ACOUSTIC ECHO CANCELLATION USING NLMS-NEURAL NETWORK STRUCTURES

A. N. Birkett, R. A. Goubran
Department of Systems and Computer Engineering
Carleton University, 1125 Colonel By Drive
Ottawa, Canada, K1S 5B6
Tel: (613) 788-5747, Fax: (613) 788-5727
e-mail: birkett@sce.carleton.ca

ABSTRACT

One of the limitations of linear adaptive echo cancellers is nonlinearities which are generated mainly in the loudspeaker. The complete acoustic channel can be modelled as a nonlinear system convolved with a linear dispersive echo channel. Two new acoustic echo canceller models are developed to improve nonlinear performance. The first model consists of a time-delay feedforward neural network (TDNN) and the second model consists of a memoryless neural network followed by an adaptive Normalized Least Mean Square (NLMS) structure. Simuations demonstrate that both neural network based structures improve the Echo Return Loss Enhancement (ERLE) performance compared to a linear NLMS acoustic echo canceller. Experimental results using the TDNN improved the ERLE by 10 dB at low to medium loudspeaker volumes.

1.0 INTRODUCTION

Limitations of echo cancellers [5][7] include (a) acoustic, thermal and DSP related noise, (b) under-modelling of the room impulse response (c) slow convergence and dynamic tracking, (d) nonlinearities in the transfer function caused mainly due to the loudspeaker, and (e) resonances and vibration in the plastic enclosure.

In this paper, a tapped delay line feedforward neural network and a cascaded neural network/NLMS structure are employed in an attempt to model the system nonlinearities and acoustic path in a hands-free environment. Since there is no feedback in the network, the backpropagation algorithm [6] is used to train the networks.

A typical handsfree terminal is illustrated in Figure 1 and normally consists of two Adaptive Filters (AF). The first AF is used to remove acoustic echos and the second AF is used for cancelling echoes from an imperfect hybrid as well as reflections from the line. In this paper, only the acoustic echo canceller (AEC) is considered.

1.1 Distortions in the Loudspeaker

A loudspeaker has several sources of nonlinearity including non-uniform magnetic field and nonlinear suspension system [1][3]. A loudspeaker consists of an electrical part and a mechanical part. The electrical part is the voice coil and the mechanical part consists of the cone, the suspension system and the air load. The two parts interact through the magnetic field resulting in a nonlinear force deflection characteristic f_M of the loudspeaker cone suspension system, usually approximated [3] by;

$$f_M = \alpha x + \beta x^2 + \delta x^3 \qquad (1)$$

where α, β and δ are modelling constants and x is the displacement of the voice coil. Suspension system nonlinearity manifests itself as soft clipping at the loudspeaker output and results in odd-order harmonics under large signal conditions. The nonlinear distortion consists mainly of cubic terms and can easily be 5 to 10 percent of the total output, especially when dealing with small loudspeakers that have low power ratings.

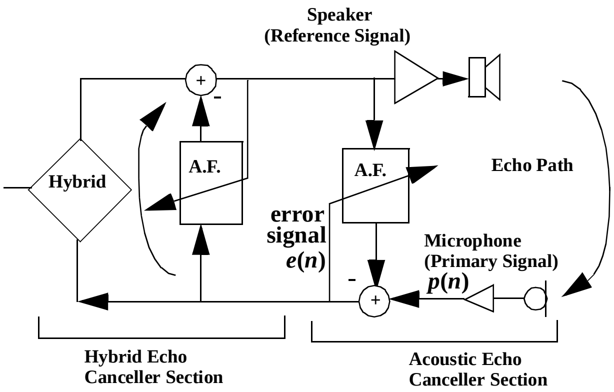

FIGURE 1. Adaptive Acoustic Echo Canceller Structure. The hybrid echo canceller is also shown for reference. Variables *p*(*n*) and *e*(*n*) are the primary and error signals.

2.0 CONVENTIONAL ADAPTIVE ECHO CANCELLER MODELS

Conventional AECs utilize a linear adaptive transversal filter to model the room impulse response and cancel the echo signal. The NLMS algorithm [9] is the baseline by which performance of alternative models is measured but it is incapable of reducing nonlinear distortion. A measure of the AEC performance is the Echo Return Loss Enhancement (ERLE) which is defined as;

$$ERLE(dB) = \lim_{N \to \infty}\left[10\log\frac{E[p^2(n)]}{E[e^2(n)]}\right] \cong 10\log\left[\frac{\sigma^2_p}{\sigma^2_e}\right] \qquad (2)$$

where σ^2_p and σ^2_e refer to the variances of the primary and error signals respectively and E is the statistical expecation operator.

