Sindh Univ. Res. Jour. (Sci. Ser.) Vol.46 (4):433-438 (2014)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Robust Automatic Speech Recognition in Real Life Environment using Active Noise Control

N. AHMAD, Y. S. AFRIDI, J. S. AFRIDI*, B. BALOCH**

Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar

Received 13th August, 2014 and Revised 5th November 2014

Abstract- This paper presents the implementation of a robust automated speech recognition system when operated in real life environment by using active noise control. Filtered-x Least Mean Square Algorithm is used as a primary tool for performing Active Noise Control. Previous researches carried out on ANC using FxLMS were having distinct applications. Whereas, this research evaluates the performance of FxLMS algorithm specifically for the application of Automated Speech Recognition. Experiments were conducted and it was found that the performance of the ASR system showed considerable improvements in the presence of the ANC, when tested in noisy environments.

Keywords: Automatic Speech Recognition, Active Noise Control, Robust ASR, Human Computer Interaction

1. INTRODUCTION

Attempts are being made to develop a human-machine interface to facilitate the communication among humans and machine. To develop speech operated computers which are able to understand the spoken message and can generate speech from textual inputs is the most suitable options for this purpose. The development of automatic speech recognition (ASR) and text-to-speech (TTS) systems has put this idea into a reality but the ultimate goal of achieving such natural communication with machines in the real life environment is still very much an inception. The speech recognition by humans is a very spontaneous process and is carried out subconsciously, however the key to such recognitions is the extensive processing carried out by an ultimate machine "The Human Brain". Although researchers have worked hard to imitate the processes and functionality of human brain to facilitate the automatic recognition of speech but this has posed itself to be a challenging task.

In recent times ASR has achieved an acceptable level of accuracy in controlled environment; however, its performance degrades drastically in real world environments due to distortion of the speech signal (Nefian, *et al.*, 2002). The sources for such distortion can mainly be categorized into *Additive Noise* and *Channel Distortion*. Additive noise is caused by a background noise such as engine sound, voice of other speakers, sound of a fan etc. whereas, channel distortion is caused by either reverberation or microphone's frequency response or an electric filter present in A/D circuitry. Both types of distortion contaminate the speech signal, thereby causing a mismatch in the acoustic realization of same speech among training and test conditions (Narayanan and Wang, 2014). The recognition of such noisy speech commonly encountered in real life situations has gained more attention of the researchers recently.

The various approaches adopted to address the issue of noise are, the extraction of noise robust audio features (Raj and Stern, 2005), noise adaptive models (Kalinli *et al.*, 2010), noise filtering and audio-visual ASR (Ahmad N., 2011). A popular technique among the noise filtering approaches is the Active Noise Control (ANC) technique. The ANC technique removes noise from the speech signal by using the principle of destructive interference (Elliot, 2001). The recent advancements in the digital signal processors have opened ways for the use of ANC in different application such as noise reduction headsets (Kuo *et al.*, 2006). Many distinct algorithms proposed in the literature are Lattice-ANC Systems (Park and Sommerfeldt, 1996), Filtered-u Recursive LMS (Eriksson *et al.*, 1987) and Filtered-v algorithms (Crawford and Stewart, 1997). Other such algorithms include Filtered-x RLS (FxRLS) (Kuo and Morgan, 1996) and Filtered-x Fast-Transversal-Filter (FxFTF) (Bouchard and Quednau, 2000) among the Recursive Least Square (RLS) algorithms, and the frequency-domain ANC (Kuo and Tahernezhadi, 1997).

This paper presents the application of noise filtering technique for the ASR in the real world (noisy) environment using Active Noise Control. The rest of paper is organized as follows. Section 2 discusses the active noise control. Section 3 provides a description of the Filtered X-LMS algorithm used in this study. Section 4 describes the experimental setup while

++Corresponding Author: N. Ahmad, n.ahmad@nwfpuet.edu.pk, Ph. No +92-91-9216590

*NUST College of Electrical and Mechanical Engineering, Rawalpindi, Pakistan

**Department of Statistics, Sindh Agriculture, University, Tando Jam

section 5 provides the results and analysis. Section 6 concludes the findings of this research and gives the directions for the future work.

2. MATERIAL AND METHODS

Active Noise Control

Beside a number of variabilities in the utterance of speech by different speakers, the speech is a highly structured signal and its recognition in controlled environments is no more a big issue due to the better feature extraction techniques and excellent language modeling. The real problem arises when dealing with speech recognition in real life environments where testing conditions are different from the training data due to the background noise.

