# NONLINEAR ECHO CANCELLATION FOR HANDS-FREE SPEAKERPHONES

*Bryan S. Nollett and Douglas L. Jones*

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
1308 W. Main St.
Urbana, IL 61801
Email: d-jones@csl.uiuc.edu

## ABSTRACT

Nonlinearities in the amplifier and loudspeaker of hands-free speakerphones limit the performance of linear adaptive acoustic echo cancellers, necessitating the use of nonlinear cancellation schemes. A nonlinear acoustic echo canceller based on the Wiener-Hammerstein model structure of a cascade of linear, memoryless nonlinear, and linear elements is proposed. By modeling the true structure of the nonlinear system, the proposed canceller requires relatively few adaptive parameters, offering significantly lower storage and computational requirements than more general nonlinear adaptive filtering techniques. Experimental results on measured loudspeaker signals indicate that the proposed nonlinear echo canceller provides as much as an 8.4 dB improvement in Echo Return Loss Enhancement (ERLE) over a linear Normalized LMS canceller with little additional computation.

## 1. INTRODUCTION

The performance of linear adaptive acoustic echo cancellers is limited by nonlinearities in the signal path. Nonlinear cone suspension and uneven magnetic flux densities in the loudspeaker introduce nonlinear distortion at large cone displacement levels. Additionally, at high volume settings, saturation effects may occur in the loudspeaker power amplifier, producing gross nonlinearities in the system that greatly impair the performance of linear acoustic echo cancellers. Because these nonlinear phenomena are likely to appear in hands-free speakerphones, adaptive nonlinear schemes are required to achieve adequate echo cancellation.

A number of methods have been proposed for nonlinear echo cancellation. Volterra series based filters [1] and neural networks [2] are two popular nonlinear adaptive filter configurations. One of the primary disadvantages of the Volterra filter is the large number of parameters it typically requires to characterize a nonlinear system and the correspondingly large computational load required to adapt such a large number of parameters. The number of parameters in a Volterra expansion goes up exponentially with the order of the nonlinearity. Since it is widely accepted that the nonlinearities in loudspeaker systems are primarily of third and higher orders, adaptive Volterra filters are not computationally feasible in this application. While significant gains have been made recently in improving the efficiency of adaptive Volterra filters [3], these advances have been primarily in the use of second-order Volterra representations, limiting their usefulness in the hands-free environment.

Acoustic echo cancellers using Time Delay Neural Networks (TDNNs) have been constructed [4]. They have the advantage that—unlike the Volterra structures which are of a certain fixed order—TDNNs are general structures that do not require an explicit characterization of the nonlinearity by the user. Disadvantages of TDNNs include their tendency to converge to local minima, slow convergence and great computational requirements.

At least two hybrid nonlinear/linear echo cancellers have been proposed using neural networks to cancel the nonlinear component and linear Normalized LMS (NLMS) to cancel the long room echo response [4], [5]. Both structures have the disadvantage of still requiring that a significant proportion of their parameters be the computationally demanding neural network canceller. Additionally, the echo canceller structure described in [4] has the significant shortcoming of requiring a second training signal from a second microphone placed near the loudspeaker.

In this work we propose a nonlinear acoustic echo canceller (NLAEC) that is composed of a cascade of linear and memoryless nonlinear elements. The proposed

This work was supported by the Joint Services Electronics Program, contract no. N00014-96-1-0129.

canceller, while similar in structure to that given in [4], does not require a second training signal and typically can be adapted with far less computation for a given response length. Simulations and experimental results suggest that the structure presented here retains many of the performance advantages of the previous proposals.

## 2. PROPOSED STRUCTURE

The nonlinear adaptive echo canceller we present here is constructed under the assumption that the nonlinearity present in typical hands-free speakerphones is of a localized nature; i.e., the chain of components in the echo path made up of the D/A, power amplifier, loudspeaker, room response, microphone, amplifier and A/D can be modeled as the cascade of a linear filter, a memoryless, amplitude-limiting nonlinearity, and a second linear filter as shown in Figure 1. Measurements of internal signals in actual loudspeaker devices indicate that this assumption is likely valid for most speakerphone systems. Consequently, we let our nonlinear echo canceller have this same structure. Training this nonlinear canceller is the nonlinear system identification problem shown in Figure 2.

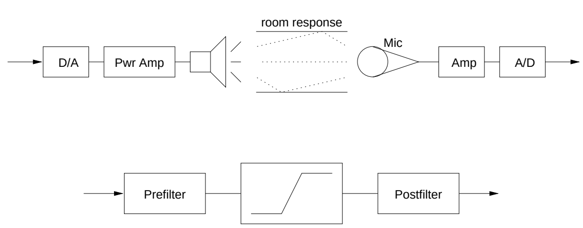


Figure 1: Typical speakerphone signal chain (top) and its model (bottom).

