Oriental Journal of Computer Science & Technology Vol. **4**(1), 165-169 (2011)

Speech Enhancement Using Neural Network

SYED MINHAJ ALI[1] and BHAVNA GUPTA[2]

[1]M-TECH (Computer Science) RGPV University, Bhopal (India).
[2]Department of Computer Science, All Saint`s College of Technology, Bhopal (India)

(Received: March 15, 2011; Accepted: April 30, 2011)

ABSTRACT

This paper describes a neural network speech enhancement system using Multilayer Perceptron (MLP) network and trained using the back propagation algorithm (BPA). Speech enhancement is generally refers to map noisy speech into cleaner speech. Noisy speech signals are obtained by adding random noise to the clean signals .Speech enhancement is then performed on the noisy signals by using the ADALINE. Here we show that neural nets can be used to significantly boost recognition accuracy, without retraining the speech recognizer.

Key words: Speech processing, adaptive neural network, MLP for Noise Reduction.

INTRODUCTION

The main objective of this system is to enhance the speech signal to obtain a clean signal with higher quality. The signal-processing problem of noise reduction and speech enhancement has received considerable attention within the adaptive filter community. Such system has been widely used in long distance telephony applications. One of the most challenging areas of this research is the development of adaptive algorithms for hearing aids. A closely linked problem, which has been the focus of research in recent years from the artificial neural network community (ANN), is that of blind separation of sources and in particular the convolutive blind source separation problem.

The choice of Adaptive Linear Neuron (ADALINE) to perform the task of noise cancellation is made based on two considerations, that is, its ability to act as an adaptive filter and the high processing speed that it can provide. ADALINE is certainly an approach worth to be explored based on the fact that it is the most widely used neural network approach in practical applications today . Its fast processing speed is contributed by its simple network architecture with minimum elements. Performance of the ADALINE system will be best evaluated if it can be compared to the performances of other systems. Thus, an alternative speech enhancement system is constructed using the well-known Multilayer Perceptron (MLP) network and trained using the backpropagation algorithm (BPA).

Neural network have existed for a long time and have recently enjoyed a resurgence of interest in their application to speech processing.

Adaptive linear neuron(adaline) as an adaptive filter

Adaptive Linear Neuron (ADALINE) is most commonly use in the field of signal processing. Based on the learning principal of the ADALINE, weight and bias of the network are adapted by using the Least Mean Squares (LMS) or Widrow-Hoff rule. The architecture of the network is relatively simple as compare to larger network such as the famous backpropagation network. It only consists of a single neuron with a single connecting weight and bias.

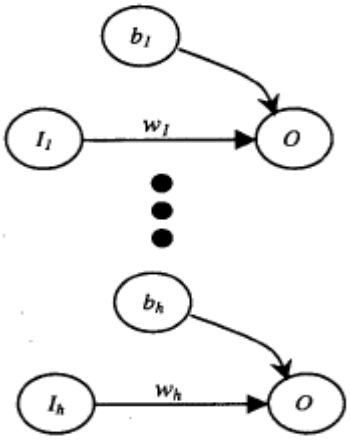

Fig. 1: Structure of the ADALINE for noise cancellation

However, to perform the task of noise cancellation, the simple architecture of the ADALINE described above is slightly modified. Instead of using a single weight and a bias to compute each output of a time sequence, each samples of a time sequence will be computed by using its own set of weight and bias. The architecture described is shown in fig. 1, where *h* is the length of the sequence.

The noise cancellation algorithm is presented below

Step **1** **:**Set learning rate, H - The learning rate is set at 0.35, which was determined experimentally.

Step 2 :Set weights {w(i)} and biases {b(i)) at 0.1015 and-0.0527 respectively.

Step 3 : Set the target signal {T } as the corrupted speech signal.

Step 4 **:**Set the input signal,{ **I }** as the noise signal.

Step 5 **:** For each time index, i, compute the output { 0) and error {E) of the network by using the following
equations.