Adaptive volterra filtering can be utilized to deal with loudspeaker nonlinearities [3][8], however, filter orders greater than 3 are required to effectively model the speaker transfer function and this very quickly leads to an unmanageably huge model [8]. Neural networks offer an alternative method of dealing with high order system nonlinearities.

3.0 NEURAL NETWORK ECHO CANCELLER MODELS

Two adaptive AEC networks were constructed. The first model utilizes a fully adaptive 3 layer feedforward Time Delay Neural Network as shown in Figure 2. The inputs are obtained from a tapped delay line. This model is referred to as the TDNN model.

.

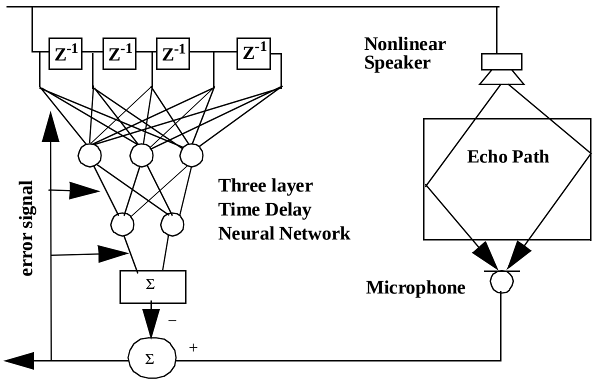

FIGURE 2. Tapped Delay Line Neural Network Adaptive Echo Canceller Structure (TDNN).

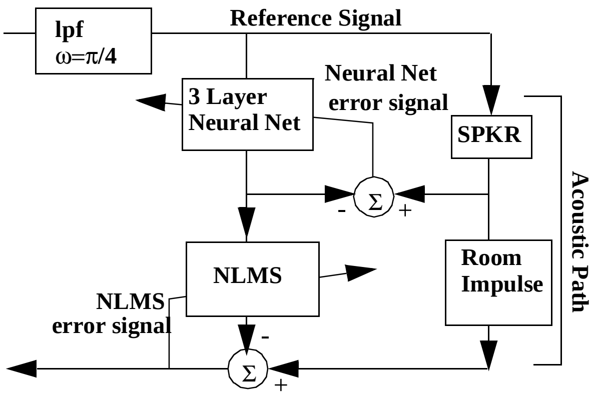

FIGURE 3. Cascaded neural network/NLMS acoustic echo canceller structure (CASC).

The second model shown in Figure 3 consists of a cascaded neural network and linear transversal filter. The transversal filter part in this model is trained by the NLMS algorithm. This model requires an intermediate training signal which represents a microphone placed directly in front of, and in close proximity to the loudspeaker. This model is referred to as the CASC model

In both neural network models a piecewise linear-sigmoid activation function is used in order to mimic the soft clipping effect and is shown in Figure 4 along with its corresponding delta function.

.

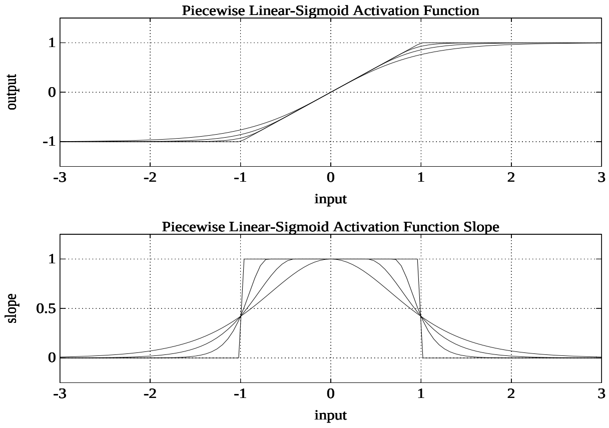

FIGURE 4. .Piecewise linear-sigmoid activation function and corresponding delta. The linear section with a value of ± 0.2 gave the best results in this study.

The transfer function is linear below a user definable point and then follows a compressed hyperbolic tangent sigmoid beyond this point such that the output is squashed between ± 1.0. The linear region was set to ± 0.2 in the TDNN model since it was found that this gave good results. For the CASC simulation, it was left at zero.

