# INCREASING THE ROBUSTNESS OF CELP-BASED CODERS BY CONSTRAINED OPTIMIZATION

*Mohamed Chibani, Philippe Gournay and Roch Lefebvre*

University of Sherbrooke, Sherbrooke, Quebec, Canada
Mohamed.Chibani@USherbrooke.ca

## ABSTRACT

The adaptive codebook used in CELP-like speech coders is extremely effective on voiced signals. Unfortunately, it is also the main source of error propagation at the decoder when a frame is lost. In this paper, we study several ways of limiting the energy contribution of the adaptive codebook to the synthesized speech signal. We show that a constrained search of the adaptive and innovative codebooks significantly improves the recovery time of the decoder after a lost frame, at the cost of only minor quality degradation in clear channel. When applied to a standard codec such as the AMR-WB, this constraint only affects the encoder, and the modified codec remains fully interoperable with the standard codec.

## 1. INTRODUCTION

Modern speech coders are generally based on the CELP paradigm: the decoded speech signal is obtained by passing an excitation signal through a synthesis filter (a short-term linear predictor). The excitation signal is the sum of the contributions of two codebooks: the adaptive codebook ACB (a long-term predictor) and the innovative codebook ICB (the algebraic code in the specific case of the AMR-WB codec [1]). The parameters for those two codebooks are determined using the analysis-by-synthesis principle: they are chosen such as to minimize the energy (in a perceptually-weighted domain) of the error signal.

The adaptive codebook is represented by a delay and gain. During stationary voiced segments, this gain is generally high, and the ACB contribution provides most of the energy of the synthesis signal. Unfortunately, since the ACB introduces inter-frame dependency, it also becomes the main source of error propagation at the decoder when a frame is lost.

Several concealment techniques have already been proposed in order to mitigate the effect of lost frames, especially in the context of voice over packet networks [2]. However, very little research has been dedicated towards improving the recovery of the decoder after a lost frame [3, 4]. In this paper, we propose to bring in some constraints into the codebooks search procedure, so that the parameters selected by the encoder do not produce too much error propagation in the decoder after a lost frame. Those constraints aim at limiting the energy contribution of the adaptive codebook to the synthesized speech. At an equivalent bit rate, as one may expect, the constrained search leads to a small quality degradation in the error-free transmission scenario. We will show however that the gain in terms of recovery speed after a frame loss clearly offsets the quality degradation in more realistic scenarios (1 to 3% of frame losses are typical in cellular telephony, and 10 to 20% are often considered for IP networks).

The paper is organized as follows. In section 2, we briefly present the standard AMR-WB codebooks search procedure, focusing on the elements where a constraint can be applied. Several constrained search procedures are then proposed in section 3. Finally, a performance comparison between the standard and the constrained AMR-WB coder is presented in section 4.

## 2. EXCITATION SEARCH IN THE AMR-WB CODEC

The search for the ACB and ICB excitations is done sequentially. First, the delay and the gain of the ACB are determined, and then the shape (or pulse positions) of the ICB is found by analysis-by-synthesis. The search for the pulse positions of the ICB excitation is performed on the updated target (i.e. target with removed ACB contribution). The gains of the ACB and ICB are jointly quantized in closed loop only after both excitation shapes have been found. We recall that the AMR-WB uses a 20 ms frame, and that most parameters, including the ACB and ICB parameters, are estimated on a sub-frame basis (a frame being divided in 4 sub-frames of 5 ms each) [1].

### 2.1 ACB parameters search

The ACB parameters are those that minimize the quadratic error in the perceptual domain:

$$D_1 = \sum_n \left(x_1(n) - g_0 y_\alpha(n)\right)^2 , \qquad (1)$$

where $x_1(n)$ is the target signal, $g_0$ is the ACB gain, and $y_\alpha(n)$ is the past filtered excitation at the delay $\alpha$ given by:

$$y_\alpha(n) = u(n-\alpha) * h_w(n) , \qquad (2)$$

The authors wish to thank Claude Laflamme for suggesting this research.

with $h_w(n)$ being the impulse response of the weighted synthesis filter, and $u(n-\alpha)$ the past excitation signal.

