A Functional Link Based Nonlinear Echo Canceller Exploiting Sparsity

Danilo Comminiello, Michele Scarpiniti, Raffaele Parisi and Aurelio Uncini
INFOCOM Dpt., "Sapienza" University of Rome,
Via Eudossiana 18, 00184 Rome.
Email: comminiello@infocom.uniroma1.it

***Abstract*—The presence of loudspeaker distortions in the echo path affects the performance of a conventional acoustic echo canceller (AEC). In order to address this problem we introduce a nonlinear acoustic echo cancellation (NAEC) architecture based on the adaptive combination of linear and nonlinear filters. The nonlinear filtering is performed by a *functional link network* (FLN) whose task is to compensate for the loudspeaker distortions. The linear contribution is given by a *variable step size improved proportionate normalized least mean square* (VIP-NLMS) algorithm which takes into account the impulse response sparsity thus further improving the system performance. Experimental results show the effectiveness of this system in the presence of loudspeaker distortions.**

I. Introduction

The main assumption that underpins conventional acoustic echo cancellers (AECs) is that the echo path is linear; unfortunately, when nonlinearities occur, AEC performance shows a decrease of the overall achievable quality. In order to model a nonlinear echo path and to tackle this problem, nonlinear acoustic echo cancellation (NAEC) systems are employed. The necessity of using NAECs is increasingly pressing due to the growing spread of low-cost loudspeakers for commercial hands-free communication systems. These devices cause significant nonlinearities in acoustic impulse responses (AIRs) leading to communication quality degradation [1], [2].

The strength of the NAEC architecture introduced in this paper lies in the combination of three fundamental features. The first feature is the nonlinear filtering carried out by a functional link network (FLN) [3]. Due to the absence of hidden layers, FLN allows to perform a cost-effective adaptive filtering that can be compared with a nonlinear finite impulse response (FIR) filtering. For this reason, in order to compensate for nonlinearities, FLN has been taken into account in several works, such as [4], [5]. Nonlinearities that affect communication quality are mainly nested in the early reflections of the AIR; therefore, it is reasonable to restrict the nonlinear filtering only to the first part of the AIR. The second feature is the linear adaptive filtering carried out by the *variable step size improved proportionate normalized least mean square* (VIP-NLMS) algorithm. The VIP-NLMS exploits the sparsity of the AIR and allows to consider AIR under-modeling situations. Both these peculiarities have been successfully used in AEC applications [6], [7]. The third feature stems from the observation that non-negligible gradient noise can result from the adaptation of a nonlinear filter for low levels of nonlinear echo contribution. This bother can decrease the NAEC performance to the point of bringing it below that of a conventional AEC. In order to avoid this drawback we adaptively combine linear filtering with nonlinear filtering. This approach has been used both in AEC applications [8] and in several other adaptive signal processing areas [5], [9], and allows the NAEC to achieve at least the same results of the best performing filtering.

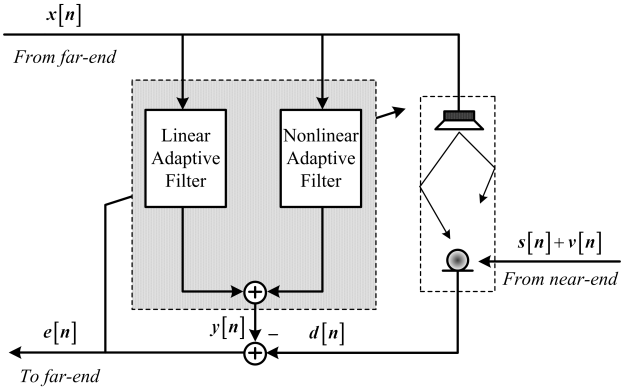

Fig. 1. Block diagram of the proposed NAEC.

This paper is organized as follows: the proposed architecture is described in Section II. Section III contains the evaluation of the effectiveness of the proposed system. The conclusions can be found in Section IV.

