A DERIVED ROBUST STATISTICS APPROACH FOR ADAPTIVE VOLterra FILTERS APPLIED IN NONLINEAR ACOUSTIC ECHO CANCELLATION SCENARIOS

Cristian CONTAN, Marina Dana TOPA, Botond Sandor KIREI, Ingrid Maria KOVACS
Bases of Electronics Department, Technical University of Cluj-Napoca, Romania
{cristian.contan,marina.topa,botond.kirei}@bel.utcluj.ro, ikovacs@mail.utcluj.ro

Abstract: The paper proposes a novel updating concept for adaptive Volterra kernels that relies on a robust statistics approach. The optimization of a certain cost function leads to the update equations for the linear and quadratic kernels with respect to an optimum error threshold, set according to the relative local noise power. Thereby, for absolute error samples larger than the threshold, which occur mainly in the convergence stage of the adaptive filter, the adaptation is achieved using a sign version of the Normalized Least-Mean-Square algorithm. In the saturation stage of the filter, when most of the absolute values of the error samples are smaller than the particular threshold, the adaptation is carried out based on the Normalized Least-Mean-Square algorithm. This technique tends to eliminate the influence of large valued outliers by improving the convergence of the conventional Volterra filters and maintaining the same misadjustment in acoustic echo cancellation setups. The acoustic system is designed using measured Volterra kernels. The method is tested in terms of Echo Return Loss Enhancement for different probability density functions of the source signal. Also, for the white Gaussian noise case as excitation, a new convergence attribute is introduced for a more precise comparison of the adaptive methods.

Keywords: Acoustic echo cancellation, Volterra filters, adaptive algorithms, robust statistics.

I. INTRODUCTION

In many acoustic applications that involve hands-free devices such as mobile communications or teleconferencing setups, an important issue is the cancellation of the undesired echoes that are fed back to the far-end speaker. Acoustic echoes arise in the Loudspeaker-Enclosure-Microphone (LEM) configuration from the acoustic coupling between the loudspeaker and the microphone [1], [2]. The general scheme for acoustic echo cancellation (AEC) is presented in Fig.1.

The principle behind the AEC procedure is to generate a replica of the actual echo signal $y[k]$ by using an adequate adaptive filter. This is achieved by subtracting the output of the adaptive filter $\hat{y}[k]$ from the observed microphone signal $d[k]$ in order to minimize the error signal $e[k]$. If a proper echo canceller is used then eventually, the steady-state error should resemble the local noise from the acoustic system (\( \lim_{k \to \infty} e[k] \approx n[k] \)). In many AEC configurations, assuming a linear acoustic echo path, linear adaptive filters had an important role in establishing some theoretical standards [3], [4]. However, in some practical situations, adaptive nonlinear models represent a more accurate identification device that takes into consideration the nonlinear behavior of the available acoustic hardware. A simplified illustration of a nonlinear acoustic echo path can be observed in Fig. 2 where different sources of nonlinear distortions are underlined.

![Image of Acoustic echo cancellation setup.](image1)

![Image of Nonlinear echo path.](image2)
We are interested only in the type of the nonlinearities encountered in the echo path: with memory, generated by a small loudspeaker driven at high volume [5] and memoryless nonlinearities due to overdriven amplifiers [6]. Certain polynomial models were used to incorporate the type of nonlinearities from the LEM setup like Wiener and Volterra models as in [7] and cascaded adaptive filtering involving a memoryless nonlinearity in [8]. In this work we propose a novel AEC method based on adaptive Volterra filters starting from the minimization task of a different cost function than the mean-square error used for the conventional Normalized Least-Mean-Square (NLMS) algorithm.

The paper is organized as follows: In Section II we briefly present the Volterra model together with the deduction of the new update equations. Some simulation results are given in Section III, while Section IV concludes this work.

