

A Variable Step Size RPE Algorithm for Echo Cancellation

Yuan Hongxing

School of Electron and Information Engineering
Ningbo University of Technology
Ningbo, China, +86-57487081230
yuanhx@mail.ustc.edu.cn

Zhu Renxiang

School of Electron and Information Engineering
Ningbo University of Technology
Ningbo, China

Guo Li

Department of Electronic Science and Technology
University of Science and Technology of China
Hefei, China, +86-5513601955
lguo@ustc.edu.cn

Wu Shaoqun

School of Electron and Information Engineering
Ningbo University of Technology
Ningbo, China

Abstract—Acoustic echo cancellation (AEC) algorithms are widely used in communication devices. One of the main design problems is double-talk issue. The recursive prediction error (RPE) algorithm based on autoregressive (AR) model has been proven to be more attractive to traditional approaches. However, it adapts too fast in double-talk which may cause the algorithm to diverge. A variable step size (VSS) approach is developed to tackle the problem. On the assumption that prediction error is more than 10 times smaller than microphone signal when the adaptive filter has converged, the step size is adjusted adaptively according to the energies of near-end, far-end and prediction error signals. The experimental results show that our algorithm, compared with original RPE algorithm, the performance in double-talk is improved more than 8 dB while no any performance losses in single-talk and echo path change cases.

Keywords—acoustic echo, AR model, recursive prediction error, double-talk, step size, misalignment

I. INTRODUCTION

The acoustic echo is caused by feedback of the far-end voice through the near-end loudspeaker-microphone path, i.e. room impulse response (RIR). The communication quality is interfered by acoustic echo seriously. Acoustic echo cancellation (AEC) is one of the best solutions for this problem. Thus, AEC is an essential part in modern multimedia communication systems. To cancel acoustic echo in mobile communication, Terminal Coupling Loss weighted (TCLw) must reach 46 dB at least during single-talk and 26 dB at least during double-talk. AEC is based on adaptive filter techniques. An FIR filter is commonly used to estimate the RIR. The filter output, which estimates the acoustic echo, is subtracted from the near-end's microphone signal to cancel the echo. Ideal AEC should have fast convergence and tracking abilities on condition that misalignment is low. These two features are both dependent on the adaptive algorithms. The existing adaptive algorithms can get satisfactory results during single-talk but not the case in double-talk. Due to the interference of near-end voice, the adaptive algorithms will diverge and deviate from

the actual RIR during double-talk. According to the statistical data, double-talk happens at a time about 20%. So, AEC's performance during double-talk is very important. Common solution is to introduce a double-talk detector (DTD). When double-talk is declared by the DTD, the adaptive filter update will be stopped. The disadvantage of this solution is that the adaptive filter may have diverged before double-talk is checked. To solve this problem, a variety of VSS adaptive algorithms have been proposed. The principle of these algorithms is to adjust the adaptive filter update step according to characteristics of near-end and far-end voice so that the update step is large in single-talk and small in double-talk. Paleologu proposed the VSS affine projection algorithm (APA) and VSS normalized least mean squares algorithm (NLMS)^[1,2]. However, the performance of these approaches will be affected during single-talk and echo path change. The results of our preliminary study^[3] show that the adaptive algorithm based on recursive prediction error which proposed by Toon^[4] outperforms the literature [1, 2] and other adaptive step size control algorithms^[5-7]. Its idea is to model near-end voice by adaptive autoregressive model so as to accelerate the convergence rate. The advantage of this approach is the convergence rate will be improved by AR model. But the filter may diverge in double-talk due to fast convergence rate.

To resolve AEC problems in double-talk, a VSS AEC is proposed which inspired by Toon's approach^[4]. The update step is derived taken into account the near-end, far-end and prediction error signal statistical properties of energy.

II. PRINCIPLES OF RPE ALGORITHM FOR AEC

Assuming L-tap FIR filter $\mathbf{h} = [h(0), h(1), \dots, h(L-1)]^T$ can estimate RIR, near-end signal and far-end signal at sample index i are represented by $x(i)$ and $v(i)$ respectively. Microphone signal at sample index i is $y(i)$. Thus, following equations will be got,

