# NOVEL SCHEMES FOR NONLINEAR ACOUSTIC ECHO CANCELLATION BASED ON FILTER COMBINATIONS

Luis A. Azpicueta-Ruiz<sup>1</sup>, Marcus Zeller<sup>2</sup>, Jerónimo Arenas-García<sup>1</sup>, and Walter Kellermann<sup>2</sup>

 <sup>1</sup> Department of Signal Theory and Communications\*
 Universidad Carlos III de Madrid 28911 Leganés-Madrid, Spain {lazpicueta, jarenas}@tsc.uc3m.es

# ABSTRACT

Nonlinear acoustic echo cancellers (NLAEC) are becoming increasingly important in hands-free applications. However, in some situations, an NLAEC is inferior to a linear AEC, especially when the channel generates a negligible (or no) nonlinear echo. In general, the ratio of the linear to nonlinear echo signal power is unknown *a priori*, and will vary over time, thus making it difficult to know if an NLAEC would improve or degrade the cancellation. In this paper, we present two novel solutions to this problem based on the adaptive combination of linear and nonlinear echo cancellers. Both solutions perform efficiently regardless of the level of nonlinear echo. The benefits and robustness of both schemes are illustrated by experiments using Laplacian colored noise and speech input signals.

*Index Terms*— Adaptive filters, Volterra filters, nonlinear acoustic echo cancellation, combination of filters.

#### 1. INTRODUCTION

In recent years, nonlinear acoustic echo cancellation (NLAEC) schemes have become increasingly important, not at least due to the popularity of hands-free devices and mobile phones that use low-cost amplifiers and loudspeakers introducing significant nonlinearities into the acoustic echo path. Adaptive Volterra filters are widely used for NLAEC because of their generic structure, which can be considered as a straightforward generalization of linear adaptive filters [1].

Although Volterra filters decrease the residual nonlinear echo, they may not always be superior to a plain linear filter: For instance, if the echo cancellation scenario presents a low level of nonlinear echo, non-negligible gradient noise produced by the adaptation of second and higher order kernels degrades the performance of the NLAEC, so that the use of a simple linear adaptive filter would be more efficient. Note that, in general, the power of nonlinear echo is unknown, and will be time-varying for nonstationary signals like speech. Thus, <sup>2</sup> Multimedia Communications and Signal Processing\* University of Erlangen-Nuremberg Cauerstr. 7, 91058 Erlangen, Germany {zeller,wk}@LNT.de

the selection of the most effective adaptive filter, linear or Volterra, is not a trivial problem since it requires  $a \ priori$  knowledge about the echo channel and the signal statistics.

Combinations of filters constitute an interesting way to mitigate different kinds of compromises involving adaptive filters [2, 3]. In this approach two or more adaptive filters adaptively combine their outputs obtaining a combined scheme that performs at least as well as the best contributing filter. Due to their simplicity, these schemes have been used in several areas of adaptive signal processing in communications and control applications, including blind equalization [4] and signal characterization [5], among others.

In this paper, we present two novel schemes for modelling nonlinear systems employing the principle of combining several filters. The first one consists of a combination of a linear filter and a Volterra filter, while the second one is a more elaborate scheme based on a combination of kernels –a generic concept for Volterra filters. Both schemes fulfill their promises independently of the linear-to-nonlinear power ratio (LNLR) of the echo signal, obtaining the desired effectiveness of the Volterra filter when necessary while performing like the linear filter for low levels of nonlinear echo.

The rest of the paper is organized as follows: In Section 2 both schemes for improved nonlinear echo cancellation are presented. The experiments that corroborate the benefits and robustness of both solutions are included in Section 3. Finally, conclusions are presented in Section 4.

# 2. PROPOSED SCHEMES

In this section, we present two novel schemes for NLAEC which are robust to different levels of the LNLR. Both systems include second-order Volterra filters, with a linear and a quadratic kernels, and could straightforwardly be extended to generic *N*th-order Volterra filters.

