POSTPRINT VERSION. The final version is published here :
Merabti, H., Massicotte, D. et Zhu, W. P. (2019, 10-12 Dec. 2019). Nonlinearity-Robust Linear Acoustic Echo Canceller Using the Maximum Correntropy Criterion. Communication présentée 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman. doi: 10.1109/ISSPIT47144.2019.9001737

Nonlinearity-Robust Linear Acoustic Echo Canceller Using the Maximum Correntropy Criterion

Hocine Merabti[1], Daniel Massicotte[1], Wei-Ping Zhu[2]

[1] *Dept. of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, Québec, Canada*
E-mail: {hocine.merabti, daniel.massicotte}@uqtr.ca
[2] *Dept. of Electrical and Computer Engineering, Concordia University, Québec, Canada*
E-mail: weiping@ece.concordia.ca

***Abstract*—For the problem of acoustic echo cancellation (AEC) with nonlinear distortions, we propose to use a linear adaptive filter that maximizes the Correntropy similarity measure instead of the conventional minimization of the mean squared error (MSE) criterion. The maximum Correntropy criterion (MCC) offers robustness to outliers and impulsive noise, which is interesting for the case of speech signal coupled with nonlinearities. To assess the performance of the algorithm, we consider a hard-clipping memoryless saturation nonlinearity. Our simulation results show very interesting performance of the normalized MCC-based linear adaptive filter for the echo return loss enhancement (ERLE) and misalignment measures compared to the MSE-based normalized least mean squares (NLMS) approach. Furthermore, the NMCC adaptive filter has a similar computational complexity as the NLMS algorithm, which makes it very attractive in practical implementations**

Keywords*— *acoustic echo cancellation, nonlinear systems, adaptive filtering, correntropy, mean squared error

I. Introduction

We use speech communication systems very often in our daily life via services operating over cellular networks and the internet. A very annoying type of distortion in these systems is the acoustic echo, which occurs due to the reflection of acoustic waves on the surfaces of the room and surrounding objects. Linear acoustic echo cancellation (AEC) systems are widely used, and aim to cancel the echo effect by adaptively identifying the time varying room impulse response (RIR), which is generally modeled by a linear finite impulse response (FIR) filter [1]. Linear adaptive filtering algorithms like the least mean squares (LMS) and its variants, and the recursive least squares (RLS) are commonly used [2]. However, other components of the echo-path like the analog amplifier and the loudspeaker exhibit non-negligible nonlinear behavior when driven near their saturation region, thereby introducing nonlinear distortions, leading to poor echo cancellation performance of linear adaptive filtering algorithms due to their inability to model nonlinearities.

To address the problem of nonlinearities in the echo-path, nonlinear acoustic echo cancellation (NL-AEC) methods are used. The basic idea is to introduce a nonlinear model in cascade with the linear filter modeling the RIR, and adaptively identify the models parameters. The nonlinear model is expected to approximate the behavior of the real nonlinear components to yield good practical results. Nonlinear models based on Volterra expansions [3], polynomial expansions [4][5], saturation-like nonlinearities [6]-[9] and others [10]-[12] have been proposed. In other approaches, additional hardware components like voltage/current sensors [13] and accelerometers [14] are used to capture the nonlinearities from the loudspeaker and use it to feed the linear adaptive filter.

It is important to mention that most of these methods are computationally complex to implement in practice, or require the use of additional hardware, which is not compatible with most of the already available speech communication systems in the market.

In this paper, we approach the problem of nonlinear distortions as a problem of presence of outliers in the optimization process, which are caused by high amplitude samples of the speech signal being distorted by the nonlinear components of the system. In this perspective, the process of identifying the linear RIR using a linear adaptive filter robust to outliers would provide interesting improvements compared to the conventional linear adaptive filters. We will compare the performance of the correntropy similarity measure against the popular mean squared error (MSE) criterion [15] used by traditional adaptive filters.

The rest of this paper is organized as follows: the acoustic echo cancellation problem with nonlinear distortions, and the hard-clipping memoryless saturation are presented in Section II. Section III covers the selected outliers-robust linear adaptive filter. In Section IV, simulation results and discussions are given. Finally, we draw our conclusion in Section V. Nonlinear Modeling and Inverse Filter

II. Acoustic Echo Cancellation Problem

The block diagram of the linear acoustic echo cancellation system in the presence of nonlinear distortions is shown in Fig. 1. The main sources of nonlinearities are the analog amplifier and the loudspeaker in the downlink path [5]. The far-end input signal $x(n)$ is first affected by a nonlinear transformation $f(\bullet)$ before being linearly convolved with the room impulse response $h(n)$. At the microphone, the echo path signal $d(n)$ is superimposed with a corrupting background noise $e_0(n)$. The resulting signal $y(n)$ is used by the canceller as a reference entry with the goal of estimating the echo path which is commonly achieved by minimizing the power of the error signal $e(n)$, thereby aim to cancel the return of the far-end signal and its delayed versions to the source.

