Applied Mechanics and Materials
ISSN: 1662-7482, Vols. 719-720, pp 982-987
doi:10.4028/www.scientific.net/AMM.719-720.982

Submitted: 2014-09-02
Accepted: 2014-10-16
Online: 2015-01-13

A VAD-based Variable Order Affine Projection Algorithm for Acoustic Echo Cancellation

F. R. Liu[1,a], Y. Zhou[2,b*], S. S. Yin[3,c]

[1,2]School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications Chongqing, China
[3]Dept. of Electronic and Information Engineering, Anhui University of Finance and Economics Bengbu, China
[a]frliu@hotmail.com, [b]zhouy@cqupt.edu.cn, [c]ssyin@eee.hku.hk

Keywords: AEC; VAD; APA

Abstract: This paper proposes a new variable order affine projection algorithm for acoustic echo cancellation. By exploiting an efficient voice activity detection technique, the proposed algorithm can distinguish the significant and insignificant input data periods. It can thus switch between the higher and lower algorithm orders in these two situations, respectively. Hence, input data reusing can enhance convergence with more input excitation and the computation cost can be saved when the input is relatively weak. Simulations results verify the effectiveness of the proposed algorithm.

Introduction

Acoustic echo cancellation (AEC) is the key solution to the echo problem widely encountered in hands-free speech communication. The principle of AEC is to use an adaptive filter to estimate the echo path of the loudspeaker-enclosure-microphone (LEM) and generate a replica of the echo so as to cancel it. The efficacy of AEC has been widely proved [1]. The echo path of the LEM can be usually modeled by a FIR filter with hundreds to thousands of coefficients, which depends on the room size and acoustic environment. The computational cost for AEC in big rooms is quite demanding as the adaptive filter will have to update thousands of coefficients at every iteration [2]. The normalized least mean square (NLMS) algorithm [3] consumes a low computational complexity of $O(L)$, where L denotes the adaptive filter order. However, its convergence rate is unsatisfactorily slow. Another well-known algorithm, the affine projection algorithm (APA) [4], is a NLMS extension which gains an enhanced convergence rate by reusing the input data. However, its computational complexity increases to $2LN + C_{\text{inv}}N^2$ where N stands for the APA order and C_{inv} is a constant for matrix inversion manipulation. The higher the order N is, the faster the algorithm converges and the heavier computation load is required. With the fast growing arithmetic capability of processors, APA and other less parsimonious algorithms are becoming ideal choices for AEC. However, the power consumption is still critical for mobile devices. Reducing the computation costs without degrading the algorithms' performances is always a hot research topic [5].

This paper proposes a new variable order APA (VOAPA) for AEC. The principle is to switch the APA between its two operating orders, say N_{th} and $(N-i)_{\text{th}}$, $i = 1, 2, \ldots, N-1$. Since the order variation is within the same APA structure, it is stable and easy to implement. The resultant algorithm is desired to work in N_{th}-order mode when significant excitation data can be reused to enhance convergence. On the other hand, it will work in lower order mode to save computational cost when the input excitation is insufficient. The key to the realization of the VOAPA is the order switching mechanism. Voice activity detection (VAD) technique plays an important role in AEC. The method based on envelop tracking [6] tracks the fast and slow envelops of the input speech signal to distinguish between significant and insignificant excitation and control the adaptive algorithm. In this paper, we exploit these trackers' outputs to construct a new stable variable and corresponding threshold. Switching decision can be made by comparing the values of this variable and the threshold. With the variable value going below (above) the threshold, the APA is switched from Nth-order to lower order (lower order to Nth-order). The resultant VOAPA can work stably

with variable orders given the threshold is reasonably designed. The rest of the paper is organized as follows. VAD and APA for AEC are described in Section II. In Section III, a switching mechanism is designed and the new VOAPA is derived. Experiments of the new algorithm with real AEC setup are conducted in Section IV. Finally, conclusions are given in Section V.

