NONLINEAR ACOUSTIC ECHO CONTROL USING AN ACCELEROMETER

*Tushar Gupta[1], Seth B. Suppappola[*2], and Andreas Spanias[1]*

[1]SenSIP Center, Arizona State University, Tempe, AZ 85281 USA
[2]Acoustic Technologies, Inc., Mesa, AZ 85204 USA
tgupta1@asu.edu, seth.suppappola@acoustictech.com, spanias@asu.edu

ABSTRACT

A typical echo canceller uses an FIR adaptive filter to estimate the impulse response of the near-end echo path. However nonlinear echo components due to the loudspeaker and driving amplifier are not captured by such an arrangement. We propose a novel scheme for nonlinear acoustic echo cancellation where an accelerometer sensor captures loudspeaker vibration and this sensor signal is used along with the error signal to adapt the adaptive filter coefficients. The accelerometer sensor is used in lieu of explicit nonlinear modeling in the echo cancelling loop and hence it is compatible with typical NLMS and RLS echo cancellation algorithms. Experiments with an accelerometer mounted on the magnet of a hands-free kit show an echo reduction improvement by as much as 15 dB in situations of nonlinear loudspeaker distortion.

Index Terms— Nonlinear acoustic echo cancellation, accelerometer, NLMS, RLS, total non-coherent distortion.

1. INTRODUCTION

Nonlinear effects in real-time acoustic echo cancellation result in significant loss of performance particularly in compact hands-free kits (HFKs). Nevertheless the increasing demand for slimmer HFKs requires new algorithms and enclosure solutions. Much of the problems associated with HFKs have been attributed to the inexpensive, low-quality loudspeakers that are generally poorly decoupled from the outer enclosure. When such a loudspeaker is driven with a high sound pressure level (SPL) signal, saturation effects associated with the loudspeaker and its amplifier distort speech in a nonlinear manner. The acoustic echo signal in such a case contains a mixture of linear and nonlinear distortion components. An acoustic echo canceller (AEC) with a typical NLMS or RLS adaptive filter estimates only the linear acoustic impulse response (AIR) of the loudspeaker enclosure microphone (LEM) system. The remaining nonlinear echo signal components could be large and audible particularly at high volume levels.

Several approaches have been used to cancel the nonlinear echo [1-10]. One approach is nonlinear preprocessing of the loudspeaker signal. This method involves pre-distortion of the loudspeaker signal aimed at compensating for its nonlinear behavior. A coherence function criterion for the identification and compensation of nonlinear distortion in a Hammerstein system is proposed in [1]. This approach could degrade the quality of speech in practical applications. It also assumes that the loudspeaker system is memoryless. Another approach is to employ a nonlinear adaptive filter such as the Volterra or the orthogonal polynomial filters [2-4]. Different realizations of the Volterra filters have been presented in [5-7]. The nonlinear echo path has been modeled by a cascade of memoryless nonlinear preprocessor function followed by FIR adaptive filter in [8]. The use of a nonlinear transformation based on the raised cosine function for reducing the nonlinear echo components has been proposed in [9]. Neural network based structures have also been proposed in [10]. Most or all of the aforementioned techniques have high computational cost and are associated with slow convergence which makes them unsuitable for most real-time embedded applications.

Our objective is to reduce the nonlinear echo to a sufficiently low level with reasonable computational complexity. We propose an approach to canceling nonlinear echo that simply involves measuring the nonlinearly distorted speech signal from the loudspeaker and then using it along with the error signal to adapt the FIR echo canceller. The proposed acoustic echo canceller uses an accelerometer sensor mounted on the magnet of a loudspeaker to capture the nonlinearly distorted speech signal. This is then used as a reference signal to the conventional adaptive FIR filter for compensating the linear and nonlinear components of the acoustic echo. By using an accelerometer, we avoid estimating the nonlinear parameters of the LEM system. A detailed physical simulation is carried out to demonstrate the effectiveness of the proposed method. We show that with the small additional cost of an accelerometer and a second analog-to-digital converter, the proposed technique enhances the performance of traditional algorithms considerably in the presence of nonlinear distortion.

