



# A New Voice Controlled Noise Cancellation Approach

Ali O. Abid Noor<sup>1</sup>, Salina Abdul Samad<sup>1</sup>, Aini Hussain<sup>1</sup> and Aishah Fauthan<sup>2</sup>

<sup>1</sup> *Department of Electrical, Electronic and System Engineering,  
Faculty of Engineering and Built Environment  
University Kebangsaan Malaysia  
Bangi, 43600, Malaysia*

<sup>2</sup> *Institute of Micro Engineering and Nano Electronics  
Faculty of Engineering and Built Environment  
University Kebangsaan Malaysia  
Bangi, 43600, Malaysia*

Tel.: +603-8921 6317, E-mail: [ali511@vlsi.eng.ukm.my](mailto:ali511@vlsi.eng.ukm.my)

**Abstract**— This paper presents a new approach to control the operation of adaptive noise cancellers (ANCs). The technique is based on using the residual output from the noise canceller to control the decision made by a voice activity detector (VAD). Threshold of full band energy feature is adjusted according to the residual output of the noise canceller. In variable background noise environment, the threshold controlled VAD prohibits the reference input from containing some components of actual speech signal during adaptation periods. The convergence behavior of the adaptive filter is greatly improved, since the reference input will be highly correlated with the primary input. In addition, the computation power will be reduced since the output of the adaptive filter will be calculated only during non- speech periods. The threshold controlled noise canceller achieves a cleaner output in about 50% of the time required by a non-controlled noise canceller.

**Keywords**— *voice activity detector; adaptive noise cancellation; threshold control.*

## I. INTRODUCTION

In adaptive noise cancellation ANC applications, two-sensor models are often used for speech enhancement. In this technique, it is assumed that the two sensors are physically separated and isolated from each other, so that no substantial speech leakage into the reference input occurs, otherwise intelligibility of the speech signal will be degraded by the adaptive process. In practice, the two microphones should be located within few centimeters [1]. Traditionally, directional microphones and acoustic barriers are used to prevent speech leakage into the reference input. Voice activity detectors VAD are offered in more advanced systems nowadays [2],[3],[4] and [5]. The primary function of a voice activity detector is to

provide an indication of speech presence in order to facilitate speech processing as well as providing delimiters for the beginning and end of a speech segment. In this work, a variable threshold VAD is used to improve the operation of an adaptive noise canceller ANC in variable background environments. The use of VAD in this context has two advantages, first, the convergence behavior of the adaptive filter will be improved since the reference input will be highly correlated with the primary input, and second, the computation power will be reduced since the output of the adaptive filter will be calculated only during non-speech periods. This power saving is of great importance in hand free communications, where processing power should be kept as low as possible, due to size and weight limitations. Fig. 1 depicts a typical one-end speech of a general telephone conversation. It is clear that

pauses or non-voiced intervals are quite long in speech communications; therefore this property shall be used as an advantage to improve the performance of the noise canceller as well as reducing the computational costs and hence the power consumption of the system.

In the absence of speech, the primary input of the adaptive filter could be used as a reference signal for the present noise signal to adapt the filter coefficients in the normalized least square NLMS system. Detailed discussion of the NLMS algorithm can be found in [6]. This noise should be a very close estimate of the component in the signal. If speech signal is then detected, the VAD switches the reference input back to the reference sensor. The adaptive filter in the NLMS should now have the same characteristics as the noise path, to reduce the noise to a minimum. Furthermore, the VAD freezes the filter adaptation when speech is present. In literature, several VAD schemes have been introduced, each providing a solution to a certain aspect. The main issues of VAD are threshold control [7], computational complexity [8] and robustness [9].

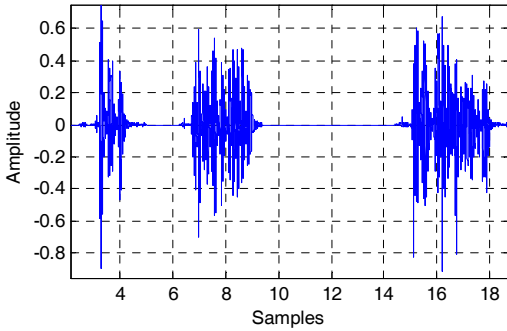


Fig. 1 A typical one-end telephone conversation

In the current work, a VAD and an adaptive noise canceller are made to have a mutual control, so that improved performance is obtained in varying background noise. The paper is organized as follows. In addition to this introductory section, section II describes the features used in the voice activity detector, section III explains the mutual control between the VAD and the noise canceller, section IV gives a performance evaluation of the system and section V concludes the paper.

