International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:10, 2008

A Modified Speech Enhancement using Adaptive Gain Equalizer with Non linear Spectral Subtraction for Robust Speech Recognition

C. Ganesh Babu, and P. T. Vanathi

***Abstract*—**In this paper we present an enhanced noise reduction method for robust speech recognition using Adaptive Gain Equalizer with Non linear Spectral Subtraction. In Adaptive Gain Equalizer method (AGE), the input signal is divided into a number of subbands that are individually weighed in time domain, in accordance to the short time Signal-to-Noise Ratio (SNR) in each subband estimation at every time instant. Instead of focusing on suppression the noise on speech enhancement is focused. When analysis was done under various noise conditions for speech recognition, it was found that Adaptive Gain Equalizer method algorithm has an obvious failing point for a SNR of -5 dB, with inadequate levels of noise suppression for SNR less than this point. This work proposes the implementation of AGE when coupled with Non linear Spectral Subtraction (AGE-NSS) for robust speech recognition. The experimental result shows that out AGE-NSS performs the AGE when SNR drops below -5db level.

***Keywords*—**Adaptive Gain Equalizer, Non Linear Spectral Subtraction, Speech Enhancement, and Speech Recognition.

I. Introduction

AN early and fundamental method for noise reduction was to use the theory of the optimum wiener filter [1]. Given a desired signal and an input signal, the wiener filter produces the minimum mean square error estimate of the desired signal. The Wiener filter can also be adapted to a non stationary noise environment. Adaptive algorithms such as Least Mean Square (LMS) and Recursive Least Squares (RLS) are well known examples and widely used [2], [3].

Today, a frequently used digital method for effective noise reduction in speech communication is spectral subtraction [4], [5]. This frequency domain method is based on Fast Fourier Transform and is a non linear, yet straight forward way of reducing unwanted broadband noise acoustically added to the signal. The noise bias is estimated in frequency domain during speech pauses and then subtracted from the noisy speech spectra. The quality of the noise bias estimate is crucial and VAD is required to detect the voice activity. Multi microphone techniques such as microphone arrays, have also been used in order to suppress a disturbance both spatially and temporally by means of adaptive beamforming and signal separation [6], [7], [8].

C. Ganesh Babu is with (Research Scholar, PSG College of Technology, Coimbatore) Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India (e-mail: ganeshbabuc@bitsathy.ac.in).

P. T. Vanathi is with Department of ECE, PSG College of Technology, Coimbatore, Tamil Nadu, India (e-mail: ptvani@yahoo.com).

II. Adaptive Gain Equalization

The Adaptive Gain Equalization (AGE) method for speech enhancement, introduced by Westerlund et al., [9] separates itself from the traditional methods of improving the SNR of a signal corrupted by noise, through moving away from noise suppression and focusing primarily on speech boosting. Noise suppression traditionally, like spectral subtraction, looks at subtracting an estimated noise bias from the signal corrupted by noise. Whereas speech boosting aims to enhance the speech part of the signal by adding an estimate of the speech itself, thus boosting the speech part of the signal. The difference between noise suppression and speech boosting is presented in Fig. 1. Fig. 1 (a) shows a noise estimate being subtracted form a noise corrupted signal. While in Fig. 1. (b), an estimate of the speech signal is used to boost the speech in the noise corrupted.

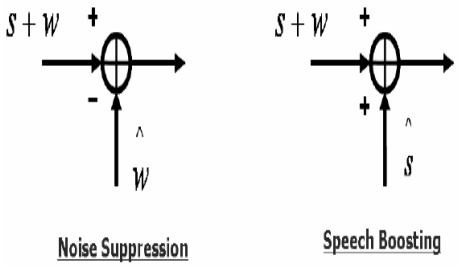

Fig. 1 Difference between Noise Suppression and Speech Boosting

III. Concept of Adaptive Gain Equalization

The concept of obtaining a speech bias estimate to perform speech boosting may seem like a daunting task. But it does not need to be, the AGE method of speech enhancement relies on a few basic ideas. The first of which is that a speech signal which is corrupted by bandlimited noise can be divided into a number of subbands and each of

these subbands can be individually and adaptively boosted according to a SNR estimate in that particular subband.

