FUNCTIONAL LINK BASED ARCHITECTURES FOR NONLINEAR ACOUSTIC ECHO CANCELLATION

D. Comminiello[†*], *L. A. Azpicueta-Ruiz*[‡], *M. Scarpiniti*[†], *A. Uncini*[†] *and J. Arenas-García*[‡]

[†] Dept. Information Eng., Electronics and Telecomms.
"Sapienza" University of Rome, Italy

[‡]Dept. Signal Theory and Communications
Univ. Carlos III de Madrid, Spain

ABSTRACT

This paper introduces a collaborative architecture for nonlinear acoustic echo cancellation. This system is achieved linearly combining a standard linear filter and a *functional link* (FL) nonlinear filter. The FL filter is composed of merely nonlinear elements thus acting like a pure nonlinear kernel and modeling at best a nonlinear path. A more robust architecture is further introduced in which the FL filter is adaptively combined with an *all-zero kernel* (AZK) avoiding any gradient noise when nonlinearities are negligible. Experimental results show that proposed architectures present improved performance compared with other nonlinear echo cancellers notwithstanding the nonlinearity level in the echo path.

Index Terms— Nonlinear Acoustic Echo Cancellation, Functional Links, Combination of Filters

1. INTRODUCTION

Acoustic echo cancellation (AEC) has been widely studied in recent years; however, many difficulties linger on satisfying the communication quality requirements. An important issue occurs when nonlinearities affect the acoustic path demanding the employment of nonlinear acoustic echo cancellers (NAECs). The necessity of using NAECs is increasingly pressing due to the growing spread of low-cost loudspeakers for commercial hands-free communication systems, that cause significant nonlinearities in the echo path and lead to communication quality degradation [1], [2]. However, when the echo path is roughly linear or contains negligible nonlinearities an NAEC could perform worse than a conventional AEC due to the gradient noise introduced by the nonlinear filter. Moreover, the ratio between linear and nonlinear echo signal power is unknown *a priori* and it is time-varying for nonstationary signals like speech. Thereby, it is not possible *a priori* to know if an NAEC will improve or deteriorate the cancellation.

A possible solution to this problem is the use of collaborative structures [3]. These systems are based on the adaptive combination of two or more filter outputs, thus performing at least as well as the best contributing filter. This approach has been used in AEC applications [4], as well as in other adaptive signal processing areas [5], [6].

In [7], collaborative schemes are investigated for NAEC employing Volterra filters and kernels, which are frequently employed as nonlinear solutions [8]. These collaborative schemes offer improved performance over the use of a single linear or nonlinear filter when the nonlinearity level is unknown or time-varying. However, the computational cost remains expensive due to the employment of Volterra kernels. An alternative collaborative architecture for NAEC is introduced in [9], in which a *Functional Link Network* (FLN) is adopted for the nonlinear filtering achieving good performance results. Due to its structure, FLN performs a cost-effective adaptive filtering that can be compared with a nonlinear finite impulse response (FIR) filtering. Another similar functional link based architecture is used in [10], as a nonlinear equalizer, however resulting in a more complex structure based on a neural approach.

In this paper, novel NAECs are introduced based on the work in [9], and on collaborative structures for adaptive filtering. Unlike the algorithm in [9], the presented NAECs completely decouple linear and nonlinear terms, using independent step sizes and memories for them. Thus, the proposed NAECs obtain their outputs as the mere sum of a linear filter and a pure nonlinear *functional link* (FL) filter, which are addressed to identify the linear and nonlinear parts of the echo path, respectively. Apart from a very effective performance, this scheme entails a cheaper cost compared to Volterra filters and the FLN combination of [9]. Moreover, a further architecture is proposed in which the FL nonlinear filter is adaptively combined with an *all-zero kernel*. This combination yields a better robustness to the gradient noise which could be introduced by the nonlinear filter.

The paper is organized as follow: the pure nonlinear FL filter is described in Section 2. In Section 3, both architectures for NAEC are presented and, in Section 4, the experimental results are shown. Finally, in Section 5 our conclusions are presented.

