Fast Converging Nonlinear Echo Cancellation Based On Optimum Step Size

Chia-Sheng Shih and Shih-Fu Hsieh*

Department of Electrical Engineering
National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China
Tel: 886-3-5731974, E-mail: joinson.cm95g@nctu.edu.tw, sfhsieh@mail.nctu.edu.tw

Abstract *Adaptive Volterra filter is generally known to track nonlinear echo path at slow convergence rate. An optimum time- and tap– variant step-size for second-order Volterra filter is proposed to speed up its convergence rate. The step-size is derived based on the MMSE criterion. As the optimum step-size needs to know the real echo path coefficient, an exponential echo-path model is proposed for practical implementations. Computer simulations are provided to validate the proposed algorithms.*

Keywords *- Adaptive filtering, Nonlinear Identification*

1. INTRODUCTION

In these years, hands-free system telephone and teleconference systems are widely used. However, these systems usually suffer from the annoying acoustic echo problem because the far end speech is transmitted back to the microphone at the near end. Many adaptive algorithms that have been proposed [1] for echo cancellation. The least-mean-square (LMS) algorithm is famous for its low computational cost in linear acoustic echo cancellation (AEC). However, linear AEC is not sufficient to estimate the acoustic echo path for many hand-held mobile phones due to their limited physical size constraint and high-volume loudspeaker/amplifier requirements [2]. To overcome this nonlinear acoustic echo problem, there have been many popular methods such as polynomial functions, Hammerstein model [3], Volterra filter [4]-[5], and so on. In this paper, we use the second-order Volterra filter [6] to model the nonlinear echo path. Although these nonlinear echo models have less modeling error, they require to estimate a significant amount of parameters unfortunately. As a result, the convergence rate of the nonlinear adaptive filters becomes slow because the number of nonlinear parameters is much more than that required in linear AEC.

To overcome its slow convergence rate, we propose an optimum time- and tap– variant step-size control for the second-order Volterra filter by introducing an optimum MMSE criterion between coefficients errors of real echo path kernel and adaptive coefficients [6]. However, this ideally optimum step size is impractical because it needs to know the true echo path which remains unknown to us. For practical implementation, we turn to exploit the recursive relation of coefficient error variance, and propose an exponentially tapped model function to model the real linear echo path. Computer simulations will validate our proposed algorithms.

2. SYSTEM MODEL

The general second-order Volterra model is summarized in Fig 1; the microphone signal $y(k)$ is composed of the echo signal $y'(k)$, the noise signal $n(k)$, and the speech signal $s(k)$ of the near-end talker. The output of the Volterra filter is given by:

$$z(k) = z^{(1)}(k) + z^{(2)}(k) \quad (1)$$

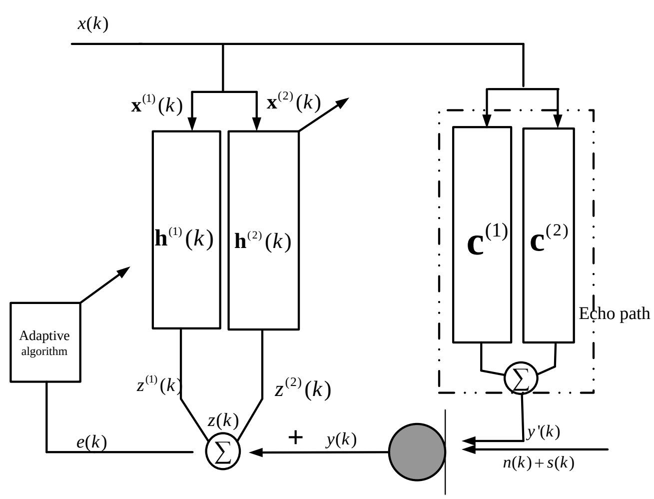


Fig 1: Second-order Volterra filter

where $z^{(1)}(k)$ and $z^{(2)}(k)$ can be formulated as

$$z^{(1)}(k) = \sum_{l=0}^{M-1} h_l^{(1)}(k)x(k-l) \quad (2)$$

$$z^{(2)}(k) = \sum_{l_1=0}^{N_2-1} \sum_{l_2=l_1}^{N_2-1} h_{l_1,l_2}^{(2)}(k)x(k-l_1)x(k-l_2) \quad (3)$$