## 2.1. 'Combination of filters'-Scheme (CFS)

The first scheme consists of a straightforward convex combination of an adaptive linear filter,  $\mathbf{w}(n)$ , and an adaptive Volterra filter including linear and quadratic kernels,  $\mathbf{h}(n)$ and  $\mathbf{H}(n)$ , respectively, where the triangular representation has been used for the latter [1]. The outputs of the linear and the Volterra filters, can be expressed as

$$y_L(n) = \mathbf{w}^T(n)\mathbf{u}(n) \tag{1}$$

<sup>\*</sup>This work was partly supported by the Spanish Ministry of Education and Science under grant CICYT TEC-2005-00992, by Madrid Community grant S-505/TIC/0223, and by the Deutsche Forschungsgemeinschaft (DFG) under contract number KE 890/5. This work has been performed while the first author was a visiting researcher at the Chair of Multimedia Communications and Signal Processing.

$$y_V(n) = y_{LK}(n) + y_{QK}(n)$$
  
=  $\mathbf{h}^T(n)\mathbf{u}(n) + \mathbf{u}^T(n)\mathbf{H}(n)\mathbf{u}(n)$  (2)

where  $\mathbf{u}(n)$  denotes the vector of the input signal samples, and  $y_{LK}(n)$  and  $y_{QK}(n)$  represent the outputs of the linear and quadratic kernels, respectively. The output of the combined filter reads:

$$y(n) = \lambda(n)y_L(n) + [1 - \lambda(n)]y_V(n), \qquad (3)$$

where  $\lambda(n)$  is an adaptive weighting parameter that controls the combination.

According to [3], for a good performance of the combination scheme, the contributing filters should update their coefficients following their own rules, in order to minimize the power of their own error signals. When using standard gradient descent rules, this results in

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_L e_L(n)\mathbf{u}(n),$$
  

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu_{VL} e_V(n)\mathbf{u}(n),$$
  

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \mu_{VQ} e_V(n)\mathbf{u}(n)\mathbf{u}^T(n),$$
(4)

where  $\mu_L$ ,  $\mu_{VL}$  and  $\mu_{VQ}$  are step sizes, and  $e_L(n) = d(n) - y_L(n)$  and  $e_V(n) = d(n) - y_V(n)$  are the errors produced by the linear and the Volterra filters, respectively, and d(n) is the reference signal to be approximated by the adaptive filters.

The mixing parameter  $\lambda(n)$  can also be updated using a gradient descent method with the aim of minimizing the square of the error produced by the combined filter, e(n) = d(n) - y(n). However, instead of directly adapting  $\lambda(n)$ , we will rely on the adaptation of another parameter a(n), which defines  $\lambda(n)$  via a sigmoïdal activation function<sup>1</sup>,  $\lambda(n) =$  $\operatorname{sgm}[a(n)] = [1 + e^{-a(n)}]^{-1}$ .

Recently, a new update rule for a(n) has been presented in [6]. By normalizing the adaptation of a(n), this rule allows an easier selection of the step size  $\mu_a$ , and provides improved performance in scenarios with time-varying signal-to-noise ratio (SNR). This normalized rule reads can be expressed as:

$$a(n+1) = a(n) + \frac{\mu_a}{p(n)} \lambda(n) [1 - \lambda(n)] e(n) [e_V(n) - e_L(n)],$$
(5)

where  $p(n) = \beta p(n-1) + (1-\beta)[e_V(n) - e_L(n)]^2$  is an estimate of the power of  $[e_V(n) - e_L(n)]$  (see [6] for more details).

The functionality of the presented scheme can be described as follows. When the LNLR is low (i.e., there is a significant level of nonlinear echo), the Volterra filter represents an effective model of the channel, and minimization of the overall error yields  $\lambda(n) \to 0$ , so that the  $y(n) \approx y_V(n)$ . The opposite occurs for high LNLR, with  $\lambda(n) \to 1$  and  $y(n) \approx y_L(n)$ , so that the combination is equivalent to a linear filter, avoiding the gradient noise caused by the adaptation of the Volterra quadratic kernel. Figure 1(a) summarizes our first proposal for NLAEC, to which we will refer in the following as combination of filters scheme (CFS).