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

At the echo path level, the system equations are described as follows:

The time-variant FIR filter which models the linear RIR $\mathbf{h}(n)$, and the input vector $\mathbf{x}(n)$ is denoted as

$$\mathbf{h}(n) = [h_0(n), h_1(n), ..., h_{L-1}(n)]^T \quad (1)$$

$$\mathbf{x}(n) = [x(n), x(n-1), ..., x(n-L+1)]^T \quad (2)$$

where L is the length of the truncated RIR,

The nonlinear mapping of the input signal yields

$$\mathbf{v}(n) = \left[f\left(x(n)\right), f\left(x(n-1)\right), ..., f\left(x(n-L+1)\right) \right]^T \quad (3)$$

The received signal at the microphone is expressed as

$$d(n) = \mathbf{h}^T(n)\mathbf{v}(n) + e_0(n) \quad (4)$$

At the canceller level

$$\mathbf{w}(n) = [w_0(n), w_1(n), ..., w_{L-1}(n)]^T \quad (5)$$

is the FIR weights vector used to estimate $\mathbf{h}(n)$, and the filter output is expressed as

$$\hat{d}(n) = \mathbf{w}^T(n)\mathbf{x}(n) \quad (6)$$

and the estimation error is given by

$$e(n) = y(n) - \hat{d}(n) \quad (7)$$

This error signal is used by the adaptive filtering algorithm to update the coefficients vector $\mathbf{w}(n)$. The LMS algorithm adapts the coefficients vector as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \quad (8)$$

where μ is the learning rate.

As mention in Section I, different models have been proposed to model the nonlinear distortion introduced by the amplifier and the loudspeaker. Here, we will consider the hard-clipping saturation model, which is a memoryless nonlinearity defined as follows [5][6]:

$$f\left(x(n)\right) = \begin{cases} -a, & x(n) \le -a \\ x(n), & -a < x(n) < a \\ a, & x(n) \ge a \end{cases} \quad (9)$$

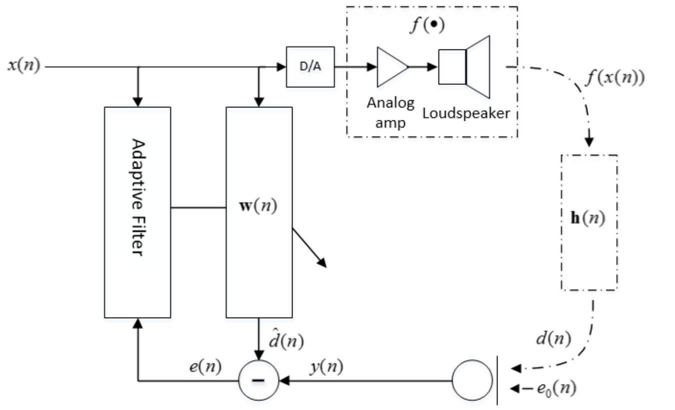

Fig. 1. Linear Acoustic echo cancellation system in the presence of nonlinear distortions.

where a is the clipping threshold. This memoryless function mimics the behavior of amplifiers, where a linear response is obtained when the instantaneous input signal level is within a given operating range, and an abrupt clipping is observed when the clipping threshold is exceeded. This behavior is also observed in miniature loudspeakers.

III. THE MAXIMUM CORRENTROPY CRITERION

The LMS algorithm and variants are widely used linear adaptive filtering algorithms. They are based on the minimization of the MSE cost function, which is not robust against outliers and impulsive noise in the error signal. A very interesting alternative similarity measure is the Correntropy [15], which measures the localized similarity between two arbitrary random scalar variables, it is defined by:

$$V_\sigma(X,Y) = E\left[k_\sigma(X-Y)\right] \quad (10)$$

where $k_\sigma(.)$ is a positive definite kernel, and σ being the width of the kernel.

By taking a normalized Gaussian kernel with standard deviation σ :

$$V_\sigma(X,Y) = \frac{1}{\sqrt{2\pi}\sigma} E\left[\exp\left(-\frac{(X-Y)^2}{2\sigma^2} \right) \right] \quad (11)$$

When compared to the MSE measure, the authors [16] obtained improved performance for the application of system identification in the presence of an additive impulsive noise, and adaptive noise cancellation with a corruptive non-stationary noise. The following development is reproduced from the same paper.