*Corresponding author e-mail: danilo.comminiello@uniroma1.it. This work has been performed while the first author was a visiting Ph.D. student at the Department of Signal Theory and Communications. The work of J. Arenas-García and L.A. Azpicueta-Ruiz was partly funded by MEC project TEC 2008-02473.

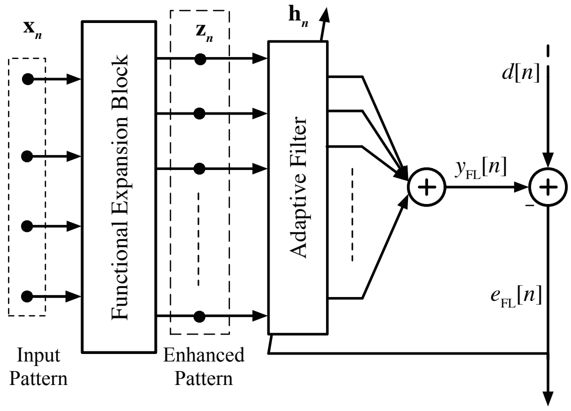

Fig. 1. The Functional Link nonlinear filter.

2. THE FUNCTIONAL LINK NONLINEAR FILTER

In [11] the *Functional Link Network* (FLN) is described as an *artificial neural network* (ANN) with a single layer.
FLN receives an input pattern formed by one or more samples of the input signal and processes it by means of a functional expansion block. This block consists of a series of linearly independent functions that accept the input pattern as argument. The functional expansion projects the input pattern in a higher dimension space. This process derives directly from the machine learning theory, more exactly from Cover's Theorem on the separability of patterns [12].

The functional expansion used in [9] is a trigonometric series expansion composed of both linear and nonlinear elements. However, in order to implement a purely nonlinear kernel, we will consider that the enhanced pattern $\mathbf{z}[n]$ is composed just by some nonlinear transformations of the original variables:

$$\mathbf{z}[n] = \left[1,\ \mathbf{z}_0^T[n], \ldots,\ \mathbf{z}_{\mathrm{L}_{in}-1}^T[n]\right]^T \tag{1}$$

where the sub-vectors $\mathbf{z}_i[n]$ are:

$$\mathbf{z}_i[n] = \left[z_{i,0}[n], \ldots,\ z_{i,\mathrm{Q}-1}[n]\right]^T \tag{2}$$

and whose elements $z_{i,j}[n]$ assume the following values:

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

in which $p = 0, 1, \ldots, P-1$, being P the expansion order, and $x[n]$ represents the input sample 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 $Q = 2P$ is the number of functional links for the i-th input. The first element of (1) represents a unitary bias; therefore, the enhanced pattern is composed of $L_{en} = 2QL_{in} + 1$ elements. Consequently, L_{en} is also the length of the functional link weight vector $\mathbf{h}[n] = \left[h_0[n], h_1[n], \ldots, h_{\mathrm{L}_{en}-1}[n]\right]^T$. The resulting enhanced pattern will be the input of a FIR filter. The adaptive algorithm chosen for our FL filter is the *normalized least mean square* (NLMS) yielding the FL output $y_{\mathrm{FL}}[n]$ and the FL error signal $e_{\mathrm{FL}}[n]$, as can be seen in Fig. 1.

The main advantages of the FL nonlinear filter are its low computational cost compared to ANNs or Volterra filters and the possibility to be used as a nonlinear kernel in order to model nonlinear echo paths.

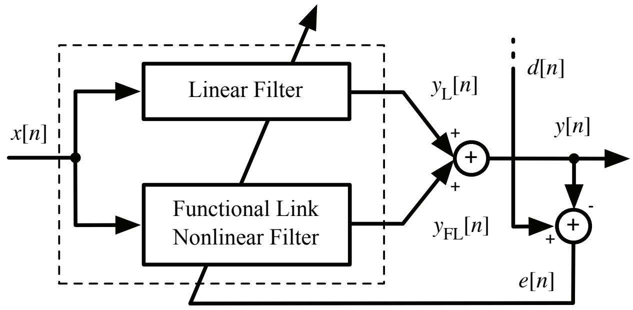

Fig. 2. "NLMS-FL" architecture: linear combination between a NLMS linear filter and a FL nonlinear filter.