## II. VAD FEATURE EXTRACTING

Features used here are the full band energy measurement and zero crossing rate. A logic circuit is used to decide if speech is present or not. In the following, formulations as well as possible realization of these features are given.

### A. Full Band Energy

The full band energy  $E_f$  is the logarithm of the normalized first autocorrelation coefficient  $A(0)$  and can be determined by the following .

$$E_f = 10 \times \log_{10} \left( \frac{1}{10} A(0) \right) \quad (1)$$

Based on background noise level, a silence flag,  $f_{e-sil}$ , is set according to the following equation.

$$f_{e-sil} = \begin{cases} 1, & \text{if } E_f < T_e \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where  $T_e$  is a noise threshold. The full band energy algorithm is implemented as shown in Fig. 2.

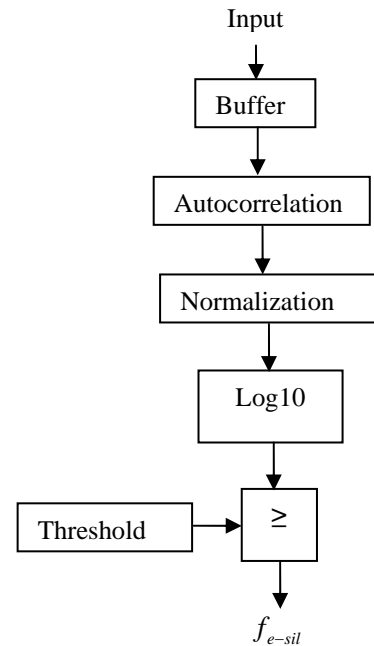


Fig. 2 Full band energy block diagram

### B. Zero Crossing Rate

Zero- crossing rate  $Z_x$  is a measure of how often a signal crosses the zero value in a given time. The zero-crossing rate of background noise is often constant. However, if speech is present the  $Z_x$  decreases. Zero crossing  $Z_x$  can be found in the time domain by comparing the sign of adjacent speech samples. The zero crossing  $Z_x$  of a sampled speech  $s(n)$  is defined as

$$Z_x = \frac{1}{2} \sum_{n=1}^N \left| \text{sgn}[s(n)] - \text{sgn}[s(n-1)] \right| \quad (3)$$

where  $\text{sgn}(s)$  is 1 for  $s > 0$ , -1 for  $s < 0$ . Two flags  $f_{z-vce}$  and  $f_{z-unv}$  stand for voiced and unvoiced speech respectively are set according to the following

$$f_{z-vce} = \begin{cases} 1, & \text{if } Z_x < T_{z1} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

and

$$f_{z-unv} = \begin{cases} 1, & \text{if } Z_x > T_{z2} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $T_{z1}$  and  $T_{z2}$  are thresholds for voiced and unvoiced speech respectively. These two thresholds are determined using empirical procedure. Fig. 3 shows the zero crossing calculation block diagram, and Figure 4 depicts the output of the zero crossing detector for a signal corrupted with noise. It can be seen from Fig. 4, that the crossing rate per time frame decreases if speech is present.

