

# A CENTRALIZED ACOUSTIC ECHO CANCELLER BASED ON PERCEPTUAL PROPERTIES

*H. Gnaba-Daassi*<sup>1</sup>, *P. Scalart*<sup>2</sup>, and *C. Marro*<sup>1</sup>

<sup>1</sup> France Télécom – TECH/SSTP, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France

<sup>2</sup> ENSSAT – LASTI, 6 Rue de Kerampont, B.P. 447, 22305 Lannion Cedex, France

E-mail: hela.daassi@math-info.univ-paris5.fr; pascal.scalart@enssat.fr; claude.marro@francetelecom.com

## ABSTRACT

The use of a centralized acoustic echo canceller in the mobile switching center of a GSM network allows the operator to enhance the audio quality before transmission to all subscribers. However, the main problem is the impact of the speech coder/decoder nonlinearities along the echo path. In this paper, we propose a combined acoustic echo canceller/post-filter based on perceptual properties to reduce the coding noise. Theoretical and experimental comparisons against classical solutions are given to evaluate the performance of the proposed approach in a centralized context of echo cancellation.

## 1. INTRODUCTION

Acoustic echo is a major problem in telecommunications in the GSM network where echo delay is especially annoying for speakers. There is an evident need for an acoustic echo canceller (AEC) to overcome this problem, particularly when poor quality mobile phones are used and in case of hands-free communication. The classical solution is to implement AEC in local user terminals. However, recent studies and applications of AEC [1, 2, 3, 4] propose a centralized AEC, i.e. located in the mobile switching center (MSC). This attractive configuration can significantly cancel the echo for all the subscribers and improve the communication interactivity for terminals having deficient AEC.

Low-bit-rate speech coder/decoder -such as GSM-EFR, AMR-NB,..- used in GSM network introduce distortions in the far-end and the near-end microphone signals [2, 4]. The main problem of a centralized AEC is the nonlinearity introduced by the speech codec, which are cascaded along the echo path as shown in Figure 1. In such application where an adaptive linear filter is used to identify the acoustic echo path, a conventional AEC is not sufficient and does not guarantee an imperceptible echo. In the same way, a centralized AEC based on classical nonlinear adaptive filtering [5] cannot handle nonlinear characteristics of the GSM network.

As a first solution to this problem, we have proposed in [2] a centralized combined AEC/post-filter system: a classical linear AEC identifies roughly the echo path, and a time domain post-filter which operates as a noise reduction system is applied to the residual echo. To improve the performance of this technique, we propose in this paper an original post-filter that incorporates perceptual properties.

Section 2 describes the problem of a centralized AEC in mobile communication. The AEC/post-filter structure is reviewed and the aspect of its application to a centralized context is discussed in Section 3. Theoretical and experimental comparisons with the proposed AEC/perceptual post-filter

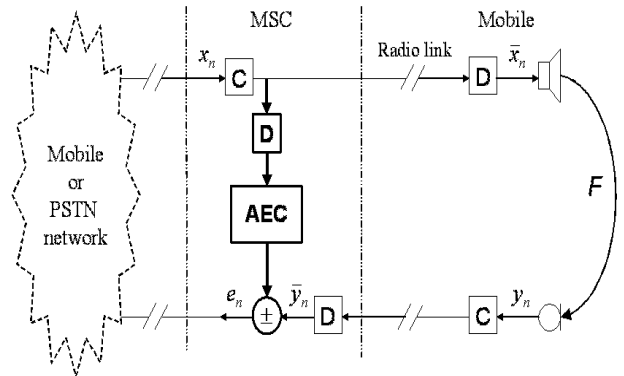


Figure 1: Centralized AEC for a communication with a mobile.

are presented in Section 4.

## 2. CENTRALIZED AEC PROBLEM IN MOBILE COMMUNICATION

Assuming that the radio channel is ideal, this section addresses the problem of centralized AEC when a communication with a mobile is considered (see Figure 1). In order to cancel the echo generated by the mobile terminal, the MSC seems to be a judicious location to implement a centralized AEC.

