An Efficient Switching Mechanism Using Voice Activity Detection for Acoustic Echo Cancellation

A.A.M.Muzahid, S.I.M.M. Raton Mondol, K.M.R. Ingrid and Y. Zhou

Chongqing University of Posts and Telecommunications,

Chongqing, China

muzahid07@yahoo.com, {simmraton, becky.kim1432, yzhou_7}@gmail.com

Keywords: *AEC, VAD, APA, NLMS.*

Abstract. Adaptive algorithms are widely used for acoustic echo cancellation (AEC). In this paper, an efficient switching tactic with advantageous features of NLMS and APA is proposed for AEC application. Switching points are decided by employing a smart voice activity detection scheme based on speech utterance. The speech signal is distinguished as voiced, unvoiced and silent periods then switch mechanism will automatically assign algorithm, either NLMS or APA, to the corresponding period. The prime motive of this proposed switch algorithm is to reduce the computational complexity and get better output over the NLMS but close to APA. Simulation results show the performance of proposed scheme is significantly improved.

1. Introduction

In hands-free communication systems, acoustic echo is induced by the coupling between loudspeaker and microphone. Adaptive filter is broadly used in the acoustic echo cancellation problem. Various kinds of adaptive filtering algorithms have been proposed over the years such as the normalized least mean square (NLMS) [10] and the affine projection algorithm (APA) [9]. The computational complexity increases with the improvement of algorithm performance. Switching mechanism is generally used to reduce the computational cost and improve the convergence property. A variable step size affine projection adaptive algorithm (VSS-APA) was proposed in [1-2] which improved the convergence but the computational complexity was high. Voice activity detection (VAD)-based APA and fast LMS/Newton algorithm are proposed [3-4] by classifying speech into significant and insignificant data periods which are distinguished by short-term energy estimation. Both of these new techniques gained a better accuracy to reduce the amount of residual echo and aimed to reduce the computational complexity.

A new switching mechanism for AEC is proposed in this paper, where the speech signal is classified into voiced, unvoiced and silent regions. The ramification of speech signal into voiced and unvoiced periods provides an initiatory acoustic partitioning for speech processing applications [5]. A speech sound is produced when a periodic pulses of air oscillate the vocal cords and form pitch frequency that is called voiced speech. Approximately about two thirds of speech is voiced. The unvoiced speech sound is non-periodic in nature which causes no vocal cords oscillation and also has no pitch structure. The typical nature of the voiced sound is that it is nearly periodic and conveys high energy. Moreover, unvoiced sound occurs randomly, containing lower energy. If the energy of a period of speech is much lower or negligible then it can be considered as noise. In this paper, by distinguishing the speech status, a switch algorithm is designed which uses the fast but more expensive algorithm in the voiced speech case and the slow but parsimonious algorithm in the unvoiced case. APA and NLMS adaptive algorithms are correspondingly employed to implement the switching mechanism. The computational complexity of APA and NLMS are respectively $2LN+K_{inv}N^2$ and $3M+1$multiplications per sample period where K_{inv} is a constant. L is the filter length and N is the length of input data ($L>N$) [6]. APA provides faster convergence than the NLMS algorithm but it has a higher computational complexity. In order to use the switching mechanism, computational complexity will be decreased significantly. The proposed VAD based AEC has been illustrated in Fig. 1. The far-end $\boldsymbol{x}(n)$ is the reference input signal to VAD block. This $\boldsymbol{x}(n)$ reflects back to the near-end microphone through the loudspeaker is called the echo signal $d(n)$.

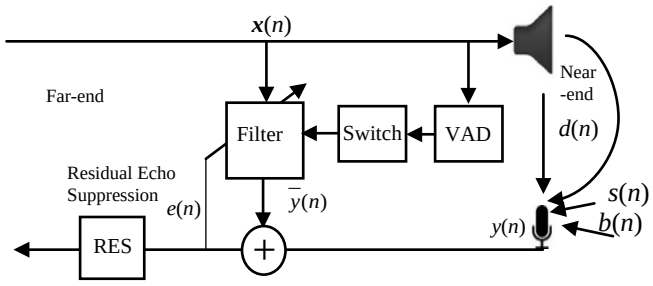

Fig. 1.The proposed VAD based AEC block diagram

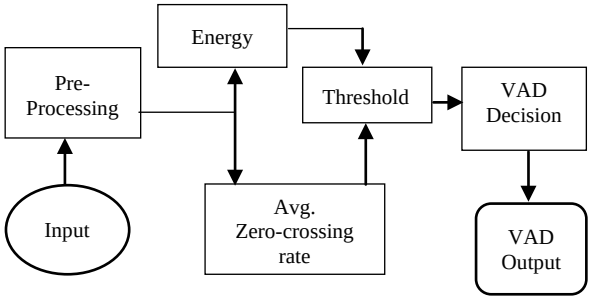

Fig. 2. The block diagram of VAD

2. Voice Activity Detection

Various VAD methods have been developed over the past a few years [4, 7]. In this paper, we combine zero-crossing rate (ZCR) and energy calculation to build up a new VAD scheme. The ZCR is an important parameter of classification of voiced/unvoiced periods and the energy of speech indicates the active speech and silent periods. The counting of average ZCR is defined as