II. PROPOSED STRUCTURE

In the following we assume that the input/output dependency of the LEM setup can be adequately modeled by an N-th order Volterra filter described as in [9]:

\[
y[k] = \hat{y}_{NVF}[k] = \sum_{p=0}^{N} \sum_{n_1=0}^{M_1} \sum_{n_2=0}^{M_2} \hat{h}_{p,n_1,n_2} \prod_{i=1}^{p} x[k-m_i],
\]

where \( x[k] \) represents the source signal, \( \hat{y}_{NVF}[k] \) is the output of the Volterra structure while \( \hat{h}_{p,n_1,n_2} \) is the \( p \)-th order Volterra kernel of memory length \( M_p \). The Volterra kernels employed in (1) present a general symmetry which consists in using only terms with non-decreasing indexes \( (m_p \geq m_{p-1}) \). As the accuracy of the Volterra filter increases with the order of the filter, the computational cost for implementation also grows.

In this paper we consider the second-order Volterra filter as a sufficient tool to enclose the nonlinear distortions produced by the electro-dynamic loudspeakers and by the overdriven amplifier. The output of a second-order Volterra filter follows the expression:

\[
y_{2VF}[k] = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} \hat{h}_{2,m_1,m_2} \prod_{i=1}^{2} x[k-m_i].
\]

For simplicity, equation (2) can be rewritten in vector notation as:

\[
y_{2VF}[k] = \hat{h}^T[k] x[k] + \hat{h}_2^T[k] x_2[k],
\]

using the associated vector definitions:

\[
x[k] = (x[k], x[k-1], \ldots, x[k-M_1+1])^T;
\]

\[
\hat{h}[k] = (\hat{h}_{0,0}[k], \hat{h}_{0,1}[k], \ldots, \hat{h}_{M_1-1}[k])^T;
\]

\[
x_2[k] = (x^2[k], x[k-1], \ldots, x[k-M_2+1], \ldots, x[k-M_2+1])^T;
\]

\[
\hat{h}_2[k] = (\hat{h}_{2,0,0}[k], \hat{h}_{2,0,1}[k], \ldots, \hat{h}_{2,0,M_2-1}[k], \hat{h}_{2,1,0}[k], \hat{h}_{2,1,1}[k], \ldots, \hat{h}_{2,2,M_2-1}[k])^T,
\]

where \((\cdot)^T\) denotes the transposing operation.

To obtain an optimum design for the LEM setup, besides the polynomial model, an adaptive algorithm for the kernels’ update is also needed to minimize the residual error signal \( e[k] \) defined as:

\[
e[k] = d[k] - \hat{y}_{2VF}[k].
\]

For stability and robustness, the NLMS adaptive algorithm is widely used in echo acoustic cancellation scenarios [10]. The update equations of the two Volterra kernels using the NLMS algorithm are presented in (6):

\[
\begin{align*}
\hat{h}_1[k+1]^\text{NLMS} & = \hat{h}_1[k]^\text{NLMS} + \mu_1 x_1^T[k] e[k] x_1[k] + \phi, \\
\hat{h}_2[k+1]^\text{NLMS} & = \hat{h}_2[k]^\text{NLMS} + \mu_2 x_2^T[k] e[k] x_2[k] + \phi,
\end{align*}
\]

where \( \phi \) is a positive constant introduced to prevent division by zero or a very small value. The step-size parameters \( \mu_1 \) and \( \mu_2 \) control the convergence rate and steady-state error of the adaptive filter. To ensure a stable convergence in the mean of the adaptive filter, the step-size parameters should be selected in the range \((0, 2)\).

The functionality of the second-order Volterra filter that uses the NLMS adaptation algorithm to update the kernels consists in the minimization of a cost function defined as:

\[
J_{NLMS}[k] = \frac{1}{2} \left( \frac{e[k]}{\|x[k]\|} \right)^2, \quad i = \{1, 2\},
\]

where \( \| \cdot \| \) is the Euclidean norm.