The input of this AEC scheme (see the bold style drawing in Figure 1) corresponds to a decoded signal  $\bar{x}_n$  and the reference signal is the uplink decoded output  $\bar{y}_n$ . The copy of the downlink codec operation in the AEC path enables the compensation of the downlink channel nonlinearity. The adaptive identification scheme equivalent to this centralized AEC is depicted in Figure 2. The entire nonlinear echo path to be identified is a cascade of an acoustic echo path, which is described by an impulse response  $F$  with  $Q$  coefficients, followed by a nonlinear function that represents the GSM codec.

In this section, we study the theoretical performance of a classical linear adaptive AEC in the centralized context. In the following, the time evolution of the adaptive filter  $H_n$  is controlled by the Normalized Least Mean Squares (NLMS) algorithm which is expressed as:

$$e_n = \bar{y}_n - H_{n-1}^T \cdot \bar{X}_n \quad (1)$$

$$H_n = H_{n-1} + \frac{e_n}{\bar{X}_n^T \cdot \bar{X}_n} \cdot \bar{X}_n \quad (2)$$

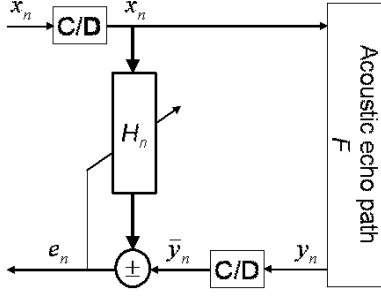


Figure 2: Equivalent adaptive identification scheme of the centralized AEC given in Figure 1.

where,

- $e_n$  is the residual echo signal (*a priori* error).
- $\bar{y}_n$  represents the microphone signal passed through the speech codec.
- $\bar{X}_n = [\bar{x}_n, \dots, \bar{x}_{n-P+1}]^T$  is a vector of the  $P$  past samples of the decoded far-end signal  $\bar{x}_n$ .
- $H_n = [h_n^0, \dots, h_n^{P-1}]^T$  represents the adaptive filter vector of length  $P$ .

In the case ( $P = Q$ ), the adaptive filter  $H_n$  corresponds to the optimal Wiener filter given by:

$$H_{opt} = R_{\bar{X}\bar{X}}^{-1} r_{\bar{X}\bar{y}} \quad (3)$$

where  $R_{\bar{X}\bar{X}}^{-1}$  is the inverse auto-correlation matrix of the vector  $\bar{X}_n$  and  $r_{\bar{X}\bar{y}}$  is the cross-correlation vector between signals  $\bar{x}_n$  and  $\bar{y}_n$ .

In a first approach, we can assume that the codec operation is a simple addition of a quantization noise  $b_n^y$ . Therefore, we can consider the decoded microphone signal as:

$$\bar{y}_n = y_n + b_n^y$$

Then Equation (1) becomes:

$$e_n = F^T \bar{X}_n + b_n^y - H_n^T \bar{X}_n \quad (4)$$

Furthermore, since  $F = R_{\bar{X}\bar{X}}^{-1} r_{\bar{X}\bar{y}}$ , Equations (3) and (4) can be written as

$$H_{opt} = F + R_{\bar{X}\bar{X}}^{-1} r_{b^y\bar{X}} \quad (5)$$

and

$$e_{opt} = b_n^y - (R_{\bar{X}\bar{X}}^{-1} r_{b^y\bar{X}})^T \bar{X}_n, \quad (6)$$

respectively. Hence, the optimal residual echo is audible and still too high to achieve the goal of echo cancellation. As a consequence, the use of a traditional conventional AEC is not sufficient in the presence of codec nonlinear effects.

In order to enhance the performance of the classical AEC, we propose in the next section a centralized AEC that incorporates a post-filter which is able to reduce this residual echo.