In each subband, a short term average is calculated simultaneously with an estimate of a slowly varying noise floor level. By using the short term average and floor estimate, a gain function is calculated per subband through dividing the short term average by the floor estimate. This gain function is multiplied with the corresponding signal in each subband to form an output per subband. The sum of the outputs from each subband forms the final output signal, which should contain a higher SNR when compared to the original noisy signal.

The proposed method of the AGE acts as a speech booster, which is adaptively looking for a subband speech signal to boost. Fig. 2 shows this underlying concept behind the AGE. Outlining that speech energy is a highly non-stationary input amplitude excursion, if there is no such excursions no alteration to the subband will be performed, the AGE will remain idle, as a result of the quotient between the short term magnitude average and the noise floor estimate being unity, with them being approximately the same. If speech is present the short term magnitude average will change with the noise floor level remaining approximately unchanged, thus amplifying the signal in the subband at hand due to the quotient becoming larger than unity.

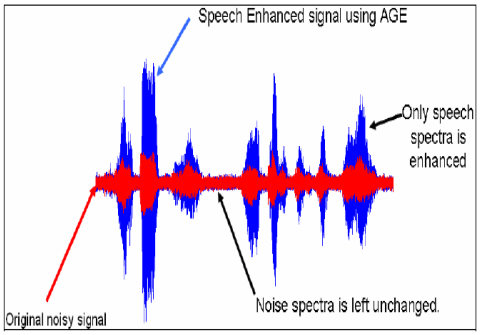

Fig. 2 Speech signal enhanced using AGE

IV. Some Insight about Algorithm Age

During periods of no speech activity, using the AGE provides distortion free background noise during speech activity due to masking effects. This results in increased speech quality with the output signal having a natural sound with minimum distortion and artifacts.

- The AGE algorithm can be implemented either on digital or analogue circuits proving to be versatile and flexible.
- The speech enhancement is performed continuously in each subband, which means no voice activity detectors are required.
- The method is Stand-Alone; it works independently of different speech coding schemes or other adaptive algorithms.
- Using the AGE requires minimum amendment for good performance

V. An Illustration of the Adaptive Gain Equalizer

The previous two section of the paper have outlined the fundamentals of the Adaptive Gain Equalizer is used in speech enhancement. To show the potential of using the AGE method a brief example will be demonstrated in this chapter. A speech signal which is corrupted by white noise is presented in Fig. 3.

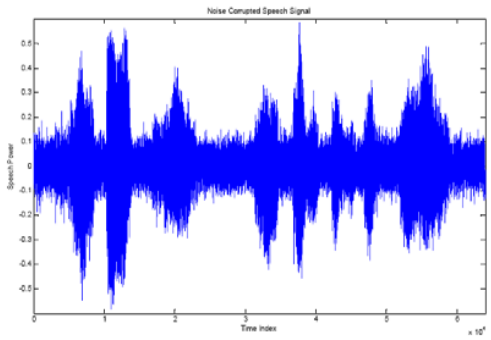

Fig. 3 Speech signal corrupted by white noise

The first step requires the signal to be filtered into number of subbands. In this example, number of subbands is chosen to be eight. The signal which is sampled at 16KHz is filtered into eight subbands which are shown in the Fig. 4. From the Fig. 4, it is clear majority of speech is concentrated in first subbands, which is expected since human speech is generally assumed to be between 300 Hz and 3400 Hz and is expected to dominate in the subbands corresponding to this frequency range. Short term exponential magnitude average and noise floor is taken simultaneously and is shown in the Fig. 5. Using the short term exponential magnitude average and noise floor the gain is calculated, in this example the gain is limited to 5 dB and is displayed in Fig. 6. It is evident from Fig 6 that the AGE algorithm amplifies only the components of the signal which contain speech and remains idle when there is no speech component.