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{v}, \quad (1)$$

This work was supported by the National Natural Science Foundation of China (Grant No. 61071173), Zhejiang Provincial Natural Science Foundation of China (Grant No. LY12F01001) and Ningbo Natural Science Foundation (Grant No. 2012A610043)

where,

$$\mathbf{y} = [y(i), y(i-1), \dots, y(i-M+1)]^T, \quad (2)$$

$$\mathbf{v} = [v(i), v(i-1), \dots, v(i-M+1)]^T, \quad (3)$$

$$\mathbf{X} = \begin{bmatrix} x(i) & x(i-1) & \dots & x(i-L+1) \\ x(i-1) & x(i-2) & \dots & x(i-L) \\ \vdots & \vdots & \ddots & \vdots \\ x(i-M+1) & x(i-M) & \dots & x(i-L-M+2) \end{bmatrix}, \quad (4)$$

With least squares, RIR \mathbf{h} estimation $\hat{\mathbf{h}}$ is defined as follows,

$$\hat{\mathbf{h}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (5)$$

To minimize estimation error variance, covariance matrix of near-end signal is introduced and we have,

$$\hat{\mathbf{h}} = (\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y}, \quad (6)$$

Above equation is hard to be directly applied to AEC as covariance matrix of near-end signal is unknown. In order to solve the above Eq., assume that near-end signal $v(i)$ is generated by a M order AR model $\mathbf{a} = [a(0), a(1), \dots, a(M-1)]^T$ which input is a white noise $b(i)$.

$$\mathbf{y}^T \mathbf{a} = d(i) + b(i), \quad (7)$$

where,

$$d(i) = \mathbf{x}^T \mathbf{h}, \quad (8)$$

$$\mathbf{x} = [x(i), x(i-1), \dots, x(i-L+1)]^T, \quad (9)$$

According to the best one-step-ahead predictor^[8], estimated microphone signal at sample index i can be got from Eq. (7),

$$\hat{y}(i) = y(i) - \mathbf{y}^T \mathbf{a} + \mathbf{d}^T \mathbf{a}, \quad (10)$$

where,

$$\mathbf{d} = [d(i), d(i-1), \dots, d(i-M+1)]^T, \quad (11)$$

Eq. (10) tells that estimated microphone signal $y(i)$ is dependent on RIR \mathbf{h} and AR model coefficients \mathbf{a} . Solution for \mathbf{h} and \mathbf{a} will be got by optimal estimation for $y(i)$ as follows,

$$\varepsilon(i) = \mathbf{y}^T \mathbf{a} - \mathbf{d}^T \mathbf{a}, \quad (12)$$

According to literature [9], estimation for \mathbf{h} and \mathbf{a} should be got by minimizing estimation error variance. Further, iterative solution is got according to literature [8] as follows,

$$\hat{\mathbf{h}}(i) = \hat{\mathbf{h}}(i-1) + \frac{1}{\sigma_e^2(i)} \mathbf{R}_h^{-1}(i) \boldsymbol{\Psi}_h(i) \hat{\varepsilon}(i), \quad (13)$$

$$\hat{\mathbf{a}}(i) = \hat{\mathbf{a}}(i-1) + \frac{1}{\sigma_e^2(i)} \mathbf{R}_a^{-1}(i) \boldsymbol{\Psi}_a(i) \hat{\varepsilon}(i), \quad (14)$$

where $\mathbf{R}_h(i)$, $\mathbf{R}_a(i)$ represents estimation for the Hessian matrix $\frac{1}{2} \frac{\partial E\{\varepsilon^2(i)\}}{\partial \mathbf{h}^2}$ and $\frac{1}{2} \frac{\partial E\{\varepsilon^2(i)\}}{\partial \mathbf{a}^2}$ respectively. They define the searching direction for fir filter update. $\boldsymbol{\Psi}_h(i)$ and $\boldsymbol{\Psi}_a(i)$ is gradient vector of \mathbf{h} and \mathbf{a} respectively. $\sigma_e^2(i)$ is estimation energy for prediction error signal at sample index i . Substitute \mathbf{h} in Eq. (8) with $\hat{\mathbf{h}}$ and substitute \mathbf{a} in Eq. (12) with $\hat{\mathbf{a}}$, $\hat{\varepsilon}(i)$ will be got. Prediction error signal energy, gradient vector and Hessian matrix are estimated as follows,

$$\sigma_e^2(i) = \lambda_a \sigma_e^2(i-1) + (1 - \lambda_a) \hat{\varepsilon}^2(i), \quad (15)$$

$$\boldsymbol{\Psi}_h(i) = \mathbf{x} + \mathbf{X}^T \mathbf{a}, \quad (16)$$

$$\boldsymbol{\Psi}_a(i) = \mathbf{X} \hat{\mathbf{h}} - \mathbf{y}, \quad (17)$$

$$\mathbf{R}_h(i) = (1 - \lambda_h) \mathbf{R}_h(i-1) + \frac{1}{\sigma_e^2(i)} \boldsymbol{\Psi}_h(i) \boldsymbol{\Psi}_h^T(i), \quad (18)$$

$$\mathbf{R}_a(i) = (1 - \lambda_a) \mathbf{R}_a(i-1) + \frac{1}{\sigma_e^2(i)} \boldsymbol{\Psi}_a(i) \boldsymbol{\Psi}_a^T(i), \quad (19)$$