II. The Proposed NAEC Architecture

The proposed architecture depicted in Fig. 1 is composed of two main components: the linear adaptive filter and the nonlinear adaptive filter. At the time index n, the far-end signal $x\,[n]$ is fed through a loudspeaker-enclosure-microphone system (LEMS) activating a nonlinear echo path. The microphone also captures a near-end signal $s\,[n]$ and a possible near-end additive noise $v\,[n]$, thus generating the desired signal $d\,[n]$. The two adaptive filters take the reference signal $x\,[n]$, create an estimate $y\,[n]$ of the signal passed through the LEMS and remove it from the desired signal $d\,[n]$ yielding the error signal $e\,[n]$ to deliver to far-end. In this section the two components are first described and then adaptively combined to form the final structure.

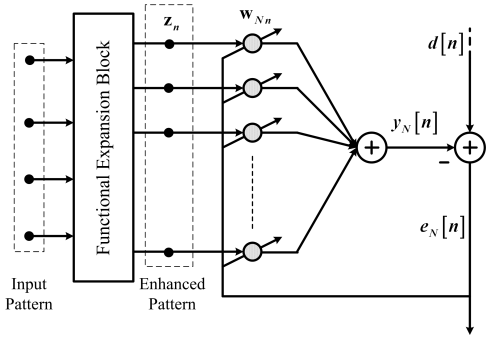

Fig. 2. The Functional Link Network.

A. Nonlinear Adaptive Filter: FLN

The *functional link network* (FLN) is an artificial neural network (ANN) with a single layer originally proposed by Pao [3] as an alternative to the well-known *multi-layer perceptron* (MLP) networks. FLN is able to generate complex decision regions by setting nonlinear decision boundaries [3]. As can be seen in Fig. 2, FLN receives an input pattern formed by one or more samples of the input signal. Unlike MLP, the FLN processes the input pattern directly by means of a functional expansion block consisting of a series of linearly independent functions that accept the input pattern as argument. Performance wise, the most commonly used functional expansion is the trigonometric series expansion. This functional expansion can be described by the enhanced pattern vector $\mathbf{z}_n$, whose elements are:

$$z[n-j] = \begin{cases} 1 & \text{for} \quad j = 0 \\ x[n-i] & \text{for} \quad j = 1 \\ \sin(j\pi x[n-i]) & \text{for} \quad j = 2, 4, \ldots, p \\ \cos(j\pi x[n-i]) & \text{for} \quad j = 3, 5, \ldots, p+1 \end{cases} \tag{1}$$

where $x[n-i]$ represents the i-th input at n-th instant; $i = 0, 1, \ldots, L_{in}-1$ is the input pattern index and L_{in} is the length of the input pattern; $j = 0, 1, \ldots, Q-1$ is the enhanced pattern index and $Q = 2p+1$ is the number of functional links for the i-th input; p is the expansion order. The first element of this series, $z[n]$, represents an unitary bias; therefore, the enhanced pattern is composed of $L_{en} = L_{in}(2Q+1) + 1$ elements. Consequently, L_{en} is also the length of the weight vector $\mathbf{w}_{N_n} = [w_N[n], w_N[n-1], \ldots, w_N[n-L_{en}+1]]^T$. We choose the length of the input pattern L_{in} equal to the length of the AIR segment bearing the largest energy contribution in order to restrict the nonlinear filtering only to the early reflections of the AIR. The functional expansion projects the input pattern in a higher dimension space.

Each element of the resulting enhanced pattern is linearly combined after being appropriately weighted. The weight vector $\mathbf{w}_{N_n}$ is adaptively updated according to an iterative learning rule.

The estimated output of the FLN is:

$$y_N[n] = \sum_{k=0}^{L_{en}-1} z[n-k]\, w_N[n-k] \tag{2}$$

The adaptive algorithm used in our system is the *normalized least mean square* (NLMS). The updating rule of the weight vector is described by the following equations:

$$e_N[n] = d[n] - y_N[n] \tag{3}$$

$$\mathbf{w}_{N_n} = \mathbf{w}_{N_{n-1}} + \frac{\mu_N \mathbf{z}_n e_N[n]}{\mathbf{z}_n^T \mathbf{z}_n + \delta_N} \tag{4}$$

where $e_N[n]$ is the FLN error signal, μ_N is a fixed step size and δ_N is the regularization parameter.