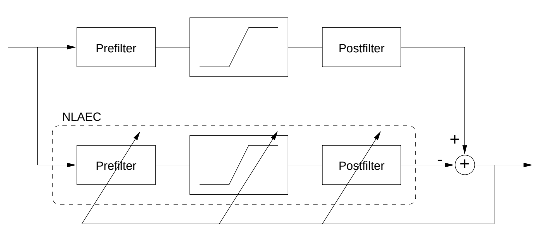


Figure 2: Equivalent model of system and identification of that system with the NLAEC.

This cascade of a linear filter, memoryless nonlinearity and a second linear filter is referred to as a Wiener-Hammerstein model or G-model ("General model") in the field of control theory. It is the cascade of a Wiener model (a linear filter followed by a memoryless nonlinearity) and a Hammerstein model (a memoryless nonlinearity followed by a linear filter) and represents the logical generalization of these two component structures. While several proposals have been made regarding the identification of Wiener and Hammerstein systems with general nonlinearity under various inputs ([6],[7], for example), most identification schemes for the cascade Wiener-Hammerstein model operate under the requirement that the input to the system be Gaussian white noise, an obviously unacceptable requirement for practical hands-free speakerphones [8], [9]. Below we derive a simple technique for identifying the parameters of a Wiener-Hammerstein system that does not impose such strict constraints on the system input and is thus more appropriate for the application of acoustic echo cancellation.

## 3. PARAMETER UPDATE

It is assumed that the saturator's clipping amplitude is its only adaptable parameter so that the parameters requiring adaptation in the cascade system are the linear weights of the filters preceding and following the nonlinearity (referred to here as the "prefilter" and "postfilter" for convenience), and this saturation level. All parameters are updated using a stochastic gradient algorithm minimizing the expected value of the squared error signal.

The adaptive structure and signals internal to the structure are shown in Figure 3. For known input sequence $x(n)$ and desired sequence $d(n)$, at time $k$ we have

$$\bar{s}(k) = \sum_{\ell=0}^{N_w-1} w_k(\ell)x(k-\ell) \quad (1)$$

$$\hat{s}(k) = \rho_{\gamma_k}(\bar{s}(k)) \quad (2)$$

$$y(k) = \sum_{m=0}^{N_h-1} h_k(m)\hat{s}(k-m) \quad (3)$$

$$e(k) = d(k) - y(k) \quad (4)$$

where $w_k(i)$ are the prefilter taps at time $k$, $h_k(i)$ are the postfilter taps at time $k$, and $\rho_{\gamma_k}(\cdot)$ is the amplitude clipping nonlinear function with adaptive saturation level $\gamma_k$ at time $k$.

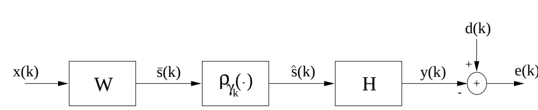


Figure 3: Relevant signals in the nonlinear adaptive structure.

Updating the nonlinear model parameters using stochastic gradient optimization to minimize

$E\{|e(k)|^2\}$ and invoking the assumption that the adaptive parameters are slowly varying produces the parameter update equations

$$w_{k+1}(\ell) = w_k(\ell) + 2\mu e(k) \sum_{n=0}^{N_w-1} x(k-\ell-n) h_k(n) \rho'_{\gamma_k}(\bar{s}(k-n)) \quad (5)$$
$$\text{for} \quad \ell = 0, \ldots, N_w - 1$$

$$\gamma_{k+1} = \gamma_k + 2\mu e(k) \sum_{m=0}^{N_h-1} \frac{\partial}{\partial \gamma_k} \rho_{\gamma_k}(\bar{s}(k-m)) \quad (6)$$

$$h_{k+1}(m) = h_k(m) + 2\mu e(k) \hat{s}(k-m) \quad (7)$$
$$\text{for} \quad m = 0, \ldots, N_h - 1$$

Note that the update of the postfilter coefficients $h(i)$ is simply the commonly used LMS update for a linear adaptive filter. In an attempt to increase the rate of convergence, a Normalized LMS (NLMS) update may be used for these taps. From Equations (5)–(7) it is clear that updating the prefilter requires $O(N_w^2)$ operations, while the remainder of the algorithm requires only $O(N_h)$ operations. However, since the update equations for the $w(i)$ are decoupled, they may be easily updated in a round-robin manner, one tap per data point, if a lower rate of convergence can be tolerated. In practice this is rarely required because, as is discussed below, the prefilter for acoustic echo cancellation applications is typically very short, and is certainly much shorter than the postfilter so that a full-rate update of the prefilter taps is usually feasible.