The inverse of Hessian matrix \mathbf{R}_h and \mathbf{R}_a have to be computed when using Eq. (13) and (14) to estimate \mathbf{h} and \mathbf{a} respectively. The results of our preliminary study^[3] show that 1-order AR model is superior to high order AR model in estimating near-end signal. So 1-order AR model is adopted in this paper. Thus, \mathbf{R}_a becomes a scalar and inverse operation in Eq. (14) is just a simple division. To avoid inverse operation in Eq. (13), \mathbf{h} is estimated using stochastic gradient method as follows,

$$\hat{\mathbf{h}}(i) = \hat{\mathbf{h}}(i-1) + \mu \frac{\boldsymbol{\Psi}_h(i) \hat{\varepsilon}(i)}{L \sigma_e^2(i) + \boldsymbol{\Psi}_h^T(i) \boldsymbol{\Psi}_h(i) + \Delta}, \quad (20)$$

where Δ is regular factor to prevent dividing by zero. Once using Eq. (20) to estimate \mathbf{h} , the Hessian matrix \mathbf{R}_h and its inverse are no more need.

III. VARIABLE STEP SIZE RPE ALGORITHM

Using a white noise as excitation to estimate near-end signal, the above approach reduces correlation between near-end signal and far-end signal so as to speed up convergence rate. However, too fast convergence rate in double-talk may cause the fir filter to diverge. To solve the issue, a VSS approach is proposed as showed in the Fig.1.

In VSS approach, μ in Eq. (20) will be adjusted according to communication scenarios, and it is defined as follows,

$$\mu = \begin{cases} \text{step}_1, & \text{if } \sigma_x^2(i) > \sigma_1^2(i) \\ \text{step}_2, & \text{otherwise} \end{cases}, \quad (21)$$

where $\sigma_x^2(i)$ is far-end signal energy estimation at sample index i and $\sigma_1^2(i)$ is a temporary variable considering prediction error and near-end signal,

$$\sigma_x^2(i) = \lambda_x \sigma_x^2(i-1) + (1-\lambda_x) x^2(i), \quad (22)$$

$$\sigma_1^2(i) = \begin{cases} w_2 \sigma_e^2(i), & \text{if } \sigma_y^2(i) > \frac{w_2}{w_1} \sigma_e^2(i) \\ w_1 \sigma_y^2(i), & \text{otherwise} \end{cases}, \quad (23)$$

$\sigma_y^2(i)$ is energy estimation for microphone signal,

$$\sigma_y^2(i) = \lambda_y \sigma_y^2(i-1) + (1-\lambda_y) y^2(i), \quad (24)$$

In the experiments, step_1 and step_2 are set to 1 and 0.01 respectively which means step size will be reduced by 100 times during double-talk. In fact, step_2 can be set to any small value which will not affect its performance. The step size is reduced once $\sigma_x^2(i) \leq \sigma_1^2(i)$ which means double-talk happens. $\sigma_1^2(i)$ reflects energy changes of microphone and prediction error signals. w_1 and w_2 in Eq. (23) are set to 5 and 50 respectively in all experiments. The choice of w_1 and w_2 makes sure that prediction error for microphone signal is 10 times smaller at least. $\sigma_1^2(i)$ is scaled up by 5 to improve double-talk checking sensitivity. In single-talk and echo path change scenarios, $\sigma_1^2(i)$ will be smaller than far-end signal energy $\sigma_x^2(i)$ and the step size will not be changed which preserves the original algorithm's fast tracking abilities. In double-talk, $\sigma_y^2(i)$ will raise up and quick adaptation will make the filter apart from the true echo path which let $\sigma_e^2(i)$ becomes larger and larger. The two will make $\sigma_1^2(i)$ becomes larger than $\sigma_x^2(i)$ and the second case in Eq. (21) happens.