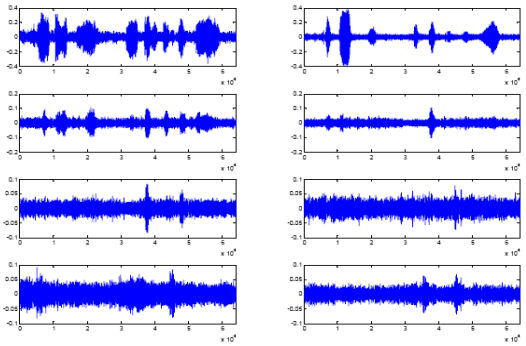

Fig. 4 Speech signal corrupted by white noise and is filtered into eight subbands

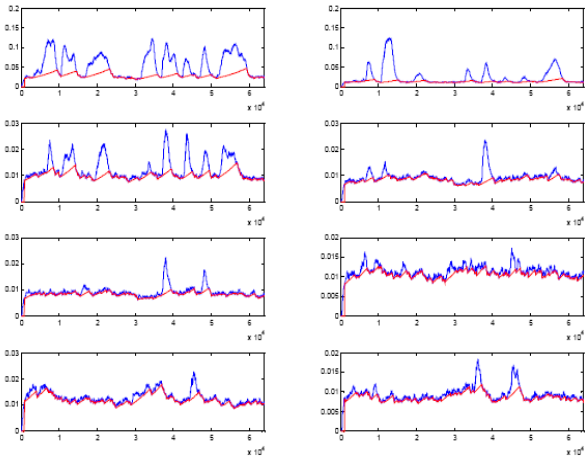

Fig. 5 Short term exponential magnitude average and noise floor per subband

The gain per subband is shown in Fig. 6 and is multiplied with its corresponding and then summed to form a speech enhanced signal and is shown in Fig. 7.

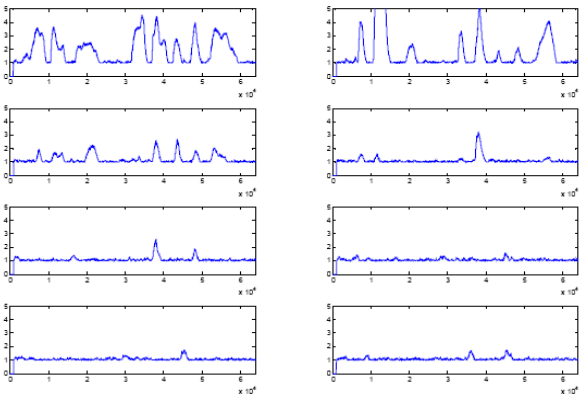

Fig. 6 Gain calculated per subband

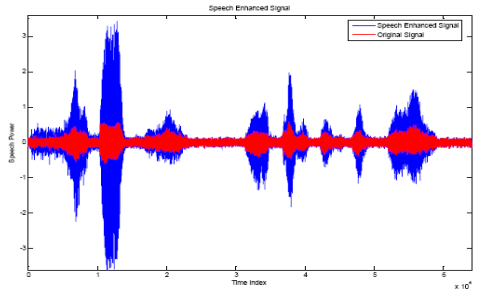

Fig. 7 Speech enhanced signal using 8 subbands

VI. Drawbacks of Adaptive Gain Equalization

When analysis was done under various noise conditions for robust speech recognition it was identified that the algorithm has an obvious failing point for a SNR of -5 dB, with inadequate levels of noise suppression for SNR less than this point. This was shown as result of the short term average failing to track the speech spectra of a speech signal which is heavily corrupted by noise.

VII. Nonlinear Spectral Subtraction

The basics of nonlinear spectral subtraction techniques (NSS) reside in the combination of two main ideas [10]:

- The noise-improvement model is used which is obtained in the course of a speech pause.
- The nonlinear subtraction is used when a frequency-dependent signal-to-noise ration (SNR) is obtained. This means that in spectral subtraction a minimal subtraction factor is high SNR is used in turn.

VIII. Proposed Speech Enhancement Algorithm

AGE when coupled with Non linear Spectral Subtraction (AGE-NSS) performs better than AGE when SNR drops below -5db. The first step requires the signal to be filtered into number of subbands. In this paper, number of subbands is chosen to be eight. The signal which is sampled at 16KHz is filtered into eight subbands. Non linear spectral subtraction is applied to each of the subband. Short term exponential magnitude average and noise floor is taken simultaneously. Using the short term exponential magnitude average and noise floor the gain is calculated and it is multiplied with the spectra.

