

ADAPTIVE NON-LINEAR NETWORK FILTER ESTIMATION ERROR FOR STEREO ECHO CANCELLATION IN HOME THEATRE 9.1 SURROUND SOUND SYSTEM

*Sunisa Kunarak

Department of Electrical Engineering, Srinakharinwirot University, Nakhonnayok, Thailand 26120

* Corresponding Author, Received: 02 Dec. 2017, Revised: 20 Dec. 2017, Accepted: 11 Jan. 2018

ABSTRACT: In this paper, an adaptive non-linear network filter (ANLNF) approach based on Radial Basis Function Neural Networks (RBFNNs) is proposed for the stereo echo cancellation that is a necessary process for reducing undesired signal owing that the audiences can receive the apparent sound signal. The Gaussian activation function is suitable in used to model the characteristic of room transfer function. The samples of the direct sound and the echo sound signal in home theatre are applied as the input for the adaptive non-linear network filter. Finally, the simulation results illustrate the predicted error between the actual sound and direct sound, the Echo Return Loss Enhancement (ERLE) and Mean Square Error (MSE) in order to guarantee the clarity sound signal. We observe that the proposed algorithm outperforms compared with the other methods as Adaptive Filter with Gain and Time-Shift, Wiener Adaptive Filter, Feedforward Network and Average Recursive Least Square, respectively.

Keywords: Adaptive Non-Linear Network Filter, Echo Return Loss Enhancement, Mean Square Error, Radial Basis Function Neural Networks, Stereo Echo Cancellation

1. INTRODUCTION

The echo is a phenomenon that is delayed and distorted of the original signal which is reflected back to the source as the voice signals as shown in Fig. 1. Stereo Echo Cancellation (SEC) is a common approach in removing the echo and noise path. The normalized least mean square (NLMS) algorithm is the most popular method since this one is a simple structure and lower computational time complexity. However, the performance of this algorithm is essentially deteriorated for correlated input speech signals [1-2].

To overcome in echo cancellation, the affine projection algorithm (APA) that reuses the input vector to accelerate the convergence speed in [3]. Unfortunately, the performance of this method is depend on the step size i.e. if the step size is small, the approach can get a smaller misalignment but has a slower convergence speed. In the other hand, when the step size is large, the algorithm can achieve a faster convergence but has a bigger misalignment. The Average Recursive Least Squares (ARLS) is an adaptive filter which recursively finds the coefficients that minimize a weighted linear least squares cost function relating to the input signals in the frequency domain. The goal of this approach is to decrease the mean square error. As specified the previously the memory of the ARLS algorithm, the input signals are considered deterministic so that the ARLS exhibits extremely fast convergence but has the limited number of

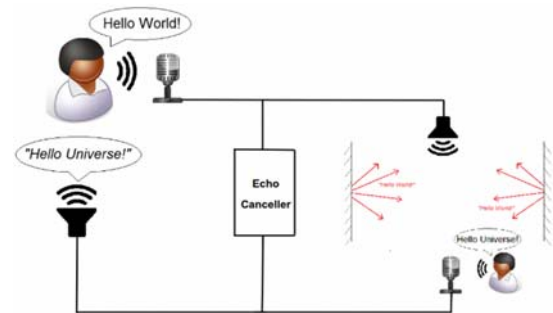


Fig.1 Direct and echo sound signal paths

values, relating to the order of the filter tap weight vector [4].

Additionally, the frequency-domain adaptive filter [5-6] can reduce the distinction between the expected and actual output filter is also known as error signal but this method is not suitable for non-statistical information situation due to unpredictable environments. To enhance the performance, we propose the adaptive non-linear network filter that is based on radial-basis function neural networks. This proposed algorithm can control the tendency of mean square error (MSE) and echo return loss enhancement (ERLE).

This paper is organized as follows: the adaptive non-linear based on radial-basis function network filter is summarized in the next section. After that, the stereo echo cancellation using adaptive radial-basis function network filter is explained. Next, the simulation environment and simulation results is

presented. The conclusion is finally given in the last section.