The main advantage of the FLN compared to other ANNs consists in an increase of the learning rate and in a decrease of the computational cost [3].

B. Linear Adaptive Filter: VIP-NLMS

The choice of the linear adaptive filter is paramount as this can prominently increase the performance of the NAEC. In our system we adopt the VIP-NLMS filtering algorithm. VIP-NLMS takes into account the sparsity constraints of proportionate algorithms; it is a variant of IPNLMS algorithm [6]. Proportionate algorithms are based on the main assumption that the AIR is sparse, as often is the rule in AEC applications. For this reason this class of algorithms provides a step size weight proportional to each filter coefficient thus converging more rapidly than NLMS. Besides exploiting these sparsity constraints, VIP-NLMS comprises the use of a variable step size useful in under-modeling situations [7], i.e. when the length of the adaptive filter is shorter than the length of the AIR. The under-modeling of the AIR limits the achievable performance of an AEC [1] and that is why it is important to consider under-modeling scenarios in AEC applications.

The first step in the algorithm description consists in the definition of the AIR vector $\mathbf{h}_n = [h[n], h[n-1], \ldots, h[n-M_A+1]]^T$ and the filter vector $\mathbf{w}_{L_n} = [w_L[n], w_L[n-1], \ldots, w_L[n-M+1]]^T$ which estimates the AIR in an under-modeling case. In these equations M_A is the length of the AIR and $M < M_A$ is the adaptive filter length. The input pattern $\mathbf{x}_n = [x[n], x[n-1], \ldots, x[n-M+1]]$ will consequently be an $M \times 1$ vector.

Sparse adaptive algorithms [6] can be generalized by the following set of equations:

$$e_L[n] = d[n] - y_L[n] \tag{5}$$

$$y_L[n] = \mathbf{x}_n^T \mathbf{w}_{L_n}. \tag{6}$$

The updating rule of the adaptive filter is:

$$\mathbf{w}_{L_n} = \mathbf{w}_{L_{n-1}} + \frac{\mu_L[n]\, \mathbf{G}_{n-1}\mathbf{x}_n e_L[n]}{\mathbf{x}_n^T \mathbf{G}_{n-1}\mathbf{x}_n + \delta_p} \tag{7}$$

where $\mu_L[n]$ is the variable step size, δ_p is the regularization factor and $\mathbf{G}_{n-1}$ derives from the following $M \times M$ diagonal matrix $\mathbf{G}_n = diag\{g[n], \ldots, g[n-M+1]\}$, which defines the step size weight for each filter coefficient and depends on the proportionate algorithm used.

In the VIP-NLMS algorithm, the diagonal elements of $\mathbf{G}_n$ are defined as in the IPNLMS case [6]:

$$g[n-l] = \frac{1-\alpha}{2L} + (1+\alpha)\frac{|w_L[n-l]|}{2\|\mathbf{w}_{L_n}\|_1 + \xi} \tag{8}$$

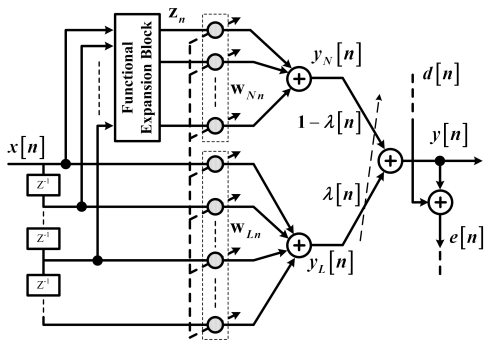

Fig. 3. The proposed NAEC architecture.