The residual error signal is defined as

$$e(k) = y(k) - \mathbf{h}^{(1)T}(k)\mathbf{x}^{(1)}(k) - \mathbf{h}^{(2)T}(k)\mathbf{x}^{(2)}(k) \quad (4)$$

In vector representation, we define:

$$\mathbf{x}^{(1)}(k) = [x_1^{(1)}(k), x_2^{(1)}(k), ...x_M^{(1)}(k)]^T$$
$$= [x(k), x(k-1), ...x(k-M+1)]^T \quad (5)$$
$$\mathbf{x}^{(2)}(k) = [x_1^{(2)}(k), x_2^{(2)}(k), ...x_{L_2}^{(2)}(k)]^T$$
$$= [x^2(k), x(k)x(k-1), ...x^2(k-N_2+1)]^T \quad (6)$$
$$\mathbf{h}^{(1)}(k) = [h_1^{(1)}(k), h_2^{(1)}(k), ...h_M^{(1)}(k)]^T \quad (7)$$
$$\mathbf{h}^{(2)}(k) = [h_1^{(1)}(k), h_2^{(1)}(k), ...h_{L_2}^{(1)}(k)]^T$$
$$= [h_{0,0}^{(2)}(k), h_{0,1}^{(2)}(k), ...h_{N_2-1,N_2-1}^{(2)}(k)]^T \quad (8)$$

where M and N_2 represent memory lengths of linear and quadratic kernel, and the lengths of $\mathbf{x}^{(2)}(k)$ and $\mathbf{h}^{(2)}(k)$ are both equal to $L_2 = N_2(N_2+1)/2$.

3. OPTIMUM STEP SIZE

We want to find out the step size at time k which can minimize each tap coefficient error variance at time $k+1$, i.e. MSE for each iteration step. Hence, we use diagonal matrices to replace the step size of conventional LMS algorithm [1], thus the corresponding LMS algorithm can be rewritten as

$$\mathbf{h}^{(1)}(k+1) = \mathbf{h}^{(1)}(k) + \mathbf{U}^{(1)}(k)e(k)\mathbf{x}^{(1)}(k) \quad (9)$$
$$\mathbf{h}^{(2)}(k+1) = \mathbf{h}^{(2)}(k) + \mathbf{U}^{(2)}(k)e(k)\mathbf{x}^{(2)}(k) \quad (10)$$

where $\mathbf{U}^{(1)}(k)$ and $\mathbf{U}^{(2)}(k)$ denote linear and quadratic step size matrices of our interest:

$$\mathbf{U}^{(1)}(k) = \begin{bmatrix} \mu_1^{(1)}(k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_M^{(1)}(k) \end{bmatrix},$$
$$\mathbf{U}^{(2)}(k) = \begin{bmatrix} \mu_1^{(2)}(k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{L_2}^{(2)}(k) \end{bmatrix}$$

where the l^{th} element of the step size matrices is chosen to minimize l^{th} coefficient error variance at time $k+1$. The criterion is summarized as

$$\mu_l^{(i)}(k) = \underset{\mu_l^{(i)}(k)}{\arg\min} E\left\{[h_l^{(i)}(k+1) - c_l^{(i)}]^2\right\}$$

where $i \in 1,2$, and the coefficient error is defined as

$$\mathbf{v}^{(i)}(k) = \mathbf{h}^{(i)}(k) - \mathbf{c}^{(i)} \quad (11)$$

Using (9), (11), and (4), we may rewrite the linear kernel coefficient error $\mathbf{v}^{(1)}(k)$ as