### 3. CENTRALIZED COMBINED AEC/POST-FILTER

#### 3.1 Principle

Most of noise reduction methods are based on short-term spectral attenuation techniques. Such methods introduce an extra delay due to the time to frequency domain analysis/synthesis. We propose in this section to use an adaptive

noise reduction filter  $H_n^2$  implemented in the time domain, whose coefficients are updated with the NLMS algorithm. A copy of this post-filter is applied to the residual echo  $e_n^1$  available at the output of the classical AEC  $H_n^1$ . Figure 3 presents the structure of this combined AEC/post-filter.

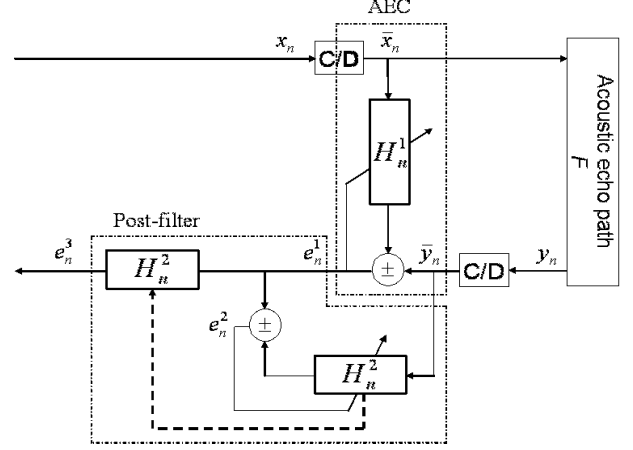


Figure 3: Centralized combined AEC/post-filter.

Similar systems have been used in speech enhancement [6] to reduce the residual echo related to long impulse response of the echo path. In our application, we do not only take into account the later phenomenon but we also reduce the quantization noise generated by the codec operation.

#### 3.2 Experimental results

We have studied theoretically in [2] the importance of the post-filter in a centralized context for a mobile-land call. The analysis of this approach is extended here to a mobile to mobile context and we propose to validate the results presented in [2] by some simulations.

The input signal is a speech encoded with the codec GSM-EFR (Enhanced Full Rate). The length  $P$  of the AEC adaptive filter ( $H_n^1$ ) and the one of the echo path  $Q$  are equal to 300 and 2000, respectively. Actually,  $F$  is room impulse response of 570 ms reverberation time. The adaptive filter  $H_n^2$  updated by the NLMS algorithm, is designed to minimize the mean square error  $E[(e_n^2)^2]$ . The time evolution of  $H_n^2$  is described by:

$$e_n^2 = e_n^1 - H_{n-1}^2 \bar{Y}_n \quad (7)$$

$$H_n^2 = H_{n-1}^2 + \frac{e_n^2}{\bar{Y}_n^T \cdot \bar{Y}_n} \bar{Y}_n \quad (8)$$

where  $\bar{Y}_n$  is the vector of the  $M$  past observations of the signal  $\bar{y}_n$ . Notice that  $M$  corresponds to the length of the post-filter which is fixed to 80 in the following. As depicted in Figure 3, the adaptive filter  $H_n^2$  is duplicated (represented by a dashed arrow) and applied to the residual echo  $e_n^1$ .

The local speech is not considered in this case, and the Echo Return Loss Enhancement (ERLE =  $10 \log \frac{E[(\bar{y}_n)^2]}{E[(e_n^1)^2]}$ ) is used to evaluate the echo reduction. Figure 4 represents the time evolution of the ERLE related to two cases, the solid line is the classical AEC (without post-filter:  $e_n = e_n^1$ ) and the

dashed line is AEC/post-filter (with post-filter:  $e_n = e_n^3$ ).

