Nonlinear Loudspeaker Compensation for Handsfree Acoustic Echo Cancellation

A. N. Birkett and R. A. Goubran

Department of Systems and Computer Engineering,

Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada

Indexing Terms: Acoustic Echo Cancellation, Nonlinear Adaptive Filters, Neural Networks.

ABSTRACT

Loudspeaker nonlinearity at high volumes limits the achievable echo cancellation performance in linear acoustic echo cancellers. A new nonlinear adaptive filter for improving the echo cancellation performance at high volumes for handsfree telephones is proposed. Experimental measurements show that an echo cancellation improvement of over 8 dB can be obtained at high volumes as compared to a linear adaptive filter.

1.0 INTRODUCTION

Acoustic echo cancellers (AEC) for handsfree telephony must be capable of identifying and tracking a changing Acoustic Impulse Response (AIR) which includes a room transfer function, a nonlinear loudspeaker and other components as shown in Figure 1. Conventional AECs utilize a linear adaptive transversal filter to model the AIR and cancel the echo signal, however, this architecture is incapable of reducing nonlinear distortion. This letter proposes a new structure to mitigate the effects of loudspeaker nonlinearity. A commonly used AEC performance metric is Echo Return Loss Enhancement (ERLE) which is defined as [1];

$$ERLE(dB) = \lim_{N \to \infty}\left[10\log\frac{E[p^2(n)]}{E[e^2(n)]}\right] \cong 10\log\left[\frac{\sigma^2{}_p}{\sigma^2{}_e}\right] \qquad (1)$$

where $\sigma^2{}_p$ and $\sigma^2{}_e$ refer to the variances of the primary and error signals respectively and E is the statistical expectation operator. Non-uniform magnetic fields and a nonlinear suspension system in loudspeakers will cause soft clipping under large signal conditions resulting in odd-order harmonics [2] at the output, however other factors such as room noise, DC offsets, rattling and vibration [3], dynamic changes in the AIR, and undermodelling of the acoustic transfer function [4] also limit the ERLE.

2.0 NONLINEAR ARCHITECTURE

The proposed structure shown in Figure 2 consists of both nonlinear and linear sections. The nonlinear section consist of a two layer neural network that cancels the first part of the AIR where most of the energy is contained. The weight update equations for the nonlinear portion are based on the gradient backpropagation algorithm [5] with a normalized adaptive step size. The nonlinear node consists of a linearized hyperbolic tangent function which is linear for inputs below a user definable amplitude *a*, where $0 \le a \le 1$. Based on measurements reported in [6], the parameter *a* was set to 0.2 since it was found that this produced an ERLE approximately 1.5 dB higher than with a conventional (i.e. *a*=0) sigmoid. The node activation function *f* ( ) is defined by;

$$f(s) = \begin{cases} s & ;|s| \le a \\ \mathrm{sign}(s)\left[(1-a)\cdot\tanh\left(\frac{|s|-a}{1-a}\right) + a\right] & ;|s| > a \end{cases} \qquad (2)$$

where *s* is the input. In Figure 2, the output $y(k)$ of the neural network portion at time *k* is defined by;

$$y(k) = w^{(2)}(k)x^{(2)}(k) + w_b{}^{(2)}(k) \qquad (3)$$

$$x^{(2)}(k) = f(s(k)) \qquad (4)$$

$$s(k) = \mathbf{w}^{(1)}(k)^T\mathbf{x}^{(1)}(k) + w_b{}^{(1)}(k) \qquad (5)$$

where $\mathbf{x}^{(l)}(k)$ represents the input vector to layer l, $\mathbf{w}^{(l)}(k)$ represents the weight vector in layer l, $w^{(l)}{}_b(k)$ represents the single bias weight for layer l, $s(k)$ represents the input to the nonlinear node and T is the transpose operator. The weight update equations are described by;