$$\mathbf{v}^{(1)}(k+1) = \mathbf{v}^{(1)}(k) - \mathbf{U}^{(1)}\mathbf{v}^T(k)\mathbf{x}(k)\mathbf{x}^{(1)}(k) + \mathbf{U}^{(1)}\left[n(k)+s(k)\right]\mathbf{x}^{(1)}(k)$$
$$= [\mathbf{I} - \mathbf{U}^{(1)}(k)\mathbf{x}^{(1)}(k)\mathbf{x}^{(1)^T}(k)]\mathbf{v}^{(1)}(k) - \mathbf{U}^{(1)}\mathbf{x}^{(1)}(k)\mathbf{x}^{(2)^T}(k)\mathbf{v}^{(2)}(k)$$
$$+ \mathbf{U}^{(1)}n(k)\mathbf{x}^{(1)}(k) + \mathbf{U}^{(1)}s(k)\mathbf{x}^{(1)}(k) \quad (12)$$

where the cascaded representations of input and coefficient error are denoted as

$$\mathbf{x}(k) = [\mathbf{x}^{(1)T}(k) \quad \mathbf{x}^{(2)T}(k)]^T \quad (13)$$
$$\mathbf{v}(k) = [\mathbf{v}^{(1)T}(k) \quad \mathbf{v}^{(2)T}(k)]^T \quad (14)$$

We can derive the autocorrelation matrix of the linear kernel coefficient errors by the direct-average method [1]:

$$\mathbf{R}_{\mathbf{v}^{(1)}}(k+1) \approx [\mathbf{I} - 2\mathbf{U}^{(1)}(k)\mathbf{R}_{\mathbf{x}^{(1)}}(\mathbf{k})]\mathbf{R}_{\mathbf{v}^{(1)}}(k)$$
$$+[\mathbf{U}^{(1)}(k)]E\{\mathbf{x}^{(1)}(k)\mathbf{x}^{(1)T}(k)\mathbf{v}^{(1)}(k)$$
$$\times \mathbf{v}^{(1)T}(k)\mathbf{x}^{(1)}(k)\mathbf{x}^{(1)T}(k)\}[\mathbf{U}^{(1)}(k)]^T$$
$$+[\mathbf{U}^{(1)}(k)]E\{\mathbf{x}^{(1)}(k)\mathbf{x}^{(2)T}(k)\mathbf{v}^{(2)}(k)$$
$$\times \mathbf{v}^{(2)T}(k)\mathbf{x}^{(2)}(k)\mathbf{x}^{(1)T}(k)\}[\mathbf{U}^{(1)}(k)]^T$$
$$+[\mathbf{U}^{(1)}(k)]\sigma_n^2\mathbf{R}_{\mathbf{x}^{(1)}}(k) + \sigma_s^2\mathbf{R}_{\mathbf{x}^{(1)}}(k)[\mathbf{U}^{(1)}(k)]^T \quad (15)$$

In Eq. (15), $E\{\ .\ \}$ denotes expectation operator. By the assumption of mutual independence among $x(k)$, $n(k)$ and, $s(k)$, and the assumption that the probability density function of $x(k)$ is an even function, we have

$$E\left\{x^3(k)\right\} = 0.$$

Thus the cross products terms $[\mathbf{I} - \mathbf{U}^{(1)}(k)\mathbf{x}^{(1)}(k)\mathbf{x}^{(1)T}(k)]\mathbf{v}^{(1)}(k)$, $\mathbf{U}^{(1)}\mathbf{x}^{(1)}(k)\mathbf{x}^{(2)T}(k)\mathbf{v}^{(2)}(k)$, $\mathbf{U}^{(1)}n(k)\mathbf{x}^{(1)}(k)$, and $\mathbf{U}^{(1)}s(k)\mathbf{x}^{(1)}(k)$ in (12) can be neglected.

