12

Collaborative Architectures for NAEC

Contents

Adaptive combination of filters may result very useful in setting the critical parameters of a filter during the adaptation, as shown in the previous chapter. However, the adaptive combination might result non-optimal when the goal is to exploit the capabilities of different models, or adaptive filters having different modelling tasks. In fact, in such situations in order to reach a desired performance the contribution of each filter might be necessary to reach a goal. This is the reason why affine and convex constraints might be not appropriate, since the sum of the mixing coefficients could be larger than one. In this chapter we use the adaptive combination of filters in a different way, in order to develop not combined but collaborative filtering architectures through the introduction of a virtual filter. We apply such collaborative architectures for the nonlinear acoustic echo cancellation. Experimental results show that proposed architectures show a more robust behaviour compared with other nonlinear echo cancellers aside from the nonlinearity level in the echo path[1].

12.1 A SERIUOS PROBLEM IN NAEC

In immersive speech communications, 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 [18], [147]. 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. This trouble, along with the expensive computational cost of a NAEC, affects the strategies of many companies that provide teleconferencing services, which often choose to drop the use of nonlinear echo cancellers even at the expense of communication quality.

A possible solution to this problem is the use of collaborative filtering architectures. Collaborative filtering architectures are based on the convex combination of an adaptive filter with an *all-zero kernel* (AZK), i.e. a *virtual* kernel whose coefficients are set to zero and do not need adaptation [10]. Such convex combination is depicted in Fig. 12.1, where it is possible to see that, while the adaptive filter is updated according to its own error signal

[1]The work in this chapter has been partly performed while the author was a visiting Ph.D. student at the Department of "Teoría de la Señal y Comunicaciones", at "Universidad Carlos III de Madrid".

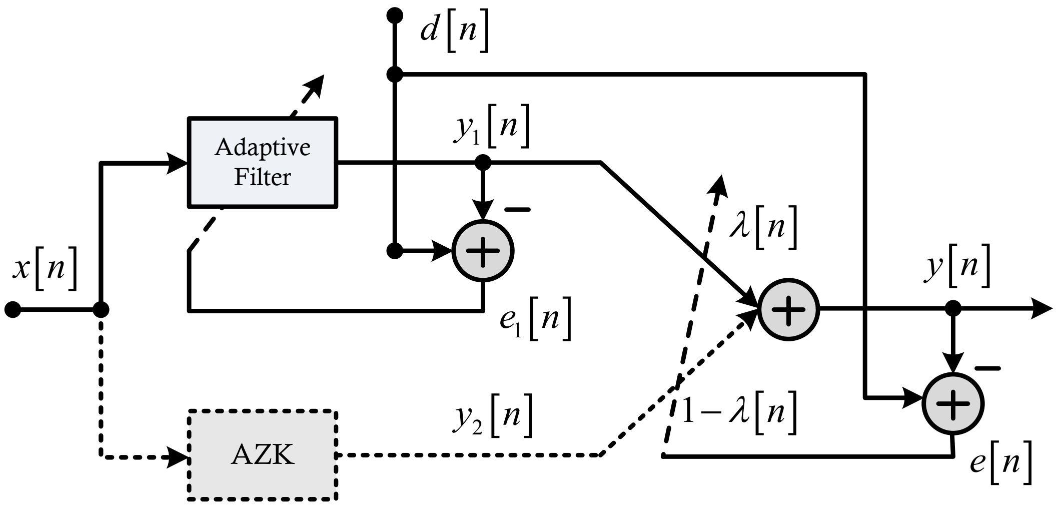

Fig. 12.1: *Intelligent switching circuit. The structure is composed of a convex combination between an adaptive filter and an all-zero kernel.*

$e_1[n]$, the AZK is not adapted since it is a vector with null coefficients. As a consequence, the output signal of the AZK $y_2[n]$ is a null contribution. This scheme is nothing but an *intelligent switch circuit*. In fact, according to the cost function chosen for the adaptation of the mixing parameter $\lambda[n]$, the circuit can automatically activate or deactivate the adaptive filter. Such switching is performed by the convex combination: according to equation (10.5), when the mixing parameter $\lambda[n]$ is close to 1 the circuit output $y[n]$ will bear the adaptive filter contribution $y_1[n]$, while when $\lambda[n] \to 0$ the circuit selects the AZK output, thus resulting in a null output signal for the overall circuit.