Describing a parabolic characteristic dependent on the error samples, the cost function exhibits high values for large error samples. Hence, for these ranges of large error values, a new cost function that depends only on the absolute value of the error is proposed:

\[
J_{SIGN}[k] = \frac{1}{\|x[k]\|} \|e[k]\| = \|e[k]\| = \frac{1}{\|x[k]\|^2} \text{sign}(e[k]) e[k].
\]

Starting from the steepest-descent recursion described in [10] as:

\[
\hat{h}_i[k+1] = \hat{h}_i[k] - \mu \nabla_{\hat{h}_i} J[k]
\]

and also taking into account the update equations of the linear and the second-order Volterra kernels, a different set of update equations for the kernels can be deduced when the adaptive filter has the role of minimizing the proposed cost function. Next, we will compute the gradient operators of the cost function for the second-order Volterra scenario. The two gradient vectors are defined as:
By employing equations (13) and (15) in the steepest-descent recursion for updating the kernels (9), the following set of equations is derived:

\[
\begin{align*}
\hat{h}_1[k + 1]^{\text{SIGN}} &= \hat{h}_1[k]^{\text{SIGN}} + \frac{\mu_1}{x_1[k] x_1[k]} \text{sign}(e[k]) x_1[k], \\
\hat{h}_2[k + 1]^{\text{SIGN}} &= \hat{h}_2[k]^{\text{SIGN}} + \frac{\mu_2}{x_2[k] x_2[k]} \text{sign}(e[k]) x_2[k].
\end{align*}
\]

The SIGN update procedure of the Volterra kernels (16) reduces the influence of outliers with large amplitude. In this work, a nonlinear acoustic echo cancellation procedure is proposed based on a derived robust statistics (RS) approach like the one presented in [11]. A system identification routine of the LEM features is outlined using a NLMS second-order Volterra filter for reduced error values, while the SIGN version of the Volterra filter is applied for large error samples. For implementation, an error threshold \( e_{\max} \) is set. According to this threshold, the sequent cost functions associated with the methods are minimized:

\[
J_{RS}[k] = \frac{1}{2} \left( \frac{|e[k]|}{|x_1[k]|^2} \right)^2, \quad 0 \leq |e[k]| < e_{\max}
\]

\[
J_{NLMS}[k] = \frac{1}{2} \left( \frac{|e[k]|}{|x_1[k]|^2} \right)^2, \quad |e[k]| \geq e_{\max}.
\]

Following (17), the output of the new second-order adaptive Volterra filter is computed as:

\[
\hat{y}_L[k] = \hat{y}_{NLMS}[k] + \hat{y}_{NLMS}[k] = h_1^T[k] y_{NLMS} + h_2^T[k] y_{NLMS} x_2[k], \quad 0 \leq |e[k]| < e_{\max}
\]

\[
\hat{y}_L[k] = \hat{y}_{NLMS}[k] + \hat{y}_{NLMS}[k] = h_1^T[k] y_{NLMS} x_2[k] + h_2^T[k] y_{NLMS} x_2[k], \quad |e[k]| \geq e_{\max}.
\]

The block-diagram of the setup is depicted in Fig. 3:
III. EXPERIMENTAL RESULTS

To highlight the convergence rate and the steady-state error of the proposed adaptive filter, the following representation of the microphone signal is used:

\[ d[k] = h_1^l x[k] + \alpha h_2^l x_2[k] + \beta n[k], \]  

(19)

where \( h_1 \) and \( h_2 \) are the linear and the quadratic kernels written in vector notation as in (4). These kernels are depicted in Fig. 4 and were obtained from measurements conducted in a low reverberant room with low-cost acoustic components, at a sampling rate of 8 kHz.

The memory lengths of the kernels are equal to 320 and 50 taps, to include all the coefficients with significant nonzero values.

The Linear-to-Nonlinear Ratio (LNLR) and the Signal-to-Noise Ratio (SNR) quantities are set to 10 dB and 30 dB throughout simulations, by choosing proper values for parameters \( \alpha \) and \( \beta \). The efficiency of the RS method is compared to the one of the conventional nonlinear echo cancellation setup that employs the NLMS second-order Volterra filter. The achieved echo reduction of the methods is evaluated in terms of Echo Return Loss Enhancement (ERLE):

\[ \text{ERLE}[k] = 10 \log_{10} \frac{\text{E}[d^2[k]]}{\text{E}[e^2[k]]} \text{ [dB]}, \]  

(20)

where \( \text{E}[] \) denotes statistical expectation.

The lengths of the Volterra kernels are identical to the ones that describe the nonlinear acoustic enclosure. In simulations, white Gaussian noise and nonstationary audio signals are used as source signals. Additive white Gaussian noise is used as local noise in each of the conducted experiments. For updating the Volterra kernels, the following step-size parameters are chosen: \( \mu_1 = 0.01 \) and \( \mu_2 = 0.005 \).