$$\mathbf{w}^{(l)}(k+1) = \mathbf{w}^{(l)}(k) - \mu_{TDNN}(k)\delta^{(l+1)}(\dot{k}) \cdot \mathbf{x}^{(l)}(k) \quad (6)$$

$$w_b^{(l)}(k+1) = w_b{}^{(l)}(k) - \mu_{TDNN}(k)\delta^{(l+1)}(\dot{k}) \quad (7)$$

$$\delta^{(l+1)}(k) = \begin{cases} -2e_1(k) & ;l=2,\text{output layer} \\ f'(s(k))\delta^{(l+2)}(k)w^{(l+1)}(k) & ;l=1,\text{hidden layer} \end{cases} \quad (8)$$

where $e_1(k) = p(k) - y(k)$, $f'(\,)$ represents the derivative of the activation function at the input value $s(k)$, $\delta^{(l+1)}(k)$ represents the local gradient "delta" term in layer $l+1$, and $\mu_{TDNN}(k)$ is the normalized step size parameter defined by;

$$\mu_{TDNN}(k) = \frac{\alpha}{2 + \mathbf{x}^{(1)}(k)^T\mathbf{x}^{(1)}(k) + [x^{(2)}]^2} \quad (9)$$

The parameter α is a number between 0 and 2, and is set to 0.5. The weights in the linear portion of the proposed structure are updated using the Normalized LMS [7] algorithm;

$$\mathbf{w}_{FIR}(k+1) = \mathbf{w}_{FIR}(k) - \left[\frac{\alpha}{1 + \mathbf{x}_{FIR}(k)^T\mathbf{x}_{FIR}(k)}\right]e_2(\dot{k}) \cdot \mathbf{x}_{FIR}(k) \quad (10)$$

$$w_b(k+1) = w_b(k) - \left[\frac{\alpha}{1 + \mathbf{x}_{FIR}(k)^T\mathbf{x}_{FIR}(k)}\right]e_2(k) \quad (11)$$

3.0 MEASUREMENT SETUP

Measurements are performed in a low-noise, furnished conference room. A handsfree telephone which has been modified to allow access to the primary and reference electrical signals is placed on top of the conference table. The reference source signal consists of white noise which is bandlimited from 300 Hz to 3400 Hz. The filtered reference signal is then amplified such that the loudspeaker produces a sound pressure level (SPL) from 60dB to 95dB as measured 0.5m directly above the loudspeaker. The primary and reference signals are then recorded onto a TEAC Digital Audio Recorder (DAT). The DAT signals are downloaded to a computer via an ARIEL DSP96 board sampling at 16 kHz. These samples are then applied to both the proposed structure and a 600 tap linear adaptive filter which has DC bias compensation and weights updated in the same fashion as (10) and (11). In the proposed structure, the number of taps in the nonlinear section delay line equals 200 to cover the bulk of the loudspeaker impulse response. The number of taps in the linear section is 400 for a total impulse length of 600 taps. For each SPL, both algorithms are tested with the same input data of length 80,000 to allow convergence to a steady state at which point the average ERLE is measured and plotted.

4.0 RESULTS

The experimental results shown in Figure 3 show that at low volumes in the vicinity of 60 dB SPL, the proposed structure improves the ERLE by 3 dB as compared to the linear adaptive filter even though there is little nonlinear distortion in this range. In the low volume ranges, room noise becomes a dominant limitation and the proposed structure offers some improvement. In the medium volume range from 70-75 dB SPL, the proposed structure performs about 1 dB poorer than the linear filter due to an extra bias weight variance not included in the linear filter, and also because $f(s)$ will generate some small amount of distortion for any $|s| > a$ even when the inputs are linear. However, in the vicinity of 80 to 95 dB SPL where nonlinear effects dominate, the proposed structure clearly outperforms the linear filter in terms of converged ERLE and demonstrates over 8 dB improvement at 90 dB SPL.