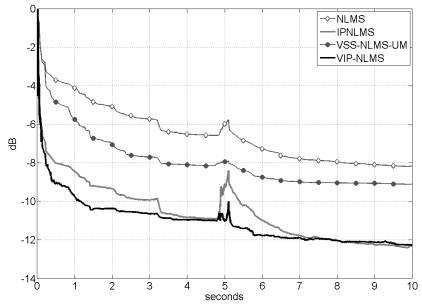

Fig. 4. Misalignment comparison with a male speech input. VIP-NLMS converges faster than other algorithms even under-modeling the AIR.

where $l = 0, \ldots, M-1$ and ξ is a small positive number which avoids divisions by zero; the parameter α balances the proportionality and its recommended value is 0 or -0.5 [6]. The regularization parameter is chosen as $\delta_p = (1-\alpha)\,\delta_L/2L$, where δ_L is the typical NLMS regularization parameter.

The variable step size parameter $\mu_L[n]$ is chosen according to [7] in order to take into account the under-modeling scenario:

$$\mu_L[n] = \begin{cases} \mu_f, & n \leq M \\ \left| 1 - \frac{\sqrt{\left|\hat{\sigma}_d^2[n] - \hat{\sigma}_{y_L}^2[n]\right|}}{\hat{\sigma}_{e_L}[n] + \xi} \right|, & n > M \end{cases} \tag{9}$$

where the general parameter $\widehat{\sigma}_\theta^2[n]$ represents the power estimate of the sequence $\theta[n]$, considering $\theta = \{d, y_L, e_L\}$:

$$\widehat{\sigma}_\theta^2[n] = \gamma \widehat{\sigma}_\theta^2[n-1] + (1-\gamma)\,\theta^2[n] \tag{10}$$

where $\gamma = 1 - 1/(kM)$ is a weighting factor with $k > 1$. The initial value is $\widehat{\sigma}_\theta^2[0] = 0$.

VIP-NLMS proves very effective in achieving high performance due to its proportionate nature and its variable step size.

C. Adaptive combination of linear and nonlinear filterings

The complete architecture of the proposed NAEC is depicted in Fig. 3; it consists of a convex combination of the nonlinear adaptive filter $\mathbf{w}_{N_n}$ and the linear adaptive filter $\mathbf{w}_{L_n}$ described above. In this way, according to [8] and [5], the two filtering processes can be seen as a single block whose output is given by the combination of the two filter outputs:

$$y[n] = \lambda[n]\, y_L[n] + (1 - \lambda[n])\, y_N[n] \tag{11}$$

where the mixing parameter $\lambda[n]$ adaptively balances the combination. As a consequence, the overall NAEC error signal is defined as $e[n] = d[n] - y[n]$.

In order to reduce the gradient noise and to keep the mixing parameter in the range $(0, 1)$, the adaptation of $\lambda[n]$ can be carried out through the adaptation of another parameter, $a[n]$, related to $\lambda[n]$ by the following equation:

$$\lambda[n] = \left(1 + e^{-a[n]}\right)^{-1}. \tag{12}$$

The update equation of $a[n]$ is given by:

$$a[n] = a[n-1] + \frac{\mu_a}{p[n]} \lambda[n]\,(1 - \lambda[n])\, e[n]\, \varepsilon[n] \tag{13}$$

where $\varepsilon[n] = e_L[n] - e_N[n]$; the term $p[n] = \beta p[n-1] + (1-\beta)\,\varepsilon^2[n]$ is the estimate power of $\varepsilon[n]$, and β is a smoothing factor.

As a result of this filter combination, when the echo path is affected by a high level of nonlinearities the NAEC benefits from the effectiveness of both filters. On the other hand, when the level of nonlinearities is low or negligible the NAEC works as the linear filter avoiding a performance decrease.

III. Experimental Results

In this section we evaluate the proposed system first showing the performance of the VIP-NLMS algorithm and then proving the effectiveness of the NAEC architecture. Both the experimental sets are conducted in a simulated teleconferencing scenario taking place in a $10 \times 6, 6 \times 3$ m room with a reverberation time of $T_{60} \approx 250$ ms. The AIR is computed by means of a Matlab tool, *Roomsim* [10], and is measured by using an 8 kHz sampling rate with $M_A = 2048$ coefficients.