If we consider the cost function for the AEC problem of figure 1 as the Correntropy between the desired signal $y(n)$ and the FIR filter output $\hat{d}(n)$, with a normalized Gaussian kernel, and replacing the expectation operator by the sample estimator, we obtain

$$J(n)=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{N}\sum_{i=n-N+1}^{n}\left(\exp\left(-\frac{\left(y(i)-\hat{d}(i)\right)^{2}}{2\sigma^{2}}\right)\right) \quad (12)$$

which is equivalent to:

$$J(n)=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{N}\sum_{i=n-N+1}^{n}\left(\exp\left(-\frac{\left(y(i)-\mathbf{w}^{T}(n)\mathbf{x}(i)\right)^{2}}{2\sigma^{2}}\right)\right) \quad (13)$$

Taking the gradient $J(n)$ of with respect to $\mathbf{w}(n)$:

$$\nabla J(n)=\frac{1}{\sqrt{2\pi}\sigma^{3}}\frac{1}{N}\sum_{i=n-N+1}^{n}\left(\exp\left(-\frac{e^{2}(i)}{2\sigma^{2}}\right)e(i)\mathbf{x}(i)\right) \quad (14)$$

Maximizing the cost function iteratively is performed by taking a small step μ in the direction of the gradient:

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\frac{\mu}{\sqrt{2\pi}\sigma^{3}}\frac{1}{N}\sum_{i=n-N+1}^{n}\left(\exp\left(-\frac{e^{2}(i)}{2\sigma^{2}}\right)e(i)\mathbf{x}(i)\right) \quad (15)$$

By taking the approximation $N=1$ (inspired by the stochastic gradient), the final weights vector updating becomes:

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\frac{\mu}{\sqrt{2\pi}\sigma^{3}}\exp\left(-\frac{e^{2}(n)}{2\sigma^{2}}\right)e(n)\mathbf{x}(n) \quad (16)$$

From the last equation, we can see that obtained algorithm is basically an LMS with variable step size, which is controlled by the error and the kernel width σ . The exponential function penalizes the weighs vector updating process when outliers and impulsive behaviours of the desired signal are observed. We expect this interesting feature to be very useful in the context of acoustic echo cancellation with nonlinear distortions which are caused by high amplitude samples of the speech signal being distorted by the nonlinear components of the system.

Another interesting characteristic of the MCC-based adaptive filter over many other robust algorithms is its low computational complexity which is equivalent to the LMS linear complexity *O(L),* making it very attractive in practical implementations.

IV. Simulation Results

We consider the linear acoustic echo cancellation problem with presence of nonlinear distortions, depicted in Fig. 1. We will compare the performance of the normalized LMS (NLMS) and the normalized version of MCC-based linear adaptive filter presented in Section III (NMCC), by introducing the same normalization factor $\frac{1}{\mathbf{x}^{T}(n)\mathbf{x}(n)+\delta}$ for both algorithms, where δ is a small regularization parameter.

A length of $L=128$ is considered for both the RIR [17] and the linear adaptive FIRs. The input signal is a male speech sampled at 8 kHz sampling rate.

The performance of the algorithms is evaluated for the echo return loss enhancement (ERLE) and the normalized misalignment, defined respectively as

$$ERLE(n)=10\log_{10}\frac{\sum_{i=0}^{l-1}y^{2}(n-i)}{\sum_{i=0}^{l-1}e^{2}(n-i)} \ (dB) \quad (17)$$

$$\varepsilon(n)=10\log_{10}\frac{\left\|\hat{h}(n)-h\right\|_{2}^{2}}{\left\|h\right\|_{2}^{2}} \ (dB) \quad (18)$$

The results are smoothed for visual clarity, with $l=5000$.

We will assess the ERLE performance for different levels of nonlinear distortions, which is measured by the signal-to-distortion ratio:

$$\text{SDR} = 10\log_{10}(\frac{\sigma_{x_n}^{2}}{\sigma_{f(x_n)-x_n}^{2}}) \ (dB) \quad (19)$$

The tuning parameters of the adaptive filtering algorithms have been obtained empirically for the best average ERLE for $\text{SDR}=\infty$, then kept unchanged for the remaining simulations. A step size of $\mu_{NLMS}=0.2$, $\mu_{NMCC}=1.02\times10^{-3}$, $\sigma=0.08$, and a regularization parameter of $\delta=10^{-3}$ for both methods were used during all simulations. A fixed clipping threshold of $a=0.8$ and a variable power of the speech signal were considered to achieve the desired SDR. The background noise is modeled with an additive white Gaussian noise (AWGN) with a SNR=40 dB.

