AN ADAPTIVE CONTROL METHOD FOR REGRESSION COEFFICIENT IN FREQUENCY DOMAIN NON-LINEAR ECHO SUPPRESSOR

Kenji Nakayama *Yuki Ofuji* *Akihiro Hirano*

Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, 920-1192, Japan
E-mail: nakayama@ec.t.kanazawa-u.ac.jp

ABSTRACT

An echo canceller, which consists of a linear echo canceller and a non-linear echo suppressor, has been proposed for mobile phones. This approach provides an efficient non-linear echo suppressor with low computational complexity. In this method, the non-linear echo is estimated in the frequency domain from the linear echo replica. A ratio of the non-linear echo and the linear echo replica is called 'Regression Coefficient'. In this approach, it is very important how to estimate and how to update the regression coefficient by using the residual signal after the linear echo canceller, which includes the residual non-linear echo and the near-end signal, and the linear echo replica. In this paper, we propose a new adaptive control method for the regression coefficient. The ratio of the residual signal and the replica quickly changes along frame, which is an interval of FFT. Therefore, first the average of the ratio is estimated. Second, the average is amplified in order to suppress the residual non-linear echo. In the proposed method, the regression coefficient is not updated in the double talk intervals, because effects of the near-end signal is large. The double talk intervals are detected based on the correlation coefficient between the linear echo replica and the microphone input signal, which includes the linear and non-linear echo components and the near-end signal. Simulation results show the proposed method can improve echo reduction and the segmental SNR by 13 dB and $1.5 \sim 2.5$ dB, respectively.

Index Terms— Adaptive filter, Mobile phone, Echo canceller, Non-linear, Spectral suppression, Segmental SNR

1. INTRODUCTION

In these days, mobile phones are widely used in a variety of environments. In the mobile phones, however, communication quality is suffered from echo. The echo caused by the mobile phone includes linear and non-linear components. Especially, to cancel the non-linear echo is very complicate, and requires very high computationally complexity [4],[5],[6], [9]. For this purpose, an echo canceller has been proposed, which combine a linear echo canceller and a non-linear echo suppressor [7],[8],[10]. Some relation is assumed, that is, the linear echo is mainly cancelled by the linear echo canceller, and the residual non-linear echo is proportional to the linear echo replica in the frequency domain. This relation is roughly held. Their ratio is called as 'Regression Coefficient' in this paper. In this approach, it is very important how to estimate and update the regression coefficient, by using the residual signal after the linear echo canceller and the linear echo replica can be used.

In this paper, we propose a new adaptive control method for the regression coefficient. Simulation results obtained by using many kinds of non-linear noise and speeches will be shown.

2. AN ECHO CANCELLER FOR MOBILE PHONES

Figure 1 shows the echo canceller proposed for mobile phones [7],[8]. $x(n)$ is the far-end signal. $y(n)$ and $q(n)$ are linear and non-linear echo components, respectively. $s(n)$ is the near-end signal. First, a linear echo canceller, which realized by an FIR adaptive filter controlled by NLMS algorithm, is used in the time domain in order to cancel mainly the linear echo $y(n)$. $\hat{y}(n)$ is the linear echo replica for mainly $y(n)$. $d(n)$ includes the residual echo, which is mainly non-linear

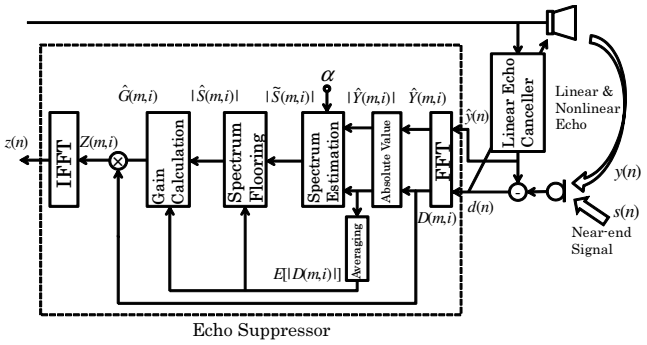

Fig. 1. Echo cnaceller for mobile phones [7],[8].

echo $q(n)$, and the near-end signal $s(n)$. In the echo suppressor, the residual echo, which is mainly $q(n)$, is suppressed.