Adaptive schemes using such intelligent switching circuit are introduced in [10] for NAEC employing Volterra filters and kernels, which are frequently employed as nonlinear solutions [138]. 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 effective collaborative architecture for NAEC is introduced in this chapter using the intelligent switching circuit in combination with a functional link adaptive filter (FLAF) (see Chapters 8 and 9). The resulting collaborative NAEC exploits the capabilities of FLAF-based NAECs introduced in Chapter 9 and, in addition, shows robustness against the variations of nonlinearity degree in an acoustic path.

12.2 COLLABORATIVE FLAF

Changes proposed in SFLAF (see Section 9.2), compared to the standard FLAF in Chapter 8, gives robustness to the flexibility of an NAEC based on functional links, due to the possibility to make the right choice for the critical parameters of the filter. However, some drawbacks may linger on when the nonlinearity degree varies in time. In particular, a non-optimal filtering may occur when the nonlinearity degree changes from a medium/high level to a very low one, such that the nonlinearity can be considered as irrelevant. It is well known [10, 30], indeed, that NAEC performance may result inferior than that of a conventional linear AEC when the desired signal is not affected by any nonlinearity, or when the nonlinearity degree is negligible. In that case, the nonlinear filter only brings some gradient noise in the filtering process, thus NAEC performance is subjected to a decrease. This is also the reason why conventional AEC devices are more commercially available than NAECs.

In order to design an NAEC robust to the changes of nonlinearity degree, we propose a collaborative architecture based on the convex combination of adaptive filters (see Section 10.2). Using the convex combination it is possible to exploit the capabilities of the individual filters, thus performing at least as well as the best contributing filter. Convex combination may result very useful in setting the critical parameters of a filter, as it is shown in [6, 126, 4]. However, convex combination might result non-optimal when the goal is to exploit the capabilities of different models, or adaptive filters having different modelling tasks, as in our case. As a matter of fact, the convex combination of a nonlinear FLAF with a linear filter might not fully exploit the linear filter capability to model the acoustic echo path when the desired signal is affected

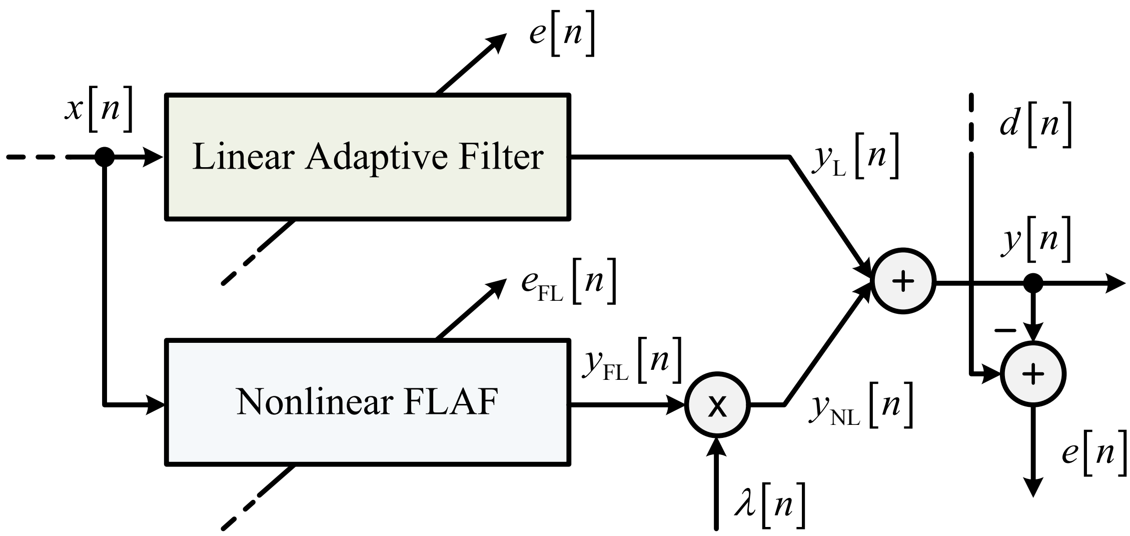

Fig. 12.2: *Collaborative functional link adaptive filter.*

by any nonlinearity.