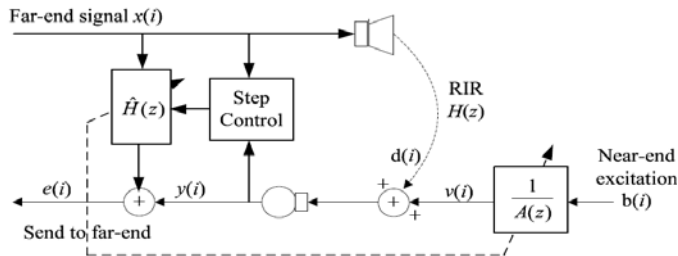


Figure 1. VSS RPE algorithm for AEC

IV. EXPERIMENTAL RESULTS AND COMPARISON

Three simulation experiments will be performed. For simulation case, the RIR is known, so the performance is evaluated in terms of misalignment (in dB). Misalignment reflects how the adaptive FIR filter impulse response approximates the real echo path. The lower the misalignment is, the better the performance of adaptive filter is. The misalignment is defined as follows^[10],

$$\text{misalign}(i) = 10 \log_{10} \left\{ \frac{\|\mathbf{h} - \hat{\mathbf{h}}(i)\|^2}{\|\mathbf{h}\|^2} \right\}, \quad (25)$$

The order of the AR model is set to 1 and the length of the adaptive filter is set to 256 coefficients in all experiments.

Two measured RIR are plotted in Fig. 2 and Fig. 3 respectively (the sampling rate is 8 kHz); their entire length has 256 coefficients.

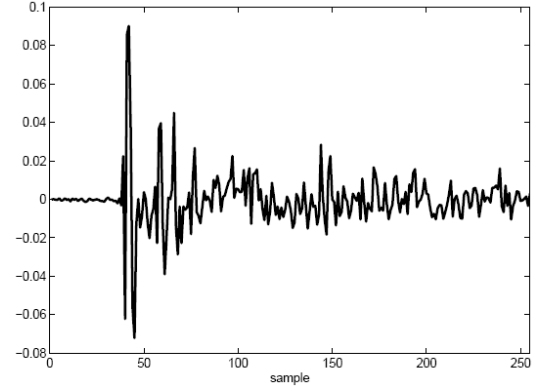


Figure 2. Measured RIR 1

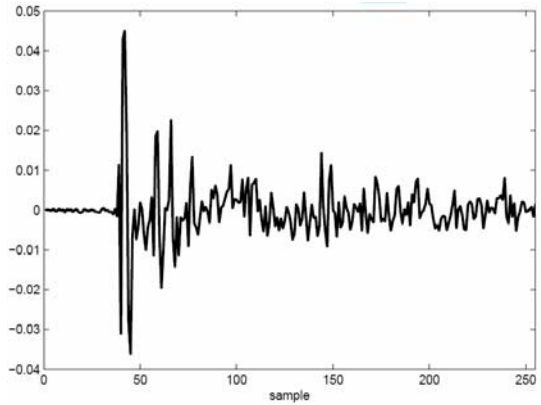


Figure 3. Measured RIR 2

A. Single-talk scenario

Convolve far-end signal with RIR showed in Fig. 2 and add 60 dB gaussian white noise to get microphone's speech signal in single-talk. The misalignment curves for our approach and Toon's approach are plotted in Fig. 4. It can be noticed that there is no any performance loss thanks for the carefully choice of step size in Eq. (21).

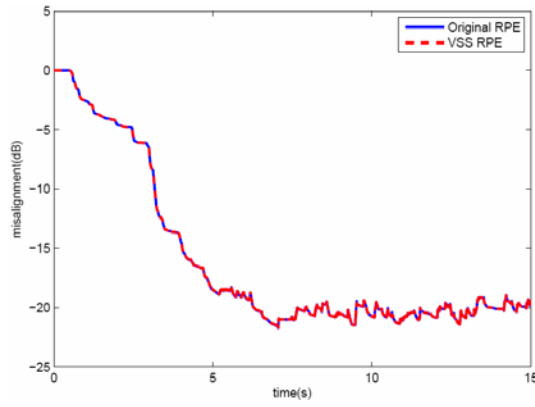


Figure 4. Misalignments in single-talk

B. Double-talk scenario

Add a speech to near-end signal in single-talk case between 8s and 9s to generate microphone's speech signal in double-talk. The misalignment curves are plotted in Fig. 5. It can be noticed that compared with Toon's approach, the misalignment of the proposed algorithm achieves at least 8 dB improvement thanks for the right identification of double-talk.