The optimal delay is the one that maximises:

$$C_\alpha = \frac{\sum_n x_1(n) y_\alpha(n)}{\sqrt{\sum_n y_\alpha^2(n)}}, \tag{3}$$

and the corresponding gain, which is found by setting the derivative of $D_1$ with respect to $g_0$ to 0, is:

$$g_0 = \frac{\sum_n x_1(n) y_\alpha(n)}{\sum_n y_\alpha^2(n)}. \tag{4}$$

Once the ACB parameters are found, the ACB contribution is subtracted from the target signal:

$$x_2(n) = x_1(n) - g_0 y_\alpha(n), \tag{5}$$

and $x_2(n)$ becomes the target signal for the innovative codebook search.

### 2.2 Joint quantization of the gains

The ACB and ICB gains are quantized jointly after applying a 4th order MA predictor to the ICB gain. Their quantized value $\hat{g}_p$ and $\hat{g}_c$ is found by searching through a table for the entry that minimizes the quadratic error in the perceptual domain between the target and the synthesized speech:

$$D_2 = \sum_n \left(x_1(n) - \hat{g}_p y_\alpha(n) - \hat{g}_c y_c(n)\right)^2, \tag{6}$$

where $y_\alpha(n)$ and $y_c(n)$ are the ACB and ICB filtered excitations.
For most modes (12.65 kb/s and higher), a 7-bit quantization table is used. Half of the table (i.e. 64 entries) is explored around the ACB gain value calculated in (4), and the gain pair that minimizes the quadratic error $D_2$ is selected.

## 3. CONSTRAINED SEARCH ALGORITHM

The key idea behind our work is to decrease the energy contribution of the ACB to the synthesized speech signal in order to reduce inter-frame dependency. In other words, we will force the total excitation to contain more contribution from the ICB during subframes where the ACB contribution is otherwise dominating. The ICB being memoryless, the decoder should then take less time to recover after a lost frame.
A constraint can be applied at two stages: when calculating the target signal for the ICB search, and when quantizing the ACB and ICB gains.

### 3.1. Target modification for the innovative codebook search

At this stage of the encoder, we use the following ratio as a measure of the ACB contribution:

$$R_1 = \frac{E_p}{E_x} = \frac{g_0^2 \sum_n y_\alpha^2(n)}{\sum_n x_1^2(n)}, \tag{7}$$

where $E_p$ is the energy of the ACB contribution and $E_x$ the energy of the target, both calculated in the perceptual domain.

The constraint is achieved by comparing the energy ratio $R_1$ to a pre-defined threshold $R_{th1}$. If $R_1$ is below $R_{th1}$, the ACB gain is kept as in equation (4). Otherwise, it is modified as follows:

$$g_0 = \sqrt{R_{th1} \frac{\sum_n x_1^2(n)}{\sum_n y_\alpha^2(n)}}. \tag{8}$$

It is important to note that the value of the ACB gain in itself is not limited. Actually, it's the energy coming from the ACB that is limited. We will show in Section 4 that, for suitable values of $R_{th1}$, this constraint affects only minimally the overall quality of the coded speech in clear channel. However, it improves significantly the recovery after a missing frame.
The constrained gain given in equation (8) is such that a fraction of the ACB contribution, proportional to ($1-R_{th1}$), will be left in the target signal $x_2(n)$. The ratio of the ACB contribution to the synthesized speech is however not guaranteed to be under any predefined threshold for three reasons. First, because the ICB contains only sparse vectors and is therefore not able to properly reproduce the pitch excitation. Then, because the ratio $R_1$ is calculated on the target signal, not on the synthesized speech. This point will be discussed in section 3.2. Finally, because the quantized ACB gain is in general different from the value calculated in (8).
At this stage, it might be legitimate to question about the optimality of the ACB delay determined by (3). The constrained gain is obviously no longer optimal in the sense that it does not necessarily minimize the quadratic error $D_1$. However, one can easily prove that the ACB delay is in fact still optimal.