Contrariwise, in designing an NAEC, it is desirable to enable the nonlinear modelling only if necessary. This is the reason why the proposed collaborative architecture exhibits a linear filtering always active and a nonlinear filtering which can be adaptively enabled and deactivated by means of an *intelligent switching circuit*, as depicted in Fig. 12.2. Such a collaborative architecture avoids the nonlinear contribution, and consequently the introduction of any gradient noise, when the echo path is almost linear, and the nonlinear FLAF is unnecessary.

The *collaborative FLAF*-based NAEC, that we denote as CFLAF, is depicted in Fig. 12.2, in which it is possible to notice that the overall output signal results as:

$$y\left[n\right] = y_{\mathrm{L}}\left[n\right] + \lambda\left[n\right] y_{\mathrm{FL}}\left[n\right] \tag{12.1}$$

where the *mixing parameter* $\lambda\left[n\right]$ allows to either keep or remove the output of the nonlinear FLAF as required by the filtering scenario. In equation (12.1) we omit the term weighted with $\left(1-\lambda\left[n\right]\right)$ and related to the AZK, as its contribution is null.

Due to the fact that linear and nonlinear filterings have different tasks, each filter is updated using different error signals in order completely to exploit the collaborative structure. In particular, the linear filter $\mathbf{w}_{\mathrm{L},n}$ pursues the minimization of the overall error signal $e[n] = d[n] - y[n]$, as the output contribution of the linear filter is always present. Differently, the nonlinear FLAF $\mathbf{w}_{\mathrm{FL},n}$ is updated using the local error $e_{\mathrm{FL}}[n]$ from which the linear output $y_{\mathrm{L}}[n]$ is subtracted, as it is always taken into account by the linear filtering:

$$e_{\mathrm{FL}}[n] = d[n] - (y_{\mathrm{L}}[n] + y_{\mathrm{FL}}[n]). \tag{12.2}$$

The mixing parameter $\lambda[n]$ can be adapted in a convex way assuming that $0 \leq \lambda[n] \leq 1$ through the adaptation of an auxiliary parameter, $a[n]$, related to $\lambda[n]$ by means of a sigmoidal function defined as (10.8). Therefore, $\lambda[n]$ is computed adapting $a[n]$ through a gradient descent rule as $a[n+1] = a[n] + \Delta a[n]$, where $\Delta a[n]$ results from a *normalized least mean squares* (NLMS) adaptation (see Paragraph 10.3.2):

$$\begin{aligned}
\Delta a[n] &= -\frac{1}{2}\mu_a \frac{\partial e^2[n]}{\partial a[n]} \\
&= -\frac{\mu_a}{r[n]} e[n] \frac{\partial (d[n] - y_{\mathrm{L}}[n] - \lambda[n] y_{\mathrm{FL}}[n])}{\partial \lambda[n]} \frac{\partial \lambda[n]}{\partial a[n]} \\
&= \frac{\mu_a}{r[n]} e[n] y_{\mathrm{FL}}[n] \lambda[n] (1 - \lambda[n])
\end{aligned} \tag{12.3}$$

where

$$r[n] = \beta r[n-1] + (1-\beta) y_{\mathrm{FL}}^2[n] \tag{12.4}$$

is a rough low-pass filtered estimate of the power of the signal of interest [9]. The parameter β is a smoothing factor which ensures that $r[n]$ is adapted faster than any filter component. The value of $a[n]$ is kept within $[4, -4]$ for practical reasons [6] (see Section 10.3).

The proposed CFLAF architecture is robust against any nonlinearity level, since when the echo path is merely linear $\lambda[n]$ converges towards 0 and the whole scheme behaves like a purely linear filter, thus avoiding any gradient noise from the nonlinear FLAF. On the other hand, when the echo path conveys nonlinearities the mixing parameter approaches 1 according to the nonlinearity level in the echo path. Note that when $\lambda[n] = 1$ the CFLAF architecture performs like the SFLAF.