In this work, two forms of the clipping nonlinearity are proposed, a hard clipping saturator and a soft clipping nonlinearity. Both were designed to exhibit a slope of one at zero input so as to produce an input/output relation with unity gain for small signals. Also, both are antisymmetric and have a maximum output amplitude that is denoted by $\gamma_k$ at time $k$. The hard clipping saturation is given by the piecewise linear function

$$\rho_{\gamma_k}(v) = \begin{cases} -\gamma_k & v < -\gamma_k \\ v & -\gamma_k \le v \le \gamma_k \\ \gamma_k & v > \gamma_k \end{cases} \quad (8)$$

and the soft clipping saturation by

$$\rho_{\gamma_k}(v) = \frac{\gamma_k v}{\sqrt[\alpha]{|\gamma_k|^\alpha + |v|^\alpha}} \quad (9)$$

in which alpha is a positive, nonadapting parameter that determines the softness or sharpness of the nonlinearity. Plots of these saturating functions are given in Figure 4.

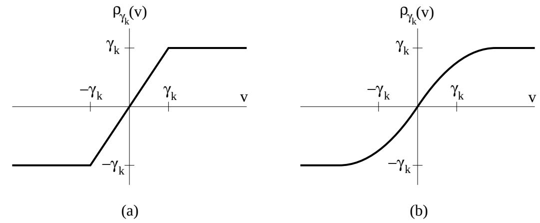


Figure 4: (a) Hard clipping nonlinearity. (b) Soft clipping saturator with $\alpha = 2$.

## 4. INITIALIZATION

Simulations indicate that, as presented above, the squared error of the adaptive system does not always converge to the global minimum under all initializations. Instead, local minima in the error surface are found. However, for the specific application of nonlinear echo cancellation in hands-free speakerphones, an initialization technique that experimentally performs very well has been developed.

Assuming that the localized nonlinearity in a hands-free speakerphone is present in either the power amplifier or the speaker itself, we see from Figure 1 that the prefilter response in a nonlinear echo canceller corresponds primarily to the cascaded response of the D/A, power amplifier, and possibly part of the speaker response. A reasonable assumption for this aggregate response is that it is generally bandpass with a wide passband (relative to the sampling rate). Under these circumstances, the bulk of the prefilter impulse response energy is well concentrated in time. Similarly, the postfilter corresponds to the room impulse response and microphone-related input components so we expect the postfilter to be significantly longer than the prefilter.

Given this knowledge of prefilter and postfilter characteristics, a sensible initialization of these filters may be made that will ensure convergence in typical applications. Since the prefilter should converge to a response that is well concentrated in time, it is likely that it can be well approximated by a scaled and shifted unit pulse. After initializing the prefilter to a shifted unit pulse located at the "center" of the available prefilter taps, an initialization for the postfilter is obtained by temporarily removing the nonlinearity and adapting only the postfilter as in Equation (7), performing a linear identification. Simultaneous adaptation of all components in the nonlinear system by iterative evaluation of Equations (5)–(7) begins after this initialization of the postfilter fails to further decrease the error energy.

## 5. SIMULATION RESULTS

A commonly used measure of echo canceller performance is Echo Return Loss Enhancement (ERLE) defined by

$$ERLE\ (\mathrm{dB}) = 10 \log_{10} \left( \frac{\sigma_d^2}{\sigma_e^2} \right) \qquad (10)$$

where $\sigma_d^2$ and $\sigma_e^2$ are the variances of the desired and error signals, $d(k)$ and $e(k)$, shown in Figure 3.

Simulations were conducted to test the initialization technique and compare the performance of the NLAEC to that of the commonly used NLMS echo canceller. The system to be identified was a cascade of a length-11 bandpass prefilter, a hard clipping saturation nonlinearity, and a random, length-21 postfilter. The NLAEC used to identify the system contained 15 prefilter taps, 43 postfilter taps and used a hard clipping saturator with adjustable saturation level. For comparison, a linear NLMS adaptive filter with 58 taps was also applied. The system to be identified was excited with Gaussian white noise, both the NLAEC and NLMS cancellers were allowed to adapt until converged, and then the converged ERLE was recorded. Measurements were made with various saturation thresholds in the model system, from severe clipping at twice the standard deviation of the model prefilter output, to minimal clipping at four times the prefilter output standard deviation. The measured ERLE for the NLAEC and NLMS averaged over 10 experiments is shown in Figure 5.

