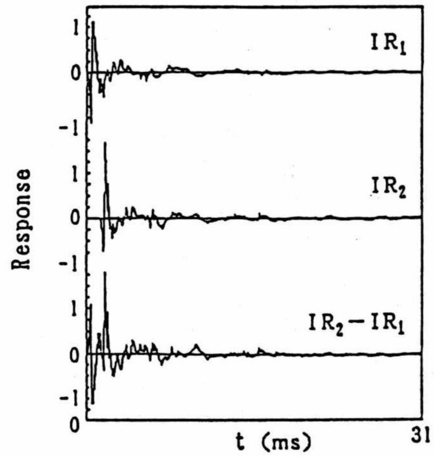
Fig. 1 Configuration of an acoustic echo canceller.

1. Variation of an impulse response

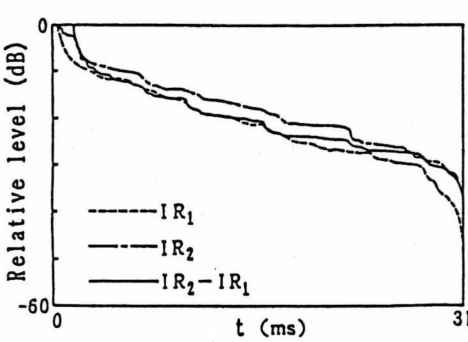
An adaptive Finite Impulse Response filter updated with the gradient algorithm[1],[2] is applied to an acoustic echo canceller shown in Figure 1. An adaptive algorithm suitable for the variation characteristics of an acoustic echo path is expected to improve convergence performance. The impulse response variation is studied to determine common impulse response properties. Here, for simplicity, the case is discussed where the distance between a loudspeaker and a microphone changes.

Impulse responses are measured for loudspeaker-microphone distances for 0.6 m (IR_1) and 1 m (IR_2) in a conference room (reverberation time at 500 Hz is 160 ms). These wave forms IR_1 and IR_2, and a subtraction (IR_2-IR_1) are shown in Figure 2(a).

The reverberent energy decay curves of IR_1, IR_2 and IR_2-IR_1, calculated by integrating squares of these wave forms[3], are shown in Figure 2(b). Figure 2 shows that impulse responses IR_1 and IR_2 attenuate exponentially and that the subtraction, which is the variation in these impulse responses, attenuates by the same exponential ratio.



(a) Impulse response



(b) Reverberent energy decay curve

Fig. 2 Impulse response and reverberent energy decay curve. IR_1 and IR_2 are impulse responses for loudspeaker-microphone distance 0.6 m and 1 m respectively. Reverberation time at 500 Hz is 160 ms.

2. Exponential step (ES) algorithm

A step gain matrix α, which is a scalar step gain α extended to a diagonal matrix, is introduced here to update large error coefficients in large steps and small error coefficients in small steps.

$$\alpha = \mathrm{diag}\,[\,\alpha_1, \alpha_2, \ldots, \alpha_N\,] \tag{1}$$

where

$$\alpha_i = (\alpha_{max} - \alpha_{min}) \exp(-6.9\,(i-1)\,T_S / T_R) + \alpha_{min} \quad (i=1, 2, \ldots, N).$$

T_S : sampling interval, T_R : reverberation time,
N : number of taps.

The ES algorithm for learning identification method[2] is proposed by the following equation.

$$\hat{h}(k+1) = \hat{h}(k) + \alpha\, e(k)\, X(k) / \|X(k)\|^2 \tag{2}$$

where

ASSP WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS

IEEE

Mohonk Mountain House
New Paltz, New York
October 15-18, 1989

$\hat{h}(k)$: tap coefficients of a canceller, $e(k)$: residual echo, $X(k)$: received input, $\|X\|$: norm of X.

Multiplying diagonal matrix α to the adjustment vector means modifying the magnitude and direction of the adjustment vector of the conventional method. For a practical system constructed with Digital Signal Processor chips, step gain α_i is set in discrete steps with one constant value per DSP chip as shown in Figure 3. In this case the number of calculations per iteration is equal to that of conventional algorithm.

The ES algorithm is different from other algorithms[4] in that it is based on room acoustic characteristics and that fast convergence speed is obtained without increasing the number of calculations per iteration.

