LIMITATIONS OF HANDSFREE ACOUSTIC ECHO CANCELLERS DUE TO NONLINEAR LOUDSPEAKER DISTORTION AND ENCLOSURE VIBRATION EFFECTS

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

The limitations of adaptive echo cancellers (AEC) for handsfree telephony include noise, finite precision and truncation effects, undermodelling of the acoustic impulse response, vibration of the plastic enclosure, loudspeaker nonlinearities, dynamic tracking and convergence and double-talk. This paper examines the effect that the loudspeaker nonlinearity and enclosure vibration have on the steady state ERLE performance and concludes that enclosure vibration is an important limitation which, although rarely mentioned in the literature, is probably the major limitation in terms of sound quality. Experimental measurements indicate that enclosure vibration at medium to high loudspeaker volumes limits the achievable ERLE and the perceived audio quality more than any other limitation mentioned above.

1.0 INTRODUCTION

A typical handsfree terminal normally consists of two Adaptive Filters (AFs). The first AF is used to remove acoustic echoes and is referred to as an Acoustic Echo Canceller (AEC). The second AF is used for cancelling echoes from an imperfect hybrid as well as reflections from the line. The AEC structure must be capable of identifying and tracking not only the reflected signals from the room, i.e. its Acoustic Impulse Response (AIR), but also of modelling the plastic enclosure vibrations and nonlinear loudspeaker response, as shown in Figure 1.

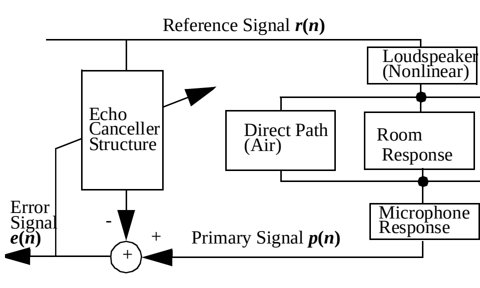

FIGURE 1. Acoustic Echo Canceller Structure. The AEC must identify not only the AIR but nonlinear and vibration effects as well.

Conventional AECs utilize a linear adaptive transversal filter to model the room impulse response and cancel the echo signal. The **NLMS** algorithm [1] 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;

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

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

2.0 OVERVIEW OF LIMITATIONS OF AECs

The limitations of AECs include the following:

1. ***Noise, Finite Precision and Truncation***: This includes acoustic noise from fans and air conditioning in the room as well as thermal and impulsive circuit noise from amplifiers, and DSP related noise such as truncation, finite word lengths and characteristics of the particular algorithm being used. Caraiscos and Lui [6] have done an analysis of roundoff effects in the LMS family of algorithms and point out that there exists an optimum value of step size μ which obtains the minimum mean square error. However, this step size is generally too small to maintain a reasonable convergence rate. Room noise is probably the largest contributor to this limitation. Our experiments have shown that wall coverings, carpets etc. can make a significant decrease in the amount of room noise picked up by the microphone.
2. ***Undermodelling of the Acoustic Transfer Function***: This occurs when the number of taps or variables in the AEC adaptive filter is less than the AIR of the room. The remaining uncancelled tail portion of the AIR manifests itself as a finite error at the output of the AEC. Blindly increasing the number of taps results in added complexity, greater algorithmic noise and slower convergence. The achievable ERLE has been shown to be determined (in the absence of other limitations) to be equal to the ratio of powers of the Total Impulse Power (TIP) of the AIR to the uncancelled Tail Portion (TP) of the AIR. Knappe and Goubran [9] have shown that the TIP/TP ratio defines the ERLE up to a ratio of about 20 dB. Beyond this point, other effects dominate.

3. ***Enclosure Vibration Effects***: A major part of the AIR is due to loudspeaker/microphone/enclosure coupling which is stationary in nature and larger in amplitude than the echoes. The particular adaptive algorithm used will devote a portion of its computation to adapt these AIR coefficients which may be better modelled by another method. Whistling can occur in small orifices in sealed enclosures. This whistling is essentially chaotic in nature and can be a problem if it occurs close to the microphone. Our experiments have shown that vibration, especially in the lower voice frequencies (when speech is the reference signal) causes significant vibration an noise to be picked up in the microphone, even though to the listener, the vibration from the set is barely audible.
4. ***Nonlinearities in the Transfer Function***: Generated mainly in the loudspeaker, nonlinear distortion effectively puts a limit on the achievable ERLE of algorithms based on linear mechanics. In addition to the direct loudspeaker effects, secondary nonlinear effects such as rattling can be considered nonlinear in nature. Rattling is very difficult, if not impossible to model. However, the loudspeaker nonlinearity is weak and can therefore be modelled accurately with nonlinear state-space (see [3][4]) and neural network models (see [11] [12]). Our experimental results have shown that this nonlinearity is a less serious problem than the enclosure vibration.
5. ***Dynamic Tracking in Nonstationary Conditions***: The initial convergence of a particular algorithm identifies the room configuration, however as objects move and the input characteristics become nonstationary, the tracking ability of the algorithm becomes important. For example, although RLS based algorithms have fast convergence, it has been found that algorithms based on instantaneous gradient estimates like the exponential step size LMS family actually outperform RLS algorithms in AEC applications [7].
6. ***Double Talk:*** During periods when the far end speaker and near end speaker are simultaneously talking in a full duplex system, it is often necessary to freeze the adaptive filter coefficients such that divergence of the tap weights does not occur. For more details on double talk detector, refer to [15].