$$Z_{\text{avg}} = \tfrac{1}{P} \sum_{m=n-P+1}^{n} \left|\text{sgn}[x(m) - \text{sgn}\, x(m-1)]\right| w(n-m), \tag{1}$$

where $w(n)$ is the P length Hamming window and the sgn function is defined as

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

The short-time energy equation is given by

$$E_{\text{ST}} = \sum_{m=n-P+1}^{n} [x(n)w(n-m)]^2. \tag{2}$$

It is known voiced speech periods contain higher energy and lower zero-crossing rate in comparison with unvoiced speech periods [8]. The thresholds are thus placed to get decision from iterative resultant values for both zero-crossing and energy calculation. The VAD algorithm will be applied on the far-end signal $\boldsymbol{x}(n)$. The average ZCR counter value Z_{avg} will be compared with a predifined (or empirically obtained) constant threshold TH_Z. At the same time, the short-time energy E_{ST} of the same sample speech will be calculated and compared with a constant threshold TH_E. The VAD decision is therefore decided by the following rule, where σ_b^2 indicates the backgroud noise power.

$$\boldsymbol{x}(n) = \begin{cases} \text{voiced}, & \text{if } Z_{\text{avg}} \leq TH_Z \text{ and } E_{\text{ST}} \geq TH_E \\ \text{unvoiced}, & \text{if } Z_{\text{avg}} \geq TH_Z \text{ and } E_{\text{ST}} \leq TH_E \\ \text{silence}, & \text{if } Z_{\text{avg}} \geq TH_Z \text{ and } E_{\text{ST}} \simeq \sigma_b^2 \end{cases} \tag{3}$$

3. Implementation of VAD Based Switch Adaptive Algorithm for AEC

In this section we propose a switch block with NLMS and APA in AEC. The NLMS algorithm provides moderate convergence rate with low computational complexity. However, APA algorithm achieves an enhanced convergence rate by reusing input data [9]. Higher projection order gives faster convergence but increases the computational burden as well. The APA is defined as below

$$\boldsymbol{e}(n) = \boldsymbol{y}(n) - \bar{\boldsymbol{y}}(n). \tag{4}$$

$$\bar{\boldsymbol{y}}(n) = \boldsymbol{X}_{\text{ap}}^{T}(n)\boldsymbol{W}(n). \tag{5}$$

$$\varepsilon(n) = [\boldsymbol{X}_{\text{ap}}^{T}(n)\boldsymbol{X}_{\text{ap}}(n) + \gamma \boldsymbol{I}]^{-1}\boldsymbol{e}(n), \tag{6}$$

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) + \mu \boldsymbol{X}_{\text{ap}}^{T}(n)\varepsilon(n). \tag{7}$$

where $\boldsymbol{I}$ represents the $N \times N$ identity matrix, $\boldsymbol{W}(n)$ is the weight vector of order L. $0 \leq \mu \leq 2$ is the step size. $\boldsymbol{X}_{\text{ap}}$, $\boldsymbol{y}(n)$ and $\boldsymbol{e}(n)$ are respectively the input, desired and error signals. $\bar{\boldsymbol{y}}(n)$ is the adaptive filter output. When the order of APA is 1, it will reduce to the NLMS algorithm as follows

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) + \mu \frac{e(n)\boldsymbol{x}(n)}{\| \boldsymbol{x}(n) \|^2 + \gamma}, \tag{8}$$

where γ is a small positive constant avoiding division by zero. In this paper, the speech is classified by VAD block into the voiced and unvoiced sections. The silent periods will also be treated as unvoiced. According to (3), a VAD flag $\xi(n)$ is generated to be the indicator of the voiced speech:

$$\xi(n) = \begin{cases} 1, & \text{if } Z_{\text{avg}} \leq TH_Z \;\&\&\; E_{ST} \geq TH_E \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