3. NAEC ARCHITECTURES BASED ON FL KERNELS

3.1. Combination of FL with a linear filter

The first proposed architecture is composed of a standard NLMS linear filter which sums its output to that of an FL nonlinear filter (see Fig. 2). Denoting by $\mathbf{w}[n] = [w_0[n], w_1[n], \ldots, w_{\mathrm{L}-1}[n]]^T$ the linear filter vector, where L is the filter length, it is possible to define the system output as the sum of the linear and the nonlinear outputs:

$$y[n] = y_{\mathrm{L}}[n] + y_{\mathrm{FL}}[n] \tag{4}$$

and thereby the overall error signal is:

$$e[n] = d[n] - y[n] \tag{5}$$

where $y_{\mathrm{L}}[n] = \mathbf{x}^T[n]\,\mathbf{w}[n]$ and $y_{\mathrm{FL}}[n] = \mathbf{z}^T[n]\,\mathbf{h}[n]$ are the linear and FL output signals respectively, and $d[n]$ is the desired signal.

Minimizing the overall square error $e^2[n]$ by using the NLMS adaptation rule leads to:

$$\mathbf{w}[n] = \mathbf{w}[n-1] + \mu_{\mathrm{L}} \frac{\mathbf{x}[n]\,e[n]}{\mathbf{x}^T[n]\,\mathbf{x}[n] + \delta} \tag{6}$$

$$\mathbf{h}[n] = \mathbf{h}[n-1] + \mu_{\mathrm{N}} \frac{\mathbf{z}[n]\,e[n]}{\mathbf{z}^T[n]\,\mathbf{z}[n] + \delta} \tag{7}$$

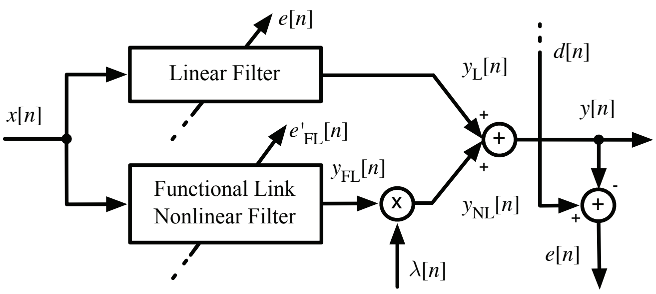

Fig. 3. "NLMS-FL/AZK" architecture: linear combination between an NLMS linear filter and a nonlinear filter composed of a convex combination between an FL nonlinear filter and an all-zero kernel.

where δ is a regularization constant, μ_{L} and μ_{N} are step sizes, and where it is remarkable that the normalizing factors in (6) and (7) take only into account the inputs to each filter, similarly to what is done for the Volterra filter adaptation in [7].

Compared to the algorithm in [9], which uses a convex combination of the outputs of a linear filter and a FLN component with weights $\lambda[n]$ and $(1-\lambda[n])$, the new architecture, that will be referred to in the following as "NLMS-FL", presents the advantage of using the whole linear filtering notwithstanding the nonlinearity level in the echo path. Moreover, the system can model the nonlinear path independently from the linear filtering due to the pure nonlinear nature of FL. Furthermore, FL is able to model the nonlinear path even with a small expansion order, entailing a very small nonlinear filter length. This property affects the computational complexity of the system, which results inferior than the system in [9] and Volterra filters.

3.2. Robust architecture employing an all-zero kernel

The main possible disadvantage of the "NLMS-FL" is that in presence of a pure linear path or of a path with negligible nonlinearities, it could suffer from gradient noise introduced by the nonlinear filter. In order to overcome this drawback we introduce a more robust architecture in which the FL filter is adaptively combined with an *all-zero kernel* (AZK), i.e. a *virtual* kernel whose coefficients are set to zero and do not need adaptation [7]. This architecture avoids the nonlinear contribution when the echo path is nearly linear, and the FL filter is unnecessary. This scheme is depicted in Fig. 3.

In this case, the overall output signal in (3) becomes:

$$y[n] = y_{\mathrm{L}}[n] + \lambda[n]\, y_{\mathrm{FL}}[n] \tag{8}$$

where $\lambda[n]$ is a parameter in range $[0,1]$, so that the output of the FL can be either kept or removed as required by filtering scenario.