IX. Results and Discussions

The proposed framework uses a speech processing module including a Speech Enhancement Algorithm with Hidden Markov Model (HMM)-based classification and noise language modeling to achieve effective noise knowledge estimation.

In the training phase, the uttered words are recorded using 8-bit pulse code modulation (PCM) with a sampling rate of 8 KHz and saved as a wave file using sound recorder software. The Automatic speech recognition systems work reasonably well under clean conditions but become fragile in practical applications involving real-world environments.

Table I shows the recognition performance achieved by the AGE and AGE-NSS were compared.

TABLE I
Average Word Accuracy

SNR in db	AGE	AGE-NSS	% of Improvement
5	41.5	50.75	18.23
0	35.5	41.25	13.94
-5	30.25	36.25	16.55
-10	21.75	29.5	26.27
Average	**32.25**	**39.4375**	**18.22**

X. Conclusion

100 words were taken for speech recognition (using isolated word recognition with statistical modeling - Hidden Markov Model), after adding various noise environments. We have analyzed input word utterance (FOUR) under the most commonly encountered noise environments. We have proposed a method for combining the Adaptive Gain Equalizer method and Non linear Spectral Subtraction, so that improved speech recognition accuracy performance can be obtained under these noise conditions. Comparative experimental results are shown in the figure through Fig. 8. to Fig. 11 against AGE & AGE-NSS, the speech recognition accuracy for AGE-NSS performs better that AGE for all cases.

The proposed Algorithm has the advantage of providing word recognition accuracy grater than 18% in all noise environments. Better performance is achieved in the 5 db noise added in all cases.