### 3.2. Constrained joint quantization of the gains

The contribution of the ACB to the synthesized speech can also be controlled in the quantization stage by forcing the quantized ACB and ICB gains to satisfy (or almost satisfy) a constraint.
At this stage, both the ACB and ICB excitation shapes are known and we can precisely evaluate the contribution of each codebook to the synthesized speech. We can therefore use the following constraint:

$$R_2 = \frac{E_p}{E_T} = \frac{\hat{g}_p^2 \sum_n y_\alpha^2(n)}{\sum_n \left(\hat{g}_p y_\alpha(n) + \hat{g}_c y_c(n)\right)^2} \le R_{th2}, \tag{9}$$

where $E_p$ is the energy of the ACB contribution and $E_T$ the energy of the synthesized speech, both calculated in the perceptual domain. Note that the new energy threshold $R_{th2}$ is not necessarily equal to $R_{th1}$.
The most straightforward approach to control the contribution of the ACB during the joint quantization of the gains consists in rejecting systematically all the entries in the quantization table that do not satisfy constraint (9). However, an experiment described in section 4.2 showed that strictly respecting the constraint at this stage produces unacceptable artifacts, especially for low values of $R_{th2}$.
There are several possibilities to relax the constraint in the gain quantization in order to minimize the impact on quality. In a first approach, the quantization table is searched around the result of the joint constrained gains optimization (see appendix). The search range is also limited to 12 or 16 entries (instead of 64 in the standard).

In another approach, the minimization criterion for the joint quantization of the gains is modified by adding a penalty term that increases the value of the error when the constraint is not respected. The modified criterion is given by:

$$D_3 = \beta \sum_n \left(x_1(n) - g_p y_\alpha(n) - g_c y_c(n)\right)^2 \\ + (1-\beta)\left[ g_p^2 \sum_n y_\alpha^2(n) - R_{th2} \sum_n \left(g_p y_\alpha(n) + g_c y_c(n)\right)^2 \right]. \qquad (10)$$

The coefficient $\beta$ is used to set the balance between quadratic error minimization and ACB contribution reduction.

## 4. EXPERIMENTAL RESULTS

In this section, we present some experimental results for the methods for ACB contribution control described in Section 3.

### 4.1 Target modification

The most important requirement for the constrained encoder is that it must not degrade the decoded speech quality significantly in clear channel. In order to choose the appropriate value for the threshold $R_{th1}$, we calculated the segmental signal-to-noise ratio (SEG-SNR) of the local synthesis (i.e. the synthesized speech at the encoder) for different values of $R_{th1}$. All active speech frames (including unvoiced ones) are taken into account. This is summarized in figure 1. We remind that $R_{th1}$ can be interpreted as the amount of the ACB contribution to be removed from the target signal when the ACB contribution is greater than $R_{th1}$. A value of 1 for $R_{th1}$ means that all the ACB contribution has been removed from the target, which corresponds to the standard update of the target signal.

We see in figure 1 that the degradation introduced by the modification is very small for $R_{th1}$ values greater than 0.6. The choice of the threshold value to be used in practice results from a tradeoff between the increased robustness and speech quality in clear channel. Informal listening tests showed that when $R_{th1}$ is around 0.55, the robustness is significantly improved whilst the overall quality of the decoded speech remains quite acceptable in clear channel.

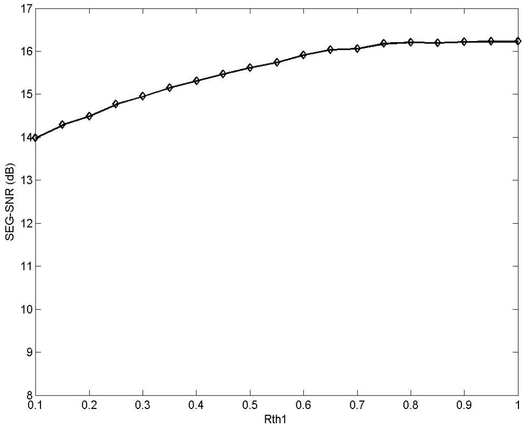

**Figure 1.** *SEG-SNR of the synthesized speech as a function of $R_{th1}$*

Figure 2 presents an example of a synthesized speech signal obtained with the standard and the constrained encoder. Recall that no modification has been made to the decoder. In particular, the standard concealment method described in [5] is used in both cases. The original signal is represented in (a). For comparison, the error signals obtained with the standard encoder and the constrained encoder are plotted in (d) and (e), respectively. Both error signals are computed with the original signal as a reference. It appears clearly that the reconvergence of the decoder is much faster with the constrained encoder than with the standard encoder. In that case, the decoder recovers almost completely at the end of the third frame after the lost frame when the encoder applies the constraint, while it takes as much as 11 frames when the encoder does not apply the constraint.