2. ADAPTIVE NON-LINEAR BASED ON RADIAL BASIS FUNCTION NETWORK FILTER

In this section, the radial-basis function neural networks (RBFNNs) based on Probabilistic Neural Networks (PNN) is applied the stereo echo cancellation (SEC) process. The radial basis networks may require more neurons than the standard of feed-forward backpropagation networks, but often we can be designed in a fraction of the time it takes to train standard feed-forward networks. We work the best when many training vectors are available. This approach is suitable for abundant non-linear sound signal.

The basic structure of probabilistic neural networks can be used for classification problems. When an input is presented, the first layer computes distances from the input vector to the input weights (i.e. transpose of the training input vectors), and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a competitive transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes. The architecture for this system is shown in Fig. 2

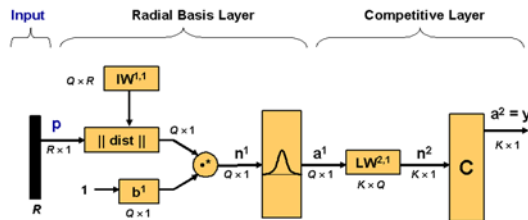


Fig.2 Probabilistic neural networks structure

where Q is the number of input/ target pairs in 1st layer,

K is the number of classes of input data in 2nd layer, respectively.

It is assumed that there are Q input vector/ target vector pairs. Each target vector has K elements. One of these element is 1 and the rest is 0. Thus, each input vector is associated with one of K classes. The first-layer input weights, $IW^{1,1}$ are set to the transpose of the Q training pairs matrix, P' .

The second-layer weights, $LW^{2,1}$ are set to the matrix T of target vectors. Each vector has a 1 only in the row associated with that particular class of input, and 0's elsewhere. The multiplication sums

the elements of a^1 due to each of the K input classes. Finally, the second-layer transfer function, produces a 1 corresponding to the largest element of n^2 , and 0's elsewhere. Thus, the network has classified the input vector into a specific one of K classes because that class had the maximum probability of being correct [7].

The speech voice that is direct sound, is sampled equal to 4,096 samples as the input nodes, the non-linear Gaussian are the transfer function in the hidden layer to connect them to all of the input nodes. Finally, the output layer consists of one node that is obtained by a linearly weighted sum of the outputs of the hidden units that are updated in the training procedure as illustrated in Fig. 3. The output of this network is the error that is minimized less than 0.001.

In order to adjust minimized the error, the network will find the optimal weight, $w = [w_1, w_2, \dots, w_N]^T$ is learned as the following steps [8].

1. Set the initial value of center (μ_{ji}) in the hidden layer for the i^{th} input node and the j^{th} hidden node and the value of span (σ_j) in the j^{th} hidden node, respectively. Also, the initial weight vector which is distributed uniformly between 0 and 1.
2. Calculate the output of hidden layer as:

$$z = \exp\left(-\|\bar{x} - \bar{\mu}_{ji}\|^2 / 2\sigma_j^2\right) \quad (1)$$

where $\bar{x} = [x_1, x_2, \dots, x_{i-1}, x_{5000}]^T$ represents the input vector.

3. Calculate the output of output layer as:

$$y_k = \sum_{j=0}^M w_{kj} z_j, \quad k = 1; M = 10 \quad (2)$$

4. Calculate the error of this network:

$$e_k = d_k - y_k \quad (3)$$

where d_k is the desired output.

5. Update the weight as:

$$w_{kj}(n+1) = w_{kj}(n) + \eta_w (e_k) z_j \quad (4)$$

where η_w is the learning rate of weight.