The silents portions are short periodic samples but not negligible. Energy of the silent portions is much close to zero in theory. In accordance to the zero-crossing theory, noise has a high zero-crossing rate that may cause a confusion with the unvoiced periods. To avoid this barrier and reduce the complexity, we made the VAD flag for just voiced indicators. In voiced periods, the average ZCR will be much lower and higher energy can be found in lower frequency. If we estimate the energy in time domain then (2) can be formed as

$$E_{\text{ST}} = \sum_{m=n-P+1}^{n} [x(n)]^2. \tag{10}$$

Every iteration, short time energy will be calculated and compared with the threshold energy TH_E. The amplitude of voiced sample would be higher in comparison with the others. The average ZCR counter in time domain can be written from (1) as

$$Z_{\text{avg}} = \tfrac{1}{P} \sum_{m=n-P+1}^{n} \left| \text{sgn}[x(m) - \text{sgn}\, x(m-1)] \right|, \tag{11}$$

The value of Z_{avg} will be compared with TH_Z. Finally, these two decision parameters will decide which adaptive algorithm is used for this running iteration. The cost effective NLMS algorithm will be functioning in the periods of unvoiced and silent segments. APA will be run over the voiced periods. For both algorithms the weight vecctor order is the same L. During the unvoiced peirods, NLMS will help reduce the computational complexity. On the other hand, the APA functioning in voiced periods can better reduce the residual echo and speed up convergence. The switching mechanism is given below

$$\delta(n) = \begin{cases} \text{APA}, & \text{if } \xi(n) = 1 \\ \text{NLMS}, & \text{if } \xi(n) = 0 \end{cases} \tag{12}$$

where $\delta(n)$ is the switching decision parameter which depends on the value of $\xi(n)$. The value of $\xi(n)$ is 1 and 0 which respectively points out the APA and NLMS active operation.

4. Simulation

In this section, we conduct the AEC experiement using the switch algorithm. The far-end signal $\boldsymbol{x}(n)$ is recorded with 8 kHz sampling rate. The room impulse response is modeled to have 1024 coefficients. The system model is shown in Fig. 1. The switching mechanism is implemented between APA and NLMS using the scheme proposed in section 3. The length of adaptive filer for both algorithms is L = 1024. μ = 0.8 is used for both algorithms. Second order APA is employed that will ensure the fast convergence and moderate complexity. The amplitude of energy flag is enhanced by 60 times for presenting a clear data. Here TH_Z = 0.3 and TH_E = 0.42 are empircially decided using abundant sample speech data.The average ZCR and energy of $\boldsymbol{x}(n)$ are plotted in Fig. 3(b).

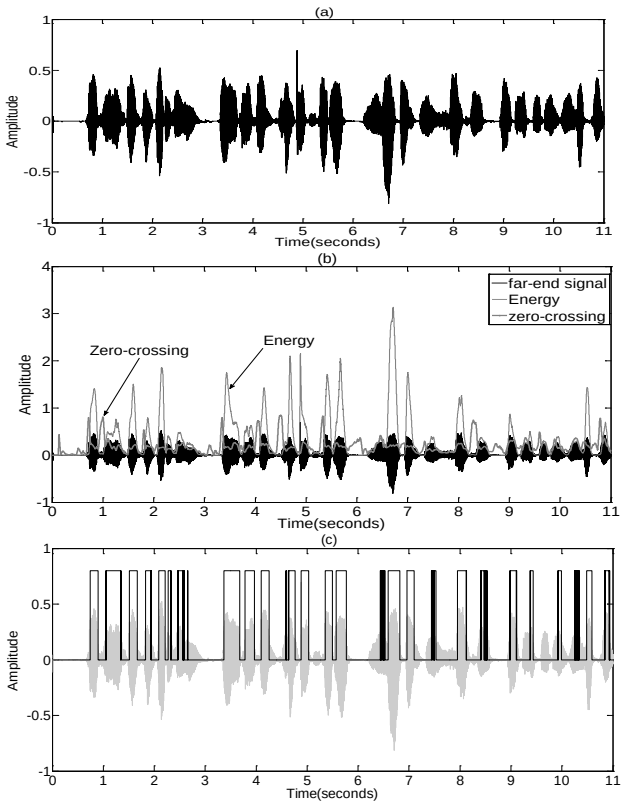

Fig. 3. (a) Far-end reference input signal $\boldsymbol{x}(n)$ (b) VAD parameters (c) Switching points with voiced flag on $\boldsymbol{x}(n)$