2. NONLINEAR DISTORTION

The total LEM system transfer characteristic has linear and nonlinear components. Therefore the acoustic echo signal can also be considered as sum of linear and nonlinear echo. The signal flow from the received far end signal to the transmit signal is shown in Figure 1. A nonlinear mapping of the speech signal could occur during D-to-A and A-to-D conversions but those are generally assumed to be linear because of modern high resolution converters used in today's telephony devices [11]. The overloaded loudspeaker amplifier causes memory-less nonlinear distortion

[*] *This work is sponsored by Acoustic Technologies Inc., Mesa, Arizona through the SenSIP consortium. A major portion of this work has been performed at Acoustic Technologies facilities.*

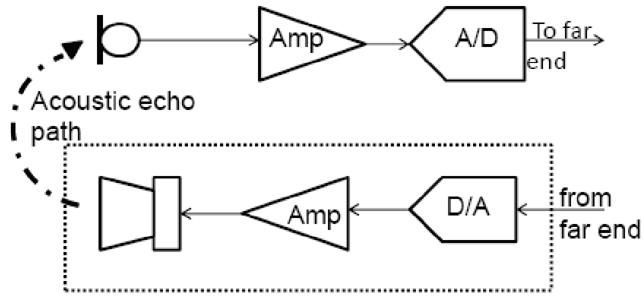

Figure 1. Nonlinear echo path model.

through creation of harmonics and inter-modulation products by the soft clipping of large amplitude values. Enclosure vibration due to mechanical coupling of the loudspeaker components with the case of the device especially at the lower voice frequencies causes significant nonlinear distortion to be picked up by the microphone. Finally the loudspeaker itself is a major source of nonlinear distortion. The nonlinearities in the loudspeaker could be due to acoustic, electromagnetic, and mechanical reasons [12]. If we could capture the signal present immediately after the section shown in Figure 1 by a dotted box, then we should be able to capture most of the nonlinear distortion and the acoustic path could then be estimated using existing AEC techniques.

One way is to capture the loudspeaker acoustic signal by placing a microphone just next to the loudspeaker. But this method is not very useful due to it also sensing the near-end speech and the background noise. An optical sensor can be used to measure the loudspeaker's cone displacement [13]. Experiments were performed using a laser position sensitive detector (PSD) sensor to capture the motion of cone of the loudspeaker and to obtain the echo source signal. However, this method requires sophisticated hardware that drives up the cost of real-time device.

An accelerometer is an electromechanical device that measures acceleration. It can be used to measure vibrations precisely. Experiments with an accelerometer mounted on the frame, the cone, and the magnet of a loudspeaker were performed. Unfortunately, the frequency response of the loudspeaker changes when an accelerometer is mounted on the cone due to the low mass of the cone of a small loudspeaker. An accelerometer mounted on the loudspeaker frame or magnet gives similar signals, and mounting an accelerometer at these positions does not affect the loudspeaker's frequency response. An accelerometer could also be placed on the device enclosure to sense the vibrations.

Nonlinear distortion depends highly on the stimulus signal's time-domain amplitude and spectral content. Total harmonic distortion (THD) measurement quantifies harmonic distortion due to single tones, but it cannot quantify intermodulation distortion (IMD) in a multitone signal caused by the nonlinear behavior of a system. To keep distortion metrics meaningful, efficient measures that can quantify nonlinear distortion due to multi-tone signals, such as speech, are required.

ITU-T Recommendation O.42 [14] measures nonlinear distortion using a 4-tone inter-modulation method, but it is designed for telephony network equipment and uses only four specific tones. If those tones do not stimulate the system's nonlinearity then the distortion cannot be measured properly. Stenger et al. [15] have proposed measuring the impulse response of a LEM with cheap analog equipment in the presence of nonlinear distortion and background noise by modeling nonlinear distortion as noise and then estimating the linear part using the Wiener solution. There is an absence of any clear definitive metric that defines the nonlinear distortion.

Non-Coherent Distortion (NCD) is a nonlinear distortion measure that provides a continuous frequency domain distortion curve for multi-tone signals such as speech or music [16]. A single nonlinear distortion measure called Total Non-Coherent Distortion (TNCD) is obtained by power summing the NCD curve across the frequency band from DC up to the Nyquist frequency. It is the ratio of the uncorrelated noise plus the nonlinear distortion power to the total output power. TNCD is calculated as

$$\lambda = \sqrt{\frac{\sum_{\omega} G_{NN}(\omega)}{\sum_{\omega} G_{yy}(\omega)}} \tag{1}$$

where

$$\left[1 - \frac{|G_{XY}(j\omega)|^2}{G_{XX}(\omega) G_{YY}(\omega)}\right] * G_{YY}(\omega) = G_{NN}(\omega) \tag{2}$$

with $G_{XX}(\omega)$ and $G_{YY}(\omega)$ being the real and positive autospectra of the input and output signals respectively. $G_{XY}(j\omega)$ is the complex input-output cross spectrum. Figure 2 shows the TNCD against different sound pressure levels (SPL) for a small loudspeaker. It is observed that the TNCD increases very sharply as the loudspeaker volume is increased.