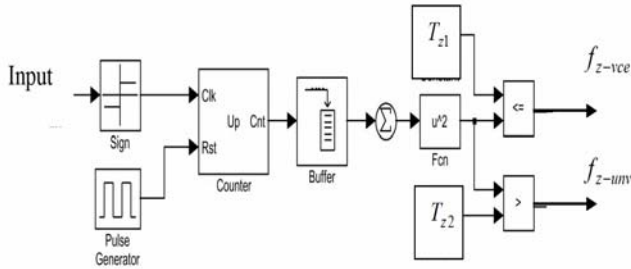


Fig. 3 Zero crossing rate calculation

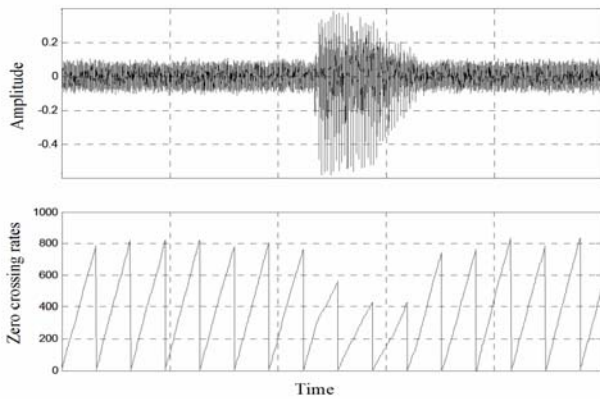


Fig. 4 Zero crossing detection of a noisy signal

### C. Decision Circuit

Using Equations (2),(4) and (5) the VAD decision  $D$  can now be represented in logic algebra as

$$D = (\bar{f}_{e-sil} + f_{z-unv}) + (f_{e-sil} * f_{z-vce}) \quad (6)$$

where  $\bar{\cdot}$ ,  $+$ ,  $*$  denote the logic operators (NOT, OR, AND) respectively. A decision circuit is constructed according to (6) as shown Fig. 5. The output of this circuit is used to control the operation of the adaptive filter in a noise cancellation system. The zero and nonzero outputs of the decision circuit are used to control the operation of the adaptation process in the noise canceller system. The adaptation stops on reception of logic zero, and it continues when receiving logic one.

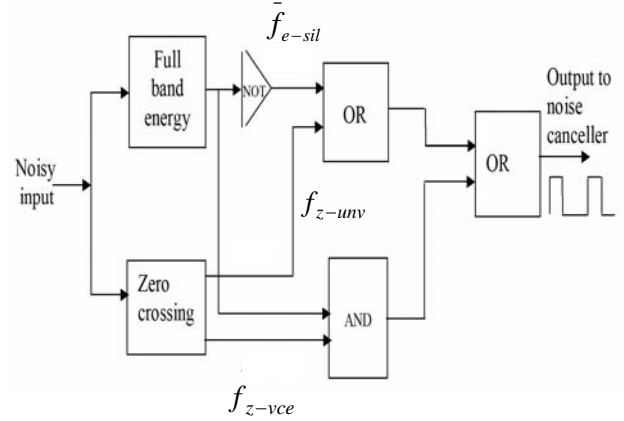


Fig. 5 Decision circuitry

### III. THRESHOLD ADJUSTMENT

The background noise can vary between different environments and situations, from a silent room to a noisy factory or fast moving car. Problems may arise if the VAD does not switch the reference input of the noise canceller back to the reference sensor. The reference sensor could record speech and adapt these characteristics to the filter. Then, the adaptive filter may reduce speech signals as well as noise from the desired signal; hence the signal to noise ratio SNR is decreased. Therefore, measurement of the power of the background noise is required. In literature, several methods are used to measure the background noise in VAD systems, examples are found in [5] and [11]. In this work, a new technique is used to adjust threshold values of the VAD. The information of the residual noise at the output of the noise canceller is used to adjust threshold of the full band energy described in section II. Schematic of this idea is shown in Fig. 6.

The residual output of the noise canceller is received by the VAD, and an outgoing prompt signal is sent to control the operation of the adaptive filter, such that the adaptive filter freezes operation when receiving logic "low" and continues to operate when receiving logic "high". The inputs to the VAD are divided into frames using frame sequencer which divides the incoming signals into frames of data comprising 256 contiguous samples. The energy of speech is considered to be relatively stationary over 15 milliseconds; therefore, frames of 32 milliseconds are used. In order to make the VAD more robust to impulsive noise, an overlap of 16 milliseconds between adjacent frames is allowed.