6. Update the center momentum with the learning rate of center as:

$$\mu_{ji}(n+1) = \mu_j(n) + \eta_\mu \frac{z_j}{\sigma_j} (x_i - \mu_{ji}) \sum e_k w_{kj} \quad (5)$$

7. Update the span with the learning rate of span (η_σ) as:

$$\sigma_j(n+1) = \sigma_j(n) - \eta_\sigma z_j \frac{2}{\sigma_j} \ln z_j \sum e_k w_{kj} \quad (6)$$

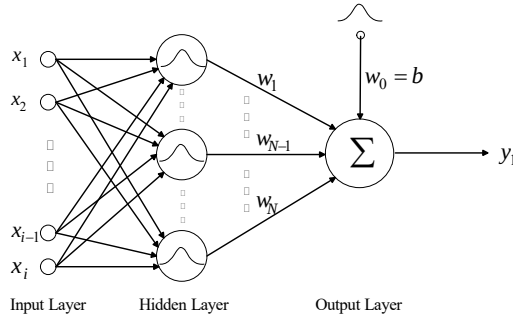


Fig.3 Adaptive radial-basis function neural network structure

In our implementation, the training process use 1,050 samples of sound and 450 samples are used to test the network. The mean square errors of actual output and desired output are calculated in each simulating epoch.

3. STEREO ECHO CANCELLATION USING ADAPTIVE RADIAL BASIS FUNCTION NETWORK FILTER

The speech direct sound is the sound signal from source to audiences which the users require its since it is obviousness without noise. Unfortunately, the resultant signal heard by audience would be low quality because the speech signal is always to combine with the noise or echo signal and reverberation as a surrounding noise. So, we use the adaptive non-linear network filter to get rid the echo sound signal in owing to remain the clarity speech voice. We can do this an adaptive non-linear network filter if we get a sample of the surroundings noise and apply it as the input to the adaptive non-linear network.

Here we adaptively train the non-linear network filter to predict the combined speech voice and environment signal as m from a surroundings noise as n . Notice that the surroundings signal does not tell the adaptive non-linear network filter anything about the speech signal contained in m . However, the direct sound signal does give the adaptive non-linear network filter information which it can use to predict the surrounding's contribution to the speech voice and environment signal as illustrated in Fig. 4.

The network will do its the best to adaptively

output m . In this case, the network can only predict the surroundings interference noise in the speech voice and environment signal. The network error (e) is equal to m , that means the speech voice and surroundings signal will be minus the predicted contaminating surrounding noise signal. Thus, the network error contains only the speech voice. Finally, the adaptive non-linear network filter adaptively learns to cancel the surroundings noise.

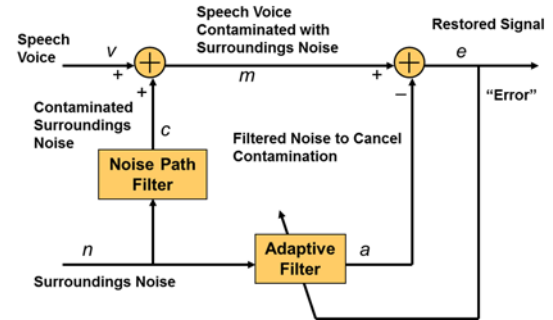


Fig.4 Echo cancellation using ANLNF process

4. SIMULATION

In our simulation, we divide into 2 sub-section i.e. the simulation environment and simulation results, respectively.

4.1 Simulation Environment

The major disadvantages of the home theatre occurs the reverberation and echo sound signal. In order to solve this problem, we propose an efficient methodology for get rid the echo and reverberation signal. In our simulation environment, the home theatre 9.1 surround sound signal characteristic is set that composes the 9+2 loudspeakers and 1 sub-woofer (SW) together with the room feature has a width $\times$ length $\times$ height size is equal to 12 meter $\times$ 20 meter $\times$ 6 meter, respectively.

The position of loudspeakers are fixed in the center (C) of the audience and under the television, the front loudspeakers in the left hand side (FL) and the right hand side (FR) are angle 30 degree. The surround in the left hand side (SL) and surround in the right hand side (SR) are arranged in the audience's axis of 60 degree. The front wide in the left hand side (FWL) and the right hand side (FWR) are managed in the middle between the front loudspeakers and surround speakers that adjust 60 degree. Furthermore, the surround back in the left hand side (SBL) and surround back in the right hand side (SBR) are set 60 degree from the listener and finally, the front high both left and right also known as FHL and FHR are fixed straight with front loudspeakers in the y-axis as shown in Fig. 5. The

sound signal that is generated 30 seconds of music, is sampled 4,096 points for each input pattern with 16 kHz sampling frequency and stored to memory for off-line processing [9-10].