A. Performance evaluation of the VIP-NLMS filter

In this simulation the echo path is supposed to be linear; for this reason we adopt only the VIP-NLMS filter in order to prove its effectiveness. We consider a noisy environment in which the target signal is a male speech signal. Furthermore, white gaussian noise with zero mean and unit variance is added as background noise with a *signal to noise ratio* (SNR) of 20 dB. In the parameter settings we choose the following values: $k = 6$, $\alpha = -0.5$, $\xi = 0.0001$, $\delta_L = 30\sigma_x^2$, where σ_x^2 is the power of the filter input signal, and $\mu = \mu_f = 0.2$. In VIP-NLMS we set the length of the filter to $M = 1024$. We evaluate the algorithm performance using the normalized misalignment:

$$\mathcal{M} = 20 \log_{10}\left(\left\| \mathbf{h}_n - \left[\mathbf{w}_{L_n}{}^T, \mathbf{0}^T \right]^T \right\|_2 / \|\mathbf{h}_n\|_2 \right) \tag{14}$$

where $\mathbf{0}$ is a row vector of $(M_A - M)$ zeros.

In Fig. 4 we compare the VIP-NLMS algorithm with NLMS, IPNLMS [6] and VSS-NLMS-UM [7]. It is possible to notice the behavior of VIP-NLMS which is better than other algorithms even under-modeling the AIR.

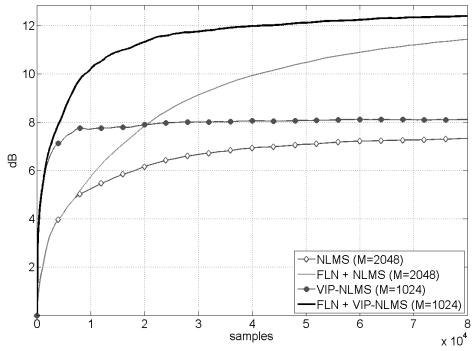

Fig. 5. ERLE comparison. The proposed NAEC (FLN + VIP-NLMS) exploits the combination of the nonlinear FLN and the linear VIP-NLMS.

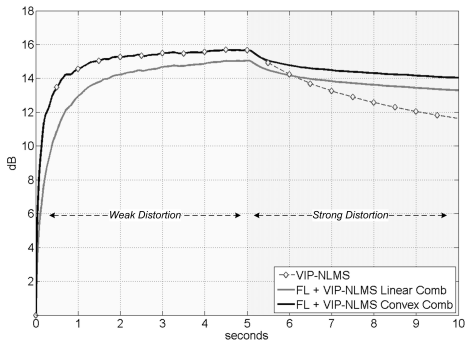

Fig. 6. The NAEC architecture with convex combination achieves at least the same results of the best performing filtering.

B. Effectiveness of the proposed NAEC architecture

In this set of experiments we consider two echo cancellation scenarios in order to show the effectiveness of the proposed NAEC. The first scenario is characterized by the presence of strong distortions of the loudspeaker achieved by a symmetrical soft clipping. We evaluate the performance using a white gaussian noise input with an independent additive noise with an SNR of 20 dB. In the parameter settings of the proposed NAEC we choose: $k = 2$, $\alpha = 0$, an expansion order of $p = 2$, $L_{in} = M_A/8$, $\mu_L = 0.2$ for the linear filter and $\mu_N = 0.4$ for the nonlinear filter, $\delta_N = 30\sigma_x^2$, a linear filter length of $M = 1024$, and $a[0] = 1$, $\mu_a = 0.5$, $\beta = 0.9$ related to the adaptive combination. Other parameters are the same used above. Performance is evaluated in terms of *Echo Return Loss Enhancement* (ERLE), which is defined as:

$$ERLE = 10\log_{10}\left(E\left\{d^2[n]\right\}/E\left\{e^2[n]\right\}\right) \qquad (15)$$

where the operator $E\{\cdot\}$ is the mathematical expectation.

