# NONLINEAR ACOUSTIC ECHO CANCELLATION USING FEEDBACK

*Pratik Shah*[⋆] *Ian Lewis*[†] *Steven Grant*[⋆] *Sylvain Angrignon*[†]

[⋆] Missouri University of Science and Technology
Department of Electrical and Computer Engineering
Rolla, MO, USA
[†]Research In Motion Ltd.
DSP Audio Team
Ottawa, Ontario, Canada

**ABSTRACT**

Acoustic echo cancellation (AEC) is a well studied problem. One of the main assumptions in existing AEC techniques is that the echo path is linear. In today's equipment, this assumption does always hold. The increasing focus on the look and feel of commercial devices and consequently the decreasing size of their loudspeakers are the major contributors to making the echo path nonlinear. This emphasizes the need for a nonlinear echo canceler to maintain the required echo return loss enhancement (ERLE). Many algorithms have been proposed to solve this problem, but they are prohibitive because of their high computational complexity. This paper proposes a hardware modification to significantly reduce the nonlinear echo. Results show that up to 6 dB of improvement in ERLE in a real device are possible.

***Index Terms***— Nonlinear echo cancellation, voltage feedback, current feedback, loudspeaker nonlinearity, back EMF

## 1. INTRODUCTION

Acoustic echo cancellation (AEC) is a well studied problem. One of the main assumptions in existing AEC techniques is that the echo path is linear. In today's equipment, this assumption does not hold well. The increasing focus on the look and feel of a commercial device and consequently decreasing size of loudspeakers are the major contributors to making the echo path nonlinear. Also, the desire to make the phones sound louder, especially in speakerphone mode, not only makes the echo louder but also adds significant nonlinearities to the downlink path. Figure 1 shows the setup of an echo canceler and highlights the sources of nonlinearities. In the figure, $x_n$ is the downlink signal which is used as the reference signal for the echo canceler. $d_n$ is the microphone signal consisting of the far-end echo and the near-end speech and background noise. $e_n$ is the output of the echo canceler and ideally must contain only the near-end speech and background noise, if any. On the downlink path, the codec contains a digital amplifier followed by a Digital to Analog Converter (DAC) and an analog amplifier on the downlink path. On the uplink path, the codec contains an analog amplifier followed by an Analog to Digital Converter (ADC) and a digital amplifier. The codec may lead to clipping, either hard or soft, if the signal levels are too high. In an attempt to make the phones sound louder, device manufacturers often set gain values that lead to clipping of the downlink signal. This clipping introduced by the codec is minimal and most of the times not noticeable for the human ear. The major sources of nonlinearity are the analog amplifier and the loudspeaker. Different loudspeakers can handle different signal levels while maintaining a linear operating point. Loud speech signals drive the small device loudspeakers into a nonlinear region of operation. Since loud speech bursts exist only for short time durations, these nonlinearities may also go unnoticed by the human ear. Another source of nonlinear distortion is the enclosure of the device and the vibrations therein. These are the most difficult to model and predict and also severely affect the performance of a linear echo canceler [1]. Linear echo cancelers are very sensitive to any level of nonlinearity. This brings about the need for nonlinear echo cancellation methods.

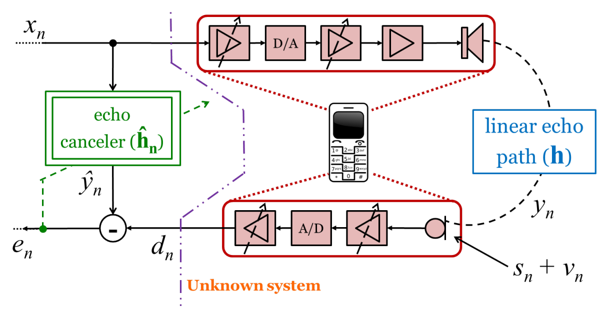



**Fig. 1**. Acoustic echo cancellation setup showing the nonlinear components.

The performance of acoustic echo cancelers is generally measured by the Echo Return Loss Enhancement (ERLE) which is defined by the following equation

$$ERLE(dB) = 10\log_{10}\frac{(\mathbf{d}^T\mathbf{d})}{(\mathbf{e}^T\mathbf{e})} \quad (1)$$

The general notation used in this paper is as follows: Scalars are italicized, vectors are bold-faced small letters. The superscript $^T$ denotes the transpose of a vector or matrix and the subscript $_n$ denotes the value at time instant n. The following section discusses the existing nonlinear echo cancellation solutions. Section 3 presents results comparing the existing algorithms. Section 4 presents the proposed method to improve the performance of existing linear echo cancelers in the presence of nonlinearities. Results using this method are presented in section 5. The last section concludes the paper with possible directions for future work.