The disadvantage of the proposed RS adaptation technique consists in the selection of the error threshold \( e_{\text{max}} \), which is set by the user. The statistical decision criterion of \( e_{\text{max}} \) is based on the evolution of the error signal discarded by the second-order Volterra filter that uses only the NLMS algorithm to adapt the kernels. By keeping trace of the absolute error’s characteristic, the threshold value is set close to the settlement amount of the filter’s error signal. In this way, a large number of samples above \( e_{\text{max}} \) can be included within the SIGN update process. Usually, these high-valued samples are found in the convergence range of the filter.

To illustrate the effectiveness of this method, a set of simulations is performed for a white Gaussian noise as input signal. In the first case, the optimum value of \( e_{\text{max}} \) is set to 0.05. The evolutions of ERLE for the linear NLMS filter, the second-order Volterra filter adapted only with the NLMS algorithm and the one of the proposed RS Volterra method are depicted in Fig. 5. Concurrently, the microphone signal \( d[k] \) is plotted below the evolutions of ERLE for each input scenario. It can be observed that for the chosen threshold, the new method achieves better convergence than the second-order NLMS Volterra filter. Also, the steady-state ERLE in this case is similar for both approaches; the proposed method does not affect the stabilization value of the error. Each nonlinear filter reaches a steady-state ERLE of 30 dB, similar to the value proposed for the SNR. The linear NLMS filter settles to an ERLE of 10 dB which equals the LNLR value.

In the following paragraph we are going to take advantage of the smooth ERLE curves to introduce a so-called convergence attribute that expresses more precisely the efficiency of the proposed method in comparison to the conventional second-order NLMS Volterra filter. This quantity will describe the gained improvement regarding the filter’s convergence rate. The convergence speed of an adaptive filter can be perceived as the settlement value of the filter in terms of ERLE related to the number of samples needed for the filter to reach this limit. For the proposed

![Figure 4. Linear kernel (top); quadratic kernel (bottom).](image)

![Figure 5. ERLE characteristics for white Gaussian noise, with \( e_{\text{max}} = 0.05 \) (top); microphone signal (bottom).](image)
nonlinear acoustic scenario where the SNR value is controlled and even more, it is kept constant throughout simulations, one can observe from Fig. 5 that the ERLE characteristics of the two nonlinear approaches settle at the imposed SNR value. As the two nonlinear filters achieve the same steady-state ERLE value, we are interested only in the number of samples required for each adaptive nonlinear filter to reach this amount. Thus, in this case $s_{NLMS}$ labels the number of samples that are required for the second-order NLMS Volterra filter to stabilize, while $s_{RS}$ stands for the number of samples needed for the proposed method to reach the steady-state ERLE value. Taking into consideration the mentioned notations, we define the relative convergence attribute (RCA) as:

$$\text{RCA} \ [%] = \left(1 - \frac{s_{RS}}{s_{NLMS}}\right) \cdot 100.$$  \hspace{1cm} (21)

This amount expresses the percentage of the convergence improvement obtained by employing the proposed method as compared to the conventional adaptation method. Thereby, for the case presented in Fig. 5, we are dealing with an RCA of 58.75%.

In the next two situations described in Fig. 6 and 7 the decision threshold between the two distinct adaptation techniques takes smaller values than in the optimum case. Therefore, by setting $e_{\text{max}} = 0.025$ (Fig.6) and $e_{\text{max}} = 0.005$ (Fig. 7), the error ranges expand, for which the SIGN adaptation procedure is applied. This will unreasonably update the Volterra kernels with the SIGN algorithm for error samples smaller than the optimum chosen error limit from Fig. 5. Analyzing the performances of the proposed adaptation method for the two selected error thresholds in comparison to the NLMS Volterra model and taking into consideration the results depicted in Fig. 5, the RS version offers a better convergence than in the case when the threshold $e_{\text{max}} = 0.05$ (optimum) was selected. But none of these occurrences provide the expected steady-state ERLE; the adaptive filter stabilizes below the established SNR value.