In Fig. 5 we compare the proposed NAEC (FLN + VIP-NLMS) with the standard NLMS, with the linear filter VIP-NLMS and with the adaptive combination of a FLN with a NLMS linear filter (FLN + NLMS). Comparing NLMS with (FLN + NLMS) it is possible to notice the performance improvement due to nonlinearities analysis by means of FLN; comparing the two linear filters NLMS and VIP-NLMS one can see the importance of considering the sparsity constraints. Finally, the (FLN + VIP-NLMS) curve shows the effectiveness of the proposed NAEC.

In order to show the advantages of the convex combination we consider another scenario in which there is a change of the nonlinearity level in the echo path. For 5 seconds a white gaussian noise input is weakly distorted by adding both quadratic and cubic nonlinearities [11]; after which the nonlinearity turns into a soft clipping. In Fig. 6 it is possible to notice the comparison between the linear and the convex combination NAEC architectures. The linear combination scheme works well for strong distortions; however, its performance is lower than the linear VIP-NLMS filtering for weak distortions. The convex combination scheme doesn't suffer this drawback; its performance equals the VIP-NLMS filtering one for low levels of nonlinearities and benefits from the two filtering for strong distortions.

IV. Conclusion

This paper introduces an architecture for nonlinear acoustic echo cancellation based on the conjunction of three features: a nonlinear filtering based on a FLN, a VIP-NLMS linear filtering which exploits the AIR sparsity constraints in under-modeling scenarios, and the adaptive combination of these filtering processes which avoids to suffer from performance decreases. In this way the proposed architecture turns out to be robust to different levels of nonlinearities. As a matter of fact for low levels of nonlinearities the adaptive combination switches on the VIP-NLMS filter which acts like the best performing filter; on the other hand, for strong distortions the system exploits the FLN nonlinear filtering achieving a considerable performance enhancement.

References

[1] N. Birkett and R. A. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *IEEE ASSP Workshop on Appl. Of Signal Processing to Aud. And Acoustics*, New Paltz, New York, October 1995.

[2] A. Uncini, A. Nalin, and R. Parisi, "Acoustic echo cancellation in the presence of distorting loudspeakers," in *Proc. of Eusipco '02*, vol. 1, Tolouse, September 2002, pp. 535–538.

[3] Y. H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989.

[4] G. L. Sicuranza and A. Carini, "On the accuracy of generalized hammerstein models for nonlinear active noise control," in *Proc. of IMTC '06*, Sorrento, April 2006, pp. 1411–1416.

[5] H. Zhao and J. Zhang, "Adaptively combined fir and functional link artificial neural network equalizer for nonlinear communication channel," *IEEE Trans. on Neural Networks*, vol. 20, no. 4, pp. 665–674, 2009.

[6] J. Benesty and S. L. Gay, "An improved pnlms algorithm," in *Proc. of ICASSP '02*, vol. 2, Orlando, FL, 2002, pp. 1881–1883.

[7] C. Paleologu, S. Ciochina, and J. Benesty, "Variable step-size nlms algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Processing Letters*, vol. 15, pp. 5–8, 2008.

[8] L. A. Azpicueta-Ruiz, M. Zeller, J. Arenas-Garcia, and W. Kellermann, "Novel schemes for nonlinear acoustic echo cancellation based on filter combinations," in *Proc. of ICASSP '09*, Taipei, April 2009, pp. 193–196.

[9] B. Jelfs and D. P. Mandic, "Signal modality characterisation using collaborative adaptive filters," in *IAPR '08*, Santorini, June 2008.

[10] D. R. Campbell, K. J. Palomaki, and G. J. Brown, "Roomsim, a matlab simulation of "shoebox" room acoustics for use in teaching and research," *Comput. and Inform. Systems*, vol. 9, no. 3, pp. 48–51, 2005.

[11] X. Y. Gao and W. M. Snelgrove, "Adaptive linearization of a loudspeaker," in *Proc. of ICASSP '91*, vol. 3, Toronto, 1991, pp. 3589–3592.