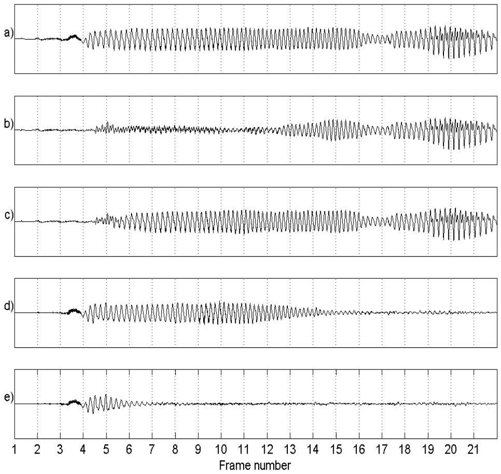

***Figure 2**. Comparison of recovery time after a frame erasure (4th frame lost): original speech signal in (a). Synthesized speech for the standard encoding in (b) and the constrained encoding in (c). Error signal for the standard encoding in (d) and modified encoding in (e).*

In some cases, we noted that the energy loss in the concealed frame is more important when the encoder applies the constraint than when it does not apply the constraint. This happens mainly when the frame erasure occurs during a stable voiced segment. In fact, this is due to the way the ACB gain is extrapolated during erased frames. The concealed ACB gain is an attenuated version of the past gain. But since the ACB gain sent by the constrained encoder is typically lower than the optimal one, the concealed ACB gain is generally too low for the ACB excitation to properly maintain the energy of voiced speech. Furthermore, the concealed ICB excitation is random (and generally with a low energy), and does not compensate for the lack of periodicity. This clearly shows the need for a concealment procedure that would take into account the constraint applied at the encoder.

Figure 3 represents the SNR of the decoded speech signal on the same speech segment. Curve C1 is for the standard encoding in clear channel. Curves C2 and C3 are for the standard and the constrained encoder, with an erasure at the third frame. We see again in this figure that the recovery is much faster with the constrained encoder than with the standard encoder.

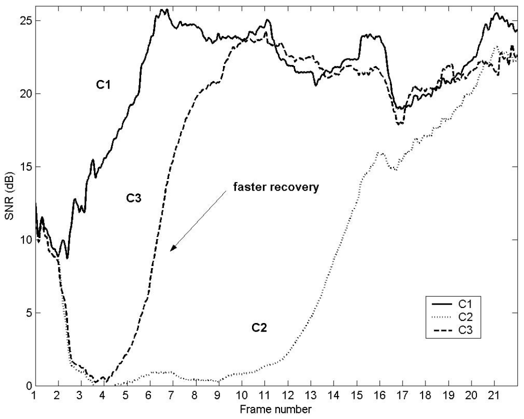

***Figure 3**. SNR comparison between the standard and constrained coders: standard encoding without frame erasure in (C1), standard encoding with frame erasure in (C2) and constrained encoding with frame erasure in (C3).*

### 4.2 Effect of the joint quantization of the gains

In a first experiment, we investigated the effect of using a gain quantization table that was originally not designed for the constrained encoder, in the coder with a constraint during the target computation and another during the gains quantization. We compared the optimal gain values obtained by the KKT method discussed in appendix, with the quantized gains obtained with the straightforward constrained quantization approach described in section 3.2. The results we obtained showed that the distortion due to using a non-optimal quantization table is negligible compared to the distortion introduced by applying the constraint. Therefore, there is no real need to modify the quantization table when a constraint is applied.