12.3 BLOCK-BASED COLLABORATIVE FLAF

A further weak spot of an FLAF-based NAEC may be a failed control over the expanded buffer. In fact, a control in that sense can be useful when the nonlinearity degree is unknown. In the previous subsection, we saw how a CFLAF is able to be robust when the nonlinearity degree varies from a negligible value to a detectable one and *vice-versa*. However, significant differences may occur when the nonlinearity degree varies between detectable levels with different intensity. As a matter of fact, a high expansion order may be necessary in order to model a high nonlinearity degree, so that the length of the expanded buffer is sufficiently large to ensure a high number of nonlinear elements. On the other hand, in case of detectable nonlinearity with a low/medium-intensity a large number of coefficients may cause an overfitting plight and, therefore, introduce some gradient noise, thus degrading filtering performance.

In order to overcome this drawback, we propose an improved CFLAF architecture featuring a block-based convex combination [4], that we name as *block-based collaborative FLAF* (BCFLAF). As we saw in the previous section, the adaptive combination in CFLAF allows to adaptively deactivate the whole nonlinear filtering whether not necessary. Similarly, the main idea which underpins BCFLAF approach is that of dividing the expanded buffer into blocks and adapting each block with its own mixing parameter, so that it is

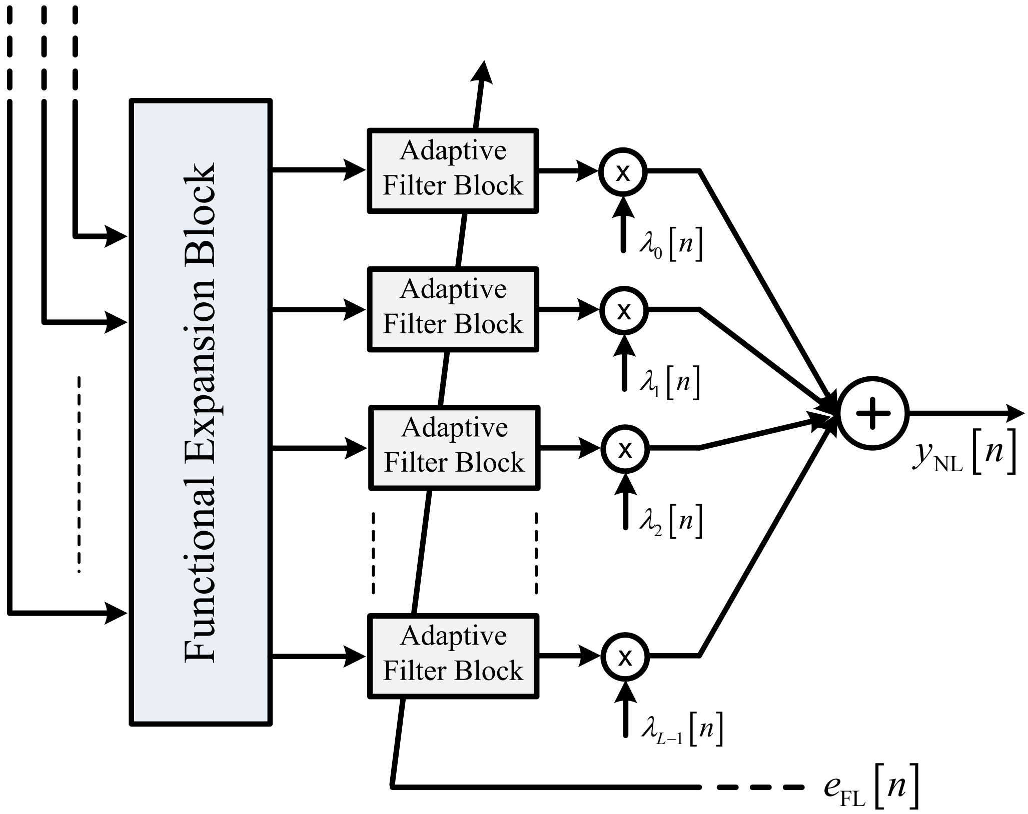

Fig. 12.3: *Nonlinear adaptive path in a block-based collaborative FLAF.*

possible to adaptively deactivate those blocks which are not useful to model nonlinearities. Due to the fact that the nonlinear filtering strictly depends on the length of the expanded buffer, and therefore on the number of nonlinear elements, it is possible to divide the expanded buffer in blocks just according to the desired accuracy. A sufficiently large number [6] of blocks may result in a high accuracy but also in an increase of the computational cost. Therefore, the block-based combination actually reduces the number of nonlinear elements selected for the nonlinear filtering and therefore avoids the introduction of any gradient noise.