O(i)= **w(i)** *I(i)+ b(i)

E(i) = **T(** i)-O(i)

Step 6 : Adjust the weights { w(i)} and biases { b(i)} of the
network.

w(i)= w(i)+ 0.0 I *(H*E(i) ***I(i))**

b(i)=b(i)+O.OI *(H*E(i))

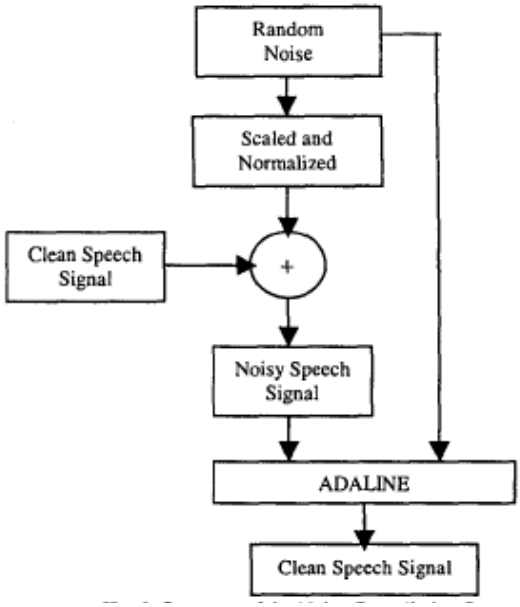

Fig. 2: Structure of the Noise Cancellation System

Implementation structure

The implementation of this experiment is carried out using the ADALINE in the time domain. A three layer multi-layer perceptrone network is used in this paper to classify one phoneme sound using a training set of isolated words spoken by different speakers. Training is done by using backpropogation algorithm. A total of 100 speech samples consisting of 10 utterances for each digit in the Malay Language are used as the test utterances. All utterances are sampled at 8KHz sampling rate. The scaled and normalized random noise is added to the clean speech signal to produce corrupted speech signal as shown in fig. 2 [2]. The corrupted speech signal is then used as the target signal to the ADALINE while raw random noise is used as the input signal. Based on the LMS rule, the ADALINE adapts to cancel out noise from the noisy signal to produce the clean speech signal. In this structure, clean speech signal is obtMLP neural network trained with BPA has been widely used in commercial applications. In this paper, the MLP is used to build a Noise Reduction Network to perform speech enhancement [4]. The MLP consists of three layers of neurons with two hidden layers. Each layers, inclusive of the input and output layers, has 40 neurons respectively.

All neurons in both of the hidden layers compute their respective outputs based on the tangent sigma transfer function. Initial learning rate and momentum of the network are set at 0.01 and 0.95 respectively. All initial weights and biases are initialized at small random values.

The recognition systems work best for high quality, "close-talking" speech and require consistent environments for training, testing and operation. Speech recognition refers to the art of producing graceful performance degradation when training and testing data set conditions differ. Performance improvements achieved with neural networks. A