## 2. BACKGROUND

A large number of nonlinear acoustic echo cancellation algorithms have been proposed. The approaches for nonlinear echo cancellation can be broadly classified into three types, viz. Time domain, Frequency domain and Subband domain. Most of the work in the field of nonlinear echo cancellation has been done in time domain. Time domain approaches have used neural networks to train for the nonlinear component [2, 3, 4]. More commonly used methods have used volterra filters to model nonlinearities with memory [5, 6, 7, 8] and without memory [9]. Approaches have also been proposed for an online estimation of memory size [10, 11]. Some approaches also account for stability in the absence of nonlinearity using a convex combination scheme [4, 7, 12].

In the frequency domain, similar techniques have been developed [13, 14, 15]. These techniques attempt to utilize the faster convergence in the frequency domain to achieve faster tracking for the rapidly changing nonlinearities in the downlink path. The obvious extension to this work is in the subband domain. In sub-band domain algorithms, nonlinearity generation is done in the time domain and processing is done using polyphase filter banks [16]. To eliminate the inherent delay problem with sub-band approaches, delayless sub-band approaches have also been explored [17].

All the algorithmic solutions suffer from the problem of parameter selection. Parameters such as memory size and order of nonlinearity have to be accurately selected to benefit from these algorithms. This information is not generally available and also difficult to predict due to the constantly varying nature of the nonlinearities. Algorithms have been proposed for an online estimation of these parameters. Online estimation methods involve adapting multiple kernels for the same order of nonlinearity. This significantly increases the complexity of the algorithm and makes them prohibitive for real-time use.

In [18], the authors propose the use of an accelerometer mounted on the magnet of the loudspeaker to record the acoustic output of the loudspeaker. The accelerometer mounted on the magnet of the loudspeaker is expected to capture the nonlinearities introduced in the downlink signal. If this signal is used as the reference signal for the linear echo canceler, the echo canceler does not have to model any nonlinearities. This approach provides nonlinear echo cancellation at no additional computational cost. It seems to provide 15 dB ERLE enhancement over a linear echo canceler in case of 40% non-coherent distortion. Mounting accelerometers on a loudspeaker used in cell phones is non-trivial and also adds distortion to the loudspeaker being used. Thus, the additional performance enhancement may be due to the additional distortion created by the presence of the accelerometer.

## 3. RESULTS FOR EXISTING ALGORITHMS

### 3.1. Simulation Results

The performance of some of the existing algorithms was tested using simulations. The amount of nonlinearity was measured in terms of Linear-to-Nonlinear Ratio (LNLR), as defined in [6]. The synthetic nonlinear speech signal was generated using a 128-tap long linear impulse response obtained from a measurement on a cell phone. Second order nonlinearity was added using a kernel derived from the linear impulse response. The memory size of this kernel was 100 samples. The sampling rate was set to 8 kHz. The signal to noise ratio was set to 30 dB. The length of the estimated linear impulse response was 128 taps. Second order nonlinearities were modeled with a memory size of 100 samples. The step size and regularization parameters were set in order to not significantly change performance across algorithms. The LNLR was varied from 40 dB (linear only) to 0 dB (highly nonlinear).

Figure 3(a) shows the comparison of the ERLE performance of all the algorithms with a linear NLMS algorithm. The parameters used were as suggested by the respective authors. In case the parameters were not explicitly specified, the values providing best and consistent performance were used. The cases of maximum interest are for the LNLR values of 20 dB, 15 dB and 10 dB, since these are realistic values of LNLR for real signals. Two algorithms, Combination of Kernels Scheme (CKS) and Combination of Filters Scheme (CFS), seem to provide the best performance in these cases. Also, the ERLE performance of these algorithms is fairly consistent (around 22 dB) for all levels of nonlinearities.

### 3.2. Experimental Results

The performance of the algorithms was also compared using real data from a cell phone. The excitation signal, shown in figure 2, was a 10 second speech signal with the device set in speakerphone mode. The sampling rate was set to 8 kHz.