# 2.2. 'Combination of kernels'-Scheme (CKS)

Rather than combining adaptive filters, our second approach for NLAEC foresees just one Volterra filter, replacing one of



(a) Combination of filters scheme (CFS)



(b) Combination of kernels scheme (CKS)

Fig. 1. Block diagrams for the proposed NLAEC schemes. Adaptation loops and error signals are omitted for simplicity.

its kernel by a convex combination of kernels. For instance, if we consider a second-order Volterra filter, and replace its quadratic kernel, the overall output of the new Volterra filter would be given by

$$y(n) = y_{LK}(n) + y_{QK}$$
  
=  $y_{LK}(n) + \eta(n)y_{Q1}(n) + [1 - \eta(n)]y_{Q2}(n)$  (6)

where  $y_{Q1}(n)$  and  $y_{Q2}(n)$  are the outputs of two kernels in the combination and  $\eta(n) \in [0, 1]$  is a mixing parameter.

To get the most out of all kernels, each should be updated using its own adaptation rules and error signal: The linear kernel should pursue the minimization of the overall error e(n) = d(n) - y(n), while the kernels inside the combination should adapt independently of each other, to minimize the square of  $e_i(n) = d(n) - [y_{LK}(n) + y_{Qi}(n)]$ , i = 1, 2.

Finally,  $\eta(n)$  can again be adapted using a gradient descent rule. Defining  $\eta(n) = \operatorname{sgm}[a'(n)]$ , and taking derivatives of  $e^2(n)$  with respect to a'(n), leads to

$$a'(n+1) = a'(n) + \mu_{a'}[y_{Q1}(n) - y_{Q2}(n)]e(n)\eta(n)[1 - \eta(n)]$$
(7)

The above expression can be interpreted as a least-meansquares (LMS) update rule, where  $[y_{Q1}(n) - y_{Q2}(n)]$  plays the role of the input signal. Using similar arguments to those in [6], a more convenient normalized adaptation rule would be

$$a'(n+1) = a'(n) + \frac{\mu_{a'}}{p'(n)}\eta(n)[1-\eta(n)]e(n)[y_{Q1}(n) - y_{Q2}(n)]$$
(8)

with  $p'(n) = \beta' p'(n-1) + (1-\beta')[y_{Q1}(n) - y_{Q2}(n)]^2$ .

The generic Eq. (6) allows that the kernels implementing  $y_{Q1}(n)$  and  $y_{Q2}(n)$  can differ in any way (e.g., they could

<sup>&</sup>lt;sup>1</sup>Introduction of parameter a(n) and the activation function is justified as an easy way to keep  $\lambda(n) \in (0,1)$  and to reduce gradient noise near  $\lambda(n) = 1$  or  $\lambda(n) = 0$ . The interested reader is referred to [3] for further details.

implement different kernel sizes or use different adaptation rules, etc). Since in this paper we are interested in schemes that work well for unknown, time-varying, LNLR, we will consider that  $y_{Q1}(n)$  implements a filter with very slow adaptation (minimizing the corresponding gradient noise). An extreme, but very interesting special case, results when all taps of  $\mathbf{H}_1(n)$  are set to zero  $\forall n$  (i.e., there is no need to adapt the coefficients of this *virtual* kernel): When using an all-zeros kernel, the overall output is given by

$$y(n) = y_{LK}(n) + [1 - \eta(n)]y_{Q2}(n)$$
(9)

so that the role of  $\eta(n)$  can be interpreted as that of deciding whether using a quadratic kernel would improve or degrade the overall cancellation performance.

Rewriting (9) as

$$y(n) = \eta(n)y_{LK}(n) + [1 - \eta(n)][y_{LK}(n) + y_{Q2}(n)]$$
(10)

shows the similarity to (3) with the notable difference that the common linear part is now used by both the linear and the Volterra filter. In computational terms this means that the number of operations needed for implementing the novel scheme of Fig. 1(b), and which we refer to as combination of kernels scheme (CKS), is only slightly larger than that for a standard Volterra filter, while CFS requires adaptation of two linear filters.

#### 3. EXPERIMENTS

In this section, we study the performance of CFS and CKS in echo cancellation scenarios with different LNLRs. Two kinds of input signals will be used: Laplacian colored noise, and real speech. The reference signal follows this model:

$$d(n) = \mathbf{h}_0^T \mathbf{u}(n) + \alpha(n) \mathbf{u}^T(n) \mathbf{H}_0 \mathbf{u}(n) + e_0(n)$$
(11)

where  $\mathbf{h}_0$  and  $\mathbf{H}_0$  are the true linear and quadratic kernels of size 320 and 64 × 64, respectively, both measured from a small low-cost loudspeaker,  $\alpha(n)$  is a variable introduced to control the LNLR, and  $e_0(n)$  is a Gaussian white noise providing 20 dB SNR in the absence of nonlinear echo.