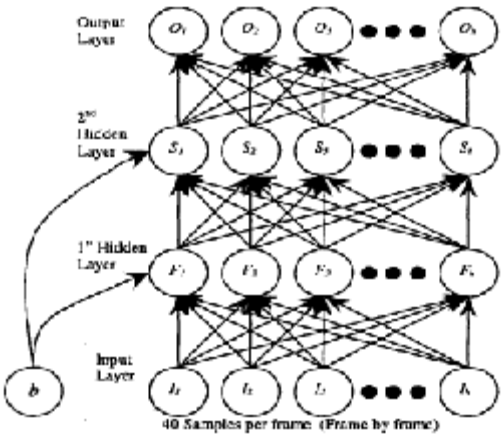

Fig. 3: MLP for Noise Reduction Network

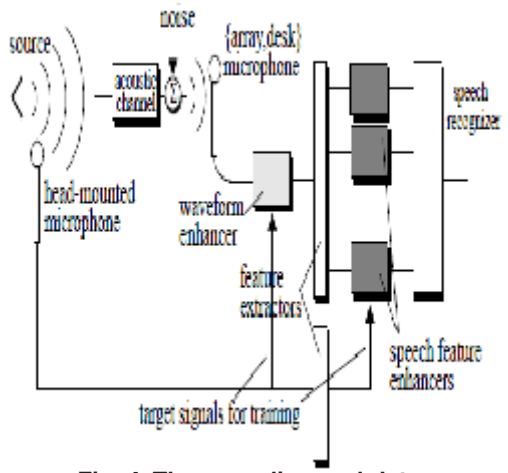

Fig. 4: The recording and data processing configuration

continuous speech recognition system, developed by Carnegie Mellon University. Before the received speech data is fed to the speech recognizer, a feature extractor converts the speech waveform signal into a cepstral vector signal, the feature extractor produces a cepstral feature vector at a frequency of 0.1 KHz. As is shown in Figure 4, there are two locations in this configuration where a speech enhancement filter can be placed.

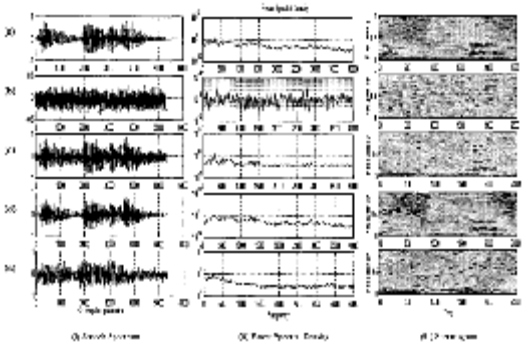

Fig. 4: Various Plots of (a) Original Signal (b) Random Noise (c) Noisy Signal (d) ADALINE Enhanced Signal and (e) MLP-Enhanced Signal

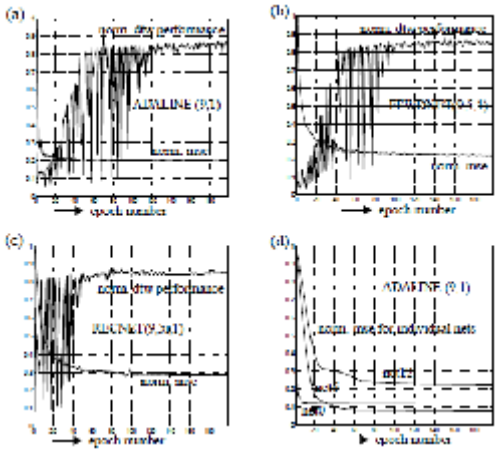

Fig. 5: Results on DTW word accuracy using different neural nets as speech feature enhancers. All curves are normalized between 0 and 1. Neural net training proceeded in batch mode, with adaptive learning rates (and no momentum term). The DTW performance curves are relative to the test set. The MSE curves are with respect to the training data set

Table 1: SNR for both Noisy and Enhanced Speech

Digit Uttered	Noisy Speech	ADALINE	MLO
0 (sifar)	4.41	34.51	1.97
1 (satu)	5.05	35.03	0.74
2 (dua)	8.05	37.11	3.44
3 (tiga)	4.70	34.71	3.17
4 (empat)	5.95	35.70	0.31
5 (lima)	6.60	36.16	4.97
6 (enam)	4.52	34.61	2.76
7 (tujuh)	6.98	36.39	4.76
8 (lapan)	6.60	36.12	1.35
9 (sembilan)	6.25	35.86	2.30

If a speech enhancement filter is placed before the feature extractor, we might anticipate that a linear equalizer would suffice. The speech enhancement filter can also be positioned after the feature extractor, a configuration which we call (speech) feature enhancement. The task of the enhancement filter seems easier than in the case of waveform enhancement. The feature extraction stage performs irreversible operations such as time Averaging as well as non-linear operations . While the ultimate goal of speech enhancement is to improve recognition performance (in terms of word recognition accuracy), we cannot use the word accuracy, a piecewise constant function, as a training cost function for the speech enhancements filters.

RESULTS

Based on visual and audio checking, results obtained from the experimental implementation are significant. From the speech enhancement system via the ADALINE method, noise from the corrupted signals has been reduced to the minimum level while maintaining all phonetic quality of the signals. Results obtained from this system are very much similar to the original signal. However, results of the MLP system are quite disappointing. It is observed that the noise level has increased and deteriorated the audio quality of the signal. The results are as shown in fig. 5 in terms of its spectrum, power spectral density and spectrogram views.