In both the proposed structures there is no feedback hence the backpropagation algorithm [6] is employed to train the networks. A normalized step size [9] is employed during the training and tracking phase. The stepsize μ is updated after each new sample is shifted into the tapped delay line. In addition, momentum is used during training such that each tap update consists of a fraction of the previous tap weight.

4.0 COMPUTER SIMULATIONS

Simulations were performed using a computer generated white noise source as the reference which was then filtered and convolved with an artificial room impulse function similar to the configuration of Figure 3. The reference and primary files were then applied to the corresponding algorithms. For each run, the reference signal is distorted by adding both quadratic and cubic distortion according to (3).

$$y = \frac{ax + bx^2 + cx^3}{|a| + |b| + |c|} \qquad (3)$$

where a, b, and c refer to the amplitude of the linear, quadratic and cubic factors, x is the input signal and y is the output signal level. The coefficients b and c were increased such that the distortion level increases relative to the primary signal level. The signal to distortion ratio is calculated by dividing the variance of the undistorted signal portion by the variance of the distorted signal portion. For each run, the algorithm was allowed to converge and then a mean converged ERLE was obtained.

4.1 Simulation Results

The converged ERLE levels of the NLMS algorithm, TDNN and cascaded network CASC are shown in Figure 5. The NLMS performs well when the signal to distortion ratio is large (i.e. little distortion). At these signal levels, there is a small amount of distortion added by the nonlinear nodes of the neural network resulting in worse performance than a purely linear algorithm. However, at higher distortion levels, the TDNN and CASC structures have better performance. The CASC structure performs better than the TDNN structure since it more closely resembles the acoustic channel model.

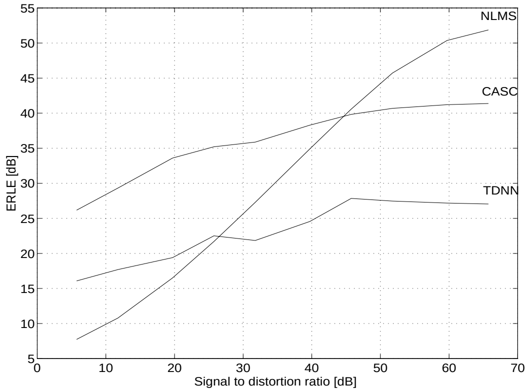

FIGURE 5. Comparison of the NLMS, TDNN and CASC algorithms for various quadratic and cubic distortion levels.

5.0 EXPERIMENTAL TEST RESULTS

A commercially available speakerphone was purchased and modified to allow access to internal signals. The modified speakerphone is placed inside an anechoic chamber. Filtered “reference” signals are applied to the loudspeaker and the microphone picks up the reflected or “primary” signal. Both the reference and primary data signals are recorded on a Digital Audio Tape and later sampled at 16 kHz and stored to disk for off-line processing.

The NLMS algorithm with 600 taps is applied to the measured data and a number of ERLE curves are obtained for various speaker volume levels. The algorithm is allowed to converge for 32000 samples and then the average ERLE is obtained from the last 8000 output values. The results illustrated in Figure 6, show that the converged ERLE is low for low speaker volumes where acoustic, thermal and DSP related noise are significant. This agrees with results presented in [5] and [7]]. The ERLE increases as the reference signal increases but reaches a plateau. Any increase in reference signal level to the loudspeaker after this point results in a decrease in the ERLE

Also shown in Figure 6 is the performance of a fully adaptive (600,2,2,1) TDNN structure. The improvement in ERLE over the NLMS case is significant in the low to medium volume ranges and is greater than 10 dB at power levels in the vicinity of 1mW.

.