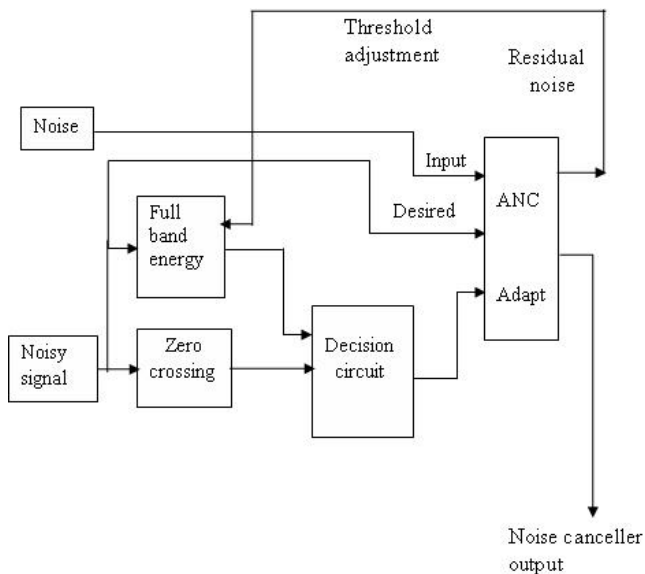


Fig. 6 Threshold adjustment using residual output Information

The residual noise  $R$  is calculated on a frame basis as the difference (in decibels) between the noisy input  $P_i$ , and the output of the adaptive noise canceller  $P_o$  as follows

$$R = 10 \log_{10} \left[ \frac{1}{M} \sum_{n=1}^M P_i(n) \right] - 10 \log_{10} \left[ \frac{1}{M} \sum_{n=1}^M P_o(n) \right] \quad (7)$$

where  $M$  is the number of samples over which the average power is calculated. The threshold  $T_e$  in Eq. (2) is calculated as follows

$$T_e = E_{\max} - R \quad (8)$$

where  $E_{\max}$  is the maximum possible input power of the desired signal. The threshold is then compared to the average energy of each frame of the input signal  $E_f$ . If the threshold is higher than  $E_f$ , the input signal to the adaptive contains speech and logic 'low' is sent to the adaptive filter deactivate the adaptation process. Otherwise, if threshold is less than  $E_f$ , then the input signals contains no speech and therefore logic 'high' is sent to the adaptive filter to activate the adaptation process. This process continues until the filter reaches a steady state.

#### IV. PERFORMANCE EVALUATION

A noisy speech signal (nspeech.wav) is applied to the VAD. The output of the VAD shows a high value if no speech is detected and a low value if speech is presents. The noisy speech signal is shown in Fig. 7. If the speech signal contains high noise levels, the VAD is not capable of measuring in an accurate way if speech is present or not if the implemented threshold is constant. Therefore, a noise measurement system

is implemented to adapt the threshold value for the full-band energy feature as explained in section III. Fig. 8 shows the error decay at the noise canceller output after using variable threshold VAD. A performance comparison between VAD controlled and non-controlled noise cancellation cases is shown in Fig. 9.

It is evident that the controlled noise canceller removes about 80% of the noise in 8000 iterations while the non-controlled noise canceller needs about 18000 iterations to reach the same level with noticeable misadjustment. When the background noise is highly variable, improvement as well as computational savings can be obtained if adaptive filtering is correctly activated only during pauses and unvoiced intervals. This improvement is aimed for limited resources DSP operations and can be very useful in applications such as audio and hearing aids.

#### V. CONCLUSIONS

A voice activity detector is developed to control the operation of an adaptive noise canceller. The VAD is based on two features, full band energy and zero crossing rates. Residual noise from the noise canceller is used to adjust threshold of full band energy feature. Results showed that an improved performance as well as reduced computational power can be achieved with this method. The VAD controlled noise canceller achieves a cleaner output in about 50 % of the time required by the non controlled noise canceller.