The l^{th} diagonal term of the autocorrelation matrix, denoting l^{th} mean-square of linear coefficient error, can be written as:

$$g_l^{(1)}(k+1) \approx (1 - 2\mu_l^{(1)}(k)\sigma_x^2)g_l^{(1)}(k)$$
$$+\mu_l^{(1)^2}(k)\sigma_x^2\left[\sigma_x^2\sum_{l=1}^{M} g_l^{(1)}(k) + \sigma_x^4\sum_{j=0}^{L_2} g_j^{(2)}(k) + \sigma_n^2 + \sigma_s^2\right] \quad (16)$$

The optimum time-& tap variant step-size can be obtained by taking derivative of (16) with respect to $\mu_l^{(1)}(k)$ and setting the result equal to zero. Thus we can get the Optimum Time-&Tap-variant step-size of the linear kernel based on the LMS algorithm (OTTLMS):

$$\mu_{l,OTTLMS}^{(1)}(k) = \frac{g_l^{(1)}(k)}{\sigma_x^2\sum_{l=1}^{M} g_l^{(1)}(k) + \sigma_n^2 + \sigma_s^2 + \sigma_x^4\sum_{l=1}^{L_2} g_j^{(2)}(k)} \quad (17)$$

Similar to the linear kernel, the optimal step size of the quadratic kernel is given by

$$\mu_{j,OTTLMS}^{(2)}(k)=\frac{g_j^{(2)}(k)}{\sigma_x^4\sum_{j=1}^{L_2} g_j^{(2)}(k)+\sigma_n^2+\sigma_s^2+\sigma_x^2\sum_{l=1}^{M} g_l^{(1)}(k)} \quad (18)$$

From the result of (17) and (18), we can see that the optimum step sizes are directly proportional to the coefficient error variance. If the coefficient error variance is large (for example, at the initial state), the optimum step sizes are large; and if the coefficient error variance is small, the optimum steps become small. The result agrees with our intuition.

The numerators in (17) and (18) account for the second moment coefficient error of linear and quadratic kernel, respectively, and the denominator for the summation of residual error power and near-end speech power. Thus our work in (17) and (18) agree with that in [6].

The above discussions are based on an LMS algorithm. However, when the input is large, the LMS algorithm suffers from a gradient noise amplification problem. In order to overcome this difficulty, we extend it to the normalized LMS (NLMS) algorithm. By the approximation of

$$\mathbf{x}^{(1)T}(k)\mathbf{x}^{(1)}(k)=M\sigma_x^2$$

and

$$\mathbf{x}^{(2)T}(k)\mathbf{x}^{(2)T}(k)=L_2\sigma_x^4 \text{ [1]},$$

the Optimum Time-&Tap-variant step-size of the linear kernel based on the Normalized LMS algorithm(OTTNLMS) can be shown to be:

$$\mu_{l,OTTNLMS}^{(1)}(k)=(M\sigma_x^2+L_2\sigma_x^4)\mu_{l,OTTLMS}^{(1)}(k) \quad (19)$$

$$\mu_{j,OTTNLMS}^{(2)}(k)=(M\sigma_x^2+L_2\sigma_x^4)\mu_{j,OTTLMS}^{(2)}(k) \quad (20)$$

Earlier, we have already derived optimum time- and tap- variant step-size for LMS and NLMS algorithms. Here, OTTLMS and OTTNLMS not only need prior statistics knowledge of the signal and noise powers, σ_x^2 , σ_x^4 , σ_s^2 and σ_n^2 , but also the prior knowledge of second moment of coefficient error $g_l^{(1)}(k)$ and $g_j^{(2)}(k)$. Thus we must know the real room impulse response $\mathbf{c}^{(1)}$ and second-order kernel caused by nonlinear loudspeaker $\mathbf{c}^{(2)}$, which are not accessible to us in general.

Now, unlike the approximation approach done by Kuech [6], we introduce the recursive formula to compute the coefficient variance. By doing so, we only need knowledge of coefficient variance at the initial time.