Settings for the NLAEC schemes are as follows: the linear filter of CFS is adapted using a normalized LMS (NLMS) rule with step size  $\mu_L = 0.3$ . The kernels of the Volterra filters of CFS and CKS use the same step size, but adapting with NLMS rules where the input power is estimated separately for each kernel (SNLMS, [7]). For CKS, an all-zeros kernel is used instead of  $\mathbf{H}_1(n)$  (i.e.,  $y_{Q1}(n) = 0$ ,  $\forall n$ ). Finally, the mixing parameters are adapted using (5) and (8), respectively for CFS and CKS, with  $\mu_a = \mu_{a'} = 0.5$  and  $\beta = \beta' = 0.9$ .

#### 3.1. Laplacian colored noise as input

Using Laplacian colored stationary noise as input signal, we will first analyze the convergence and stationary behavior of the proposed NLAECs for different LNLRs. As a figure of merit, we will estimate the resulting excess mean-squareerror,  $\text{EMSE}(n) = \text{E}\{[e(n) - e_0(n)]^2\}$ , averaging over 1000 independent realizations.

The behavior of CFS, as well as of its linear and Volterra components, is illustrated in Fig. 2(a). As expected, when only linear echo is present (LNLR =  $\infty$  dB, t < 10 s) the Volterra canceller achieves a larger steady-state EMSE than



Fig. 2. Cancellation performance of CFS and CKS. (a) EMSE evolution of CFS and CKS, as well as of the linear and Volterra components of CFS (denoted as LF and VF, respectively). (b) Time evolution of the mixing parameters  $\lambda(n)$  and  $\eta(n)$ .

the linear filter alone, and CFS retains the better performance of the linear scheme with  $\lambda(n) \approx 1$  (see also  $\lambda(n)$  evolution in Fig. 2(b)). The opposite occurs for small LNLR (see t > 20 s, with LNLR = -10 dB). In this case, the quadratic kernel of the Volterra filter is key to obtain a correct cancellation, thus  $\lambda(n) \approx 0$ , so that CFS achieves the smaller EMSE of the Volterra scheme. Note that for intermediate levels of nonlinear echo (LNLR= 8 dB during 10 s < t < 20 s) CFS can simultaneously outperform both contributing filters. This property of convex combinations of filters has been theoretically explained in [3], by a small correlation between the errors of the component filters, and by the fact that CFS averages their outputs (note that  $\lambda(n) \approx 0.5$  in this situation), thus decreasing the error variance.

Fig. 2 also shows the cancellation performance of CKS, which can be described in very similar terms to those for CFS, with  $\eta(n) \approx 0$  when there is a significant level of nonlinear echo (see (9)). For large LNLR,  $\eta(n) \approx 1$  and, in the light of (9), CKS behaves as a linear filter, thus getting rid of the gradient noise of the quadratic kernel that would degrade the cancellation. Note that in this situation the quadratic kernel can still be adapted without divergence problems, since  $e_2(n) = d(n) - [y_{KL}(n) + y_{Q2}(n)] \approx -y_{Q2}(n)$ , so that the update algorithm tries to minimize its own output.

The stationary behavior of CFS and CKS has also been studied for other LNLRs. Fig. 3 shows the steady-state EMSE of these schemes as a function of the LNLR. These results have been obtained by averaging the EMSE over 25000 iterations once the algorithms had converged, and over 200 independent realizations. It can be seen that, for all values of the LNLR, CFS performs at least as well as its components, with a significant margin  $-5 \text{ dB} \leq \text{LNLR} \leq 15 \text{ dB}$  where the combination outperforms both components. CKS offers a very similar behavior for all values of LNLR. As discussed at the end of Subsection 2.2, CKS is computationally simpler than CFS, and can therefore be considered as a more attractive scheme for NLAEC.