The traditional passive noise attenuation silencers consist of relatively large size and are costly. Moreover, their performance declines drastically when attenuating noise at lower frequencies. These limitations of the passive noise control approaches gave rise to the concept of Active Noise Control (ANC). The active noise cancellation algorithms are adaptive in nature as the noise they are dealing with is non-stationary in nature. The ANC has shown better results for the attenuation of low-frequency noise, where the passive methods were either inefficient or tend to be very expensive.

Active noise cancellation or control attenuates the noise signal by introducing a cancelling "anti-noise" signal through the secondary sources. These secondary sources are connected through an electronic system using a specific signal processing algorithm that helps generate the exact anti-noise signal. This anti-noise signal destructively interferes with the speech signal having additive noise, thereby attenuating the noise signal and leaving behind the clean speech.

ANC Systems are based either on ***Feedback Control*** which does not require an upstream reference input in order to cancel noise or ***Feed-Forward Control*** which do require a coherent reference noise input to be sensed before its propagation to the noise cancellation speaker. Feed-Forward ANC systems can further be classified as:

Adaptive Broadband Feed-Forward Control*:* This technique is based on the principal of destructive interference where the noise canceller generates an output anti-noise signal generated in reference to a prior input signal. The output anti-noise signal is of the same amplitude as the input reference signal but 180^0 out of phase. This anti-noise signal destructively interferes with the incoming noise signal causing the noise to be attenuated. The feed-forward system makes use of the propagation time delay between the input microphone and the output speaker in order to electrically introduce the cancelling noise into the field.

Adaptive Narrowband Feed-Forward Control: This technique is used where the noise signal is periodic in nature. The input microphone in such systems is replaced by a non-acoustic device such as a tachometer, an accelerometer or an optical sensor. The sensor simulates the input signal containing the fundamental frequency of primary noise and its corresponding harmonics. The synthesized reference signal is adaptively filtered in order to generate the noise canceling signal. An error microphone is used to measure the residual error.

The idea of ANC was first proposed by Lueg in 1936; however the real application of the idea were very limited until the recent years. There are a number of algorithms used for carrying out ANC based mainly on famous LMS algorithm (Widrow and Stearns, 1985). However these algorithms have got the following problems associated with them; Infinite Impulse Response (IIR) structures have got the inherited problem of instability. Substantially the Lattice-ANC is computationally very expensive. RLS based ANCs are numerically instable.

A remedy for these problems is provided by the Filtered-x Least Mean Square (FxLMS) algorithm (Kuo and Morgan, 1996). FxLMS is the modified version of the LMS algorithm which provides a quite simple but efficient solution to the typical limitations of traditional LMS algorithms due to its fast convergence. This makes it the best choice among the various versions of LMS algorithm. In this research the FxLMS algorithm has been used for Active Noise Cancellation in ASR systems when operated in real life environment.

3. FILTERED X-LMS ALGORITHM

The **(Fig. 1)** below depicts a common Feed-forward Broadband Active Noise Control when used under a framework of system identification.

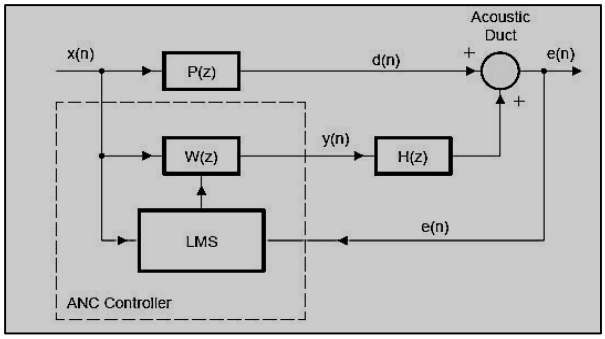

Fig. 1 .Broadband Feed-forward ANC Implementation in a System Identification Framework

In a frequency domain representation, an ideal ANC system is the one that estimates the response of a primary acoustic propagation path P(z) between the error sensor and the reference input sensor by using an adaptive filter W(z). The error signal e(n) can be expressed in z-domain by the following equation:

$$\boldsymbol{E(z) = D(z) + Y(z) = X(z)\,[P(z) + W(z)\,]} \quad (1)$$

Where;

X(z) is the input signal
W(z) is the adaptive filter
E(z) is the error signal
Y(z) is the output if the adaptive filter

Once the adaptive filter converges to its optimum value, the error signal E(z) becomes 0. Equation (1) can be re-written as:

$$\boldsymbol{E(z) = -P(z)} \quad (2)$$

Whereas, Equation (2) implies that:

$$\boldsymbol{y(n) = -d(n)} \quad (3)$$

The above equation shows that the error signal tends to become zero when the output of the adaptive filter y(n) has got the same amplitude as the primary noise d(n) but a phase shift of 180º. When both the signals are acoustically combined they interfere with each other destructively, thereby resulting in cancellation of both the signals.

