COMBINED ACOUSTIC ECHO CANCELLER FOR THE GSM NETWORK

H.Gnaba(2), *P. Scalart*(1), *M.Turki -Hadj Alouane* (2), *M.Jaidane-Saidane*(2)

(1) France Télécom R&D DIH/DIPS, Technopole Anticipa Lannion,
(2) Laboratoire des Systèmes de Communications – Ecole Nationale d'Ingénieurs de Tunis
ENIT BP 37, 1002 Tunis, Tunisia

`helagnaba@yahoo.fr– pascal.scalart@rd.francetelecom.fr`

ABSTRACT

When the implementation of an acoustic echo canceler is implemented in the Mobile Switching Center of a GSM network, it is necessary to take into account the non linearities introduced by the speech coder/decoder. A classical linear adaptive filter is not sufficient. The performances of a classical AEC are improved by noise reduction adaptive post-filter. The post-filter introduced, take into account implicitely the non linearity introduced by the GSM speech codecs cascaded with the linear echo path. It can reduce the quantization noise related to the coding/decoding operation. The proposed combined system presents better performances than that of the classic (linear) solution.

1. INTRODUCTION

The acoustic echo is an important cause of disturbance in the network GSM system for hands-free terminals. It occurs when a part of the speaking voice signal energy is reflected back to him or to her.

In order to cancel this echo to all subscribers, the operator can implement an Acoustic Echo Canceller (AEC) in the Mobile Switching Center (MSC) to ameliorate the audio quality before transmission.

The major difficulty of a centralized acoustic echo cancellation is the related to the non linearities introduced by the speech coders/decoders. Hence, the implementation of a classical linear AEC in the MSC, is insufficient and doesn't garantee a sufficient echo reduction.

To improve the performance of the classical AEC, we propose a solution based on an adaptive reduction of the residual echo. Two adaptive filters are coupled : the first one, operates to identify roughly the echo path, and the second adaptive filter, applied to the residual echo, operates as a noise reduction system. Note that, these two filters are adaptive, they work in the temporal domain, so the delay introduced is less than when the noise reduction filter works in the frequency domain as is in the case of classical noise reduction systems.

Similar systems are presented in the literature [1] [2], in order to reduce the residual echo, related to long impulse response of the echo path.

In this paper, we not only take into account the later phenomenon but also we reduce the quantization noise generated by the coding/decoding operation.

The rest of the paper is organized as follows : in section 2, we present the problematic of the AEC in GSM network. We will enhance the specific non linear aspect of the echo to cancel, and the insufficiency of a classical AEC in a such context.

In section 3, we present the proposed combined AEC, and in particular we will show theoretically the importance of the post-filter in the centralized context.

In section 4, we study by simulations the performances of the combined AEC when it is implemented in the MSC of the GSM network.

2. PROBLEMATIC OF THE ACOUSTIC ECHO CANCELLATION IN THE GSM NETWORK

There is an evident need for an acoustic echo canceller to overcome the echo problem. The classical solution is to implement an AEC in the mobile receiver [3].

However, the implementation of the AEC in the GSM network allows the operators to enhance the audio quality for speakerphones as well as cell phones [4].

The first obvious way is to implement an AEC in the MSC, hence the echo path length is increased according to the delays inherent to the GSM system. In this case, the adaptive filter has to be complex in order to handle the long echo delay (about 180 ms).

Figure 1 shows the principal structure of an acoustic echo canceller implemented in the MSC of the GSM network.

The particularity of this structure is the presence of the

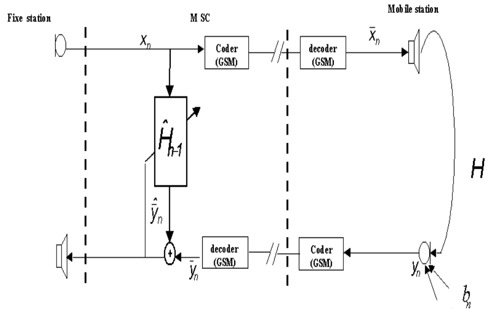

Figure 1: Acoustic echo cancellation in the GSM network

codecs in tandem. The operations of coding/decoding applied to the near end and the far end signals introduce a particular kind of non linearity. Indeed, the GSM coder/decoder is an analysis by perceptual synthesis coder based on CELP method.