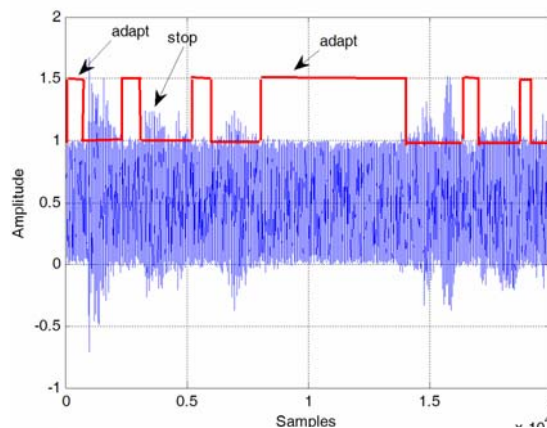


Fig. 7 Voice activity performance in the presence of noise

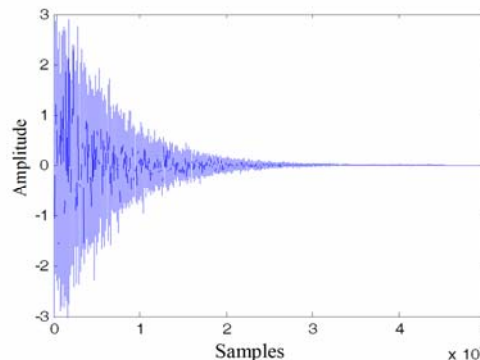


Fig. 8 Error decay after the use of controlled VAD

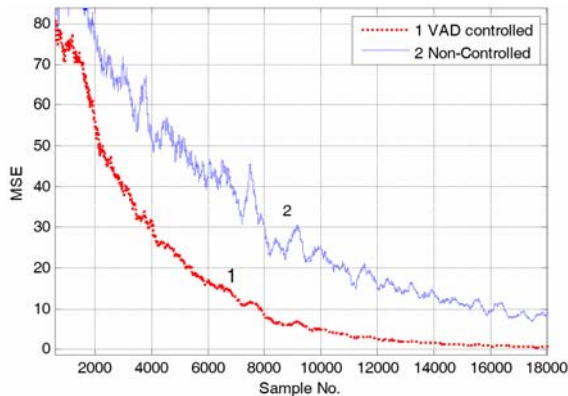


Fig. 9 Mean Square Error Comparison of controlled and non-controlled noise cancellation systems

## REFERENCES

- [1] J. C. Liberti, T.S. Pappaport and J.G. Proakis, "Evaluation of several adaptive algorithms for cancelling acoustic noise in mobile radio environments," IEEE Vehicular Technology Conference, pp126-132, May 1991. Volume 31, Number 4, August 2009. pp. 451-453.
- [2] I.C. Yoo, D. and Yook, "Robust Voice Activity Detection Using the Spectral peaks of Vowel Sounds," ETRI Journal, Volume 31, Number 4, August 2009. pp. 451-453.
- [3] W.Q. Syed and H. Chun Wu, "Speech waveform compression using robust adaptive voice activity detection for nonstationary noise," EURASIP Journal on Audio, Speech, and Signal Processing, Volume 2008. doi:10.1155/2008/639839.
- [4] D. Cournapeau, and T.Kawahara, "Voice activity detection based on high order statistics and online EM algorithm," IEICE Transactions on Information and Systems 2008 E91- D(12).pp.2854-2861.
- [5] D. L. Ying, Y. Shi, X. Lu, J. Dang and F. Soong, "Robust voice activity detection based on noise eigenspace," Acoust. Sci. & Tech. 28, 6 (2007), Publications of the Acoustical Society of Japan. pp.413-423.
- [6] A.H. Sayed, . Adaptive Filters. Wiley. NJ. 2008.
- [7] J.W. Kim, M.S. Seo, S.I. Choi and Y.G. You, "Voice activity detection algorithm for wireless communication systems with dynamically varying background noise," IEICE Trans. Communication, Vol. E83-B, No. 2 February 2000, pp.1214-1218.
- [8] B.D. Gavitt, "The VAD tree: A process-eye view of physica memory," Elsevier Digital Investigation 4S ( 2007). pp. S62-S64.
- [9] D.I. Ying, J. Dang, F. and Soong, " Robust voice activity detection based on noise eigenspace," Acoust. Sci. & Tech. 28, 6 (2007).Publications of the Acoustical Society of Japan. pp.413-423.
- [10] A. Benyassine, E. Shlomot and H. Yu Su, " ITU-T Recommendation G.729 Annex B: A Silence compression scheme for use with G.729 optimized for V.70 digital simulatenous voice and data applications," IEEE Communications Magazine, September 1997. pp.64-73.
- [11] W.G. Crouse and C.R. Knox, " Speech detector with variable threshold," United States Patent No. 4351983, Sep. 28,1982.