In a second experiment, we applied the second and third approaches of section 3.2 to constrain the gains quantization, and found that both approaches give similar results. In addition, when compared to the results obtained with the target modification only, applying another constraint during the gains quantization does not improve significantly the recovery time, even though we have in principle a better control over the ACB contribution. Hence, in the specific case of the AMR-WB, most of the gain in recovery time is achieved by modifying the target computation, as proposed in section 3.1.

## 5. CONCLUSION

We have shown that limiting the energy contribution of the adaptive codebook to the synthesized speech signal significantly improves the recovery time of a CELP decoder after a frame loss, at the cost of only minor quality degradation in the clear channel condition. Although a constraint can be introduced at various stages of the encoding process, we determined that modifying the target signal for the ICB was the simplest and most effective method. When applied to a standard codec such as the AMR-WB, the constrained search of the ACB and ICB codebooks does not require any additional bit rate or delay. It affects only the encoder, and the modified codec remains fully interoperable with the standard codec. Of course, an optimal solution would require adapting the concealment at the decoder to take into account the constraint applied at the encoder. Also, an efficient post-processing applied to the synthesized signal such as the one described in [6] would probably compensate for the small quality loss in clear channel.

## Appendix: Joint constrained optimization of the ACB and ICB gains

The aim of this appendix is to show that this constrained optimization problem has a solution, and that this solution can be expressed analytically. Only the method used to find the optimal gains is presented. Their exact analytical expressions are not given, as they are too long to fit within this paper.

The Karush-Kuhn-Tucker (KKT) conditions [7] are used to solve our constrained optimization problem. To conform to the notations in optimization theory, let us rewrite condition (9) as:

$$C(g_p, g_c) = E_p(g_p) - R_{th} E_T(g_p, g_c) \leq 0. \tag{11}$$

Using the quadratic error (4) and the constraint (8), we construct the Lagrangian function:

$$\begin{aligned} L(g_p, g_c, \lambda) = &\sum_n \left(x_1(n) - g_p y_\alpha(n) - g_c y_c(n)\right)^2 \\ &+ \lambda \left[ g_p^2 \sum_n y_\alpha^2(n) - R_{th2} \sum_n \left(g_p y_\alpha(n) + g_c y_c(n)\right)^2 \right], \end{aligned} \tag{12}$$

where $\lambda$ is the Lagrange multiplier.

The unconstrained optimal gains are obtained by solving $\partial D_2(g_p, g_c)/\partial g_p = 0$ and $\partial D_2(g_p, g_c)/\partial g_c = 0$. We then evaluate the contribution of the ACB using equation (9). If the constraint is respected, no further processing is needed. In that case the constraint is inactive. Otherwise, the solution for the constrained optimization of the ACB and ICB gains is given by solving $\partial L(g_p, g_c, \lambda)/\partial g_p = 0$, $\partial L(g_p, g_c, \lambda)/\partial g_c = 0$ and $C(g_p, g_c) = 0$.

## 6. REFERENCES

[1] B. Bessette, et al. "The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)". *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 8, pp. 620-636, Nov. 2002.

[2] B.W. Wah, et al. "A survey of error-concealment schemes for real-time audio and video transmission over the Internet", IEEE International Symposium on Multimedia Software Engineering, December 2000.

[3] P. Gournay, et al. "Improved packet loss recovery using late frames for prediction-based speech coders", In Proc. ICASSP-2003, pp. 108-111, April 6-10 2003.

[4] S.V. Andersen, et al. "ILBC - A linear predictive coder with robustness to packet losses," In Proc. 2002 IEEE Speech Coding Workshop, pp. 23-25, Tsukuba, JAPAN, 6-9 October 2002.

[5] ETSI 3GPP TS 26.191, "AMR Wideband Speech Codec; Error concealment of lost frames", V5.1.0, March 2002.

[6] M. Jelinek, et al. « Advances in source-controlled variable bit rate wideband speech coding ». Special Workshop in MAUI (SWIM): Lectures by masters in speech processing, Maui, Hawaii, January 12-14, 2004.

[7] M.S. Bazaraa et al. *Nonlinear programming, Theory and algorithms*. 2nd edition, John Wiley and Sons, Inc. 1993.