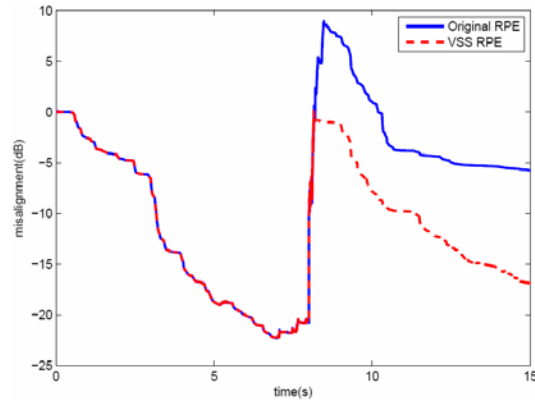


Figure 5. Misalignments in double-talk

C. Echo path change scenario

Convolve the first half of far-end signal with RIR 1 and the last half with RIR 2, then add a 60 dB gaussian white noise to generate microphone's speech signal in echo path change. The misalignment curves are plotted in Fig. 6. It can be noticed that compared with Toon's approach, there is no any performance loss which meets our desires.

V. CONCLUSION

A variable step size for recursive prediction error algorithm in acoustic echo canceller is proposed. The filter update step size will be adjusted adaptively according to near-end, far-end signal and prediction error energies. Simulation experimental results show the performance is improved at least 8 dB in

double-talk while no performance loss both in single-talk and echo path change scenarios.

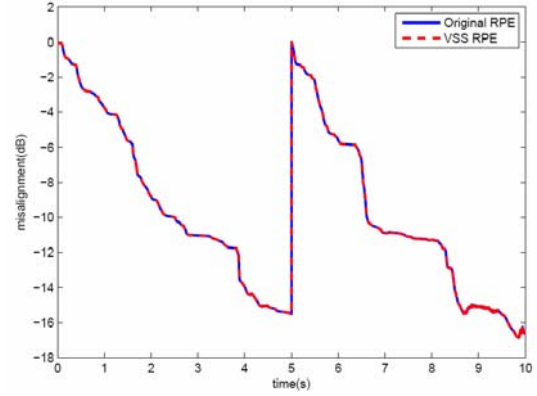


Figure 6. Misalignments in echo path change

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to Dr. Toon for many fruitful discussions and providing their RPE source code.

REFERENCES

- [1] C. Paleologu, J. Benesty, and S. Ciochina, "A Variable Step-Size Affine Projection Algorithm Designed for Acoustic Echo Cancellation," *IEEE Transactions on Audio Speech and Language Processing*, vol. 16, pp. 1466-1478, October, 2008.
- [2] C. Paleologu, S. Ciochina, and J. Benesty, "Double-talk robust vss-nlms algorithm for under-modeling acoustic echo cancellation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 245-248, 2008.
- [3] H. X. Yuan, S. Q. Wu, and C. J. Zha, "Simplified RPE Algorithm and its Fixed-Point Implementation," *Fourth International Conference on Intelligent Computation Technology and Automation*, pp. 699-701, 2011.
- [4] W. V. Toon, R. Geert, and V. Piet, "Double-talk robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Transactions on Signal Processing*, vol. 55, pp. 846-858, March, 2007.
- [5] B. Stark, "Acoustic echo cancellation: dual architecture implementation," *Journal of Computer Science and Technology*, vol. 6, pp. 101-106, February, 2010.
- [6] R. Oka, K. Fujii, and M. Muneyasu, "A step size control method steadily reducing acoustic echo even during double-talk," *International Symposium on Intelligent Signal Processing and Communications Systems*, pp. 1-4, 2009.
- [7] K. Fujii, N. Saitoh, and R. Oka, "Acoustic echo cancellation algorithm tolerable for double talk," *International Conference on Hands-Free Speech Communication and Microphone Arrays*, pp. 200-203, 2008.
- [8] L. Ljung, "System Identification: Theory for the User. Englewood Cliffs," New Jersey: Prentice-Hall, 1987.
- [9] L. Ljung, and T. Soderstrom, "Theory and practice of recursive identification," Cambridge, Massachusetts: MIT Press, 1986.
- [10] H. Li, M. M. Lin, and H. J. Hu, "A VoIP Echo Canceler Design and Algorithm Study," *Chinese Journal of Electronics*, vol.35, pp. 1774-1778, September, 2007.