5.0 CONCLUSIONS

A new structure to mitigate nonlinear loudspeaker distortion effects in AECs has been presented. The architecture is simple and the update algorithms are based on stochastic gradient methods. Experimental measurements in a conference room indicate that this new structure is capable of improving the

ERLE by over 8 dB at high volumes (where nonlinear effects are significant) and by over 3 dB at low volumes where room noise is significant.

6.0 ACKNOWLEDGEMENTS

The authors wish to thank NSERC, Carleton University, Nortel and TRIO for their financial support.

7.0 REFERENCES

[1] R. Wehrmann, J.V.D. List, P. Meissner, "A Noise Insensitive Compromise Gradient Method for the Adjustment of Adaptive Echo Cancellers", *IEEE Trans. Comm*. COM-28, No. 5, 1980, pp. 753-759.

[2] A.J.M. Kaizer, *On the Design of Broadband Electrodynamic Loudspeakers and Multiway Loudspeaker Systems*, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1986, Chapter 6.

[3] A.N. Birkett, R. A. Goubran, "Limitations of Handsfree Acoustic Echo Cancellers due to Nonlinear Loudspeaker Distortion and Enclosure Vibration Effects", in *1995 IEEE ASSP Workshop on Appl. of Sig. Proc. to Aud. and Acoustics,* New Paltz, New York, Oct. 1995.

[4] 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.

[5] D. R. Hush, B. G. Horne, "Progress in Supervised Neural Networks", *IEEE Signal Processing Magazine*, Jan. 1993, pp.8-39.

[6] A.N. Birkett, R. A. Goubran, "Acoustic Echo Cancellation Using a NLMS-Neural Network Structures", *ICASSP* 1995, Detroit, MI., Vol. 5, pp. 3035-3038.

[7] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice-Hall Information Systems Series, 1996.

8.0 ILLUSTRATIONS

.

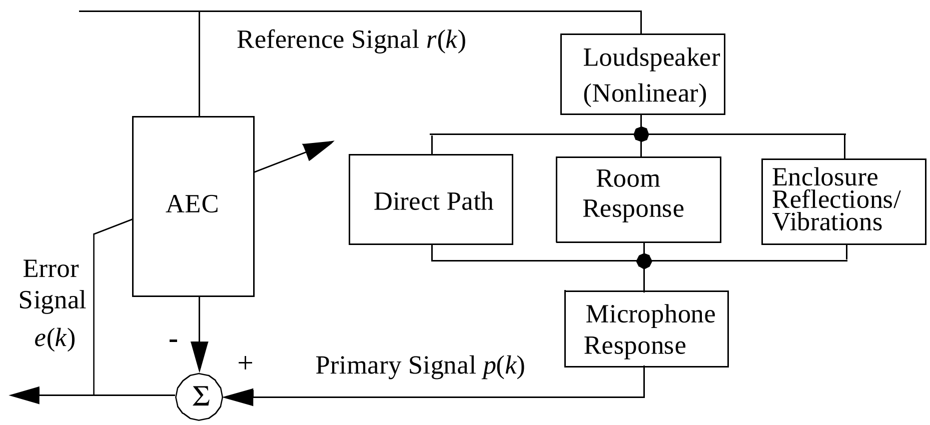

FIGURE 1. The AIR consists of both linear and nonlinear components.