4.2 Simulation Results

In the simulation results, we consider the performance of our proposed approach in the sound signal error estimation, mean square error of signal and echo return loss enhancement, respectively.

4.2.1 Sound signal error estimation

Figure 6 illustrates the estimated sound can be tracks nearly the actual signal that refer the adaptive non-linear based on radial basis function network filter can recognize the direct and echo sound.

4.2.2 Mean square error

Accordingly, we observe that the Adaptive Non-Linear Network Filter (ANLNF) which is a proposed algorithm outperforms the other algorithms as the Average Recursive Least Squares (ARLS) [4], Wiener Adaptive Filter (Wiener) [11], Feedforward Network (FF) [12] and Adaptive Filter with Gain and Time-Shift (AFGT) [13] in Mean Square Error (MSE) value as illustrated in Figs. 7-11. Figure 7 shows the speech voice or original signal (top), echo and noise signal (middle) and restored signal (button) of the proposed approach that is nearly the speech voice. Also, it has a MSE equal to 1.6206×10^{-6} which is less than the Average Recursive Least Squares, Adaptive Filter with Gain,

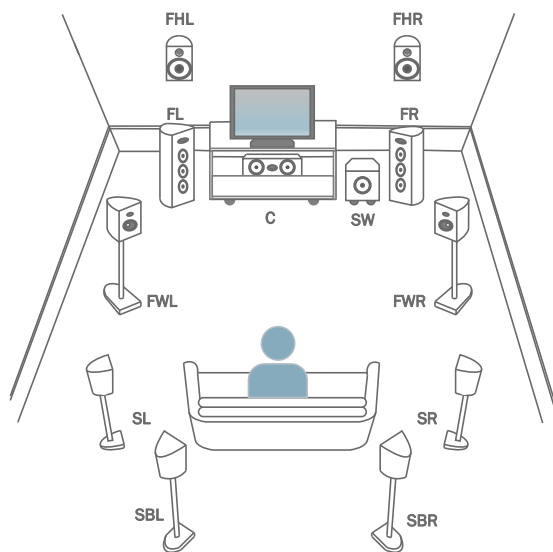


Fig.5 Home theatre 9.1 surround sound

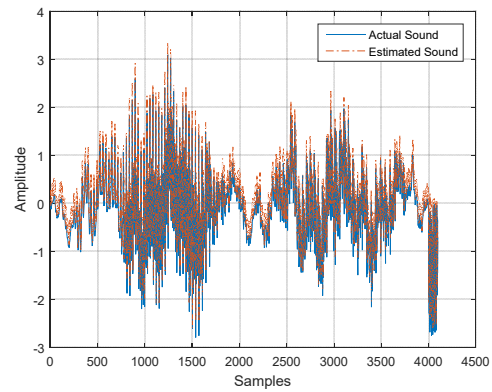


Fig.6 Estimated and actual sound signal

and Time-shift, Feedforward and Wiener methods because the proposed approach is suitable in non-stationary systems and numerous non-linear signal while the Average Recursive Least Squares remains constant coefficient factor in Fig. 8. The Wiener Adaptive Filter depends on the window size that is disadvantage of this one as shown in Fig. 9. Moreover, the Feedforward network is a one type of neural networks but this algorithm is improper to distinguish the non-linear sound signal as illustrated in Fig. 10. Finally, the Adaptive Filter with Gain and Time-shift is relatively stable systems as displayed in Fig. 11.