However, if a larger value is chosen for $e_{\text{max}}$ ($e_{\text{max}} = 0.1$) than in the first case (Fig.5), the evolutions of ERLE from Fig. 8 are achieved for the proposed adaptive Volterra methods. In this situation the steady-state ERLE for the RS scenario returns to the imposed SNR amount and the obtained convergence rate is also higher than in the NLMS second-order Volterra case. However, the convergence speed is slightly reduced from the one experienced in the initial case ($e_{\text{max}} = 0.05$). This threshold is also valid but it misses out some relatively high error values for the applicability of the SIGN adaptation technique. This offers the observed convergence rate difference in contrast to the optimum threshold scenario.

Next, we will evaluate the evolution of the proposed method in comparison with the second-order NLMS Volterra filter and also with the linear NLMS adaptive filter for an assembly of nonstationary audio input signals. Fig. 9 and 10 consider the echo cancellation configurations for a song fragment as excitation. Speech is applied as input signal for the results plotted in Fig. 11 and 12 while the last two charts from Fig. 13 and 14 characterize the resulting ERLE for instrumental music as input. The roughly selected value for the optimum error threshold $e_{\text{max}}$ equal to 0.01 remains the same in all three input signal scenarios. From Fig. 9, 11 and 13 one can notice the ERLE characteristics for this error threshold, emphasizing the performances of the Volterra-based echo cancellation approaches. Below, the microphone signal is plotted in each case. The convergence rate depicted by the adaptation technique with the RS criterion behaves better than the second-order Volterra filter that uses only the NLMS algorithm. The two ERLE patterns settle at a steady-state ERLE value of approximately 30 dB. This means that the RS procedure does not affect the stabilization value of the error. With the linear NLMS adaptive filter approximately 10 dB of discarded acoustic

![Figure 6. ERLE characteristics for Gaussian noise, with $e_{\text{max}} = 0.025$ (top); microphone signal (bottom).](image)

![Figure 7. ERLE characteristics for Gaussian noise, with $e_{\text{max}} = 0.005$ (top); microphone signal (bottom).](image)

![Figure 8. ERLE characteristics for white Gaussian noise, with $e_{\text{max}} = 0.1$ (top); microphone signal (bottom).](image)
As in the white Gaussian noise as input scenario, the convergence speed of the proposed method decreases from the previous cases when we increase the value of $e_{\text{max}}$ to 0.025. The mentioned effect can be detected in Fig. 10, 12 and 14. Even though the convergence of the RS Volterra filter slows down, it is still better than in the second-order NLMS Volterra circumstance, while the steady-state ERLE remains almost the same in all cases. Thereby, this threshold value can be also taken into consideration for speeding up the convergence of the second-order NLMS Volterra filter.

**IV. CONCLUSIONS**

Starting from the minimization procedure of a cost function dependent on the absolute values of the error, a novel set of update equations for the linear and the quadratic Volterra kernels was developed in this work. This kernel adaptation technique was used when the error samples are larger than an error threshold chosen around the local noise level. For error values below the mentioned threshold, the Volterra kernels were updated with the NLMS algorithm. Hence, by applying this interchange between the two adaptation techniques due to the mentioned threshold, a more rapid
adaptation in the convergence section of the filter was intended, without interfering too much with the steady-state error. The comparison between the two types of update structures was achieved for a nonlinear acoustic system, characterized by linear and second-order distortions with different noise conditions. The evolution of the residual error power was monitored in each case. White Gaussian noise and nonstationary audio signals were used as excitations.

In the conducted simulations, the advantages and disadvantages of the proposed method were presented with respect to the chosen values of the error threshold related to the steady-state error. In most of the cases, the proposed method offered a better convergence rate than the one of the second-order NLMS Volterra filter, reaching the suggested noise conditions after stabilization.

The drawback of this adaptation procedure is the necessity of a robust selection of the threshold. Future work will involve the optimization of the threshold selection.

ACKNOWLEDGMENT

This paper was supported by the project “Doctoral studies in engineering sciences for developing the knowledge based society-SIDOC” contract no. POSDRU/88/1.5/S/60078, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

REFERENCES