VAD and APA for AEC

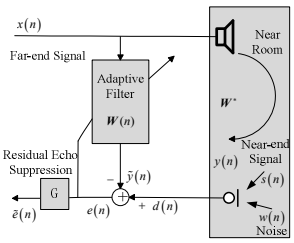


Fig.1. Diagram of AEC

Fig. 2. AEC experiment setup.

Fig.1 depicts an AEC system. $\boldsymbol{W}^* = [W_0^*, W_1^*, ..., W_{L-1}^*]^T$, $\boldsymbol{W}(n) = [W_0(n), W_1(n), ..., W_{L-1}(n)]^T$ represent the L_{th}-order echo path impulse response and adaptive filter weight vector, respectively. The far-end speech $x(n)$ travels through the LEM and results in the echo $y(n)$. The microphone signal $d(n)$ consists of $y(n)$, the near-end speech $s(n)$, and the background noise $w(n)$. $e(n)$ represents the residual echo. The AEC implemented with APA [4] can be described as:

$$\boldsymbol{X}_{\text{ap}}(n) = [\boldsymbol{x}(n), \boldsymbol{x}(n-1), ..., \boldsymbol{x}(n-N+1)], \text{ where } \boldsymbol{x}(n) = [x(n), x(n-1), ..., x(n-L+1)]^T. \quad (1)$$

$$\boldsymbol{y}_{\text{ap}}(n) = \boldsymbol{X}_{\text{ap}}{}^T(n)\boldsymbol{W}(n-1) = [y_{\text{ap},0}(n), ..., y_{\text{ap},N-1}(n)]^T, \quad (2)$$

$$\boldsymbol{e}_{\text{ap}}(n) = \boldsymbol{d}_{\text{ap}}(n) - \boldsymbol{y}_{\text{ap}}(n), \text{ where } \boldsymbol{d}_{\text{ap}}(n) = [d(n), d(n-1), ..., d(n-N+1)]^T, \quad (3)$$

$$\boldsymbol{W}(n) = \boldsymbol{W}(n-1) + \mu \boldsymbol{X}_{\text{ap}}(n)(\boldsymbol{X}_{\text{ap}}^T(n)\boldsymbol{X}_{\text{ap}}(n) + \gamma \boldsymbol{I})^{-1}\boldsymbol{e}_{\text{ap}}(n). \quad (4)$$

γ is a small positive number preventing division by zero. Algorithm order N means N input data vectors are reused to improve the convergence rate, especially when signals are highly correlated like speech. Particularly, when $N = 1$, the APA reduces to the NLMS algorithm

$$\boldsymbol{W}(n) = \boldsymbol{W}(n-1) + \mu \boldsymbol{x}(n)e(n) / (\boldsymbol{x}^T(n)\boldsymbol{x}(n) + \gamma). \quad (5)$$

The price for enhanced convergence is the increased complexity $2LN + C_{\text{inv}}N^2$,where C_{inv} is a constant associated with the matrix inversion method used in (4). In contrast, the complexity of NLMS algorithm is only $O(L)$. Some fast APAs like [5] were thus developed to reduce computational complexity. However, their implementations became more sophisticated to some extents as well.

A successful AEC system demands necessary control schemes and VAD is a critical one. It helps the adaptive algorithm distinguish the active speech periods from silent speech pauses and "freeze" coefficients update in the latter case. The advantages are two folds: calculation cost can be saved and algorithm can be protected from divergence. In this work, the VAD scheme based on fast and slow envelop tracking [6] is used. It is easy to implement and modify to adapt to advanced applications like the one studied in this work. Its diagram is depicted in Fig. 3 and key steps are summarized as follows.