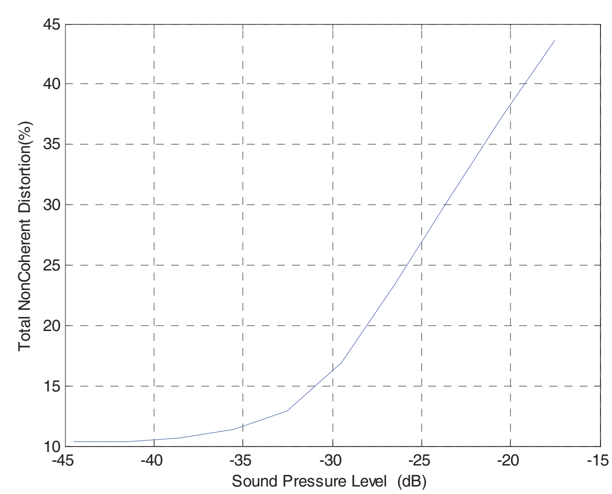

Figure 2. TNCD vs. Sound pressure level on an arbitrary scale.

3. PROPOSED AEC ARCHITECTURE

Figure 3 shows the proposed configuration for nonlinear acoustic echo cancellation where $x(n)$ denotes the far end received signal that is fed to the loudspeaker.

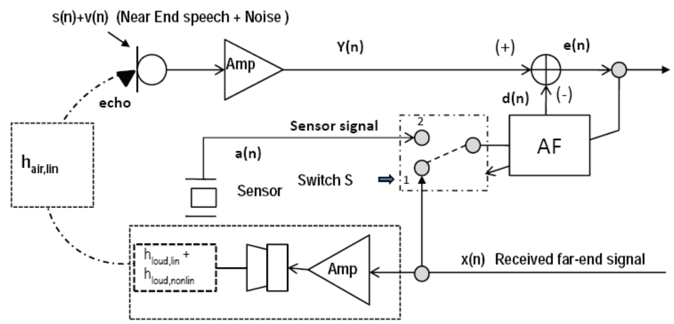

Figure 3. Block diagram of the proposed AEC.

Let the linear responses of the loudspeaker and air be $h_{loud,lin}$ and $h_{air,lin}$ respectively. The total linear response of the LEM system is then equal to:

$$h_{lin} = h_{loud,lin} + h_{air,lin} \tag{3}$$

y(n) denotes the signal recorded by the near end speech microphone. It is given by

$$y(n) = x(n) * h_{lin}(n) + x_{NL}(n) + s(n) + v(n) \tag{4}$$

with $x_{NL}(n)$ being the nonlinear distortion and *s(n)* and *v(n)* being the near end speech and background noise respectively. The accelerometer senses the nonlinearly distorted signal *a(n)* and acts as a reference signal to the adaptive filter (AF) when the switch *S* is in position 2. The adaptive filter is used to estimate the combined linear response of LEM system. The filter output *d(n)* is an estimate of *y(n)* and can be used to cancel the echo. In this study, we have used the NLMS algorithm [17,18, 20] to adapt the adaptive FIR filter of order *L*. The cost function or the error signal *e(n)* is given by the difference between the desired signal *y(n)* and the estimated signal *d(n)* as

$$e(n) = y(n) - d(n) \tag{5}$$

where

$$d(n) = \mathbf{w}_n^T \mathbf{x}_n \tag{6}$$

The adaptive filter coefficients $\mathbf{w}_n$ and the input signal vector $\mathbf{x}_n$ are given by

$$\mathbf{w}_n = [\mathbf{w}_n(0)\mathbf{w}_n(1)\mathbf{w}_n(2)...\mathbf{w}_n(L)]^T$$
$$\mathbf{x}_n = [\mathbf{x}_n(n)\mathbf{x}_n(n-1)\mathbf{x}_n(n-2)...\mathbf{x}_n(n-L)]^T$$

The update formula for the filter coefficient vector is given by

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \frac{e(n)\mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + r} \tag{7}$$

where μ is the step size, and r is a small constant called the regularization constant. However, the proposed AEC architecture can be implemented with any block LMS or RLS adaptive filter algorithm without adding any additional computational complexity to handle nonlinearities.