To show this aspect, we plot in figure 2 the coded/decoded far-end signal ($\bar{x}_n$) as function of the original far end signal (x_n) for 20000 samples of speech (sampled at 8 KHz).

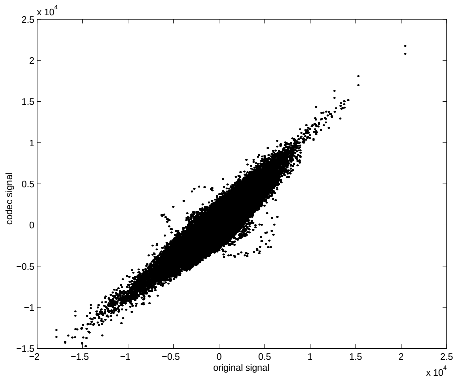

Figure 2: Non linear behavior of the GSM codecs

An adaptive linear filter $\hat{H}_n$ can not handle the non linearity introduced by the speech coding/decoding.
To show this problematic of AEC in the GSM network, we implement and simulate a linear AEC in two different cases:

In the first case, we don't consider the presence of the codecs so it is a classical AEC scheme. In the second case, the codecs are present as in figure 1. For both cases, we have considered that the far end speech signal is x_n, the length of the impulse response of the echo path (H^{opt}) is 300. The time evolution of the adaptive filter $\hat{H}_n$ is controlled by a second order Affine Projection Algorithm (APA 2).
The equations of the adaptive AEC for the two cases are described by the following equations.

$$e_n = \tilde{y}_n - \hat{H}_{n-1}^T \cdot \tilde{X}_n \quad (1)$$

$$V_n = \tilde{X}_n - \frac{\tilde{X}_n^T \cdot \tilde{X}_{n-1}}{\tilde{X}_{n-1}^T \cdot \tilde{X}_{n-1}} \cdot \tilde{X}_{n-1} \quad (2)$$

$$\hat{H}_n = \hat{H}_{n-1} + \mu \frac{e_n}{V_n^T \cdot V_n} \cdot V_n \quad (3)$$

where,

- e_n is the residual echo signal.
- $\tilde{y}_n$, is used to represent two signals : microphone signal y_n in the case without codecs, and the coded/decoded microphone $\bar{y}_n$ in the case with codecs.
- $\tilde{X}_n = [\tilde{x}_n, ..., \tilde{x}_{n-L}]^T$ is the vector of the last 300 samples of the far end signal x_n in the case without codecs, and the coded/decoded far end signal $\bar{x}_n$ in the other case.
- $\hat{H}_n$ represents the vector of the adaptative filter. The length of this filter is $L = 300$ (equal to the length $\tilde{L}$ of the impulse response of the echo path).

For the measure of the AEC performances, we compute the ERLE, (Echo Return Loss Enhancement : $ERLE = 10 \times \log \frac{E[\tilde{y}_n^2]}{E[e_n^2]}$). Simulation results are presented in figure 3. The ERLE in (-) corresponds to the case without codecs, the one in (+) corresponds to case with codecs.

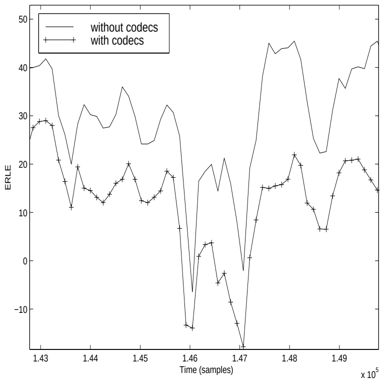

Figure 3: The ERLE for two cases

These results show that the use of the traditional approach based on linear adaptive filtering on the network side is not sufficient in the presence of the codecs. Due to quantization noise, a 35 dB degradation could be reach.