In order to completely exploit the kernel-structure in the nonlinear filtering, the linear and the FL filters are updated using different error signals [7]. The linear filter pursues the minimization of the overall error in (5) using the overall output defined in (8); however, the FL filter $\mathbf{h}[n]$ is adapted using $e'_{\mathrm{FL}}[n]$, defined as:

$$e'_{\mathrm{FL}}[n] = d[n] - (y_{\mathrm{L}}[n] + y_{\mathrm{FL}}[n]) \tag{9}$$

The mixing parameter $\lambda[n]$ can be adapted using a gradient descent rule through the adaptation of another parameter, $a[n]$, related to $\lambda[n]$ by the expression $\lambda[n] = \mathrm{sgm}(\mathrm{a}[\mathrm{n}])$. Taking derivatives of $e^2[n]$ with respect to $a[n]$, we obtain the update equation [7]:

$$a[n+1] = a[n] + \frac{\mu_a}{r[n]} y_{\mathrm{FL}}[n]\, e[n]\, \lambda[n]\,(1-\lambda[n]) \tag{10}$$

where $r[n] = \beta r[n-1] + (1-\beta)\, y^2_{\mathrm{FL}}[n]$ is the estimated power of $y_{\mathrm{FL}}[n]$, and β is a smoothing factor.

The proposed "NLMS-FL/AZK" architecture is robust to any nonlinearity level, since when the echo path is merely linear $\lambda[n]$ will converge towards 0 and the whole scheme will behave like a purely linear filter, thus avoiding any gradient noise from the nonlinear filter. On the other hand, when the echo path presents nonlinearities the mixing parameter will approach 1 according to the nonlinearity level in the echo path. Note that when $\lambda[n] = 1$ the "NLMS-FL/AZK" architecture is equal to the "NLMS-FL" one.

4. EXPERIMENTAL RESULTS

In this section we investigate the performance of both the proposed architectures "NLMS-FL" and "NLMS-FL/AZK" in three different echo cancellation scenarios. The first scenario is a simulated teleconferencing environment in which the echo path is purely linear and the impulse response length is $L = 320$ coefficients. The second scenario is similar to the first; however, a quadratic distortion is also present in the echo path. The linear and the quadratic paths produce a *linear to nonlinear ratio* LNLR $= 10$ dB. In the third scenario instead of using a quadratic distortion we simulate a symmetrical soft clipping [13] that can be caused by an increase of the volume or by the plastic enclosure vibrations of a low-cost loudspeaker and thus it represents a more realistic scenario.

Two kinds of input signal are used for these scenarios: a white Gaussian noise input with zero mean and unitary variance and a female speech input. Additive Gaussian noise is added at the output of the echo path in order to provide 20 dB of *signal to noise ratio* (SNR). The length of the experiments is $t = 10$ seconds.

The two proposed architectures are compared with other two nonlinear echo cancellers: a Volterra filter and the architecture proposed in [9]. The last one is slightly adjusted to our scenarios; in order to be fully compared with the new architectures we use a standard NLMS linear filter instead of the