The convex combination introduced in CFLAF, and described by equation (12.1), adopts the same mixing parameter for all weights. On the other hand, considering a number of L blocks, each one consisting of $M_b = M_e/L$ coefficients, it is possible to express the output of the BCFLAF as:

$$y[n] = y_{\mathrm{L}}[n] + \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} \lambda_l[n]\, g_k[n]\, w_{\mathrm{FL},k}[n-1] \tag{12.5}$$

where $\lambda_l[n]$ is the mixing parameter related to the l-th block, $w_{\mathrm{FL},k}[n-1]$ refers to the m-th coefficients of each block, and $\mathcal{I}_l = [l \cdot M_b, \ldots, (l+1)\, M_b - 1]$ is the range of indices related to the coefficients of the l-th block.

The block-based combination also affects the adaptation of the nonlinear filter $\mathbf{w}_{\mathrm{FL},n}$, which becomes:

$$\mathbf{w}_{\mathrm{FL},n} = \mathbf{w}_{\mathrm{FL},n-1} + \mu_{\mathrm{FL}} \frac{e_{\mathrm{FL}}[n] \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} \lambda_l[n]\, g_k[n]}{\delta_{\mathrm{FL}} + \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} |\lambda_l[n]\, g_k[n]|^2} \tag{12.6}$$

where μ_{FL} and δ_{FL} are respectively the step size and the regularization parameter for the all the blocks of the nonlinear filter.

The L mixing parameters can be adapted similarly to the equation (12.3) of the CFLAF case. Therefore, defining $\lambda_l[n] = \mathrm{sgm}(a_l[n])$, with $l = 0, \ldots, L-1$, the updating rule for each auxiliary parameter is given by:

$$\begin{aligned} a_l[n+1] = & a_l[n] + \frac{\mu_a}{r[n]} e[n]\, \lambda_l[n]\, (1 - \lambda_l[n]) \\ & \times \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} g_k[n]\, w_{\mathrm{FL},k}[n]\,. \end{aligned} \tag{12.7}$$

A graphical representation of the nonlinear filtering carried out by BCFLAF architecture is depicted in Fig. 12.3.

12.4 EXPERIMENTAL RESULTS

In this section we evaluate the performance of the proposed CFLAF in an acoustic echo cancellation scenario. We use the same experimental setup of Section 9.3 and also the same input signals having a length of 10 seconds. However, if the acoustic channel is nonlinear and the degree of nonlinearity remains constant, an NAEC using the CFLAF yields the same performance of the SFLAF, according to what said in Section 12.2. Therefore, in order to show the advantages of the convex combination, we consider a change of the nonlinearity level in the echo path. In fact, we start the process in linear conditions, i.e. the nonlinearities in the AIR are neglegible so that the acoustic path can be assumed as linear. After 5 seconds from the start of the process we introduce a clipping nonlinearity, the same as in Section 9.3.

In these scenario conditions, we compare the performance of three acoustic echo canceller in terms of ERLE: a conventional linear AEC, an NAEC based on the SFLAF and an NAEC based on the CFLAF. In a first experiment we consider the white Gaussian input and we use an NLMS algorithm to update the filters for all the three echo cancellers. The result is depicted in Fig. 12.4 in which it is possible to see that in the first half of the process, the best performing filter is the conventional NLMS, due to the fact that the AIR is purely linear. In this case the SFLAF shows a worse behaviour due to the gradient noise introduced by the nonlinear elements of the filter. However, it is possible to notice that, for the first 5 seconds, the CFLAF displays the same behaviour of the NLMS, and this is due to the fact that the intelligent switching circuit inside the CFLAF detects the absence of nonlinearities and selects the output contribution of the AZK; in this way the whole CFLAF reduces to be a linear filter.