In the following, signal processing in the echo suppressor is described [7],[8],[10]. $\hat{y}(n)$ and $d(n)$ are Fourier transformed resulting in $\hat{Y}(m,i)$ and $D(m,i)$, respectively. m and i mean the frame number and the frequency number, respectively. $D(m,i)$ consists of $S(m,i)$ and $Q(m,i)$, which are Fourier transform of $s(n)$ and $q(n)$, respectively.

$$d(n) = s(n) + q(n) \quad (1)$$
$$D(m,i) = S(m,i) + Q(m,i) \quad (2)$$

Furthermore, it is assumed that $|Q(m,i)|$ can be approximated by using $|\hat{Y}(m,i)|$ as follows:

$$|Q(m,i)| \simeq |\hat{Q}(m,i)| = \hat{a}(m,i)|\hat{Y}(m,i)| \quad (3)$$

$\hat{a}(m,i)$ expresses 'Regression Coefficient'.

The output signal of the echo suppressor $z(n)$ is obtained by using $D(m,i)$ and a spectral gain $\hat{G}(m,i)$, which is adjusted so as to suppress the residual echo, mainly non-linear echo.

$$|Z(m,i)| = \hat{G}(m,i)|D(m,i)| \quad (4)$$
$$Z(m,i) = |Z(m,i)|exp(j\angle D(m,i)) \quad (5)$$
$$z(n) = IFFT[Z(m,i)] \quad (6)$$

$\hat{G}(m,i)$ is calculated by the following process [7],[8].

$$\hat{G}(m,i) = \begin{cases} \beta_{GA}\tilde{G}(m,i) + (1-\beta_{GA})\hat{G}(m-1,i) \\ \quad if \quad \tilde{G}(m,i) > \hat{G}(m-1,i) \\ \beta_{GD}\tilde{G}(m,i) + (1-\beta_{GD})\hat{G}(m-1,i) \\ \quad if \quad \tilde{G}(m,i) \leq \hat{G}(m-1,i) \end{cases} \quad (7)$$

$$\tilde{G}(m,i) = \frac{|\hat{S}(m,i)|}{E[|D(m,i)|] + \sigma} \quad (8)$$

β_{GA} and β_{GD} are constants ($0 < \beta_{GD} < \beta_{GA} \leq 1$). σ is a small positive number for stabilization.

$$|\hat{S}(m,i)| = \max(\gamma_D|D(m,i)|_f, |\tilde{S}(m,i)|), \gamma_D = 1 \quad (9)$$

$$|D(m,i)|_f = \begin{cases} \beta_{FA}|D(m,i)| + (1-\beta_{FA})|D(m-1,i)|_f \\ \quad if \quad |D(m,i)| > |D(m-1,i)|_f \\ \beta_{FD}|D(m,i)| + (1-\beta_{FD})|D(m-1,i)|_f \\ \quad if \quad |D(m,i)| \leq |D(m-1,i)|_f \end{cases} \quad (10)$$

β_{FA} and β_{FD} are constants ($0 < \beta_{FA} < \beta_{FD} \leq 1$).

$$E[|S(m,i)|]^2 \simeq E[|D(m,i)|]^2 - E[|Q(m,i)|]^2 \quad (11)$$
$$\simeq E[|D(m,i)|]^2 - \hat{a}_i^2 E[|\hat{Y}(m,i)|]^2 \quad (12)$$
$$\tilde{S}(m,i)| = \sqrt{E[|D(m,i)|]^2 - \hat{a}_i^2 E[|\hat{Y}(m,i)|]^2} \quad (13)$$

$$E[|D(m,i)|] = \begin{cases} \beta_{DA}|D(m,i)| + (1-\beta_{DA})E[|D(m-1,i)|] \\ \quad if \quad |D(m,i)| > E[|D(m-1,i)|] \\ \beta_{DD}|D(m,i)| + (1-\beta_{DD})E[|D(m-1,i)|] \\ \quad if \quad |D(m,i)| \leq E[|D(m-1,i)|] \end{cases} \quad (14)$$

$$E[|\hat{Y}(m,i)|] = \begin{cases} \beta_{YA}|\hat{Y}(m,i)| + (1-\beta_{YA})E[|\hat{Y}(m-1,i)|] \\ \quad if \quad |\hat{Y}(m,i)| > E[|\hat{Y}(m-1,i)|] \\ \beta_{YD}|\hat{Y}(m,i)| + (1-\beta_{YD})E[|\hat{Y}(l-1.k)|] \\ \quad if \quad |\hat{Y}(m,i)| \leq E[|\hat{Y}(m-1,i)|] \end{cases} \quad (15)$$