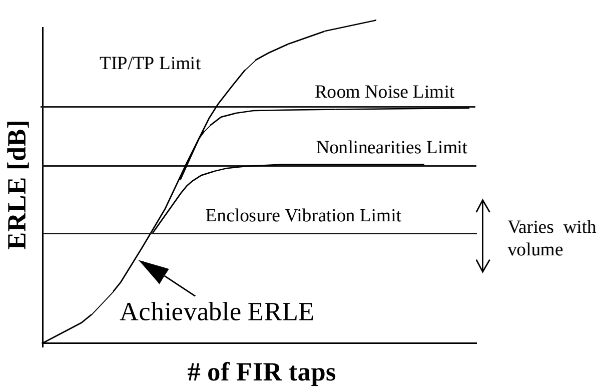

FIGURE 2. Achievable ERLE as a function of Physical Limitations.

2.1 Nonlinearity in the Transfer Function

The loudspeaker is probably the major source of nonlinearity in a speakerphone. A loudspeaker has several sources of nonlinearity including non-uniform magnetic field and nonlinear suspension system [2][4].The electrical and mechanical part of the loudspeaker interact through the magnetic field resulting in a nonlinear force deflection characteristic f_M of the loudspeaker cone suspension system, usually approximated [4] by;

$$f_M = \alpha x + \beta x^2 + \delta x^3 \tag{2}$$

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 operate at high volumes, which is generally the case for speakerphones.

Figure 3 shows the PSD of the primary signal with the loudspeaker and microphone components removed (this removes the effect of vibration). The measurements are performed in an anechoic chamber to remove room noise and echoes. Notice that there is an increase in the out-of-band signal level which is essentially nonlinear components of the original bandlimited (reference) signal..

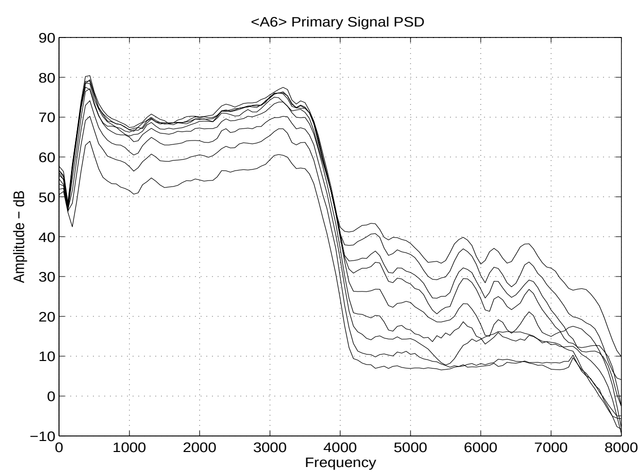

FIGURE 3. Primary signal PSD with components removed from enclosure. Volume is increased from 60 dB SPL to 100 dB SPL.

Loudspeaker nonlinearity can be modelled using a nonlinear preprocessor before the tapped delay line transversal filter using a partial adaptive time delay neural network and NLMS structure. The basic structure is shown in Figure 4. Measurements are performed in an anechoic chamber to remove the effect of room echoes.

The results shown in Figure 5 show that this method is capable of improving the ERLE by approximately 5.5 dB when operating at high volumes in the vicinity of 100 dB SPL. At lower volumes however, the distortion is not as severe and the linear NLMS model outperforms the partial adaptive TDNN.

.