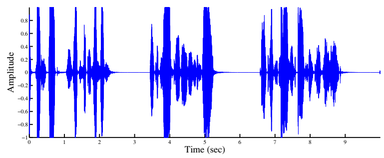
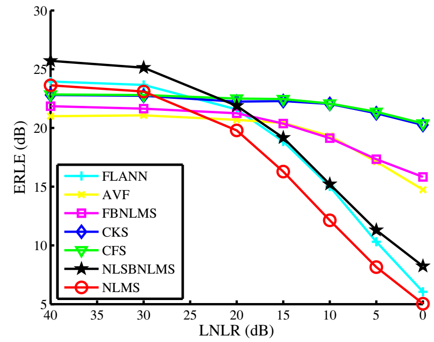


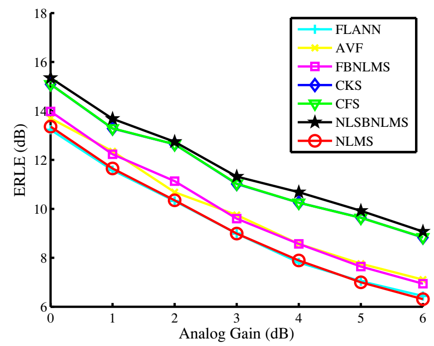

**Fig. 2**. Excitation signal used for performance comparison.

The linear impulse response was estimated using a 128-tap filter and second order nonlinearities were estimated using a second order kernel with a memory of 100 samples. The measurements were made in a semi-anechoic chamber with a high SNR of approximately 30 dB.

The results are shown in figure 3(b). As seen in the figure, the ERLE performance of all the algorithms drops with an increase in the signal level. This increase in signal level may be compared to the decrease in LNLR, i.e. increase in the amount of nonlinearities in the acoustic output of the loudspeaker. Also, decisions have to be made with respect to the order of nonlinearity to be modeled and memory size in order to obtain a considerable improvement. If these parameters are under-modeled, it often leads to a deteriorated ERLE performance. The drop in performance with real data, as compared to the case with simulations, can be attributed to the mismatch in the nonlinear model. Over-modeling of these parameters does not affect performance but forces a large real-time constraint on the implementation in terms of higher MIPS (Million Instructions Per Second). This indicates that if accurate information of the nonlinear model of the downlink path is not known, the existing algorithms do not provide improvement. Thus, alternate methods to achieve the desired performance have to be considered.




(a) ERLE performance of existing algorithms: simulations



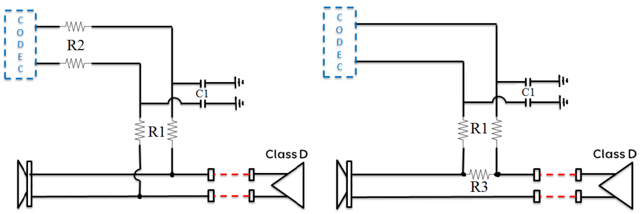

(b) ERLE performance of existing algorithms: experiments

**Fig. 3**. Comparison of ERLE performance of existing algorithms.

## 4. PROPOSED APPROACH

The path from the output of the loudspeaker to the input of the echo canceler can be considered linear. Thus, if the acoustic output of the loudspeaker could be predicted or obtained and used as the reference signal for the echo canceler, the linear echo canceler would have to only model a linear path. The easiest option would be to place a microphone in the loudspeaker cavity and use the signal recorded by this microphone as the reference signal for the linear echo canceler. In addition to providing nonlinear echo cancellation, this microphone will also pick up any background noise, add it to the reference signal and may be able to provide some noise suppression as well. Unfortunately, the sound pressure level in the loudspeaker cavity is of the order of 155 dBSPL and the microphones used in present devices support sound pressure levels upto 120 dBSPL. As a result the microphone will saturate and the reference signal will contain additional nonlinearities. Thus specialized microphones have to be manufactured in order to use this approach. Also, this microphone has the potential to pick up the near end talker (in relatively quiet environments) and suppress the near end talk to a noticeable extent. This makes it unfeasible to use a microphone in the loudspeaker cavity. In addition to the manufacturing cost, adding another microphone adds to the space requirements in cell phones that are continuously shrinking in size.