We substitute the optimum time-&tap-variant step size of the linear kernel (17) back to (16), so that we can get the recursive mean-square coefficient errors:

$$g_l^{(1)}(k+1)=(1-\mu_{l,OTTLMS}^{(1)}(k)\sigma_x^2)g_l^{(1)}(k) \quad (21)$$

$$g_j^{(2)}(k+1)=(1-\mu_{j,OTTLMS}^{(2)}(k)\sigma_x^4)g_j^{(2)}(k) \quad (22)$$

Similarly, (21) and (22) can be rewritten as:

$$g_l^{(1)}(k+1)=(1-\frac{\mu_{l,OTTLMS}^{(1)}(k)\sigma_x^2}{M\sigma_x^2+L_2\sigma_x^4})g_l^{(1)}(k) \quad (23)$$

$$g_j^{(2)}(k+1)=(1-\frac{\mu_{j,OTTLMS}^{(2)}(k)\sigma_x^4}{M\sigma_x^2+L_2\sigma_x^4})g_j^{(2)}(k) \quad (24)$$

for $l=1,....,M$, $j=1,....,L_2$.

By (21-24), we can see that we only need to know the envelope of real echo path at initial time, i.e.

$$g_l^{(1)}(0)=E\{\left[h_l^{(1)}(0)-c_l^{(1)}\right]^2\}=\left[c_l^{(1)}\right]^2,$$

thus we propose a model function to estimate those parameters for the application in nonlinear acoustics echo cancellation. Here, we will assume the real linear $\mathbf{c}^{(1)}$ and quadratic $\mathbf{c}^{(2)}$ kernel can be modeled reasonably as an exponentially decaying envelope as shown in Fig 2. Let the linear and quadratic envelope functions be modeled as:

$$w_l^{(1)}=w_0^{(1)}(r^{(1)})^l \quad \text{for } l=1\sim M \quad (25)$$

$$w_{l_1,l_2}^{(2)}=w_0^{(2)}(r^{(2)})^{(l_1+l_2)} \quad \text{for } l_1,l_2=1\sim N_2 \quad (26)$$

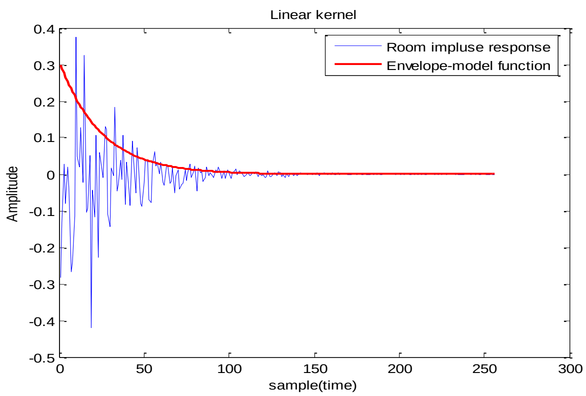
where $r^{(1)}$ and $r^{(2)}$ are the exponential decay factors of the linear and quadratic kernels, respectively. We let the initial linear and quadratic tap coefficients to be zero. i.e. $h_l^{(1)}(0)=0$ and $h_j^{(2)}(0)=0$ so that

$$g_l^{(1)}(0)=\left[c_l^{(1)}\right]^2\approx\left[w_l^{(1)}\right]^2$$

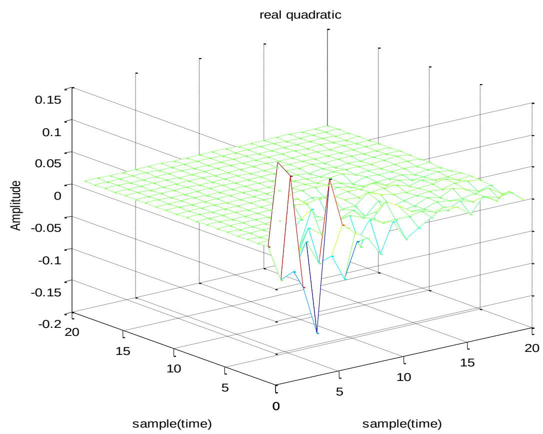
and

$$g_j^{(2)}(0)=\left[c_j^{(2)}\right]^2\approx\left[w_{l_1,l_2}^{(2)}\right]^2.$$