3. THE COMBINED ACOUSTIC ECHO CANCELLER IN THE GSM NETWORK

3.1. Justification of a post-filter in the GSM Network

In our first approach, we can consider that the operation of coding/decoding is a simple addition of a quantization noise. So we can suppose that the coded/decoded far end signal $\bar{x}_n$ is :

$$\bar{x}_n = x_n + b_n^x \quad (4)$$

also the coded/decoded microphone signal $\bar{y}_n$ is :

$$\bar{y}_n = y_n + b_n^y \quad (5)$$

Where b_n^x and b_n^y are respectively quantization noises.
The microphone signal y_n may be made up of the coded/decoded echo $\bar{u}_n$, as well as the near end speech s_n (figure 1). So the expression of the microphone signal y_n is :

$$y_n = s_n + \bar{u}_n \quad (6)$$

where $\bar{u}_n = H^{optT}\bar{X}_n = H^{optT}(X_n + B_n^x)$
We can consider that the total quantization noise of the coding/decoding operation is $b_n^c = H^{optT}B_n^x + b_n^y$. So the equation (5) becomes

$$\bar{y}_n = s_n + b_n^c + u_n \quad (7)$$

The estimated echo $\hat{u}_n$ is subtracted from $\bar{y}_n$ forming the echo compensated signal

$$e_n = s_n + b_n^c + r_n \tag{8}$$

where $r_n = u_n - \hat{u}_n$ is the residual echo. We can remark that the signal e_n is a mixture of desired signal s_n and undesired signals, b_n^c and r_n. These undesired noise signals should be removed.
The idea is to use, with the AEC, a noise reduction system able to recover the signal s_n from e_n.

3.2. Presentation of the combined AEC

Most of reduce noise methods, use Voice Activity Detector and are based on spectral subtraction techniques. Such methods lead to considerable delays, and they require a priori knowledge of the near end speech s_n considered as a reference of the system.
In this paper we use an original adaptive noise reduction filter that will operate on the residual echo e_n given by the adaptive AEC filter H_n. The figure 4 presents the structure of the proposed combined AEC. The post adaptive filter H_n^2 operates in order to extract the speech signal s_n from $\bar{y}_n$ that is a very noisy version of s_n.

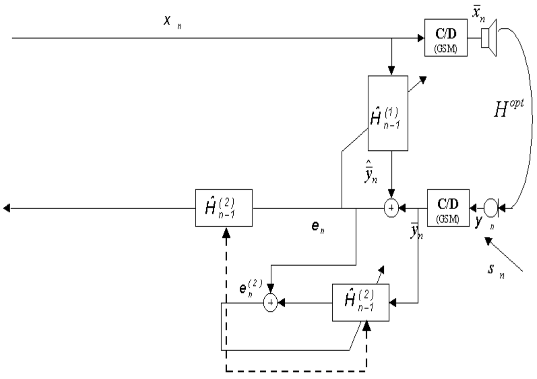

Figure 4: The combined system structure

The main idea is that, even the reference of the filter H_n^2 is e_n and not s_n, e_n is the most similar to s_n. The filtering operation applied to e_n allow us to extract s_n by a considerable attenuation of the residual echo r_n and the quantization noise b_n^c.
What's more, due to the use of this adaptive post-filter $H_n^{(2)}$, the primary filter may has a reduced length (L) compared to that of the echo path impulse response ($\tilde{L}$). Hence, the combined AEC is characterized by a reduced complexity and than a reduced delay compared to a classical AEC. Such property, lets the implementation of the combined very useful in the GSM network.