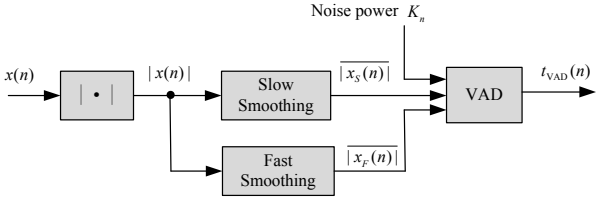


Fig. 3 Diagram of VAD technique based on envelop tracking

The slow envelop tracker adopts a 1_{st}-order IIR filter to smooth the absolute value of the input signal:

$$\overline{|x_S(n)|} = (1-\gamma_S(n))\,|x(n)| + \gamma_S(n)\overline{|x_S(n-1)|}, \tag{6}$$

where the time-varying constant $\gamma_S(n)$ assumes different values for the rising and falling signal edges respectively:

$$\gamma_S(n) = \begin{cases} \gamma_{S,r}, & \text{if } |x(n)| > \overline{|x_S(n-1)|}, \\ \gamma_{S,f}, & \text{else.} \end{cases} \quad 0 << \gamma_{S,r} < \gamma_{S,f} < 1 \tag{7}$$

which ensures the rising signal edges are tracked faster than the falling ones. Similarly, the fast envelop tracker should take the following form

$$\overline{|x_F(n)|} = (1-\gamma_F(n))\,|x(n)| + \gamma_F(n)\overline{|x_F(n-1)|}. \tag{8}$$

with

$$\gamma_F(n) = \begin{cases} \gamma_{F,r}, & \text{if } |x(n)| > \overline{|x_F(n-1)|}, \\ \gamma_{F,f}, & \text{else.} \end{cases} \tag{9}$$

The relation between the parameters in (7) and (9) is

$$\gamma_{S,r} > \gamma_{F,r}, \gamma_{S,f} > \gamma_{F,f}. \tag{10}$$

The main principle of the above envelop tracking VAD lies in the fact that within active speech periods, the fast envelop tracker output is always larger than that of the slow envelop tracker and, on the other hand, when speech signal enters silent periods, the slow envelop output will exceed that of the fast one. By comparing these two tracker output values, the VAD decision can be made. Moreover, to avoid wrong decision during speech pauses periods, a floor parameter K_n is introduced and considered together with slow envelop tracker output. Its value can be chosen some decibels above the background noise power. In [6], the following easy but effective method for estimating the background noise level is used. The short term power estimate of the microphone signal can be calculated as

$$\hat{y}^2(n) = (1-\gamma(n))y^2(n) + \gamma(n)\hat{y}^2(n-1). \tag{11}$$

where the rising and falling signal edges correspond to different constants are γ_r and γ_f ($\gamma_r < \gamma_f$)

$$\gamma(n) = \begin{cases} \gamma_r, & \text{if } y^2(n) > \hat{y}^2(n-1), \\ \gamma_f, & \text{else.} \end{cases} \tag{12}$$

To estimate the background noise level, $\hat{y}^2(n)$ and the previous estimator output is compared as follows.

$$\hat{b}^2(n) = \min\{\hat{y}^2(n), \hat{b}^2(n-1)\}(1+\varepsilon). \tag{13}$$

where ε is a small positive value to prevent the estimator from freezing at a global minimum. Considering (6)-(13), the final VAD decision can be made as

$$t_{VAD}(n) = \begin{cases} 1, & \text{if } \overline{|x_F(n)|} > \max\{\overline{|x_S(n)|}, K_n\}, \\ 0, & \text{else.} \end{cases} \tag{14}$$

The above VAD scheme can only enable adaptive algorithm update during the active speech periods and halt the algorithm update during speech pauses periods. It may be hoped more input speech signal features can be extracted from the VAD results so as to improve algorithm efficiency. Motivated by this goal, a VAD-based variable order APA is proposed in the next section. By further exploiting the VAD intermediate outputs, the resultant switch algorithm can reduce its computation complexity while maintaining the similar convergence performance as the APA algorithm.