The error signal e(n) is being measured at the error sensor located after the speaker that generates the anti-noise signal. The acoustical environment between the cancelling speaker, the error microphone and the primary noise source is represented by the summing junction in (**Fig. 2)** when the ANC system is operated in the real life environment. It is at this junction where the primary noise signal d(n) generated by the primary noise source, destructively interferes with the anti-noise signal y(n) generated by the adaptive filter. As the primary noise is being modified by the primary-path P(z), the anti-noise is also modified by the secondary-path (path between y(n) and e(n)) in the similar fashion. The secondary-path transfer function is given by H(z). The impact of secondary-path on the anti-noise is crucial and cannot be neglected. It is therefore incorporated in the ANC system as shown in Figure 2.

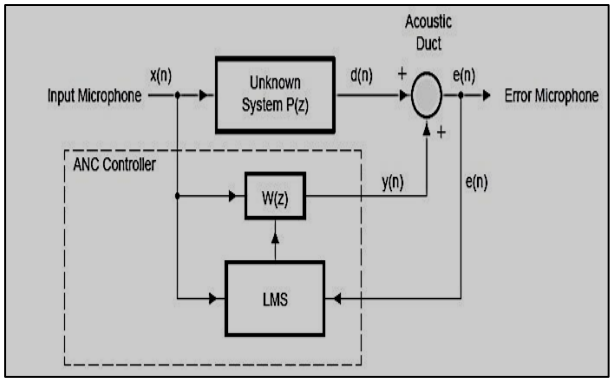

Fig. 1. Broadband Feed-forward ANC Implementation After the Incorporation of Secondary-Path Transfer Function H(z)

The modified error signal e(n) in the z-domain is given by the following equation:

$$\boldsymbol{E(z) = X(z)[\,P(z) + W(z)H(z)]} \quad (4)$$

Let us assume that the ANC system is operating in an ideal state such that the adaptive filter is converged to a point where the residual error becomes zero. Putting the value of E(z) = 0.
Equation (4) can be re-written as:

$$\boldsymbol{W(z) = \frac{-P(z)}{H(z)}} \quad (5)$$

Therefore, Equation (5) suggests that the adaptive filter W(z) has to be an exact model of the primary path P(z) and an inverse model of the secondary path H(z). This gives rise to a limitation of Broadband Feed-forward control systems that is the inherent delay caused by the secondary propagation path H(z) cannot be inverted unless and until an equal length of delay is also exhibited by the primary path P(z).

Moreover, the system exhibits unstable behavior if either of the two propagation paths contains a frequency ω such that H(ω) or P(ω) tends towards 0. This shows that the distinct characteristics of the secondary propagation path makes it an important parameter in the performance of an ANC system, hence its significance cannot be avoided.

In order to take into account the impact of secondary path transfer function, the conventional Least Mean Square (LMS) Algorithm needs to be modified. The convergence of the algorithm is ensured by inserting a filter C(z) before the input to the error microphone, where C(z) is an estimate of the secondary propagation path. This modification in the standard LMS algorithm gives rise to the Filtered-x Least Mean Square (FxLMS) Algorithm. The FxLMS Algorithm was first developed by Morgan. Burges was the first to suggest the use FxLMS Algorithm so as to compensate the secondary path effects in Active Noise Control applications.

Fig. 3 below illustrates the FxLMS Algorithm, where the output of the ANC system y(n) is computed as.

$$\boldsymbol{y(n) = w^T(n)x(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i)} \quad (6)$$

Where;