3.3. Optimal Wiener filters of the combined system

The input of the second filter H_n^2 is the coded/decoded microphone signal $\bar{y}_n$ and its reference is the signal e_n. So the frequency response of the optimal Wiener filter corresponding to $H_n^{(2)}$, is given by :

$$H^{(2)}(f) = \frac{\gamma_{e\bar{y}}(f)}{\gamma_{\bar{y}\bar{y}}(f)} \tag{9}$$

where $\gamma_{\bar{y}\bar{y}}(f)$ is the Power Spectral Density of the signal $\bar{y}_n$ and $\gamma_{e\bar{y}}(f)$ is the Cross Spectral Density of the signals e_n and $\bar{y}_n$.
Refering to the equations (7) and (8) and due to the independance of the signals, the equation (9) becomes :

$$H^{(2)}(f) = \frac{\gamma_{ss}(f) + \gamma_{b^cb^c}(f) + \gamma_{ur}(f)}{\gamma_{ss}(f) + \gamma_{b^cb^c}(f) + \gamma_{uu}(f)} \tag{10}$$

From the noise reduction point of view, the optimal noise reduction filter, has as input signal, the signal e_n and as a reference, the signal s_n. The frequency response of the corresponding optimal filter, is then given by,

$$H^{(2)}_{opt}(f) = \frac{\gamma_{se}(f)}{\gamma_{ee}(f)} \tag{11}$$

In all the following calcul, we have supposed the short term stationarity and the mutual independence of the signals s_n, b_n^c and u_n. Refering to the equation (8) :

$$\gamma_{ee}(f) = \gamma_{(s+b^c+r)(s+b^c+r)}(f) = \gamma_{ss}(f) + \gamma_{b^cb^c}(f) + \gamma_{rr}(f)$$

$$\gamma_{se}(f) = \gamma_{s(s+b^c+r)}(f) = \gamma_{ss}(f)$$

So the equation (11) will be :

$$H^{(2)}_{opt}(f) = \frac{\gamma_{ss}(f)}{\gamma_{ss}(f) + \gamma_{b^cb^c}(f) + \gamma_{rr}(f)} \tag{12}$$

To evaluate performances of the combined system we compare $|H^{(2)}(f)|$ and $|H^{(2)}_{opt}(f)|$ which corresponds to the best situation.
Simulations are made in the following context : presence of double speech, the far-end signal x_n is a white noise, s_n is a speech signal, $b_n^c = 0$, and we consider the general case where the length of the AEC (300) is smaller than the length of the impulse response echo path (4000).

In figure 5 the frequency responses of $|H^{(2)}(f)|$ and

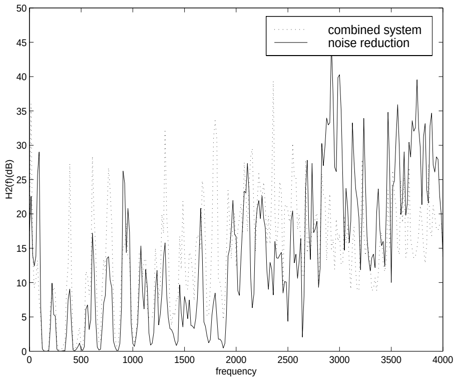

Figure 5: Frequency responses of the optimal noise reduction filter and the post-filter of the combined system

$|H_{opt}^{(2)}(f)|$ (en dB) are presented.
It is important to note the similar behaviour of the optimal post-filter in the proposed combined system and the optimal filter in the context of the noise reduction, when the reference signal is disponible.

4. PERFORMANCES ANALYSIS OF THE COMBINED AEC IN THE NETWORK GSM

The combined AEC is applied in the centralized context, where the two GSM speech codecs are present. To evaluate the post-filter performances according to those of the AEC without post-filtering, an auto-regressive input signal $x_n = ax_{n-1} + g_n$ is considered, where g_n is an additive white gaussian noise. As for speech signal, great input autocorrelation is ensured by $a = 0.9$. The impulse response echo path of the visioconference room is of length 4000. The filter $H_n^{(1)}$ is adapted typically according to the second order Affine Projection Algorithm (APA2). The adaptive filter $H_n^{(2)}$ updated by the Normalized LMS algorithm, is designed to minimize the mean square error $e_n^{(2)}$: the time evolution of $H_n^{(2)}$ is described by the following equations :