β_{DA}, β_{DD}, β_{YA} and β_{YD} are constants ($0 < \beta_{DD} < \beta_{DA} \leq 1$ and $0 < \beta_{YD} < \beta_{YA} \leq 1$).

3. CONVENTIONAL METHOD FOR UPDATING REGRESSION COEFICIENT

The ratio of the residual non-linear echo and the linear echo replica varies frame by frame and frequency by frequency. Furthermore, the non-linear echo component is depends on the far-end speech. Therefore, in this approach, it is very important how to control the regression coefficient.

One of the conventional methods is to optimally tune the regression coefficient in advance, and is fixed in operation [7],[8]. However, the tuning process is complicated, and not useful for practical applications.

Another method is to automatically control the regression coefficient [10]. The update process is given by the following equations.

$$\tilde{b}(m,i) = \frac{E[|D(m,i)|]}{E[|\hat{Y}(m,i)|]} \quad (16)$$

$$b(m,i) = \begin{cases} \beta_{b1}\tilde{b}(m,i) + (1-\beta_{b1})b(m-1,i) \\ \quad if \quad \tilde{b}(m,i) > b(m-1,i) \\ \beta_{b2}\tilde{b}(m,i) + (1-\beta_{b2})b(m-1,i) \\ \quad if \quad \tilde{b}(m,i) \leq b(m-1,i) \end{cases} \quad (17)$$

β_{b1} and β_{b2} are constant ($0 < \beta_{b1} \ll \beta_{b2} < 1$). The regression coefficient uses an independent value for each frequency. It is updated along the time axis, that is the frame. First, a ratio $\tilde{b}(m,i)$ of $E[|D(m,i)|]$ and $E[|\hat{Y}(m,i)|]$ is calculated in each frame. Since it quickly changes along the frame, it is smoothed resulting in $b(m,i)$. In the conventional method, the regression coefficient is updated in both single talk and double talk intervals. Especially, in the double talk interval, $E[|D(m,i)|]$ includes the near-end signal, and $\tilde{b}(m,i)$ is affected. In order to avoid effects of the near-end signal, $b(m,i)$ is controlled so as to traces the lower bound of $\tilde{b}(m,i)$. For this purpose, the smoothing parameters β_{b1} and β_{b2} are tuned as $0 < \beta_{b1} \ll \beta_{b2} < 1$ $b(m,i)$.

Since $b(m,i)$ is the lower bound of $\tilde{b}(m,i)$, it should be

amplified to estimate the appropriate regression coefficient.

$$\hat{a}(m,i) = vb(m,i) \tag{18}$$

v is an amplification rate. By using a large value for v, the echo can be well reduced, however, the near-end signal distortion will be increased. Therefore, v should be optimally tuned.

4. PROPOSED METHOD FOR UPDATING REGRESSION COEFFICIENT

4.1. A New Update Equation

The conventional method estimates the lower bound of $\tilde{b}(m,i)$ in order to avoid the double talk effects. This method does not require any double talk detection. However, the lower bound of $\tilde{b}(m,i)$ does not have useful information to estimate the regression coefficient. Therefore we propose a new method based on the following idea.

- The average of $\tilde{b}(m,i)$ holds useful information to express the regression coefficient.
- The regression coefficient is not updated in the double talk intervals.

The regression coefficient is updated in the single talk intervals as follows:

$$\tilde{b}(m,i) = \frac{E[|D(m,i)|]}{E[|\hat{Y}(m,i)|]} \tag{19}$$

$$b(m,i) = \alpha b(m-1,i) + (1-\alpha)\tilde{b}(m,i) \tag{20}$$

$$\hat{a}(m,i) = vE[\tilde{b}(m,i)] \tag{21}$$

$b(m,i)$ and $\hat{a}(m,i)$ are not updated in the double talk intervals. Their previous values just before the double talk are held and are fixed in the double talk intervals.