Since the output of the loudspeaker cannot be captured in the acoustic domain, an alternate method to obtain this signal may help improve the performance of the linear echo canceler. We propose the use of the voltage or the current signal that drives the loudspeaker as the reference signal. This signal will ensure that any nonlinearities introduces by the codec or the downlink amplifier are captured. In addition to this, any nonlinearities that can be captured as a result of back emf from the loudspeaker will also be captured in this reference signal. Figure 4 shows the changes required in order to obtain these signals. In order to use the voltage signal as the reference signal, the voltage at the loudspeaker terminals is applied to the input of an RC filter to eliminate any high frequency noise. This filtered signal is scaled down to the range of the codec using a potential divider circuit and then used as the reference signal for the linear echo canceler. In the case of using the current drawn by the loudspeaker as the reference




(a) Voltage feedback. (b) Current feedback.

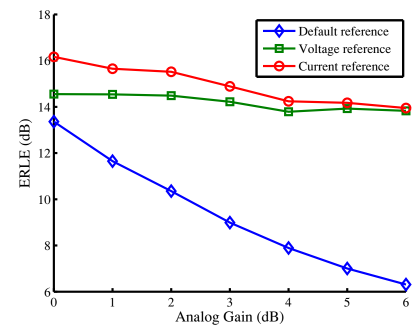
**Fig. 4**. Hardware changes for the proposed techniques.

signal, the voltage drop across a very small resistor is used as the input to an RC filter to suppress any high frequency electrical noise. A potential divider circuit is not required in this case because the voltage drop across a small resistor is in the range of the codec.

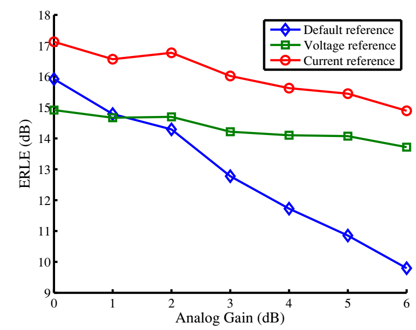
## 5. RESULTS FOR PROPOSED APPROACH

The proposed modification is a change in the hardware and thus simulating the effect is not feasible. As a result, only results from real data collected on a cell phone are presented. Figure 5 shows the ERLE performance for two different volume settings. The first one is for the maximum device volume setting and the second one corresponds to a lower device volume. In both the cases, the device was set in speakerphone mode. The same excitation signal as used in section 3.2 was used for these experiments. Linear NLMS was used to obtain the ERLE values for the three different reference signals. The advantage of using an alternate reference signal is clearly seen when the device volume is set to its maximum. For the device used, the default value of analog gain was 0 dB. Thus, for an additional 3 dB gain, the use of the current signal as the reference signal shows an ERLE enhancement of almost 6 dB. Keeping in mind that the excitation signal used was a speech signal, which includes silent periods also, this 6 dB enhancement in the ERLE performance of the linear echo canceler is a significant improvement. Another noticeable aspect is that the ERLE performance when using the suggested signals as reference does not deteriorate with the increase in analog gain. Also, the difference in ERLE performance with the current or voltage signal is not a lot. This shows that either of the signals may be used as the reference signal.

In the second case, when the device volume was set to a lower level, the ERLE performance using these reference signals was better than when using the default digital reference signal for all values of analog gain above 1 dB. For 0 dB analog gain, the ERLE performance is comparable. This proves that using these alternate reference signals not only enhances the performance of the linear echo canceler in the case of highly nonlinear signals but also does not deteriorate the performance in the case of minimal distortion.




(a) ERLE performance for maximum device volume.




(b) ERLE performance for minimum device volume.

**Fig. 5**. Comparison of ERLE performance using three different reference signals and two different volume settings.

## 6. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated that by using voltage or current feedback as the reference input to a linear echo canceler, the ERLE performance can be enhanced in the presence of nonlinearities. This is clearly due to the fact that these alternate reference signals capture some or all of the nonlinearities in the downlink path. More analysis should be done to make a selection between the current and voltage signal. A thorough analysis on how much nonlinearity is captured by these reference signals has to be done. Also, it will be useful to obtain a relation between amount of distortion and the drop in ERLE performance of a linear echo canceler.

## 7. RELATION TO PRIOR WORK

Prior work in the field of nonlinear echo cancellation has been focused on time-domain, frequency-domain and sub-band domain algorithms. Only one hardware modification has been discussed in literature, i.e. the use of an accelerometer. This paper proposes a hardware modification to use an alternate reference signal to obtain better ERLE performance in nonlinear conditions. This method provides an average ERLE enhancement of up to 6 dB over a linear echo canceler without any additional MIPS.