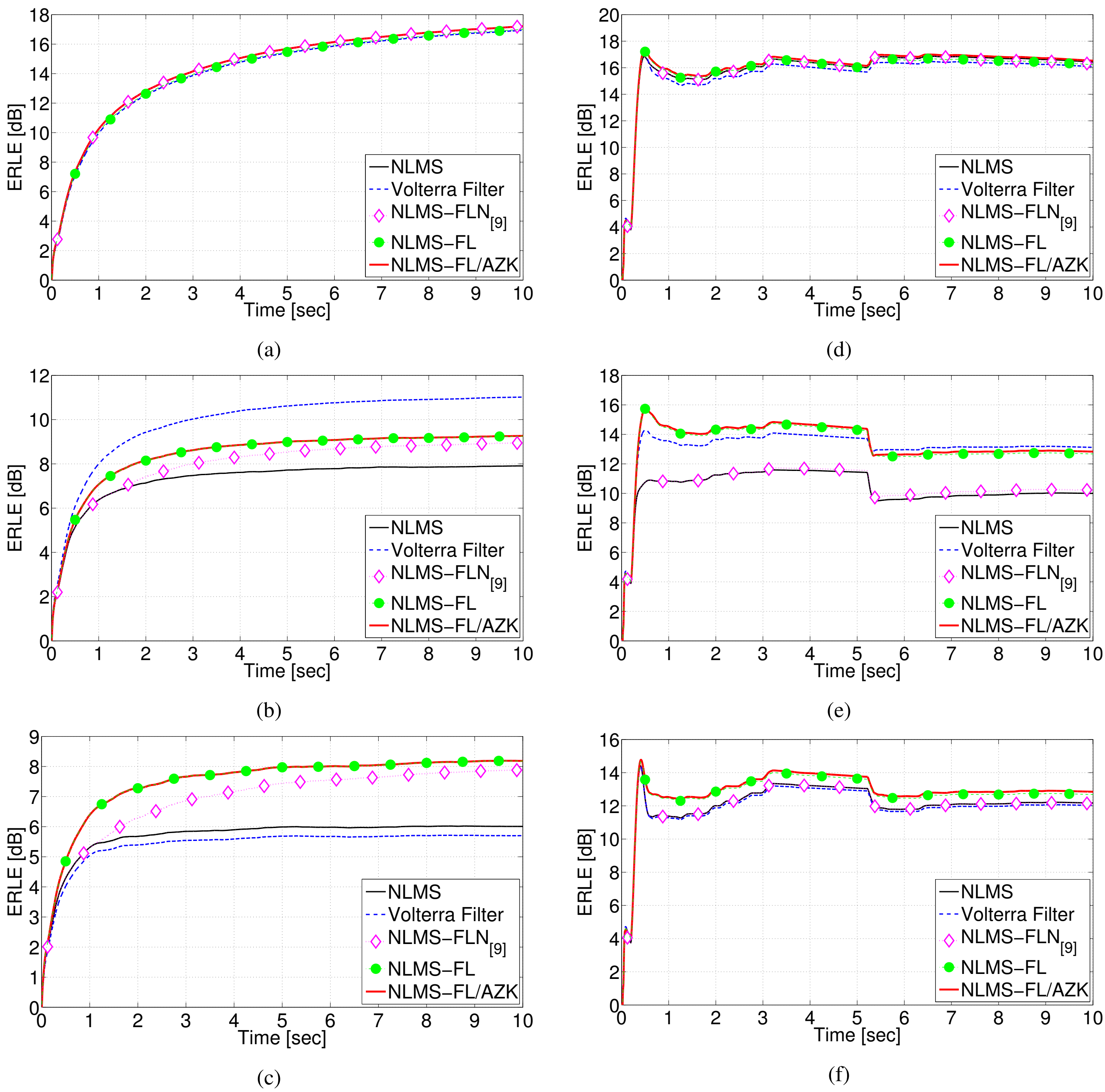

Fig. 4. Performance of the described architectures in three different scenarios: (a),(d) linear path, (b),(e) linear path with quadratic distortions, (c),(f) linear path with soft clipping. Two different signals are used as input: (a),(b),(c) white Gaussian noise and (d),(e),(f) female speech.

VIP-NLMS used in [9]. We denote this architecture "NLMS-FLN$_{[9]}$". Regarding the Volterra model we use a simplified Volterra filter having only the main diagonal. This choice is due to fact that in our experiments we consider a memoryless distortion. Therefore, in order to achieve a fair comparison with the proposed architectures, we neglect off-diagonal terms of Volterra filter, since they might introduce some gradient noise and deteriorate the performance.

The choice of parameter settings for the NAEC schemes is as follows. We choose two different step sizes for linear and nonlinear part for all the schemes. This choice is due to FL properties; since FL is a pure nonlinear filter and it does not depend on the linear path, it is possible to choose a smaller input pattern length than the linear input buffer and a different step size for the two filters. The linear step size $\mu_{\mathrm{L}} = 0.2$ is adopted for all the linear filters, while a step size $\mu_{\mathrm{N}} = 0.1$