Compared with the conventional method, the proposed method has the following advantages.

- $b(m,i)$ has correct information of $\tilde{b}(m,i)$. This means the regression coefficient can express correctly the relation between the residual non-linear echo and the linear echo replica. In other word, the non-linear echo spectrum $|\hat{Q}(m,i)|$ can be correctly expressed.
- Since $b(m,i)$ is not affected by the near-end signal, it can be amplified in order to suppress the residual echo, while maintaining low near-end signal distortion. Both high echo reduction and low signal distortion can simultaneously achieved.

4.2. Double Talk Detection

In our method, double talk detection is required. Several useful methods have been discussed [2],[3]. In this paper, we propose a new method, which employ a correlation coefficient between the linear echo replica $\hat{y}(n)$ and the microphone input signal, which includes the echo $y(n)+q(n)$ and the near-end speech $s(n)$.

$$u(n) = y(n) + q(n) \ \text{ or } \ y(n) + q(n) + s(n) \tag{22}$$

$$\boldsymbol{u}(n) = [u(n), u(n-1), \cdots, u(n-N-1)]^T \tag{23}$$

$$\hat{\boldsymbol{y}}(n) = [\hat{y}(n), \hat{y}(n-1), \cdots, \hat{y}(n-N-1)]^T \tag{24}$$

$$\rho = \frac{\hat{\boldsymbol{y}}^T(n)\boldsymbol{u}(n)}{||\hat{\boldsymbol{y}}(n)|| \cdot ||\boldsymbol{u}(n)||} \tag{25}$$

$$\rho \geq \theta \rightarrow \text{Single talk} \tag{26}$$

$$\rho < \theta \rightarrow \text{Double talk} \tag{27}$$

θ is the threshold, which will be determined by experience.

5. SIMULATION

5.1. Simulation Setup

The number of taps of the linear echo canceller is 512, the length of the linear echo path is also 512. Non-linear echo path is realized by using the 2nd-order Volterra function, with 60×60 taps. The linear and non-linear echo paths are combined in parallel. Magnitude of the impulse response of the non-linear echo path is adjusted in simulation as follows: Its mean square becomes $0 \sim 50\%$ of the mean square of the linear echo path impulse response. We want to investigate effects of the non-linearity. The sampling frequency is 8kHz.

5.2. Performance Evaluations

Echo reduction is evaluated by a ratio of the linear echo canceller input signal $u(n) = y(n) + q(n)$ and the echo suppressor output signal $z(n)$ in the single talk intervals.

$$R_{echo} = 10\log_{10}\frac{E[z^2(n)]}{E[u^2(n)]} \tag{28}$$

$E[z^2(n)]$ and $E[u^2(n)]$ are the mean square of $z(n)$ and $u(n)$, respectively.
Segmental Signal to Noise Ratio (SNR) is used to evaluate the residual echo and the near-end signal distortion.

$$SNR_{seg} = \frac{10}{L}\sum_{l=0}^{L-1}\log_{10}\frac{\sum_{n=N_l}^{N_l+N-1}s^2(n)}{\sum_{n=N_l}^{N_l+N-1}(z(n)-s(n))^2} \tag{29}$$

5.3. Simulation Results and Discussions

5.3.1. Relation between Linear Echo and Non-linear Echo

Figure 2 shows the relation between $E[|D(m,i)|]$ and $E[|\hat{Y}(m,i)|]$, where the near-end signal is not included. The frequency is

1kHz. One point means $E[|D(m,i)|]$ and $E[|\hat{Y}(m,i)|]$ at one frame. The data at all frames are plotted. As shown in this figure, their relation is different frame by frame, also frequency by frequency. The regression coefficient $\hat{a}(m,i)$ uses an independent value for each frequency, and is adaptively adjusted along the frame.