In the proposed scheme (Fig. 3, switch *S* in position 2), we are capturing the nonlinear signal by measuring the vibration of the cone that renders the speech signal after it has been nonlinearly distorted by the amplifier, the loudspeaker, and/or the vibrating enclosure. Moreover, the sensor signal captures both the memory-less nonlinearities due to soft clipping of the overdriven amplifier and the loudspeaker nonlinearities with memory. The adaptive filter is required to model only the linear acoustic impulse response since the adaptive filter reference signal is the nonlinearly distorted signal. By contrast, in a conventional AEC architecture (Fig. 3, switch *S* in position 1), the linear adaptive FIR filter has to model both the nonlinear and linear response of the system and, it will fail to model the nonlinearities.

4. EXPERIMENTS AND RESULTS

The performance of the proposed nonlinear echo canceller has been measured using the echo return loss enhancement (ERLE) measure [19]. The ERLE is defined as

$$ERLE(n) = \frac{E[y^2(n)]}{E[e^2(n)]} \tag{8}$$

where *E* is the statistical expectation. The experimental data was collected in a car with low noise. An 11s long speech signal was fed into a 2 W oval loudspeaker of dimensions 50 mm × 32 mm. The echo signal was sensed by a 6 mm omnidirectional microphone. The speech file used consists of 200 ms white noise burst appended with ITU-T Recommendation P.50 artificial male voice at the far end. Different sets of data were obtained at different sound pressure levels. The signals were recorded at 48 kHz sampling frequency with a 16-bit ADC and subsequently down-sampled to 16 kHz for analysis. A piezoelectric accelerometer weighing 8 grams with dynamic range of $\pm$150 g was mounted on the magnet of the loudspeaker to measure the vibration of the cone as the loudspeaker outputs the speech. The AEC was adapted using the NMLS algorithm with 512 taps. The step size μ was set to 0.5 and regularization factor *r* was equal to 10^{-6}. It can be seen in Fig. 4 that the residual echo power is much lower when the accelerometer is used as compared to when the far-end signal is used (i.e., the conventional method). Fig. 5 shows ERLE versus TNCD for both conventional and proposed architectures. The ERLE curve for the accelerometer signal remains almost flat as TNCD increases, whereas ERLE decreases very sharply with increasing TNCD when the accelerometer is not used. At the maximum TNCD representing a 40 dB increase in volume, the accelerometer approach improves ERLE more than 15 dB. Note that although additional hardware was used to accommodate the accelerometer, the overall computational complexity is the same as the ordinary AEC system that uses the NLMS algorithm.

As a mechanical sensor, an accelerometer is insensitive to near-end speech and background noise. If a second microphone was used for the nonlinear distortion, it would inevitably sense the near-end speech and hence the AEC would attempt to cancel it. As compared to a microphone, the accelerometer was 50 dB less sensitive to near-end speech and background noise, yet just as sensitive to echo.

5. CONCLUSION

A conventional acoustic echo canceller based on FIR adaptive filtering cannot cancel adequately echo at high sound pressure levels because of the nonlinear components generated by the audio amplifier and the loudspeaker. In this paper, we presented a nonlinear acoustic echo cancellation method that incorporates an accelerometer which captures the nonlinearly distorted speech signal. *Experiments with the NLMS AEC and with an accelerometer mounted on the magnet of a hands-free kit show an ERLE improvement by as much as 15 dB in situations of nonlinear loudspeaker distortion.* Similar results can be obtained with the appropriately tuned RLS or block LMS type algorithms. Note that although additional hardware was used to accommodate the accelerometer, the overall computational complexity is the same as the ordinary AEC system that uses the NLMS algorithm.

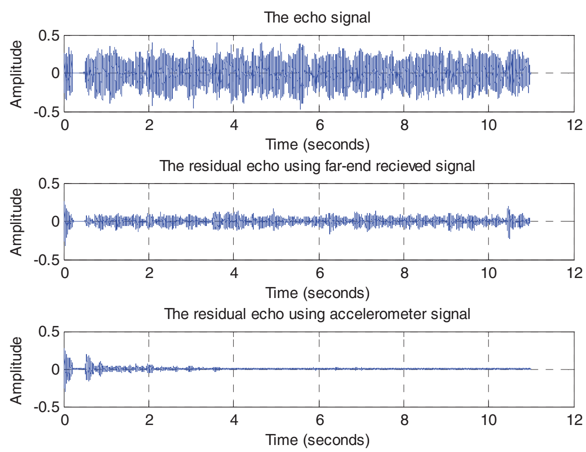

Figure 4. Residual echo signal comparison.