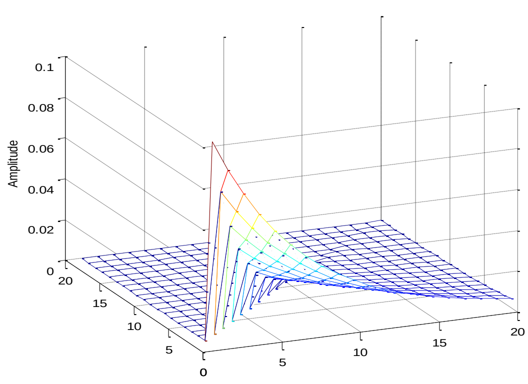
By (18) and (19), if we have $g_l^{(1)}(0)$ and $g_j^{(2)}(0)$, we can get the initial step-sizes of linear $\mu_{l,OTTLMS}^{(1)}(0)$ and quadratic kernel filter $\mu_{j,OTTLMS}^{(2)}(0)$. With these initial step-sizes plugged into (21-22) we can get $g_l^{(1)}(1)$ and $g_j^{(2)}(1)$, and so forth . Thus, we can compute $\mu_{l,OTTLMS}^{(1)}(k)$ and $\mu_{j,OTTLMS}^{(2)}(k)$, recursively.



(a)



(b)



(c)

Fig 2: Real kernel and exponentially model

The determination of $\mu^{(1)}_{l,OTTLMS}(k)$ and $\mu^{(2)}_{j,OTTLMS}(k)$, as statistical knowledge of the near-end σ_s^2 is not accessible, it is intuitive that the residual error variance $\sigma_e^2(k)$ is near to $\sigma_n^2+\sigma_s^2$ in converged condition. Thus we introduce the estimated residual error variance $\hat{\sigma}_e^2(k)$ to model the background noise and the near end speech variance by using the smoothed recursive algorithm from squared residual error as:

$$\hat{\sigma}_e^2(k)=\lambda\hat{\sigma}_e^2(k-1)+(1-\lambda)e^2(k) \qquad (27)$$

where λ is a smoothing constant.

Next, let us consider the echo path change condition. When the echo path changes, our proposed optimum step size is not robust. The reason comes from the recursive characteristic of (21-22), whether echo path changes or not, the mean-square coefficient errors of linear and quadratic kernel become smaller and smaller, and our proposed optimum step sizes become smaller and smaller, accordingly. Even if the echo path changes at some times, the step sizes are still very small at convergence.

Thus we may introduce a detector [7] to detect the echo path change. When the echo path change is detected, we re-initialize from $g_l^{(1)}(k)$ back to its initial values $g_l^{(1)}(0)$, so that our algorithm can still work in case of an echo path change.

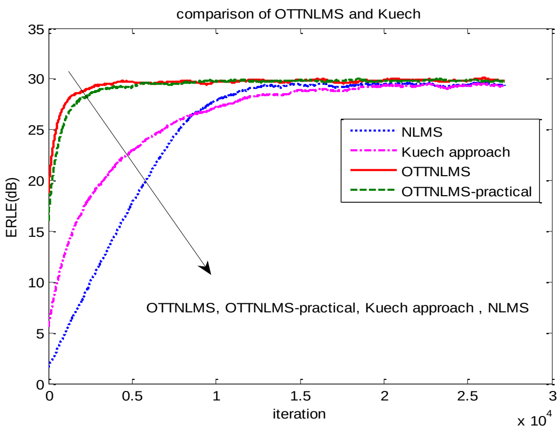
4. SIMULATION

To evaluate the performance of our proposed nonlinear AEC algorithm, computer simulations are performed. We use a white Gaussian signal as the far end input. In Fig 3, the orders of real linear and quadratic echo paths are set to 256 and 20, respectively.