3. Real time experiments

Figure 4 shows the real time experimental results on residual echo level convergence performance using a canceller which is constructed with DSP chips[5]. Reverberation time of the conference room is 300 ms and speaker-microphone distance is 2.5 m. 7 kHz frequency range is separated into two bands. A canceller samples at 8 kHz in each band. The lower band is 0.15-4 kHz, has 3840 taps, and cancels 480 ms echo. The higher band is 4-7 kHz, has 1792 taps, and cancels 224 ms echo[6]. For real time operation, α_i is set to $\alpha_{1\text{-}256}$=2.5, $\alpha_{257\text{-}512}$=1.5, $\alpha_{513\text{-}last}$=0.3. When using a white noise input signal (Figure 4(a)), convergence at -20 dB is at triple the speed and when using a speech (male) input signal (Figure 4(b)), convergence at -20 dB is at double the speed of the maximum speed (α=1.0) in the conventional method[2]. Echo return loss enhancement is over 30 dB in 0.15-7 kHz frequency range.

4. Conclusion

The ES algorithm reflecting room acoustic characteristics converges faster than previous algorithm : triple the speed for white noise signal and double the speed for speech signal without increasing the number of calculations per iteration.

References

[1] B. Widrow, et al.:"Adaptive noise cancelling : Principles and applications", Proc. IEEE, 63, pp.1692-1716 (1975).

[2] J. Nagumo, A. Noda :"A learning method for system identification", IEEE Trans., AC-12, pp.282-287 (1967).

[3] M.R.Schroeder :"New method of measuring reverberation time", J. Acoust. Soc. Am., 37, pp.409-412 (1965).

[4] R.W.Harris, D.M.Chabries, F.A.Bishop:"A variable step (VS) adaptive filter algorithm", IEEE Trans., ASSP-34, pp.309-316 (1986).

[5] T.Kaneko, H.Yamauchi, A.Iwata :"A 50 ns floating-point signal processor VLSI", ICASSP86, pp.401-404 (1986).

[6] H.Oikawa, S.Minami, T.Saeki :"A new echo canceller realized by high performance digital signal processor", ISCAS88, pp.1329-1332 (1988).

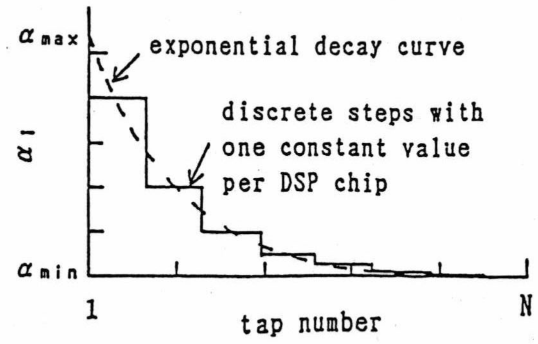
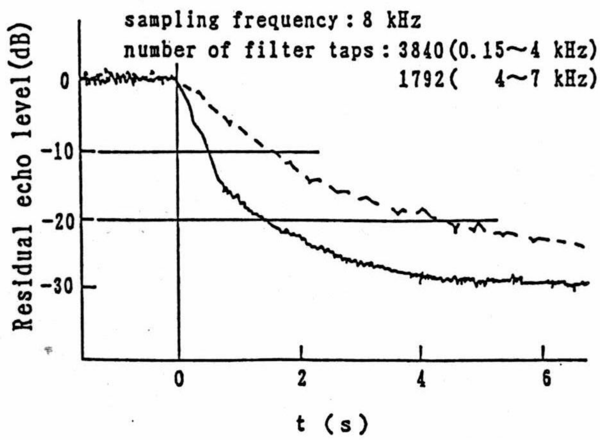
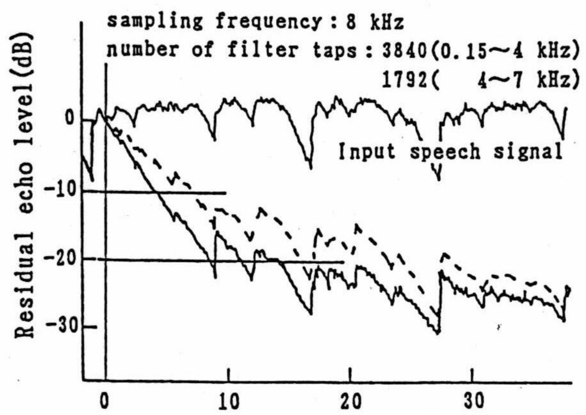


Fig. 3. Diagonal component α_i of step gain matrix α in the case of α_i is set in discrete steps with one constant value per Digital Signal Prosessor Chip.



(a) White noise input



(b) Speech (male) input

Fig. 4 Real time experimental results on convergence performance. Step gain matrix with $\alpha_{1\text{-}256}$=2.5, $\alpha_{257\text{-}512}$=1.5, $\alpha_{513\text{-}last}$=0.3 (——) and α=1.0 (- - -) in learning identification algorithm. Reverberation time at 500 Hz is 300 ms. Loudspeaker-microphone distance is 2.5 m.