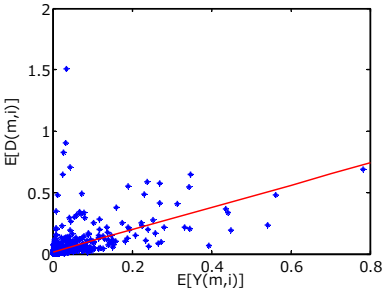

Fig. 2. Relation between $E[|D(m,i)|]$ and $E[|\hat{Y}(m,i)|]$.

5.3.2. *Estimation of Regression Coefficient*

The simulation results in the conventional method at the frequency of 1kHz are shown in Fig.3. $b(m,i)$ estimates the

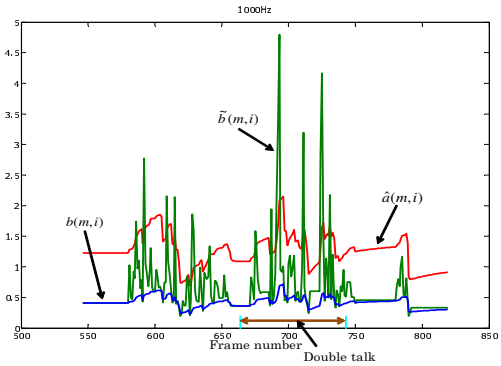

Fig. 3. Lower bound $b(m,i)$ and regression coefficient $\hat{a}(m,i)$ in conventional method. Amplifier rate is $v = 3$ and frequency is 1kHz.

lower bound of $\tilde{b}(m,i)$ and the regression coefficient $\hat{a}(m,i)$ is obtained by amplifying $b(m,i)$ by $v = 3$. They are updated in both single and double talk intervals. This method does not requires the double talk detection. However, $b(m,i)$ is affected by the near-end signal, and cannot exactly express a relation between the residual non-linear echo and the linear echo replica. Even though echo reduction can be improved by using a large value for v, at the same time, the near-end signal can be easily distorted. This is a big problem in the conventional method. As a result, both high echo reduction and low near-end signal distortion cannot be obtained.

The simulation results in the proposed method are shown in Fig.4 $b(m,i)$ estimates the average of $\tilde{b}(m,i)$. $\hat{a}(m,i)$ is

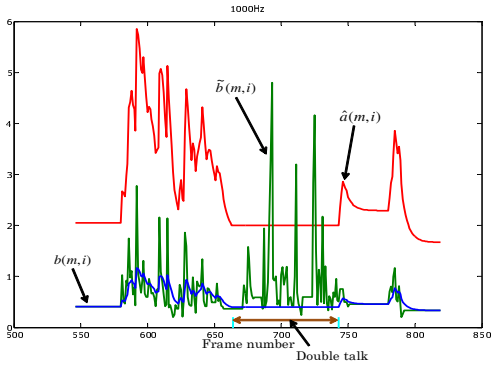

Fig. 4. Average value $b(m,i)$ and regression coefficient $\hat{a}(m,i)$ in proposed method. Smoothing rate is $\alpha = 0.8$ and amplifier rate is $v = 5$ at frequency of 1kHz.

obtained by amplifying $b(m,i)$ by $v = 5$. In the proposed method, $b(m,i)$ is not updated in the double talk intervals, in other word, it is not affected by the near-end signal. Therefore, v can be set to a relatively large value, while maintaining low near-end signal distortion. This point is a very important advantage over the conventional method. As a result, both high echo reduction and low near-end signal distortion can be achieved.

5.3.3. *Echo Reduction and Segmental SNR*

Tables 1 and 2 show the echo reduction and the segmental SNR, respectively. In this table, 'Linear+Non-linear 30%' means the mean square of the non-linear echo path impulse response is 30% of the linear echo path impulse response as described in Sec5.1. The parameters v and α are optimized so as to maximize the segmental SNR in both methods.

Table 1. Echo reduction [dB].

Echo components	Convgentional $v = 3$	Proposed $v = 5, \alpha = 0.8$
Linear	-36.81	-38.05
Linear+Non-linear 5%	-15.33	-28.67
Linear+Non-linear 10%	-12.60	-26.13
Linear+Non-linear 20%	-10.01	-23.47
Linear+Non-linear 30%	-8.62	-21.98
Linear+Non-linear 40%	-7.71	-20.82
Linear+Non-linear 50%	-7.07	-20.27

In the conventional method, v cannot be enlarged due to the near-end signal distortion. v is set to 3 in order to maximize the segmental SNR. In the proposed method, v can be

Table 2. Segmental SNR [dB].