However, in the second half of the process the nature of the AIR turns to be nonlinear and an immediate consequence is the performance decrease of the NLMS in Fig. 12.4. On the other side both the SFLAF and the CFLAF exploit the capabilities of the functional link based filtering and display better perfor-

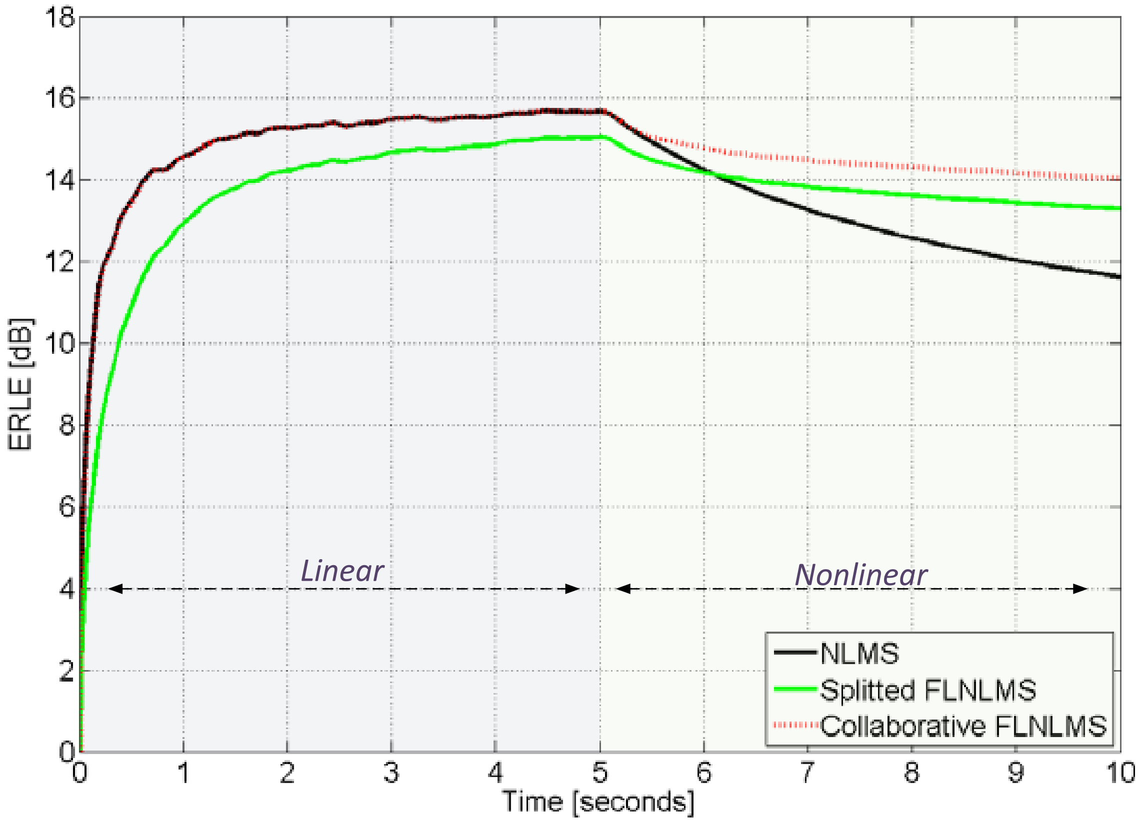

Fig. 12.4: *Performance comparison in terms of ERLE between a linear, an SFLAF-based and a CFLAF-based echo cancellers in case of white Gaussian input. All the filters are updated using an NLMS algorithm.*

mance than the linear AEC. However, due to the different initial conditions (at second 5) the CFLAF performs better than the SFLAF. Therefore, it is possible to state that, comparing to the NLMS and the SFLAF, the CFLAF is always the best performing acoustic echo canceller notwithstanding the nonlinearity degree in the echo path.

Same conclusions, even if with less evident differences, result from a second experiment using the female speech signal as input, as it is possible to see from Fig. 12.5. In this second experiment all the filters are updated using an APA with a projection order of $K = 3$.