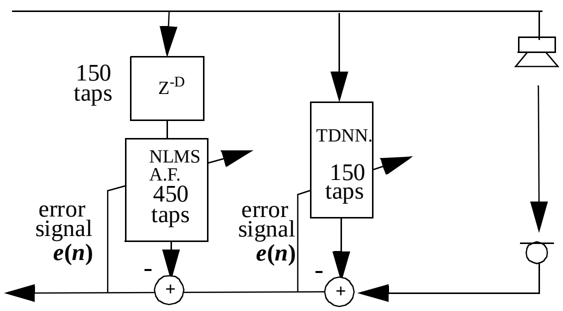

FIGURE 4. Partial adaptive structure utilizing a tapped delay line neural network pre-processor to cancel the first part of the AIR and a NLMS to cancel the tail portion of the signal.

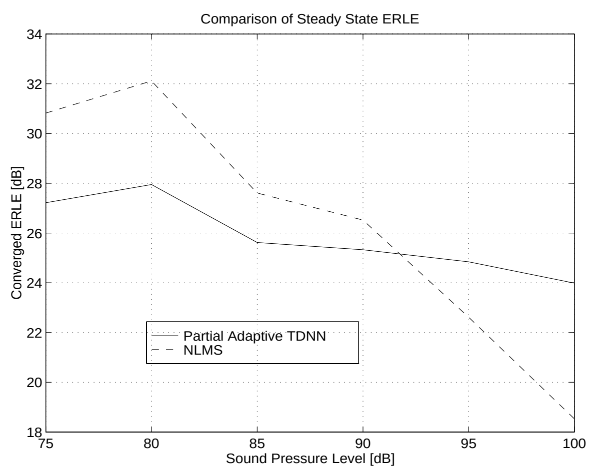

FIGURE 5. Experimental results showing converged ERLE of the partial adaptive TDNN structure compared to the NLMS for increasing volume. A 5.5 dB improvement in ERLE can be obtained at high volumes when the components are separated.

2.2 Enclosure Vibration Effects

Enclosure vibration will also greatly affect the achievable ERLE performance. Rattling of the handset and keys is nonlinear and chaotic and can be modelled as uncorrelated noise. The result is to limit the achievable ERLE, in a similar way as uncorrelated noise or as an unmodelled nonlinearity.

Measurements shown in Figure 6 show the PSD of a speakerphone primary signal, when the loudspeaker and microphone are mounted inside the plastic enclosure. Rattling and vibration cause an increase in the uncorrelated noise signal introduced into the primary path. Since this noise is uncorrelated, it essentially places a limit on the achievable cancellation, as defined in (1). In fact, the application of the partial adaptive TDNN process which gave some degree of improvement in the case where the components were separated gave no improvement when applied to signals which were obtained while the components were mounted inside the enclosure. The explanation seems to be that vibration is preventing the algorithm from properly identifying the handsfree telephone system. Further experimentation showed that noticeable vibration was being picked up by the microphone when voice signals were applied to the reference channel, especially at low frequencies.

.

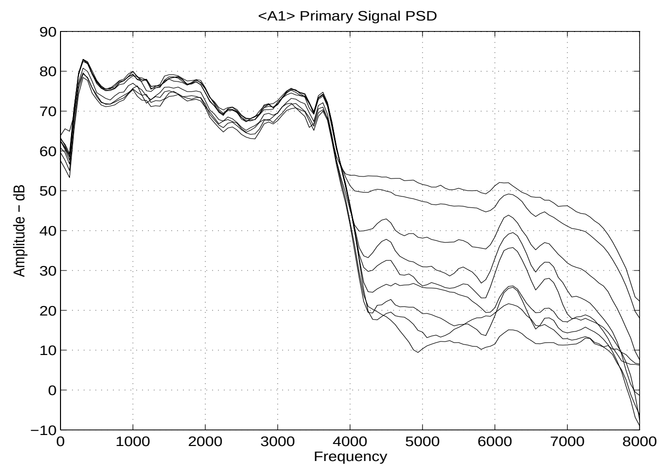

FIGURE 6. Experimental results showing the effect of increasing signal level on the power spectral density of the primary signal as the volume of the loudspeaker is increased from 60 to 100 dB SPL.

When the partial adaptve process is applied to the data collected when the components are mounted inside the plastic speakerphone enclosure, the performance improvement observed in Figure 5 are no longer seen. The results are illustrated in Figure 7

.