$$e_n^{(2)} = e_n - \hat{H}_{n-1}^{(2)T}\bar{Y}_n \qquad (13)$$

$$\hat{H}_n^{(2)} = \hat{H}_{n-1}^{(2)} + \mu\frac{e_n^{(2)}}{\bar{Y}_n^T \cdot \bar{Y}_n}\bar{Y}_n \qquad (14)$$

where $\bar{Y}_n$ is the vector of the last 80 samples of the coded/decoded near end signal $\bar{y}_n$. The length of the post filter is fixed to 80. The ERLE variations versus the $H_n^{(1)}$ length are presented in figure 6, for two cases: H_n^1 operates without the post filter H_n^2 and with it.

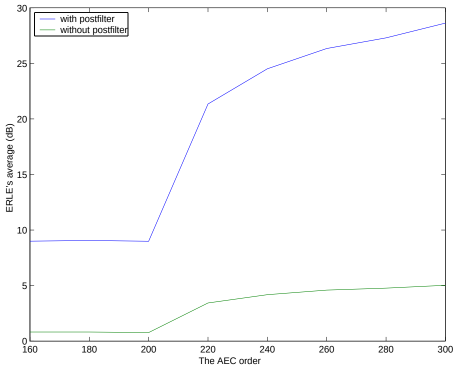

Figure 6: Reduction complexity with the combined system

First, we remark that the combined system presents better performances than the AEC without post-filtering. However, good performances of the combined system are ensured when the length of the filter H_n^1 reachs a minimum. Indeed, the ERLE's improvement is as important as the first filter is longer to take into account the first important reflections of the echo part. As shown (figure 6), 25 dB improvement is reached, when the length of $H_n^{(1)}$ filter exceeds 200 that is not really as important compared to long impulse response to identify ($\tilde{L} = 4000$).
To show the performance of the combined system in a realistic environmental noise GSM network, we have considered the presence of a local noise b_n, and we have used a real speech signal as the input signal. Hence in figure (7), we have plotted the ERLE's average as a function of the Signal to Noise Ratio ($SNR = 10\log\frac{(\sigma y)^2}{(\sigma b)^2}$).
The result shown in figure (7) is foreseeable, because the second filter is used to reduce the noise.

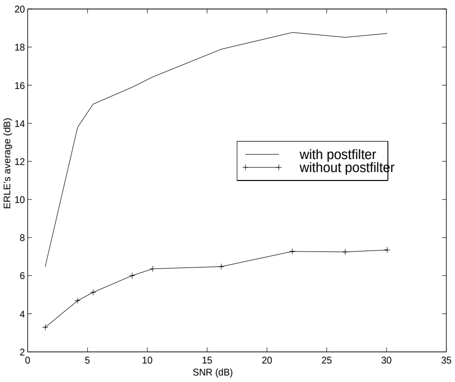

Figure 7: Performances of the combined AEC versus the SNR

5. CONCLUSION

When the implementation of an acoustic echo canceler is implemented in the Mobile Switching Center of a GSM network, it is necessary to take into account the non linearities introduced by the speech coder/decoder. A classical linear adaptive filter is not sufficient. The performances of a classical AEC are improved by noise reduction adaptive post-filter. Note that the adaptation of the post-filter is controlled by a constant step size. Such thing, causes some continuous adaptation problems and perturbations in transitions periods. A variable step size that can resolve these problems.

REFERENCES

[1] R. Martin, J. Altenhoner, “A coupled adaptive filters for acoustic echo control and noise reduction”, ICASSP 95, pp. 3043-3046, Detroit, USA.

[2] S. Gustafsson, R.Martin, P. Vary, ” Combined acoustic echo control and noise reduction for hands-free telephony ”, EURASIP, Signal processing 64, P 21-32, 1998.

[3] A. Gilloire, J.F.Zurcher, ”Achieving the control of the acoustic echo in audioterminals”, Elsiever Science Publishers, 1988, pp 491-494.

[4] D.J.Jones, S.D. Waston, K.G.Evans, ”A Network Speech Echo Canceller with Confort Noise” Eurospeech 1997.