VAD-based Variable Order APA

Extensive experiments on the VAD scheme described in section II verified its stable performance for various types of speech inputs. The fast and slow envelop trackers generate significantly distinguished curves which are accurate enough to be compared to make the VAD decision. Here we introduce a new variable $\overline{|x_D(n)|}$ which is defined as the difference of $\overline{|x_F(n)|}$ and $\overline{|x_S(n)|}$ during active speech periods. In speech silent periods the algorithm update is halted so no manipulation is needed:

$$\begin{aligned} &\overline{|x_{\mathrm{D}}(n)|}=\overline{|x_{\mathrm{F}}(n)|}-\overline{|x_{\mathrm{S}}(n)|}, \\ &\overline{|x_{\mathrm{F}}(n)|}>\overline{|x_{\mathrm{S}}(n)|}. \end{aligned} \tag{15}$$

Apparently, $\overline{|x_{\mathrm{D}}(n)|}\geq 0$. Further observation of the $\overline{|x_{\mathrm{D}}(n)|}$ curve reveals its instant values correspond to the speech signal instant amplitudes and therefore can indicate the significant periods of input excitation data which deserve being reused. This naturally motivates us to use higher order APA during the big input amplitude periods and alternatively use NLMS or other simple algorithm during the small amplitude periods. Since the NLMS algorithm is actually a special case of APA when the latter's order equals one, the resultant algorithm can thus be generally regarded as a APA whose order varies with the difference of VAD fast and slow envelop tracker outputs, i. e., $\overline{|x_{\mathrm{D}}(n)|}$. The key to the order variation control is a threshold that distinguishes the two working orders status. We can use the above introduced steps (11)-(12) to estimate the short-term power of the far-end speech signal $\overline{|x^2(n)|}$. By multiplying a constant K_v to $\overline{|x^2(n)|}$ a simple instant threshold associated with the input speech signal power is formed. The value of K_v as well as how frequent $\overline{|x^2(n)|}$ is updated are application dependent and can be empirically or elaborately designed using advanced techniques like speech linear prediction [7]. All design factors can be absorbed into a general expression $f(K_v,x)=f(K_v|\overline{x^2(n)}|)$ to explicitly represent the threshold value. Given reasonably chosen $f(K_v,x)$, the VOAPA will switch between two variable orders which guarantees the overall stability. The resultant algorithm can be summarized as follows:

For each iteration, calculating:

Step 1: $t_{\mathrm{VAD}}(n)$ by (7)-(15) ; *Step 2*: $\overline{|x_{\mathrm{D}}(n)|}$ by (16); *Step 3*: $\overline{|x^2(n)|}$ by (11)-(12) and other manipulations;

Step 4: $f(K_v,x)$;

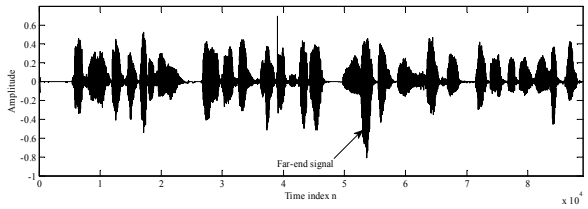
Step 5: If $t_{\mathrm{VAD}}(n)=1$, $\boldsymbol{W}(n)=\boldsymbol{W}(n-1)+\mu\boldsymbol{X}_{\mathrm{ap}}(n)(\boldsymbol{X}_{\mathrm{ap}}^{T}(n)\boldsymbol{X}_{\mathrm{ap}}(n)+\gamma\boldsymbol{I})^{-1}\boldsymbol{e}_{\mathrm{ap}}(n)$, N th -order: $\overline{|x_{\mathrm{D}}(n)|}>f(K_v,x)$, $(N-1)_{\mathrm{th}}$-order: $\overline{|x_{\mathrm{D}}(n)|}\leq f(K_v,x)$;

Else $\boldsymbol{W}(n)=\boldsymbol{W}(n-1)$.