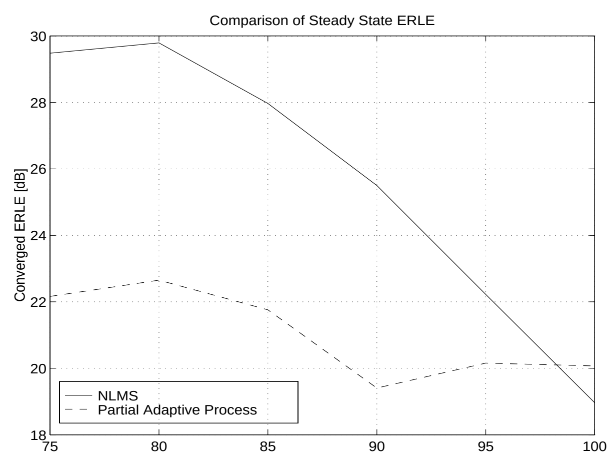

FIGURE 7. Experimental results showing converged ERLE of the partial adaptive TDNN structure compared to the NLMS for increasing volume, when the components are mounted inside the speakerphone enclosure.

The output error PSD is plotted along with the primary and reference signals for the loudest volume signal (100 dB SPL) in Figure 8. The NLMS algorithm is used in this case since the partial adaptive TDNN gave similar results, for the vibration case.

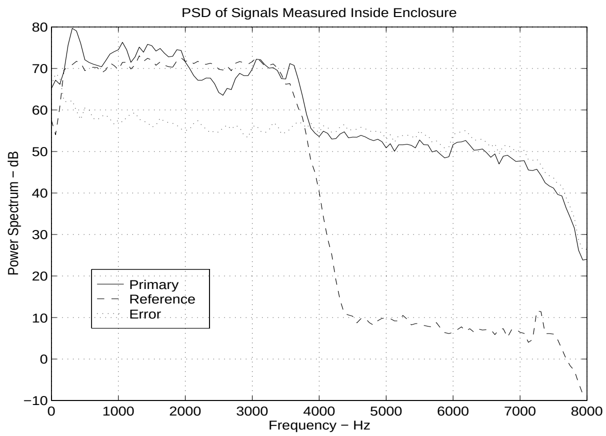

FIGURE 8. Experimental results. PSD of primary, reference and error signals shows that in the vibration case, the out of band uncancelled error is due to the noise like characteristics of the primary signal in that frequency range.

3.0 DISCUSSION

Comparison of the PSD plots of Figure 3 and Figure 6 show clearly that the vibration and rattling inside the enclosure causes an increase in the out-of-band signal components over that of the components only. In the case where the components are separated from the enclosure, a partial adaptive TDNN structure can model the loudspeaker nonlinearity at high volumes, since it is essentially a weak non-linearity. This results in an improvement in the converged ERLE over the NLMS case. In the case where the components are mounted inside a speakerphone enclosure, rattling and vibration cause an increase in the out-of-band signal, which cannot be removed by either the linear NLMS or non-linear TDNN-NLMS structure, suggesting that the signal is chaotic and noise-like. Indeed, a very complicated structure would likely be needed to model such effects as rattling keys and resonances in plastics, and is beyond the scope of this paper. In the course of this work it was noticed that when speech signals were applied to the reference, a most annoying vibration was picked up by the microphone especially on voice peaks at low frequency. This was absent when the loudspeaker was removed from the enclosure.

4.0 CONCLUSIONS

This paper has demonstrated that vibration and loudspeaker nonlinearity are limitations to the achievable ERLE in acoustic echo cancellers in handsfree telephony. It presented a method to combat nonlinear loudspeaker distortion effects in the AEC. This paper also suggests that to improve the quality of the audio in such systems, it may be more prudent to concentrate on vibration reduction before applying DSP techniques.

5.0 REFERENCES

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

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

[3] 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.

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

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

[6]C. Caraiscos, B.Liu, “A Roundoff Error Analysis of the LMS Adaptive Algorithm”, *I.E.E.E. Trans. on Acoustics, Speech and Sig. Proc*. Vol. ASSP-32, No. 1, Feb. 1984, pp. 34-41.

[7] H. Yuan, *Dynamic Behavior of Acoustic Echo Cancellation*, M. Eng. Thesis, Carleton University, Ottawa, Canada, 1994.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] J. G. Ryan, M. E. Knappe, H. Yuan, R.A. Goubran, T. Aboulnasr, “Audio Quality Enhancemenat in Noisy Environments”, *Final Report submitted to Data Network Planning Section*, Bell Canada, May 1992.

[14] S. E. Kuo, J. Chen, “Multiple Microphone Acoustic Echo Cancellation System with the Partial Adaptive Process”, *Digital Signal Processing* 3, 1993, pp. 54-63.

[15] S. M. Kuo, Z. Pan, “Distributed Acoustic Echo Cancellation System with Double-talk Detector”, *J. Acoust. Soc. Am*. No. 6, Dec.1993, pp. 3057-3060.

ACKNOWLEDGEMENTS

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