$\boldsymbol{w^T(n)}$ = Co-efficient vector of W(z) at time *n.*

$\boldsymbol{x(n)}$ = Reference input signal vector at time *n.*

The The Filtered-x LMS algorithm can be expressed as;

$$\boldsymbol{w(n+1) = w(n) - \mu e(n)x(n)h(n)} \quad (7)$$

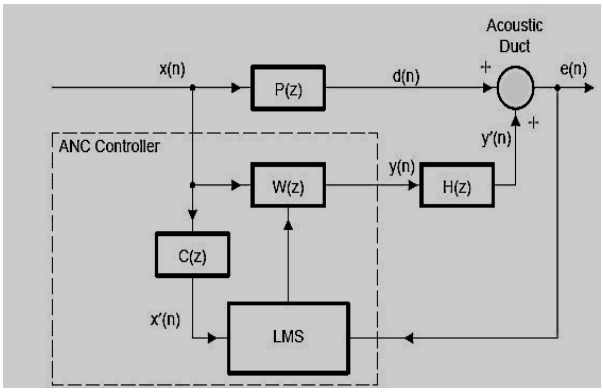

Fig.2 Block Diagram of a Filtered-x LMS Algorithm

Where $\boldsymbol{\mu}$ is the step size and $\boldsymbol{h(n)}$ is the impulse response of the secondary path H(z). The step size plays a crucial role and determines the stability and convergence speed of the algorithm. The input vector $\boldsymbol{x(n)}$ is filtered by H(z) before updating of weight vector. The secondary propagation path H(z) is not known when the ANC is implemented in real life environment, therefore it is being estimated by the filter C(z).

Now:

$$\boldsymbol{w_i(n+1) = w_i(n) - \mu e(n)x'^{(n-i)}} \quad (8)$$

Similarly:

$$\boldsymbol{w(n+1) = w(n) - \mu e(n)x'(n)} \quad (9)$$

And:

$$\boldsymbol{x'(n) = c^T x(n) = \sum_{i=0}^{M-1} c_i x(n-i)} \quad (10)$$

Where;

$\boldsymbol{x'(n)}$ is the vector for the filtered reference input signal and is computed as:

$$\boldsymbol{x'(n) = \left[x'^{(n)}\ x'^{(n-i)}\ \ldots\ldots x'^{(n-N+1)}\right]^T} \quad (11)$$

And;

$\boldsymbol{c}$ is the co-efficient vector for the estimated secondary-path, C(z) computed as:

$$\boldsymbol{c = [c_o\ c_1 \ldots\ldots c_{M-1}]^T} \quad (12)$$

4. EXPERIMENTAL SETUP

Fig. 4 depicts the block diagram of the proposed setup. The various components/units used are described below.

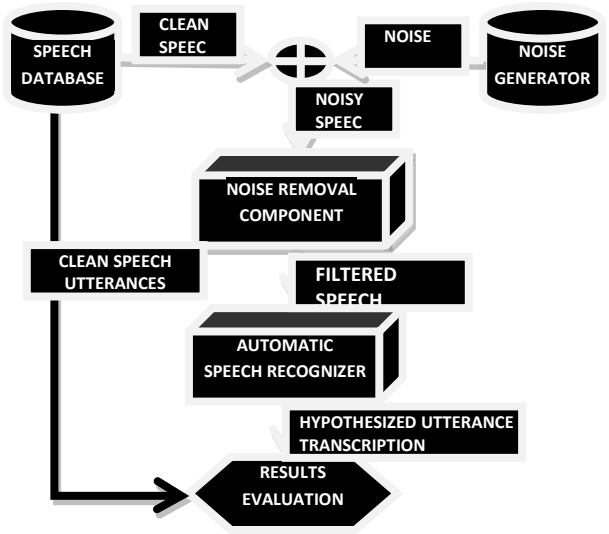

Fig. 4 Block Diagram of Experimental Setup

The **Speech Database:** Contains the pre-recorded spoken sentences and their corresponding transcriptions. The database used here is a subset of the standard TIMIT database.

Noise Generator: Generates both White and Colored Gaussian Noise.

Noise Removal Component: The Noise removal component comprises of an adaptive filter whose weights are updated by using a Filtered-x Least Mean Square (FxLMS) algorithm. This filter first estimates both the primary and secondary propagation paths and then generates an anti-noise signal that is of same amplitude as the additive noise but opposite in phase. This anti-noise signal destructively interferes with the corrupted speech signal, thereby removing noise from the speech waveform.

Automatic Speech Recognition Unit: The toolkit used for speech recognition process is the Hidden Markov Model Toolkit (HTK), developed by the Cambridge University Engi. Department (CUED).

5. RESULTS AND ANALYSIS

The experiments were conducted by first measuring the baseline performance of an automatic speech recognizer operating standalone in the presence of additive noise. The noisy speech was then filtered by the proposed FxLMS algorithm and was again input to the ASR. (**Table I)** below gives the results obtained for different SNRs and static background noise. It was observed that in case of a random background noise the algorithm showed consistent improvements in the recognition accuracy. In case of a static background noise the improvement in accuracy is even more significant at about 33%.