Fig. 2.a and Fig. 3.a shows the ERLE and misalignment performance for $\text{SDR}=\infty$, respectively. This is equivalent to a pure linear system with zero nonlinear distortion. This situation is observed when the hard-clipping saturation function is operating in the linear region. The NMCC and the NLMS archive almost similar performance with an average ERLE of 31.00 dB and 31.04 dB, respectively. An almost identical misalignment was obtained for both methods.

Fig. 2.b and Fig. 3.b presents the ERLE and misalignment performance for SDR=10 dB, respectively. The proposed NMCC plots a better ERLE curve with an average ERLE of 23.73 dB, compared to the NLMS where an average ERLE of 22.47 dB was obtained. An average gain of 1.26 dB was achieved for the ERLE and 4.15 dB in misalignment.

Fig. 2.c and Fig. 3.c covers the ERLE and misalignment results for SDR=7 dB, respectively. The NMCC outperforms the NLMS by an average gain of 1.55 dB. An average ERLE of 20.68 dB is achieved by the NMCC, while the NLMS

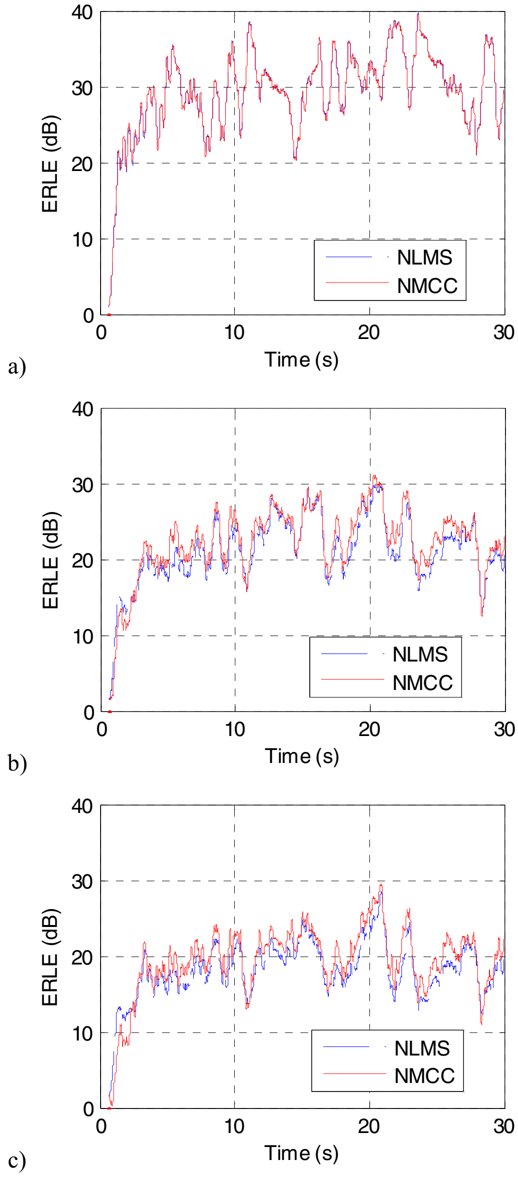

Fig. 2. ERLE performance for different SDRs: a) SDR=∞ b) SDR=10 dB, and c) SDR=7 dB.