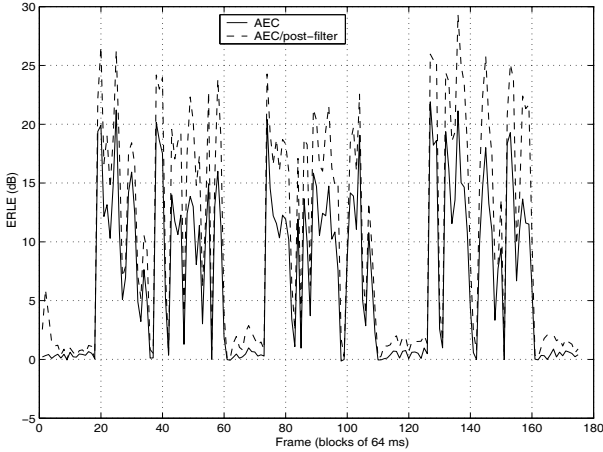


Figure 4: ERLE for AEC with and without post-filter.

The simulation results show that the combined AEC/post-filter system significantly outperforms the conventional AEC system. In average, an additional attenuation of 5 to 10 dB is introduced.

#### 4. AEC/PERCEPTUAL POST-FILTER

##### 4.1 Principle

The previous results illustrate that the AEC/post-filter system is useful in a centralized context. In this section, we propose an improvement of this technique by using a robust post-filter that exploits perceptual properties.

Our purpose is to estimate a perceptual filter giving more importance to high power frequency components, i.e. formant frequencies.

The global structure of the proposed centralized AEC/perceptual post-filter is depicted in Figure 5. The following operations are performed for each block of 160 residual echo  $e_n^1$  samples:

(1)- The signal  $e_n^1$  is analyzed to extract coefficients of an  $m$ th order linear prediction (LP) filter which is given by:

$$A(z) = 1 + \sum_{i=1}^m a_i z^{-i} \quad (9)$$

(2)- This LP filter  $A(z)$  is used to compute the desired perceptual filter

$$W(z) = \frac{A(\frac{z}{\alpha_1})}{A(\frac{z}{\alpha_2})} \quad (10)$$

where  $0 < \alpha_1 < \alpha_2 \leq 1$  are the perceptual factors.

(3)- In order to avoid filter instability related to the autoregressive moving-average (ARMA) structure, we approximate  $W(z)$  by a linear phase filter  $\hat{W}_n$  with  $m$  coefficients. Before using the signal  $\bar{y}_n$  as the input signal of the adaptive post-filter  $H_n^2$ , a copy of the linear phase filter  $\hat{W}_n$  is applied to  $\bar{y}_n$  as shown in Figure 5. For this, it is necessary to introduce a delay of  $D$  samples in the reference branch, with

$$D = \begin{cases} \frac{m}{2} & \text{if } m \text{ is even} \\ \frac{m-1}{2} & \text{if } m \text{ is odd} \end{cases} \quad (11)$$

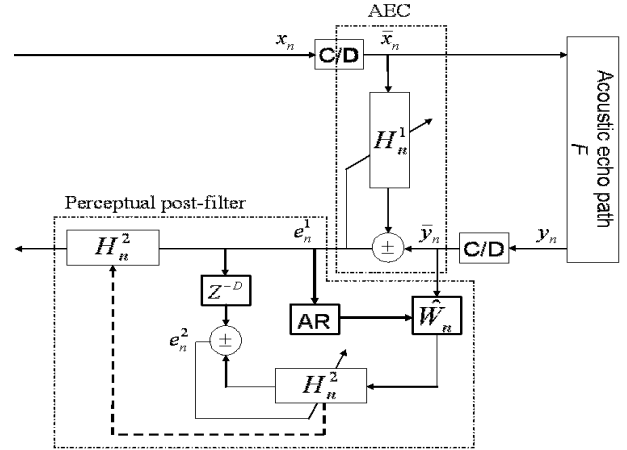


Figure 5: Centralized combined AEC/perceptual post-filter.