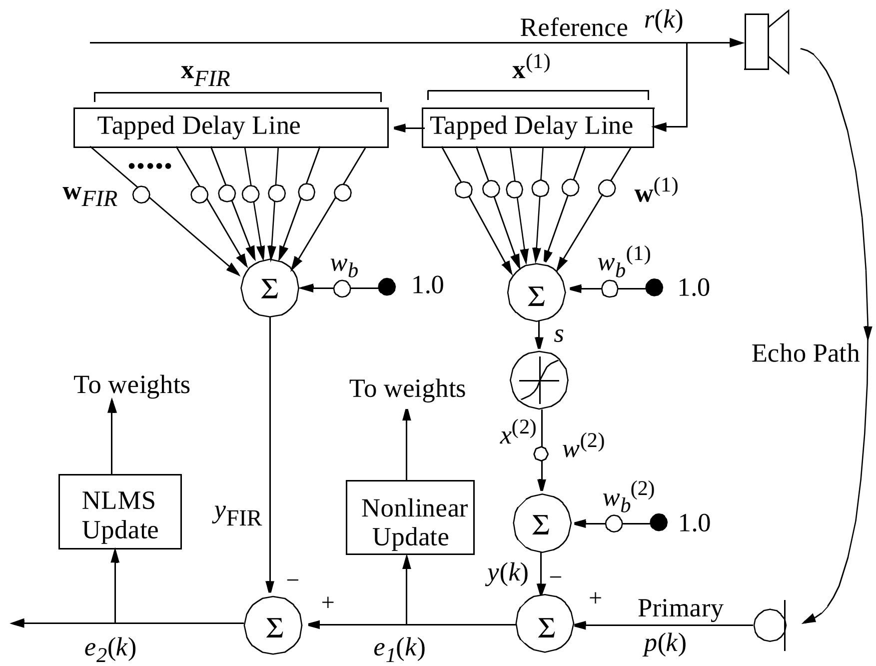

FIGURE 2. Proposed AEC structure.

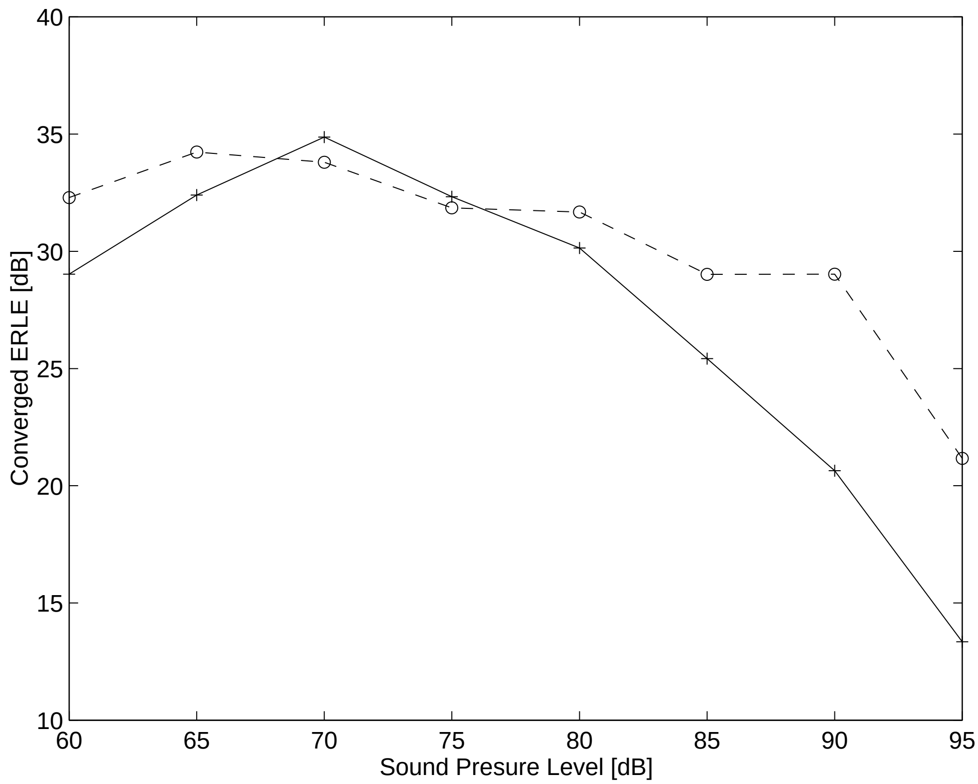

FIGURE 3. Experimental results.

—+— Linear filter, 600 taps, trained with NLMS.

--o**--** Proposed structure with 200 taps in nonlinear section and 400 taps in linear section.