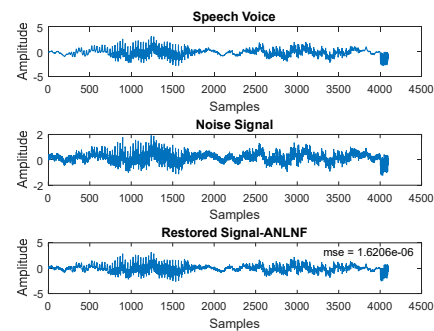


Fig.7 Restored signal with adaptive non-linear network filter

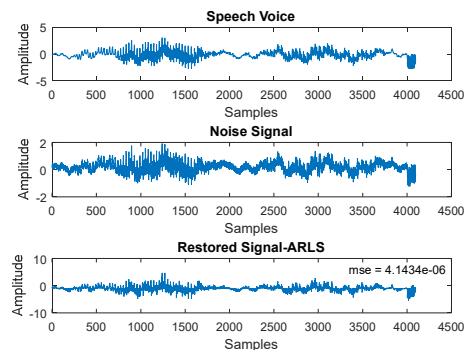


Fig.8 Restored signal with average recursive least squares

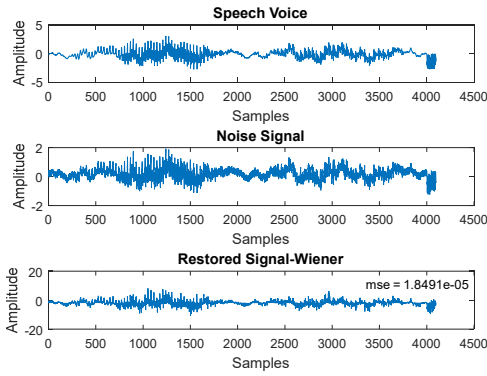


Fig.9 Restored signal with Wiener adaptive filter

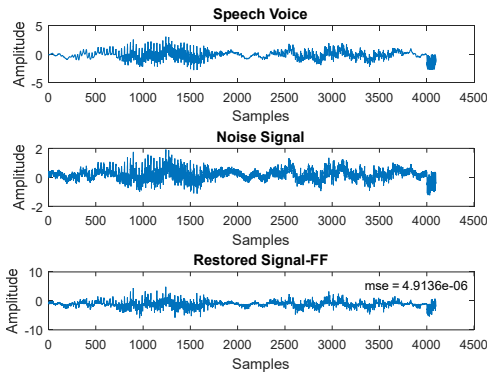


Fig.10 Restored signal with feedforward network

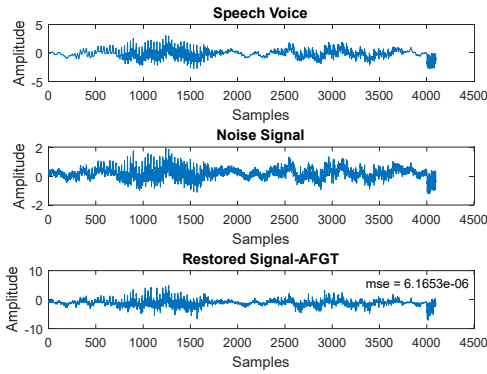


Fig.11 Restored signal with adaptive filter with gain and time-shift

4.2.3 Echo return loss enhancement

To ensure the quality of proposed method, the echo return loss enhancement (ERLE) is measured in dB that defines as the ratio of the expectation instantaneous power of the signal ($E[d(n)]$) and the expectation instantaneous power of the residual ($E[e(n)]$) as following

$$ERLE = 10 \log \frac{E[d(n)]^2}{E[e(n)]^2} \quad (7)$$

The coverage echo return loss enhancement of the ANLNF approach is increased around 10 dB with comparing the other methods. Furthermore, the proposed algorithm is not limited in high power i.e. the ERLE augments despite the high power in order that it has a recursive structure while the ERLE of the other algorithms are dropped at larger than 0.1 watt as explained in Fig. 12.

5. CONCLUSION

The Adaptive Non-Linear Network Filter (ANLNF) that is based on Radial Basis Function Neural Networks (RBFNNs) as Probabilistic Neural Networks (PNN) for home theatre 9.1 surround sound characteristic. The proposed method can recognize the speech direct sound and echo signal. The simulation results indicate the proposed algorithm that increases efficiency compared with the other four approaches by increasing the coverage echo return loss enhancement. Also, this algorithm can decrease the mean square error (MSE), respectively.