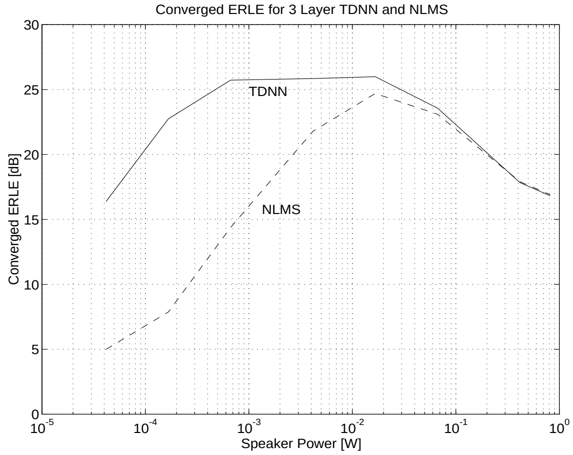

FIGURE 6. Experimental Results. Converged ERLE plot vs. loudspeaker volume using a three layer fully adaptive TDNN. The three layer network (solid line) shows over 10 dB improvement in ERLE at low to medium volumes. The NLMS algorithm (dashed line) is shown for comparison.

6.0 DISCUSSION OF RESULTS

The cascaded neural network structure shows promising potential and achieves a higher ERLE than the TDNN in the computer simulations. This is due to the fact that the cascade structure more closely resembles the acoustic channel. Both the proposed neural based structures improve the ERLE at high distortion levels at the expense of increased computational burden. For low distortion levels, the NLMS algorithm is the preferred structure. The results of Figure 6 show that the TDNN does not offer significant ERLE improvement at high speaker volumes suggesting that there still exists a deficiency in the modelling of the room/speakerphone transfer function at these volume levels. The filtering of the primary and reference signals also limits the distortion products such that the channel appears “linearized”. This will limit the amount of distortion the adaptive structure can model, and ultimately cancel. This is being considered for future study.

7.0 SUMMARY

Nonlinear distortions and undermodelling have been found to limit the converged ERLE of acoustic echo cancellers in handsfree terminals. Loudspeaker distortions include nonlinearity in the suspension system which will result in soft clipping at high volumes. A piecewise linear/tanh-sigmoid activation function has been developed to more accurately model the soft clipping effect. Two different NLMS-neural network based models have been developed. Results obtained from simulation indicate that improvements in ERLE can be achieved over that obtainable with the NLMS algorithm alone. Results obtained from experimental data using a handsfree terminal show a 10 dB improvement in converged ERLE can be obtained in the low and medium volume ranges which are usually used in handsfree phones.

8.0 REFERENCES

[1] H.F. Olsen, *Acoustical Engineering*, Toronto, D. Van Nostrand Company,Inc., 1964.

[2] X.Y. Gao, W. M. Snelgrove, “Adaptive Nonlinear Recursive State-Space Filters”, *I.E.E.E. Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 41, No. 11, Nov. 1994, pp. 760-764.

[3] X. Y. Gao, W. M. Snelgrove, “Adaptive Linearization of a Loudspeaker”, *ICASSP* 1991 Vol. 3, pp 3589-3592.

[4] O. Nerrand, P. Roussel-Ragot, L. Personnaz, G. Dreyfus, “Neural Network Training Schemes for Nonlinear Adaptive Filtering and Modelling”, *IJCNN* 1991 pp I-61 to I-67.

[5] M.E. Knappe, *Acoustic Echo Cancellation: Performance and Structures*, M. Eng. Thesis, Carleton University, Ottawa, Canada, 1992.

[6] Y. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley Publishing Company, Inc, 1989.

[7] M.E. Knappe, R.A. Goubran,”Steady State Performance Limitations of Full-Band Acoustic Echo Cancellers”, *ICASSP* 1994, Adelaide, South Australia, Vol. 2, pp. 73-76.

[8] C. E. Davila, A. J. Welch, H.G. Rylander, “A Second Order Adaptive Volterra Filter with Rapid Convergence”, *I.E.E.E. Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-35, No. 9, Sept. 1987, pp. 1259-1263.

[9] S. Haykin, *Adaptive Filter Theory*, 2nd ed., Prentice-Hall, Toronto, 1991.

[10] P. Chang, C. Lin, B. Yeh, “Inverse Filtering of a Loudspeaker and Room Acoustics Using Time-delay Neural Networks”, *Journal of the Acoustic Society of America*, Vol 95, No. 6, June 1994, pp. 3400-3408.

[11] A.N. Birkett, R. A. Goubran, “Acoustic Echo Cancellation for Hands-free Telephony Using Neural Networks”, *Neural Networks for Signal Processing 1994, IEEE Workshop Proceedings*, Sept. 1994,pp. 249-258.

9.0 ACKNOWLEDGEMENTS

The authors wish to thank NSERC, Carleton University, BNR and the Telecommunications Research Institute of Ontario (TRIO) for their financial support.