From this it can be concluded that the ADALINE showed superior ability in terms of the *SNR* results when compared to the MLP. Moreover, the ADALINE net only requires a minimal amount of processing time. Additionally, through visual and audio checking, it can be seen that the enhanced signal produced by the *ADALINE* is similar to the original signal.

CONCLUSION

The ultimate goal of speech enhancement is to improve recognition performance (in terms of word recognition accuracy), we cannot use the word accuracy, a piecewise constant function, as a training cost function for the speech enhancements filters.The suitability of a neural network based noise cancellation system using ADALINE net has been highlighted. Such implementation is a good example of a system that has taken full advantage of the adaptive nature of ADALINE. It also successfully demonstrated its fast noise cancellation process in real time implementation. In short, the ADALINE net has demonstrated its potential to perform the noise cancellation task and more research for further improvement should be performed.

REFERENCES

1. B. Widrow and M. A. Lehr.30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation. Proceedings of the IEEE, **78**(9): 1415-1422 (1990).
2. R. P. Lippman. An Introduction to Computing With Neural Nets. Acoustics, Speech and Signal Processing Magazine, vol. 4, no. 2, pp. 4-22 (1987).
3. L. Fausett, “Fundamentals of Neural Networks : Architectures, Algorithms and Applications” pp. 80-88 (1994).

4. D.P. Morgan and C.L. Scofield, "Neural Networks and Speech Processing" pp. 183-190 (1991)
5. Gong Y. and Treurniet W.C., Speech recognition in noisy environments :a survey, Technical report CRC-TN 93-002, Communications ResearchCentre, Department of Communications, Ottawa (1993).
6. Erell A. and Weintraub M., Filterbank-energy estimation using mixture and Markov model recognition of noisy speech, *IEEE transactions on speech and audio processing*, **1**(1): (1993).
7. Che C., Lin Q., Pearson J., de Vries B., and Flanagan J., Microphone arrays and neural networks for robust speech recognition, *proceedings of ARPA workshop on Human Language Technology*, Plainsboro, NJ (1994).
8. Haykin, S. Adaptive filter theory, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, (1991).
9. Toner, E., Campbell,D.R.**,** 'Speech Enhancement using Sub-Band intermittent adaptation' , *Speech Communication*, **12**: 253-259 (1993).
10. Shields, P.W., Campbell D.R.,', Multi-Microphone Sub-Band Adaptive Signal Processing For Improvement of Hearing Aid Performance: Preliminary Results Using Normal HearingVolunteers', Proc. ICASSP-97, *I.E.E.E Conference on Acoustics, Speech and Signal Processing*, **1**: 415-418 (1997).
11. Chan, D, C, B., Godshill, S, J. and Rayner, P, J, W. , Multichannel Multi-tap Signal Separation By Output Decorrelation,Cambridge University, CUED/F-INFENG/TR 250, ISSN 0951-9211 (1996).
12. Cichocki, A., Amari, S, I., and Cao, J. Blind Separation of Delayed and Convolved Signals with Self-Adaptive Learning Rate*, International Symposium on Nonlinear Theory and Applications*, pp. 229-232 (1996).
13. Charkani, N., and Deville, Y., Optimisation of the Asymptotic Performance of Time-Domain Convolutive Source Separation Algorithms., *Proc European Symposium on Artificial Neural Networks*, pp 273 – 278, ISBN 2-9600049-7-3 (1997).
14. Nguyen Thi H L., Jutten C., Blind Source Separation forConvolutive Mixtures, *Signal Processing*, **45**(2): pp 209 – 229 (1995).
15. Lee, T,W., Bell, A.J., and Orgmeister, R. Blind Source Separation of Real World Signals. In Proc. *I.E.E.E / I.C.N.N, International Conference on Neural Networks,* **4**: 2129 - 2134 (1997).