**Fig. 3**. Steady-state EMSE of CFS and CKS, and of the linear (LF) and Volterra (VF) components of CFS.



Fig. 4. Behavior of the CFS NLAEC. From up to down: speech input signal; ERLE achieved by CFS and its component filters; CFS mixing parameter evolution,  $\lambda(n)$ .

#### 3.2. Speech as input

In this subsection we show the performance of both schemes with real speech. This experiment also represents the first instant where the convergence properties of the CFS and CKS are studied for nonstationary signals. In this case the Echo Return Loss Enhancement (ERLE) is used as figure of merit:

$$\operatorname{ERLE}(n) := 10\log \frac{\operatorname{E}\{[d(n) - e_0(n)]^2\}}{\operatorname{E}\{[e(n) - e_0(n)]^2\}} \quad [\operatorname{dB}] \qquad (12)$$

Fig. 4 shows the CFS cancellation performance, compared to that of the linear and the Volterra filters, in a scenario where the LNLR changes from  $\infty$  to 2.5 dB at t = 5 s, and from 2.5 dB to -10 dB at t = 10 s. Although the performance is more irregular due to the nonstationary nature of speech, results are similar to those for Laplacian input: The combination behaves as the best component for very large or very small LNLR. For intermediate nonlinear echo levels (5 s < t < 10 s) the combination performs slightly better than both component filters. Again, CKS achieves a very similar performance, as illustrated in Fig. 5.

Table 1 shows the average ERLE calculated over each period of constant LNLR. The proposed echo cancellers achieve similar ERLE values in all cases (differences observed are not very significant) behaving at least as the best of the linear and Volterra filters, or even better than any of them (for LNLR = 2.5 dB). From Table 1, it can also be seen that the averaged values of  $\lambda(n)$  and  $\eta(n)$  increase (as expected) with the LNLR, so that this values could also be exploited as indicators of the level of nonlinear echo.



Fig. 5. ERLE comparison for CFS and CKS.

LNLR	$\bar{\lambda}(n)$	$\bar{\eta}(n)$	ERLE [dB]			
[dB]			LF	VF	CFS	CKS
$\infty$	0.85	0.78	11.9	8.6	11.9	11.9
2.5	0.56	0.45	5.2	5.1	8.7	9.1
-10	0.07	0.12	1.3	13.3	13.3	13.7

 Table 1. Cancellation performance of CFS and CKS with speech input signal (averaged results).

### 4. CONCLUSIONS

In this paper, we presented two novel NLAECs based on combination schemes. The first scheme (CFS) consists of a combination of a linear and a Volterra filter, while the second (CKS) is based on the combination of a quadratic and an allzeros kernel. Both schemes offer improved performance over the use of a single (linear or nonlinear) filter when the LNLR is unknown or time-varying. Additionally, CKS is computationally more efficient than CFS, and only slightly more complex than a standard Volterra filter, thus offering a very attractive solution to nonlinear echo cancellation.

Future work includes further developments following the combination of kernel approach, as well as the extension of the proposed schemes to higher order nonlinear echo cancellers.

## 5. REFERENCES

- V. J. Mathews and G.L. Sicuranza, *Polynomial Signal Processing*, New York: John Wiley and Sons, 2000.
- [2] M. Martínez-Ramón, et al., "An adaptive combination of adaptive filters for plant identification," in Proc. 14th Intl. Conf. Digital Signal Process., 2002.
- [3] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, 2006.
- [4] M. T. M. Silva and V. H. Nascimento, "Improving the Tracking Capability of Adaptive Filters via Convex Combination," *IEEE Trans. Signal Process.*, vol. 56, pp. 3137– 3149, 2008.
- [5] B. Jelfs and D. P. Mandic, "Signal modality characterisation using collaborative adaptive filters," in 1st IAPR Workshop on Cognitive Information Process., 2008.
- [6] L.A. Azpicueta-Ruiz, A.R. Figueiras-Vidal and J. Arenas-García, "A Normalized Adaptation Scheme for the Convex Combination of Two Adaptive Filters," in *Proc. IEEE ICASSP'08*, pp. 3301–3304, 2008.
- [7] M. Zeller and W. Kellermann, "Coefficient Pruning for Higher-Order Diagonals of Volterra Filters Representing Wiener-Hammerstein Models," in Proc. Int. Workshop on Acoustics Echo and Noise Control (IWAENC), 2008.