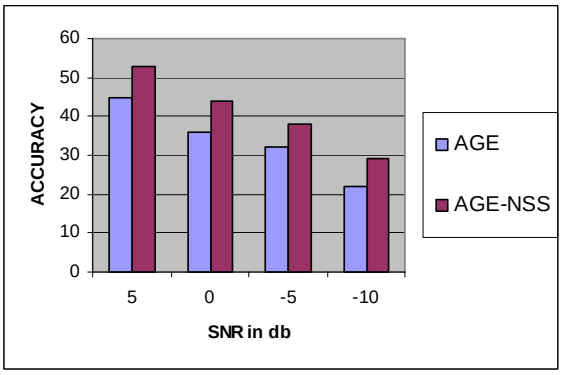

Fig. 8 Babble Noise added with speech signal

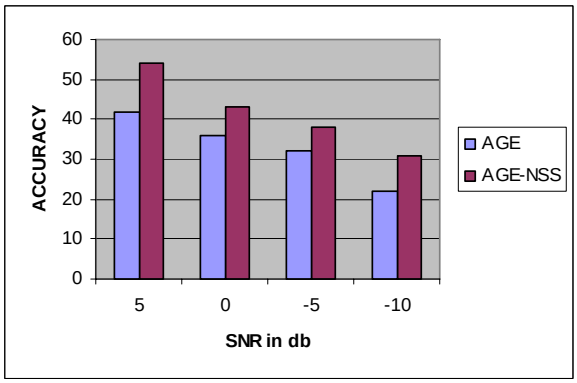

Fig. 9 Gun Noise added with speech signal

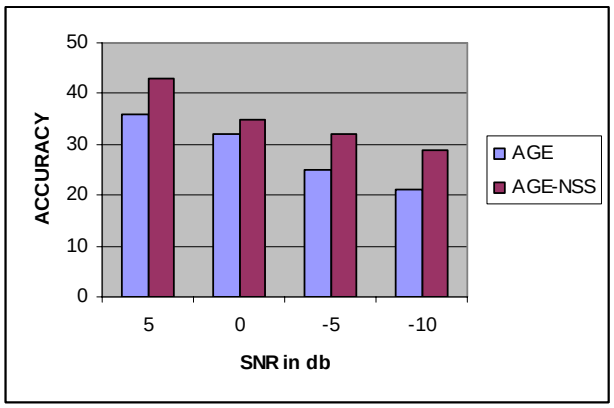

Fig. 10 Leopard Noise added with speech signal

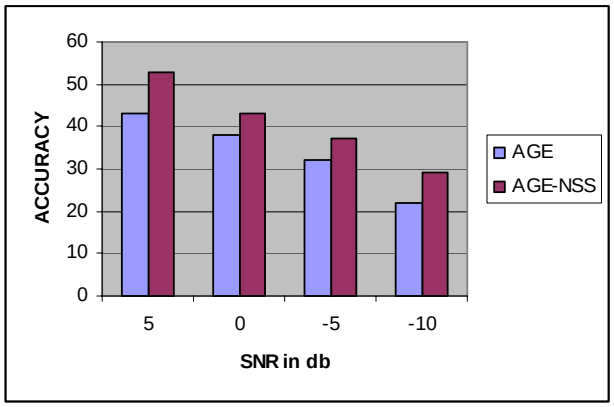

Fig. 11 Car Noise added with speech signal

Acknowledgment

Firstly, the author would like his thanks to the Supervisor, Dr. P.T.Vanathi, Professor, Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore. The author would like to express his thank to the Management and Principal of Bannari Amman Institute of Technology, Sathyamangalam. The author greatly expresses his thanks to all persons whom will concern to support in preparing this paper.

References

[1] M.H. Hayes, Statistical Digital Signal Processing and Modelling, Wiley, 1996.
[2] B.Widrow and S.D. Stearns, Adaptive Signal Processing, New Jersey, Prentice-Hall, 1985.
[3] S. Haykin, Adaptive Filter Theory, New Jersey, Prentice-Hall, 1996.
[4] S.F.Boll, "Suppression of acoustic nose in speech using spectral subtraction", IEEE Trans. Acoust. Speech and Sig. Proc., vol. ASSP 27, pp. 113-120, April 1979.
[5] J.R. D. Jr., J.G.Proakis, and J.H.L Hansen, Discrete time processing of speech signals. Macmillan Publishing Company, 1993.
[6] Y.Kaneda and J.Ohga, "Adaptive microphone-array system for noise reduction", IEEE Trans. Acoust. Speech and Sig. Proc., vol. ASSP 34,no 6, pp. 1391-1400, December 1986.

[7] M.Dahl and I.Claesson, "Acoustic noise and echo canceling with microphone array", IEEE Trans. On Vehicular Technology, Vol.48, no.5, pp. 1518-1526, September 1999.
[8] B.D.Veen and K.M Buckley, "Beamforming: A versatile approach to spatial filtering", IEEE ASSP Magazine, 1988
[9] N. Westerland, M. Dahl., and I.Classeon, "Speech Enhancement Using An Adaptive gain Equalizer", in Proceedings of DSPCS, December 2003.
[10] J.Poruba, "Speech enhancement based on nonlinear spectral subtraction", Proceedings of the Fourth IEEE International Conference on Devices, Circuits and Systems, pp T031-1 - T031-4, April 2002.

C. Ganesh Babu received the BE degree in Electronics and Communication Engineering, ME degree in Microwave and Optical Engineering in the year 1994, 1997 and 2002 respectively from PSG College of Technology, Coimbatore, and Alagappa Chettiar College of Engineering and Technology, Karaikudi Tamil Nadu, India. Currently, he is pursuing his PhD in Anna University, Chennai. His area of interest includes Speech Signal processing and Virtual Instrumentation. He is currently working as an assistant professor in the ECE department of Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India. He is having around 10 years of teaching and research experience. He is a life member of ISTE. He is having 14 national and International conference publications.

P. T. Vanathi received the BE degree in Electronics and Communication Engineering, ME degree in computer science and engineering and Ph.D in the year 1985, 1991 and 2002 respectively from PSG College of Technology, Coimbatore, Tamil Nadu, India. Her area of interest include Soft computing, Speech Signal processing and VLSI Design. She is currently working as an assistant professor in the ECE department of PSG College of Technology, Coimbatore, Tamil Nadu, India. She is having around 22 years of teaching and research experience. She is a life member of ISTE. She is having 20 National and International Journal publications and 50 national and International conference publications.