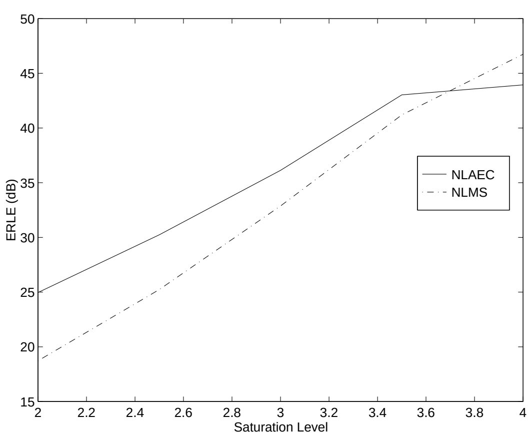


Figure 5: ERLE of NLAEC and NLMS at different levels of saturation.

The results shown in Figure 5 show that for all but the mildest clipping, the NLAEC outperforms the NLMS echo canceller. Generally, as the severity of the nonlinearity increases, the performance advantage of the nonlinear canceller increases as one would expect. At the most severe clipping levels tested, the ERLE of the proposed structure was over 6 dB higher than that of the linear canceller. At the mildest clipping levels in Figure 5, levels which correspond to an almost negligible amount of nonlinear distortion, the NLAEC performs only marginally worse than NLMS.

## 6. EXPERIMENTAL RESULTS

A very inexpensive pair of amplified loudspeakers of the type commonly used in multimedia PCs was purchased for experiments investigating the applicability of the proposed nonlinear canceller in PC hands-free telephony applications. In a typical laboratory environment, recordings were made of a single speaker playing at high volume a recording of synthesized vowels that varied in fundamental frequency and amplitude. Measurements of the voltage at the loudspeaker terminals indicated that over the range of output volumes this powered speaker could produce, the loudspeaker itself was essentially linear. However, at high volume levels, the power amplifier appeared to introduce a hardlimiting amplitude nonlinearity. Thus, the saturation nonlinearity chosen in the NLAEC was the hard clipping saturator shown in Figure 4a.

Both the NLAEC (with initialization) and NLMS algorithms were applied to the data from the synthetic vowel recordings and each algorithm was permitted to run until converged. The NLAEC prefilter contained 30 taps and the postfilter contained 200 taps. The NLMS adaptive filter trained for comparison used 230 taps. When converged, the proposed nonlinear echo canceller achieved an ERLE of 19.5 dB, significantly better than the NLMS canceller which produced an ERLE of 11.1 dB.

## 7. SUMMARY

Nonlinearities severely limit the performance of acoustic echo cancellers in hands-free speakerphones and suggest the use of nonlinear adaptive techniques. A nonlinear echo canceller based on the Wiener-Hammerstein model structure of a cascade of linear, memoryless nonlinear, and linear elements was introduced and an algorithm for its adaptation was proposed. Although the adaptive technique proposed is not guaranteed to converge to produce the global minimum mean squared error, a technique for the initialization of the adaptive parameters was suggested that experimentally performs very well in this echo cancellation problem.

Simulations and experimental data were presented that indicate that over a wide range of saturation conditions the proposed NLAEC structure can perform significantly better than linear cancellers. While cancelling an appreciable amount of the nonlinear echo components, the proposed system can be updated with an algorithm that has a computational cost that is little more than the LMS algorithm. With its low demands upon computation and coefficient storage resources, it offers a great savings over Volterra and neural-network echo cancellers and is practical for embedded real-time application.

## 8. REFERENCES


[1] V. John Mathews. Adaptive polynomial filters. *IEEE Signal Processing Magazine*, 8(3):10–26, July 1991.

[2] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, Saddle River, NJ, 3rd edition, 1996.

[3] J. Lee and V. J. Mathews. A fast recursive least squares adaptive second order Volterra filter and its performance analysis. *IEEE Transactions on Signal Processing*, 41(3):1087–1102, March 1993.

[4] A. N. Birkett and R. A. Goubran. Acoustic echo cancellation using NLMS-neural network structures. In *ICASSP '95*, pages 3035–3038, 1995.

[5] A. N. Birkett and R. A. Goubran. Nonlinear loudspeaker compensation for hands free acoustic echo cancellation. *Electronics Letters*, 32(12):1063–1064, June 1996.

[6] Torbjörn Wigren. Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, 39(11):2191–2206, November 1994.

[7] Włodzimierz Greblicki and Mirosław Pawlak. Nonparametric identification of Hammerstein systems. *IEEE Transactions on Information Theory*, 35(2):409–418, March 1989.

[8] Michael J. Korenberg. Identification of biiological cascades of linear and static nonlinear systems. In *Proceedings of the 16th Midwest Symposium on Circuit Theory*, XVIII.2.1–XVIII.2.9 1973.

[9] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.