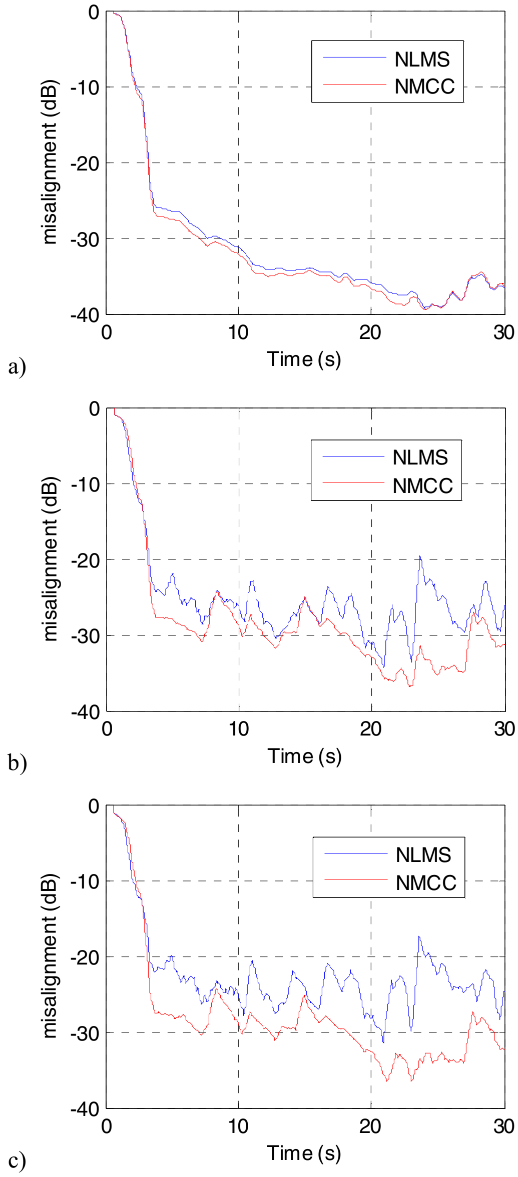

Fig. 3. Misalignment performance for different SDRs: a) SDR=∞ b) SDR=10 dB, and c) SDR=7 dB.

shows with an average ERLE of 19.13 dB. An average gain of 6.77 dB was achieved at the misalignment level.

For lower SDRs, the average performance gap increases slowly. However, the overall ERLE degrades considerably, which results in poor echo cancellation. In addition, the quality of the speech gets extremely affected by the hard-clipping saturation, which results in loud rough clipping sounds.

V. Conclusion

We have shown through simulations that the maximum Correntropy criterion is a very good alternative for the conventional mean squared error criteria in the context of acoustic echo cancellation with nonlinear distortions. Superior performance has been achieved for the hard-clipping memoryless nonlinearity with different nonlinearity levels. Another interesting characteristic of the studied method is the computational complexity order, which is similar to the baseline NLMS algorithm.

Acknowledgment

The authors wish to thank the Regroupement stratégique en microsystèmes du Québec (ReSMiQ) and the Natural Sciences and Engineering Research Council (NSERC) of Canada for financial support.

References

[1] J. Benesty et al., *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.

[2] S. Haykin, *Adaptive filter theory (3rd ed.)*: Prentice-Hall, Inc., 1996.

[3] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," *IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP)*, vol. 2, pp. 877–880, Mar. 1999.

[4] F. Kuech, A. Mitnacht, W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters, " *IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP),* March 2005.

[5] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1741–1760, Sept. 2000.

[6] S. Malik and G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 20, no. 7, pp. 2065–2079, Sep. 2012.

[7] J.-P. Costa, A. Lagrange, and A. Arliaud, "Acoustic echo cancellation using nonlinear cascade filters," *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP),* Hong Kong, vol. 5, pp. 389–392, 2003.

[8] F. Jing and Z. Wei-Ping, "A Nonlinear Acoustic Echo Canceller Using Sigmoid Transform in Conjunction With RLS Algorithm," *IEEE Trans. Circuits and Syst. II*, vol. 55, pp. 1056-1060, 2008.

[9] H. Merabti and D. Massicotte, "Robust nonlinear acoustic echo cancellation using a metaheuristic optimization approach," *IEEE Int. Conf. Digital Signal Process. (DSP)*, Singapore, pp. 297-301, 2015.

[10] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. ArenasGarcía, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1502–1512, Jul. 2013.

[11] M. M. Halimeh, C. Huemmer and W. Kellermann, "Nonlinear Acoustic Echo Cancellation Using Elitist Resampling Particle Filter," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, 2018, pp. 236-240.

[12] M. M. Halimeh, C. Huemmer and W. Kellermann, "A Neural Network-based Nonlinear Acoustic Echo Canceller," *IEEE Signal Processing Letters*, 2019.

[13] P. Shah, I. Lewis, S. Grant, and S. Angrignon, "Nonlinear acoustic echo cancellation using voltage and current feedback," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 23, no. 10, pp. 1589–1599, Oct. 2015.

[14] T. Gupta, S. Suppappola, and A. Spanias, "Nonlinear acoustic echo control using an accelerometer," *IEEE Int. Conf. Acoust., Speech, Signal Process*., pp. 1313–1316, May 2009.

[15] W. Liu, P. Pokharel, and J. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Trans. on Signal Process.,* vol. 55, no. 11, pp. 5286-5298, November 2007.

[16] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," *Int. Joint Conf. Neural Networks (IJCNN)*, pp. 2950–2955, 2009.

[17] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," *IEEE Int. Conf. on Digital Signal Process.*, pp. 1-5, 2009.