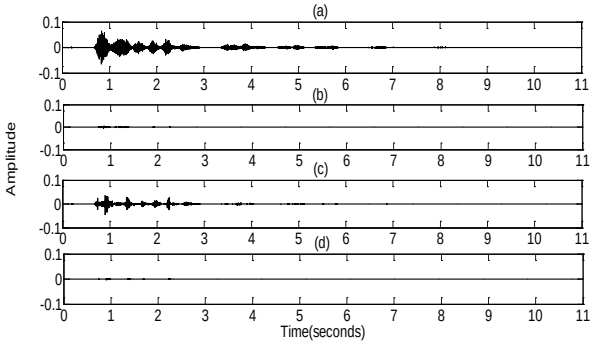

Fig. 4. Error signal (a) NLMS (b) APA (c) switch (d) switch+RES

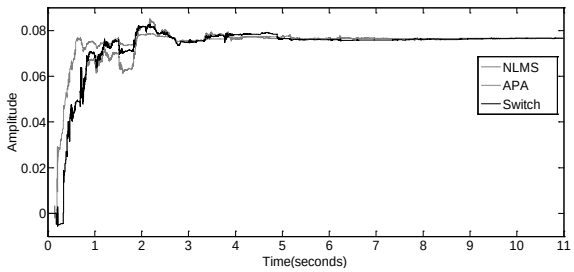

Fig. 5. The comparison of weights adaptation

The frame length of average ZCR calculation is 300. ZCR counter provide a positive integer value between 0 and 2. Since voiced portions have low ZCR and high energy in comparison with unvoiced and silent periods. Some silent periods (Fig.3(b)) indicate very low ZCR like voiced ones. In general, silent signal has similar noise-like properties that should contain high ZCR and low energy. We carefully observe the far-end signal $\boldsymbol{x}(n)$ that the signal coffiecients of that portions just lay on the zero axis meaning there is no phase changing issue. In this scenario, the detection scheme will be followed by energy estimation because energy of silent is equal ot zero ($Z_{\text{avg}} \geq TH_Z \,\&\&\, E_{\text{ST}} \simeq 0$). Fig. 3 (c) illustrates the voiced flag that indicates the switching points where amplitude of voiced flag has been decreased to about 20 percent. APA algorithm will work on voiced periods and NLMS on unvoiced and silent periods. It can be seen in Fig. 3 the voiced flag has almost occupied more than two-thirds of total periods so APA works for a long time that help to enhance the echo cancellation. On the other hand, NLMS only works for unvoiced and silent periods which saves the computational complexity. Fig. 4 depicts the residual error signal by NLMS, APA and proposed switching algorithm respectively. Fig. 4 (c) depicts the good performance of the switch-based algorithm which is just slightly inferer to the ouput of APA shown in Fig. 4(b). Furthuremore, when the switch algorithm is cascaded with a resudual echor suppersion filter [10], we can get a prominenat output which has the best performance among all algorithms. The comparison of convergence performance

in terms of adaptive filter coefficients is also illustrated in Fig. 5. The 100th coefficients of weight vector for different algorithms are plotted.

5. Conclusion

In this paper, an efficient switching mechanism for AEC is proposed. It efficiently combines the advantages of NLMS and APA with a small computational overhead. The classification of speech by ZCR and energy estimation verifies the superior performance of the new technique over the conventional algorithms in terms of error cancellation, adaptation speed and computational complexity.

Acknowledgements

This work is supported by the Chongqing Sci. and Tech. Commission (cstc2015jcyjA40027).

References

[1] Joy. J, Mathurakani. M, "A switching variable step size affine Projection adaptive algorithm for acoustic echo cancellation," *International Conference on Microelectronics, Communications and Renewable Energy (AICERA/ICMiCR)*, pp. 1-5, June 2013.

[2] Ben Jebara. S, Besbes. H, " Variable step size filtered sign algorithm for acoustic echo cancellation," *Electronics Letters* , vol. 39, no. 12 , pp. 936-938, Jun 2003.

[3] Cheng. Lu, F. Liu, H. Liu, Y. Zhou, "A new switch affine projection algorithm for acoustic echo cancellation," *Intl. Conf. on Estimation, Detection and Info. Fusion* , pp. 195-199, Jan. 2015.

[4] Yuan. W, Zhou Y., Huang. Z and Liu H. Q., "A VAD-based switch fast LMS/Newton algorithm for acoustic echo cancellation," *IEEE Intl. Conf. on DSP*, pp. 967-970, July 21-24, 2015.

[5] Bachu R.G, Kopparthi. S, Adapa. B, Barkana B.D, "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal".

[6] A. Gonzalez et. al, "An affine projection algorithm with variable step-size and projection order," *Digital Signal Processing*, vol. 22, no. 4, pp. 586-592, July 2012.

[7] Y. Wang, S. Huang, Y. Wei, "A voice activity detection algorithm with sub-band detection based on time-frequency characteristics of mandarin," 6th International Congress on Image and Signal Processing , pp. 1287-1291, Dec. 2013.

[8] Rabiner, L. R., and Schafer, R. W., Digital Processing of Speech Signals, 1978.

[9] Dewasthale. M. M., Kharadkar, R.D. "Acoustic Noise Cancellation Using Adaptive Filters: A Survey," *IEEE International Conference on Electronic Systems, Signal Processing and Computing Technologies* (*ICESC*), pp. 12-16, Jan.2014.

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