is adopted for the FL nonlinear filters and for the quadratic kernel of Volterra filter. A slightly different value, $\mu_{\mathrm{N}} = 0.05$, is chosen for the FLN filtering due to the presence of linear elements. Other settings are: an expansion order of $P = 5$, $L = 320$, $L_{in} = L/8$, $\delta = 30\sigma_x^2$. For the combination, we use $a[0] = 0$, $\mu_a = 0.5$ and $\beta = 0.9$. Performance of all architectures are evaluated in terms of *echo return loss enhancement* (ERLE), which is defined as:

$$\mathrm{ERLE} = 10\log_{10}\left(\frac{\mathrm{E}\left\{d^2[n]\right\}}{\mathrm{E}\left\{e^2[n]\right\}}\right) \tag{11}$$

where the operator $\mathrm{E}\{\cdot\}$ is the mathematical expectation which will be estimated over 1000 experiments.

Fig. 4 shows results for all the architectures in the three scenarios. Considering white noise input, as expected, "NLMS-FL/AZK" and "NLMS-FLN$_{[9]}$" perform as well as the linear filter in absence of nonlinearities [pannel (a)] while Volterra filter performs worse due to gradient noise introduced by the quadratic kernel. "NLMS-FL" performs not much worse than the best architectures due to the little amount of FL nonlinear elements thus introducing very little gradient noise. Fig. 4 (b) shows that Volterra filter is the best performing scheme in presence of quadratic distortions. This is due to the fact that Volterra filter perfectly matches the quadratic distortion and thus it is an expected result. However, it is remarkable that the proposed FL based architectures provide also a resonable cancellation performance, significally improving the ERLE of a linear filter. In Fig. 4 (c) it is possible to see that the proposed architectures display the best performance in presence of soft clipping distortions. This is an important result since it corresponds to the most realistic scenario.

Pannels (d), (e) and (f) in Fig. 4 show the performance of all studied algorithms, and for all three kinds of echo paths, in the case of female speech input. Results can be generally described in similar terms to the white input case, with a significant difference in case of quadratic distortions, shown in Fig. 4 (e): due to time-varying nature of speech, nonlinearities affect the signal in a different way compared to white Gaussian noise case. As a consequence Volterra filter is not yet the best choice and "NLMS-FL/AZK" represents the best solution in all these last three scenarios.

5. CONLUSIONS

This paper presents an NAEC architecture composed of a standard linear filter and an FL nonlinear kernel. The FL is further adaptively combined with an AZK. The strength of this system lies in the pure nonlinear nature of the FL filter and in its collaborative structure thus allowing a better nonlinear path modeling and avoiding any gradient noise. The proposed system has been evaluated in three different scenarios with different kind of nonlinearity, both for white noise and speech input signals. The proposed system shows improved performance compared with other NAECs notwithstanding the kind of nonlinearity, especially for the speech input case.

Future work includes the study of the proposed algorithms for other kind of distortions and LNLRs. Combining the coefficients of both the linear and FL filters with zeros in a block-based manner, could also be advantageous in order to exploit the sparsity that is inherent to most real echo channels.

6. 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 Audio And Acoustics*, New Paltz, New York, Oct. 1995, pp. 103–106.

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

[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. on Signal Process.*, vol. 54, no. 3, pp. 10781090, Mar. 2006.

[4] J. Arenas-García and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 17, no. 6, pp. 1087 – 1098, Aug. 2009.

[5] B. Jelfs and D. P. Mandic, "Signal modality characterisation using collaborative adaptive filters," in *1st IAPR Workshop on Cognitive Information Process*, Santorini, June 2008.

[6] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. on Signal Process.*, vol. 56, pp. 3137 – 3149, July 2008.

[7] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-García, and W. Kellermann, "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 19, no. 1, pp. 97 – 110, Jan. 2011.

[8] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. of ICASSP '99*, Phoenix, AZ , USA, Mar. 1999, vol. 2, pp. 877 – 880.

[9] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "A functional link based nonlinear echo canceller exploiting sparsity," in *Proc. of IWAENC '10*, Tel Aviv, Aug.-Sept. 2010.

[10] 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.

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

[12] S. Haykin, *Neural Networks and Learning Machines*, Pearson Prentice Hall, 3rd edition, 2009.

[13] U. Zölzer and X. Amatriain, *DAFX: Digital Audio Effects*, John Wiley & Sons, Inc., 2002.