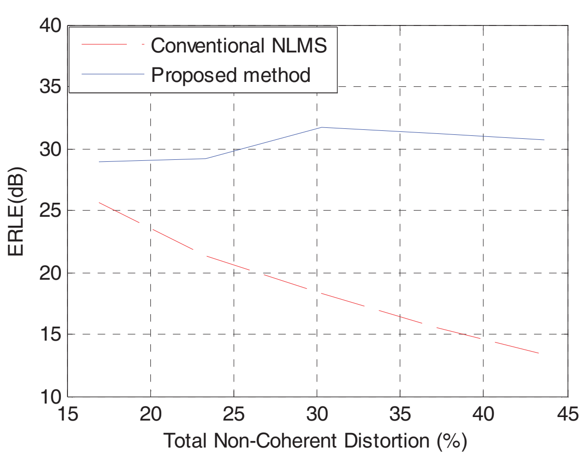

Figure 5. ERLE vs. TNCD.

REFERENCES

[1] K. Shi, G. Tong Zhou, M. Viberg, "Compensation for Nonlinearity in a Hammerstein System Using the Coherence Function With Application to Nonlinear Acoustic Echo Cancellation", *IEEE Trans. Signal Process*, Vol. 55(12), pp.5853 – 5858, Dec. 2007.

[2] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE ICASSP*, vol. 2, pp. 877–880, Mar. 1999.

[3] A. Guerin, G. Faucon, and R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters*," IEEE Trans. Speech Audio Process,* vol. 11, no. 6, pp. 672–683, Nov. 2003.

[4] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE ICASSP*, Philadelphia, vol. 3, pp. 105–108, Mar. 2005.

[5] F. Kuech and W. Kellermann, "Partitioned block frequency-domain adaptive second-orderVolterra filter," *IEEE Trans. Signal Process*, vol.53, no. 2, pp. 564–575, Feb. 2004.

[6] D. Zhou; V. Debrunner, Y. Zhai, M. Yeary**,** "Efficient Adaptive Nonlinear Echo Cancellation, Using Sub-band Implementation of the Adaptive Volterra Filter", *Proc. IEEE ICASSP*, vol. 5, pp. 277–280, May 2006.

[7] F. Kuech and W. Kellermann, "A novel multidelay adaptive algorithm for Volterra filters in diagonal coordinate representation," *Proc. IEEE ICASSP*, pp. 869–72, May 2004.

[8] A. Stenger and W. Kelermann, "Nonlinear Acoustic Echo Cancellation with Fast Converging Memoryless Preprocessor," in *Proc. IEEE ICASSP*, vol.2, pp.805-808, June 2000.

[9] H. Dai and W. Zhu, "Compensation of Loudspeaker Nonlinearity in Acoustic Echo Cancellation Using Raised-Cosine Function," *IEEE Trans. On Circuits and System II,*Vol. 53, No.11, pp.1190-1194, Nov. 2006.

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

[11] F. Kuech, W. Kellermann, "Nonlinear Acoustic Echo Cancellation," *Topics in Acoustic Echo & Noise Control*, Hänsler, Eberhard; Schmidt, (Eds.) 2006, Ch. 7 pp. 205-257

[12] Wolfgang Klippel, "Tutorial: Loudspeaker Nonlinearities—Causes,Parameters, Symptoms", *J. Audio Eng. Soc.*, Vol. 54, No. 10, pp. 907 -939, October 2006.

[13] W. Geiger, "Servo Control of Loudspeaker Cone Motion Using an Optical Linear Displacement Sensor", *J. Audio Eng. Soc.*, Vol. 53, No. 6, June 2005.

[14] ITU-T Series O - Specifications of measuring equipment, "O.42 - Equipment to measure non-linear distortion using the 4-tone intermodulation method", Date: 10/84.

[15] F. Heinle, R. Rabenstein, and A. Stenger, "A measurement method for the linear and nonlinear properties of electro_acoustic transmission systems", Elsevier Preprint, SIGNAL PROCESSING, vol. 64, no. 1, pp. 49-60, 1998.

[16] S. Temme And P. Brunet, "A New Method for Measuring Distortion Using a Multitone Stimulus and Noncoherence", *J. Audio Eng. Soc.*, Vol. 56, No. 3, pp.176-188, March 2008.

[17] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.

[18] S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, Boston, 2000.

[19] Eberhard Hänsler, Gerhard Schmidt, *"Acoustic Echo and Noise Control: A Practical Approach",* John Wiley & Sons, 2004.

[20] A. Spanias, Digital Signal Processing; An Interactive Approach, Chapter 9, ISBN: 978-1-4243-2524-5,Lulu Press On-demand Publishers, Book #2581497, Morrisville, NC , Sept. 2007.