In the following,  $m$  equals to 10, the choice of the perceptual factors  $\alpha_1$  and  $\alpha_2$  makes it possible to control the perceptual weighting function.

Indeed, if  $\alpha_1 = \alpha_2 \simeq 1$ , the perceptual filter  $W(z)$  only delays the signal of  $D$  samples: in this case,  $W(z)$  has no effect and the system corresponds to the previous case of AEC/post-filter structure (Section 3).

On the other hand if the difference between  $\alpha_1$  and  $\alpha_2$  increases,  $W(z)$  is able to give more importance to high power frequency components of the echo signal.

##### 4.2 Comparative performance

###### 4.2.1 $H_n^2$ improvement in presence of $W(z)$

To well emphasize the usefulness of  $W(z)$ , we compare the behavior of  $H_n^2$  in two cases: on the one hand, when  $H_n^2$  operates without  $W(z)$  ( $\alpha_1 = \alpha_2 = 0.9$ ) and on the other hand when  $H_n^2$  operates with  $W(z)$  ( $\alpha_1 = 0.1, \alpha_2 = 0.9$ ).

In both cases, the input signal is a simulated autoregressive (AR) order 10 stationary process. The AR parameters are derived from LPC analysis of the French vowel ‘‘i’’. The amplitude and the spectrum of this ‘‘i-like’’ signal are shown in Figure 6.

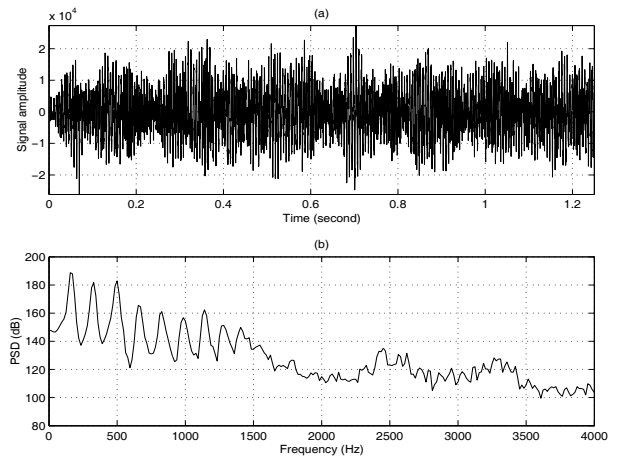


Figure 6: Artificial input signal: (a) Amplitude, (b) PSD.

Notice that in this experience the echo path ( $F$ , codecs) and the AEC/post-filter ( $H_n^1$ ,  $H_n^2$ ) parameters are the same than in Section 3.2. Figure 7(a) shows the power spectral density (PSD) of the echo signal  $\bar{y}_n$ . Figure 7(b) represents the frequency response of the post-filter  $H_n^2$  in two cases: the solid line corresponds to the classical AEC/post-filter and the dashed line to the AEC/perceptual post-filter. From Figure 7, it is clearly observed that the introduction of the perceptual filter  $W(z)$  in the post-filter computation mainly attenuates the frequency harmonic components of the echo signal. An extra attenuation of 10 dB can be observed on the overall spectrum of  $e_n^1$ .

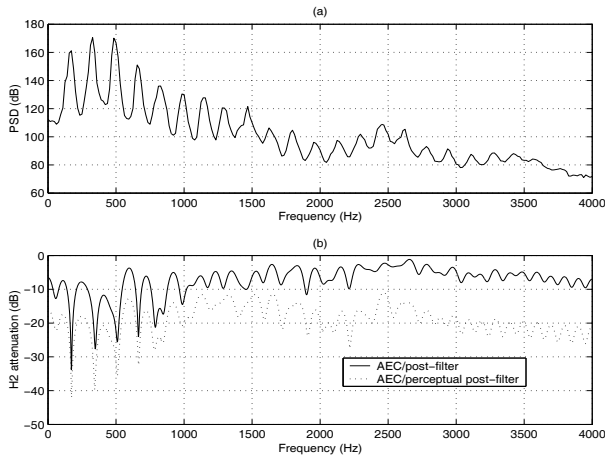


Figure 7: (a) PSD of echo signal, (b) Amplitude response of  $H_n^2$  with and without perceptual filter  $W(z)$ .