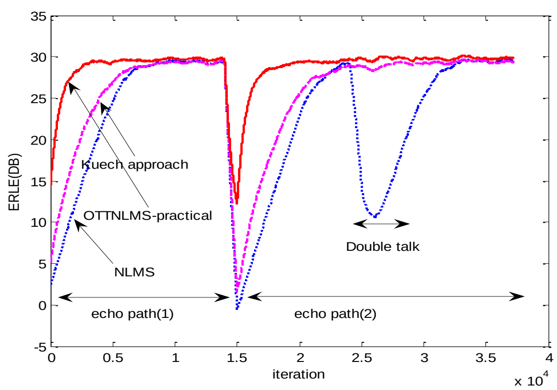
In Fig 3(a), we compare OTTNLMS, its practical implementation, Kuech approach, and NLMS. In Fig 3(b), the echo path changes at sample time 16000 and the near end talk signal exists from 25000 to 27000 sample time. For a fair comparison, parameters are chosen so that the steady-state ERLEs are equal for different algorithms. In Fig 4, the far end input is a real speech signal. These figures justify the effectiveness of our proposed algorithms.

5. CONCLUISON

In this paper, we propose an optimum time- and tap– variant step-size for Volterra filter in order to speed up convergence rate. As the optimum step-size needs to know the real echo path coefficient, we propose an exponential model for practical implementations. In case of echo path change and double talk, we incorporate an estimated model and echo path detector into our optimum step size model. The proposed optimum step size has better performance than conventional step size control of nonlinear acoustic echo canceller as shown from computer simulations.



(a)



(b)

Fig 3: Comparison of OTTNLMS and Kuech approaches

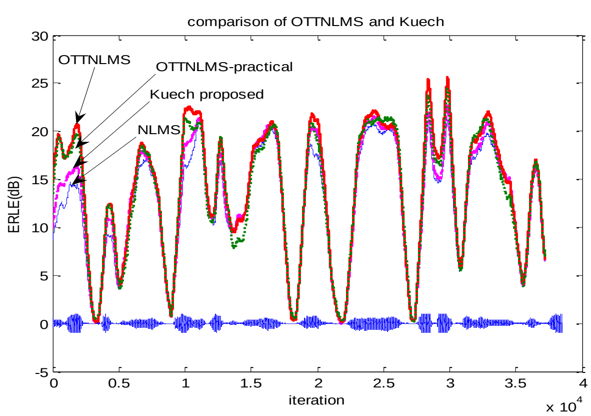


Fig 4: Comparison of OTTNLMS and Kuech approach in real speech data

ACKNOWLEDGEMENT

This work was supported by the NSC and MediaTek research center at National Chiao Tung University.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4^{th} ed, Prentice Hall, 2002.

[2] A. N. Birkett and R. A. Goubran , "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects ," *ICASSP*, pp.103-106,Oct.1995.

[3] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Automatic Control*, vol. 11, issue 13, pp. 546-550. Jul. 1966.

[4] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2^{nd} order adaptive Volterra filters," *ICASSP proceeding*, vol. 2, pp. 877–880, Nov. 1999.

[5] A. Guerin, G. Faucon, and Le Bouquin-Jeannes, R., "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech and Audio Processing* vol. 11, no. 6, pp. 672-683, Nov. 2003.

[6] F. Kuech and W. Kellermann, "Coefficient-dependent step-size for adaptive second-order Volterra filters," *EUSIPCO*, Vienna, September. 2004.

[7] M. A. Iqbal and S. L. Grant, "A novel normalized cross-correlation based on echo-path change detector." in *Proceedings of IEEE Region 5 conference*, Arkansas,April 2007.