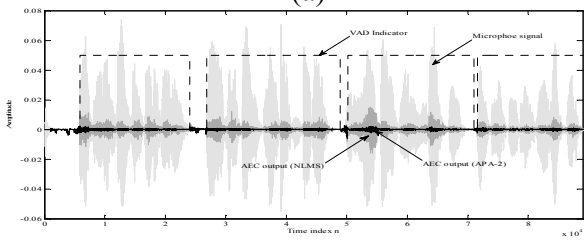
End

VAD and APA for AEC

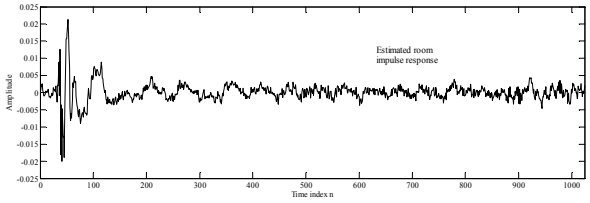
In this section, the proposed algorithm is experimentally tested in a real AEC environment. The room size is about 20 m^2 with typical simple office arrangement. As illustrated in Fig. 2, the microphone, personal computer, USB sound card, and a single multimedia loudspeaker are placed on the desk, with the microphone and speaker being spaced 0.45 m. A 11 seconds long far-end speech is played with Windows Media Player through the loudspeaker. The echo signal is picked by the microphone and recorded with Adobe Audition 1.5. The far-end and the microphone signals are then used as $x(n)$ and $d(n)$ in simulation. For simplicity, no double-talk is assumed here. Three algorithms, the NLMS, the 2_{nd}-order APA, and the VOAPA, are compared in the simulation. 2_{nd}-order APA is employed due to both its significantly improved convergence and mildly increased computational complexity as compared to NLMS algorithm. The step sizes are set as $\mu_{\mathrm{NLMS}}=\mu_{\mathrm{APA}}=0.8$. First, we test the NLMS and 2_{nd}-order APA with the real recorded speech signals. Fig. 4 (a) depicts the input signal $x(n)$, which is a male's Chinese speech "This is the far-end speech. 1-2-3-4-5-6-7, a-b-c-d-e-f-g". In Fig. 4 (b), the AEC outputs using NLMS and 2_{nd}-order APA are illustrated. The VAD scheme described in (6)-(14) is employed and the indicator output is plotted as dotted line. It can be seen the speech activity estimate is quite accurate so that the adaptive algorithm can focus on the informative excitation to avoid divergence. From the AEC results, it is obvious the echo has been significantly removed by both algorithms. However, a further look at the residual echo curves reveals the 2_{nd}-order APA has gained a much greater echo cancellation over the NLMS algorithm. Played through loudspeaker, the 2_{nd}-order APA output is hardly intelligible whereas the NLMS output is. Fig. 4 (c) illustrates the room impulse response estimated by the adaptive filter at the end of iteration.



(a)



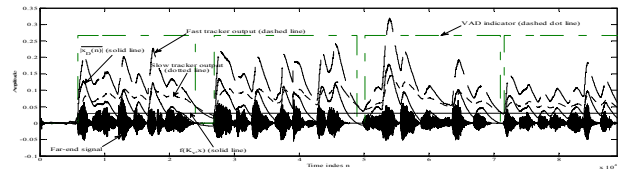
(b)



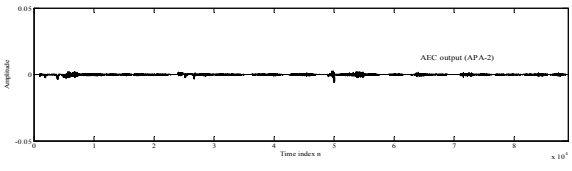
(c)

Fig. 4. AEC experiment results. NLMS vs. APA. (a) far-end speech signal $x(n)$;(b) AEC output $e(n)$ and VAD parameters; (c) Estimated room impulse response.