Echo components	Conventional $v = 3$	Proposed $v = 5, \alpha = 0.8$
Linear	12.70	17.25
Linear+Non-linear 5%	9.58	12.18
Linear+Non-linear 10%	8.38	10.53
Linear+Non-linear 20%	6.98	8.82
Linear+Non-linear 30%	6.04	8.05
Linear+Non-linear 40%	5.35	7.02
Linear+Non-linear 50%	4.86	6.26

enlarged, that is $v = 5$, at the same time, the segmental SNR is maximized. Like this, in the proposed method, the echo can be drastically reduced, while maintaining higher segmental SNR compared with the conventional method. The echo reduction is improved by 13 dB, and the segmental SNR is improved by $1.5 \sim 2.5$dB.

Simulations have been carried out by using four kinds of speakers, including male and female. Even though the optimum parameters are slightly different, the improvements from the conventional method are almost the same.

5.3.4. Experimental Results for Real Mobile Phones

The experimental results are shown in Tabel 3. The optimum parameters are different from the previous simulations, because the echo paths are different. In practical usage, it is no problem to optimally tune the parameters for real mobile phones and real usage circumstances. The proposed method can improve the echo reduction and the segmental SNR by 5.6dB and 2dB, respectively.

Table 3. Echo reduction and segmental SNR for real mobile phones.

Echo components	Conventional $v = 3$	Proposed $v = 3, \alpha = 0.9$
Echo Reduction [dB]	-12.94	-18.57
Segmental SNR [dB]	6.87	8.86

6. CONCLUSIONS

In this paper, we propose a new adaptive control method for the regression coefficient, which is used in the echo canceller based on the correlation model of the non-linear echo. The regression coefficient can express the relation between the residual non-linear echo and the linear echo replica. Simulation results shows the echo reduction and the segmental SNR can be improved by 13dB and $1.5 \sim 2.5$dB.

7. REFERENCES

[1] Y.Ephraim and D.Malah, "Speech enhancement using minimum mean-square error short-time spectral amplitude estimator", IEEE Trans. vol.ASSP-32, no.6, pp.1109-1121, Dec. 1984.

[2] Y. Wang, Y. Terada, M. Matsui, K. Iida and K. Nakayama, "Development of high quality acoustic subband echo canceller using dual-filter structure and fast recursive least squares algorithm", IEEE Proc. ICASSP2000, pp.VI-3674-677, 2000.

[3] A. Hirano, K. Nakayama and S. Ushimaru, "Double-talk resistant acoustic echo canceller with double filters," IEEE Proc. ISPACS2003, Awajishima, Japan, pp.367-370, Dec. 2003.

[4] A. Guerin, G. Faucon and R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellion based on Volterra filters", IEEE TRans. SAP, pp.672-683, 2003.

[5] K. Nakayama, A. Hirano and H. Kashimoto, "A lattice predictor based adaptive Volterra filter and a synchronized learning algorithm," EUSIPCO2004, Vienna, pp.1585-1588, Sept. 2004.

[6] K. Nakayama, A. Hirano and H. Kashimoto, "A synchronized learning algorithm for nonlinear part in a lattice predictor based adaptive Volterra filter," Proc. EUSIPCO2005, Antalya, Sept. 2005.

[7] O. Hoshuyama and A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo", IEEE Proc. ICASSP2006, pp.V269-272, 2006.

[8] O. Hoshuyama and A. Sugiyama, "Nonlinear acoustic echo suppressor based on spectral correlation between residual echo and echo replica," IEICE Trans. Fundamentals, Vol.E89-A, No.11, pp. 3254–3259, Nov. 2006.

[9] F. Kuech and W. Kellermann, "Nonlinear residual echo suppression using a power filter model of the acoustic echo path", IEEE Proc. ICASSP2007, pp.I73-76, 2007.

[10] O. Hoshuyama, "An echo canceller using smoothed coefficient filter with adaptive time constant controlled by high-pass errors", Proc. IWAENC2008, 2008.