Let us note that in this case it is difficult to comprehend the real benefits deriving from the collaborative architectures due to the fact that the ERLE does

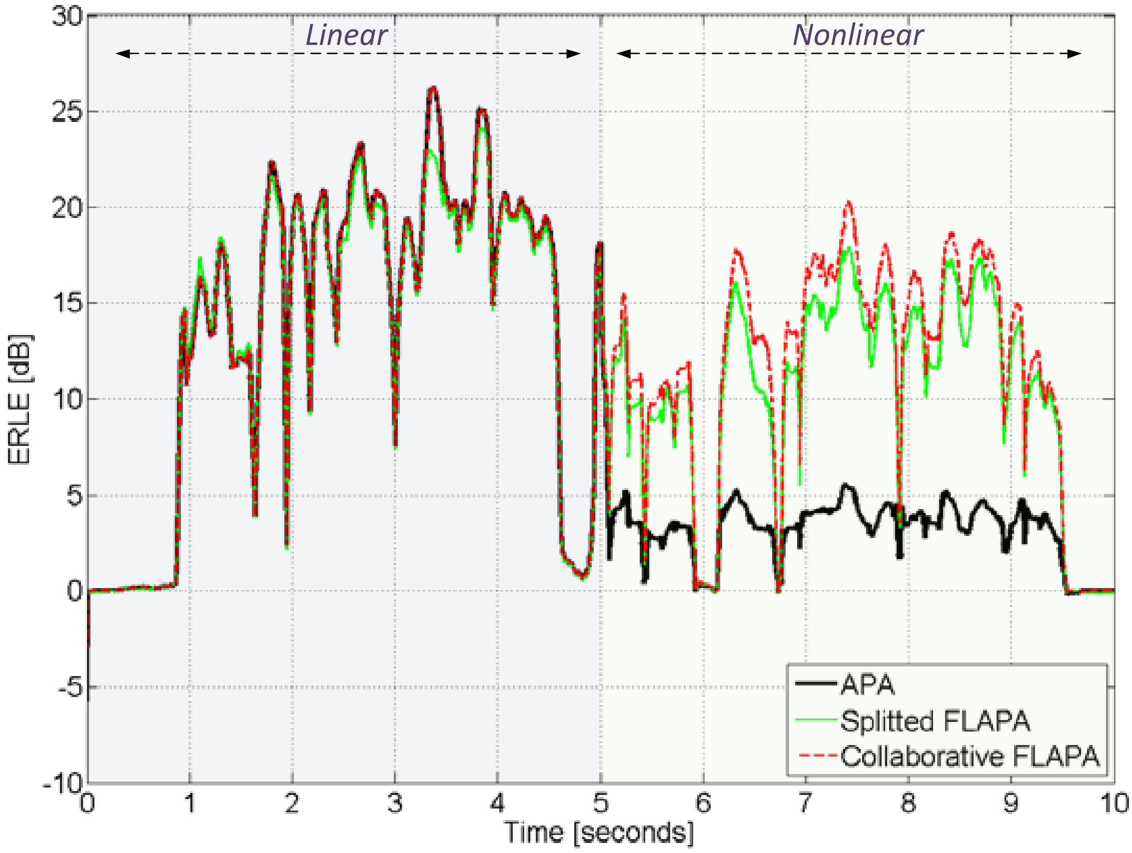

Fig. 12.5: *Performance comparison in terms of ERLE between a linear, an SFLAF-based and a CFLAF-based echo cancellers in case of female speech input. All the filters are updated using an APA.*

not reflect the perceived quality improvement of the speech, which is more evident than the ERLE improvement. In the linear case this lack is plugged by the *normalized misalignment* (see Section 3.4), however in the nonlinear case it is not possible to dispose of a similar performance measure, and it is often difficult to achieve a complete evaluation of an NAEC only using the ERLE, even if it is the most used measure in literature for the evaluation of a nonlinear echo canceller.

12.5 CONCLUSIONS

In this chapter we have introduced robust acoustic echo cancellers based on the adaptive combination of filters. In particular, we exploits the capabilities of the convex combination to develop an intelligent switching circuit which allows the combination of adaptive filters from different models. In this case, we have combined a linear adaptive filter and a nonlinear adaptive filter, thus obtaining collaborative filtering architecture that can be used for nonlinear echo cancellation. Such collaborative architectures have shown a more robust behaviour compared with other nonlinear echo cancellers notwithstanding the nonlinearity level in the echo path. This result paves the way for the development of more sophisticated architectures able to solve similar problems both for acoustic applications and also for other kinds of application.