#### 4.2.2 Simulation with real speech signal

In the previous experiment, an artificially simulated signal is used to illustrate the behavior of the proposed perceptual post-filter. In this part, the performance of the AEC/perceptual post-filter (Section 4) and the AEC/post-filter (Section 3) are compared in the case of real speech echo signals. Comparisons are made by observing in Figure 8 the ERLE time evolution for the two cases.

It can be shown that the AEC/perceptual post-filter (solid line) increases the ERLE in average of 5 to 10 dB compared to the classical AEC/post-filter technique (dashed line). These objective results are corroborated by informal listening tests where extra echo reduction is highly subjectively reduced. These results demonstrate the practical interest of the proposed centralized AEC system.

## 5. CONCLUSION

This paper has presented a centralized AEC/post-filter based on perceptual properties to reduce the coding noise. Simulations on both artificial and speech signals show that the proposed AEC system is very promising for acoustic echo suppression when low-bit-rate speech codecs are integrated in the echo path and presents robust performance. In this paper, first results are given in term of echo reduction in single talk operations (i.e. when only the far-end speech

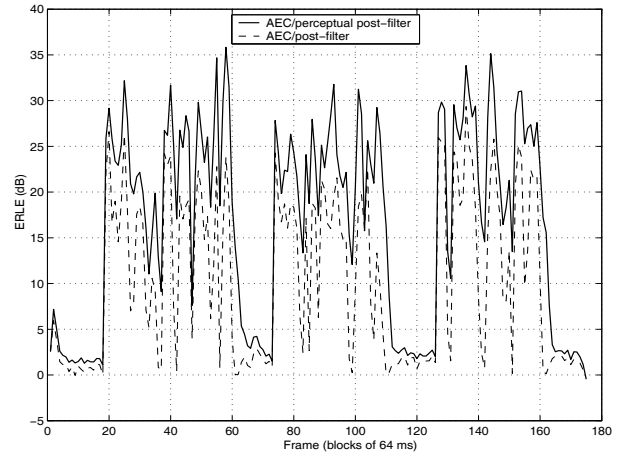


Figure 8: ERLE for AEC/post-filter with and without perceptual filter.

is active). Future work will concern the behavior of the proposed approach for double talk situations (i.e. when the far-end and the near-end speech are active simultaneously). In this case, the AEC scheme is more complicated and need an echo control strategy to switch to simple talk to double talk context and vice versa.

## REFERENCES

- [1] H. Gnaba, M. Turki, M. Jaidane, and P. Scalart, "Introduction of the CELP structure of the GSM coder in the acoustic echo canceller for the GSM network," in Proc. EUROSPEECH, pp. 1389-1392, September 2003.
- [2] H. Gnaba, P. Scalart, M. Turki, and M. Jaidane, "Combined acoustic echo canceller for the GSM network," in Proc. EUSIPCO, vol. 1, pp. 547-550, September 2002.
- [3] X. Lu, and B. Champagne, "Pitch analysis-based acoustic echo cancellation over a non-linear channel," in Proc. EUSIPCO, vol. 1, pp. 159-162, September 2002.
- [4] X. Lu, and B. Champagne, "Acoustic echo cancellation over a non-linear channel," in Proc. IWAENC, pp. 139-142, September 2001.
- [5] A. Stenger, L. Trautmann, and R. Rabenstein, "Non-linear acoustic echo canceller with 2nd order adaptive Volterra filters," in Proc. ICASSP, vol. 2, pp. 877-880, March 1999.
- [6] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," Signal Process., vol. 64, pp. 21-32, January 1998.