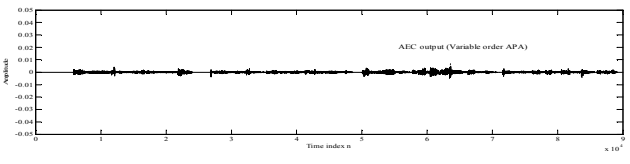
Next, we test the VOAPA performance. Due to space limitation, only the algorithm with order variation between 2 and 1 is illustrated. Other variation schemes have also been tested and the similar observation is supported. VAD relevant parameters are: $\gamma_{\mathrm{S},r}=0.999$, $\gamma_{\mathrm{S},f}=0.9997$, $\gamma_{\mathrm{F},r}=0.992$, $\gamma_{\mathrm{F},f}=0.999$, $\gamma_r=0.993$, $\gamma_f=0.997$. The dynamic threshold $f(K_v,x)$ can be chosen flexibly in various ways as described in section II. Here, we use a simple one to illustrate its effectiveness. By estimating the short term power of $\overline{|x_{\mathrm{D}}(n)|}$ using a fraction of data, say, N_{D} samples, which lie in the first detected active part of speech periods, $f(K_v,x)$ can be obtained by multiplying the power estimate with a constant K_v and then kept unchanged during the rest of convergence course. In this experiment, we set $N_{\mathrm{D}}=1000$ and $K_v=0.3$. The results are illustrated in Fig. 5, where (a) illustrates the VAD estimating curves: the fast tracker output $\overline{|x_{\mathrm{F}}(n)|}$, the slow tracker output $\overline{|x_{\mathrm{S}}(n)|}$, and that of their difference variable $\overline{|x_{\mathrm{D}}(n)|}$. The above described simple version of $f(K_v,x)$ is also plotted. Subfigure (b) and (c) plot the AEC output employing the 2nd-order APA and VOAPA, respectively. It can be clearly observed these two results differ very subtly only in those periods where $\overline{|x_{\mathrm{D}}(n)|}$ falls below the threshold $f(K_v,x)$ and thus the VOAPA exhibits slightly degraded performance. This is due to the active algorithm within these periods has been switched from 2nd-order APA to NLMS and the latter is inferior to the former in convergence rate. However, the overall stability is guaranteed thanks for the APA framework. In practical system, a residual echo suppression (RES) unit is usually appended to the AEC output to further cancel the echo. Here we employ the method introduced in [8]. This method is based on spectral subtraction theory and can be developed into real working solution. From subfigure (d) we can see the final residual echo has been almost suppressed.



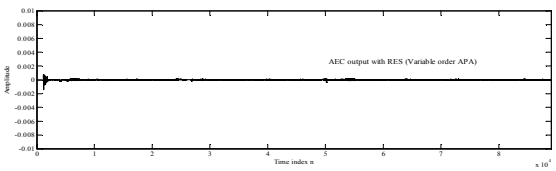
(a)



(b)



(c)



(d)

Fig. 5. AEC experiment results. APA vs. VOAPA. (a) VAD and VOAPA parameters; (b) AEC residual echo, 2nd-order APA; (c) AEC residual echo, VOAPA (*N*=2); (d) AEC+RES output VOAPA (*N*=2).

Conclusion

A new VOAPA algorithm is studied in this paper. It is based on an efficient switching mechanism developed from envelop-tracking type VAD technique. The resultant algorithm can thus distinguish the significant input data periods from the insignificant ones. Corresponding to these two statuses, the VOAPA can be respectively switched between the higher and lower orders, achieving enhanced convergence due to more input excitation and reduced computation cost when the input is relatively weak. The effectiveness of the VOAPA is verified by AEC experiments.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61102118, 61072123) and the Research Project of Chongqing Educational Commission (KJ130504).

References

[1] J. Benesty et al., Advances in Network and Acoustic Echo Cancellation, Springer Press, 2001.

[2] C. Breining et al., "Acoustic echo control: an application of very-high-order adaptive filters," IEEE Signal Processing Magazine, vol. 16, no. 4, pp. 42-69, 1999.

[3] J. I. Nagumo and A. Noda, "A learning method for system identification," IEEE Trans. Automat. Contr., vol. AC-12, pp. 282-287, June 1967.

[4] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," Electronics and Communications in Japan, vol. 67-A, pp. 19-27, 1984.

[5] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in Proc. of IEEE ICASSP'95, vol. 5, pp. 3023-3026, Detroit, MI, USA, May 1995.

[6] E. Hänsler, and G. Schmidt, Acoustic Echo and Noise Control: A Practical Approach, Wiley, 2005.

[7] R. Martin et al., Advances in Digital Speech Transmission, Wiley, 2008.

[8] S. M. Kuo, B. H. Lee, and W. Tian, Real-time Digital Signal Processing, Implementations and Applications, Wiley, 2006.