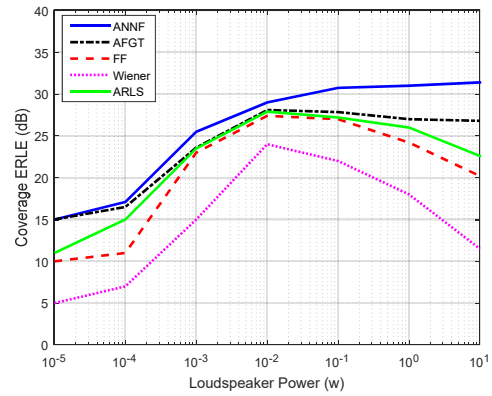


Fig.12 Coverage ERLE versus loudspeaker power

6. REFERENCES

- [1] Qin H. H. and Zhang L., Adaptive feedback cancellation based on variable step-size affine projection for hearing aids, The Journal of China Universities of Posts and Telecommunications, Vol. 22, 2015, pp. 6-10.
- [2] Strasser F., Adaptive feedback cancellation for realistic hearing aid applications, IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 23, 2015, pp. 2322-2333.
- [3] Zhu M. and Zhang L., APA with Evolving Order and Variable Step-size for Echo Cancellation, IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, 2016, pp. 841-844.

- [4] Praveen. N, Ranjitha S. and Suresh H. N., Novel based Method to Adaptive Algorithm for Acoustic Echo Cancellation of Speech Signal in an Auditorium, *International Journal of Engineering and Innovative Technology*, Vol. 4, Issue 12, 2015, pp. 164-170.
- [5] Praveen N., Ranjitha S. and Suresh H. N., A frequency-domain adaptive filter (FDAF) prediction error method and ARLS for speech echo cancellation, *2nd International Conference on Applied and Theoretical Computing and Communication Technology*, 2016, pp. 182-187.
- [6] Wu S., Qiu X. and Wu M., Stereo acoustic echo cancellation employing frequency-domain preprocessing and adaptive filter, *IEEE Transactions on Speech and Audio Processing*, Vol. 19, 2011, pp. 614-623.
- [7] Haykin S., *Neural Networks and Learning Machines*. New York: Prentice Hall, 2008, pp.68-73.
- [8] Kunarak S., Vertical Handover Decision based on RBF Approach for Ubiquitous Wireless Networks, *International Conference on Platform Technology and Service*, 2016, pp. 1-4.
- [9] Müller M., Janský J., Boháč M. and Koldovský Z., Linear Acoustic Echo Cancellation using Deep Neural Networks and Convex Reconstruction of Incomplete Transfer Function, *IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics*, 2017, pp. 1-6.
- [10] Fukui M., Shimauchi S., Hioka Y., Nakagawa A. and Haneda Y., Double-talk Robust Acoustic Echo Cancellation for CD-quality Hands-free Videoconferencing System, *IEEE Transactions on Consumer Electronics*, Vol. 60, No. 3, 2014, pp. 468-475.
- [11] Dahanayaka I. J., Kulasekara P. K. Y. N., Zihra M. N. F. and Chamindu S. H. G., Acoustic Echo Cancellation for Hands-free Applications using Adaptive Filters, *Annual Technical Conference 2010 of IET-YP Sri Lanka*, 2010, pp. 1-4.
- [12] Birkett A. N. and Goubran R. A., Acoustic Echo Cancellation for Hands-free Telephony using Neural Networks, *Proceedings of the IEEE Workshop Neural Networks for Signal Processing*, 2002, pp. 1-11.
- [13] Zhang Z. and Wu Z., An Adaptive Filter with Gain and Time-Shift Parameters for Echo Cancellation, *10th International Symposium on Chinese Spoken Language Processing*, 2016, pp. 1-5.

7. AUTHOR'S BIOGRAPHY

Sunisa Kunarak was born in Chonburi, Thailand, on December 21, 1980. She received the B.S. degree in Electrical Engineering (Electronics and Telecommunication) and M.S. degree in Electrical Engineering from King Mongkut's University of Technology Thonburi, Thailand in 2002 and 2004, respectively. She is currently a lecturer of digital logic and mobile wireless communications at Srinakharinwirot University in Thailand where she heads the digital laboratory. Her current research interests include mobile communications and broadband wireless networks.

Copyright © Int. J. of GEOMATE. All rights reserved, including the making of copies unless permission is obtained from the copyright proprietors.