Table 1: Recognition performance for different SNRs

Type Of Noise	SNR LEVEL (dB)	Recognition Rate for Noisy Speech	Recognition Rate for Filtered Speech
Random Background Noise	0	16%	19%
	10	35%	41%
	20	99%	99%
	-10	18%	20%
	-20	16%	18%
Static Background Noise	--	21%	54%

6. CONCLUSION AND FUTURE WORK

The work described in this paper is based on the use of a Filtered-x Least Mean Square Algorithm for noise removal in an automatic speech recognition system. FxLMS was used as a pre-processing tool to remove noise from the distorted speech signals before it was input in the ASR system. FxLMS was used rather than other available extensions of LMS algorithm because it takes into account the effects of secondary propagation path when dealing with removal of noise. Experiments were conducted and results were compared. It was found that ASR system showed considerable improvement in the recognition rates even in low SNR conditions.

Although multichannel feed-forward ANC algorithms have can make it possible to use the ASR systems in real life environment but their slow convergence rate has always been a constraint in their practical implementation. The high parallel data processing capabilities of Graphical Processing Units (GPUs) can be exploited to implement the computationally expensive Filtered-x LMS algorithm. The parallelization of the Filtered-x LMS algorithm can be performed to implement on a GPU.

REFERENCES:

Ahmad, N. (2011) "A motion based approach for audio-visual automatic speech recognition", Doctoral dissertation, Loughborough University, UK.

Bouchard, M., and S. Quednau, (2000) "Multichannel recursive-least-squares algorithms and fast-transversal-filter algorithms for active noise control and sound reproduction systems", IEEE Transactions on Speech Audio Processing, (8), 606–618.

Burgess, J. C., (1981) "Active Adaptive Sound Control in a Duct: A Computer Simulation" J. Acoust. Soc. Am., 70(3), 715–726.

Crawford, D. H., and R. W. Stewart, (1997) "Adaptive IIR filtered-v algorithms for active noise control", Journal of Acoustical Society of America, 101(4), 2097–2103.

Elliot, S. J. (2001) "Signal Processing for Active Control", Academic Press, London, U.K.

Eriksson, L. J., M. C. Allie, and R. A. Greiner, (1987) "The selection and application of an IIR adaptive filter for use in active sound attenuation" IEEE Transactions Acoustic Speech & Signal Processing, 35(1), 433–437.

Gan, W. S., and S. M. Kuo, (2002) "An integrated audio and active noise control headsets", IEEE Transactions Consumer Electronics, 48(2), 242–247.

Kalinli, O., M. L Seltzer,. J. Droppo, and A. Acero, (2010) "Noise adaptive training for robust automatic speech recognition", IEEE Transactions on Audio, Speech, and Language Processing,18(8), 1889-1901.

Kuo, S. M., and D. R. Morgan, (1996) "Active Noise Control Systems-Algorithms and DSP Implementations", Wiley, New York, USA.

Kuo, S. M., and M. Tahernezhadi, (1997) "Frequency-domain periodic active noise control and equalization", IEEE Transactions Speech Audio Processing, (5), 348–358.

Kuo, S. M., S. Mitra, W. S. Gan, (2006) "Active noise control system for headphone applications" IEEE Transactions on Control Systems Technology, 14(2), 331–335.

Morgan, D. R., (1980) "Analysis of Multiple Correlation Cancellation Loop With a Filter in the Auxiliary Path", IEEE Trans. on ASSP, 28(4), 454–467.

Narayanan, A., and D. Wang, (2014) "Investigation of Speech Separation as a Front-End for Noise Robust Speech Recognition", IEEE/ACM Transactions on Audio, Speech & Language Processing, 22(4), 826-835.

Nefian, A., L. Liang, X. Pi, X.Liu, C. Mao, and K.. Murphy, (2002), "A coupled HMM for audio-visual speech recognition", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Orlando, Florida, USA, 2013-2016.

Park, Y. C., S. and D. Sommerfeldt, (1996) "A fast adaptive noise control algorithm based on lattice structure", Appl. Acoust., 47(1), 1–25.

Raj, B. and R. M. Stern, (2005) "Missing-feature approaches in speech recognition", IEEE Signal Processing Magazine, 22(5), 101–116.

Widrow, B., and R. M. Stearns, (1985) "Adaptive Signal Processing", Prentice Hall, New Jersey, USA.