## 8. REFERENCES


[1] A.N. Birkett and R.A. Goubran, "Limitations of Handsfree Acoustic Echo Cancellers due to Nonlinear Loudspeaker Distortion and Enclosure Vibration Effects," in *Proceedings of the 1995 IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics*, October 1995.

[2] A.N. Birkett and R.A. Goubran, "Nonlinear Echo Cancellation Using a Partial Adaptive Time Delay Neural Network," in *Proceedings of the 1995 IEEE Workshop on Neural Networks for Signal Processing V*, August 1995, pp. 249–258.

[3] A.N. Birkett and R.A. Goubran, "Fast Nonlinear Adaptive Filtering Using a Partial Window Conjugate Gradient Algorithm," in *Proceedings of the 1996 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 1996, vol. 6, pp. 3541–3544.

[4] D. Comminiello, R.P. Scarpiniti, and A. Uncini, "A Functional Link Based Nonlinear Echo Canceller Exploiting Sparsity," in *Proceedings of the 2010 International Workshop on Acoustic Echo and Noise Control*, September 2010.

[5] A. Guerin, G. Faucon, and R. Le Bouquin-Jeannes, "Nonlinear Acoustic Echo Cancellation Based on Volterra Filters," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 672–683, November 2003.

[6] L.A. Azpicueta-Ruiz, M. Zeller, J. Arenas-Garcia, and W. Kellermann, "Novel schemes for nonlinear acoustic echo cancellation based on filter combinations," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Washington, DC, USA, 2009, pp. 193–196.

[7] L.A. Azpicueta-Ruiz, M. Zeller, A.R. Figueiras-Vidal, J. Arenas-Garcia, and W. Kellermann, "Adaptive Combination of Volterra Kernels and Its Application to Nonlinear Acoustic Echo Cancellation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 97–110, January 2011.

[8] F. Kuch and W. Kellermann, "Nonlinear Line Echo Cancellation Using a Simplified Second Order Volterra Filter," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2002, pp. 1117–1120.

[9] Alexander Stenger and Walter Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1747–1760, September 2000.

[10] Marcus Zeller, Luis A. Azpicueta-Ruiz, and Walter Kellermann, "Online Estimation of the Optimum Quadratic Kernel Size of Second-Order Volterra Filters using a Convex Combination Scheme," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Washington, DC, USA, 2009, ICASSP '09, pp. 2965–2968.

[11] M. Zeller, L.A. Azpicueta-Ruiz, J. Arenas-Garcia, and W. Kellermann, "Adaptive volterra filters with evolutionary quadratic kernels using a combination scheme for memory control," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1449–1464, April 2011.

[12] J. Arenas-Garcia, A.R. Figueiras-Vidal, and A.H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1078–1090, March 2006.

[13] F. Kuech and W. Kellermann, "Partitioned block frequency-domain adaptive second-order volterra filter," *Signal Processing, IEEE Transactions on*, vol. 53, no. 2, pp. 564–575, February 2005.

[14] Osamu Hoshuyama and Akihiko Sugiyama, "Nonlinear echo cancellation based on spectral shaping," in *Speech and Audio Processing in Adverse Environments*, Eberhard Hansler and Gerhard Schmidt, Eds., Signals and Communication Technology, pp. 267–283. Springer Berlin Heidelberg, 2008.

[15] M. Zeller, L.A. Azpicueta-Ruiz, J. Arenas-Garcia, and W. Kellermann, "Efficient Adaptive DFT-Domain Volterra Filters using an Automatically Controlled Number of Quadratic Kernel Diagonals," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 4062–4065.

[16] T.G. Burton, R.A. Goubran, and F. Beaucoup, "Nonlinear System Identification Using a Subband Adaptive Volterra Filter," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, no. 5, pp. 1389–1397, May 2009.

[17] D. Zhou, V. DeBrunner, Y. Zhai, and M. Yeary, "Efficient Adaptive Nonlinear ECHO Cancellation, Using Sub-band Implementation of the Adaptive Volterra Filter," in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2006, vol. 5.

[18] T. Gupta, S. Suppappola, and A. Spanias, "Nonlinear Acoustic Echo Control Using an Accelerometer," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2009, pp. 1313–1316.