A FAST SEARCH METHOD FOR CELP SPEECH CODING

R.S. Maia, C.B. Ribeiro, S.L. Netto and F.G.V. Resende Jr.

Programa de Engenharia Elétrica/COPPE, DEL/EE
Universidade Federal do Rio de Janeiro
PO Box 68504, Rio de Janeiro, RJ, 21945-970, Brazil

ABSTRACT

A fast search method for code-excited linear prediction (CELP) vocoders is presented. Here the candidate vectors from the adaptive codebook are not individually filtered during the analysis-by-synthesis procedure. Experiments indicate that the computational complexity is reduced in about 85%, while the speech quality is essentially sustained.

1. INTRODUCTION

Today's Internet applications and the recent evolution of the telephony systems demand speech-coding techniques with a good trade-off between speech quality and bit rate. One of the currently most employed techniques for speech coding is the code-excited linear prediction (CELP) [1] technique. CELP coders tend to present high computational complexity for determining the best excitation signal of the linear-prediction (LP) model. While most accelerating methods act in the fixed codebook, in [2] an accelerating scheme was formulated for the adaptive codebook by changing the CELP synthesis model. The present work proposes a fast search method for the adaptive codebook without such modification. Results have indicated that the proposed method greatly reduces the overall computational complexity of a CELP coder, while maintaining the speech quality of the reconstructed signal.

2. CELP CODING

In CELP coders, the reconstructed speech is obtained by passing a excitation signal $x(n)$, composed from one adaptive and one fixed codebooks, through a LP synthesis filter $H(z) = 1/A(z)$. In practice, the CELP coder is implemented as shown in Fig. 1, where the weighting filter $W(z) = A(z)/A(z/\gamma)$ is passed to the left of the adder, turning $H(z)$ into the perceptual synthesis filter $H_w(z) = 1/A(z/\gamma)$. The response of this filter is split in two parts: the zero memory response $\hat{t}(n)$ and the zero input response $\hat{s}_0(n)$. The adaptive codebook, which substitutes the pitch filter in the first CELP versions, is composed of one single sequence with excitation samples of past instants, that is,

$$\{c_a(n)\} = \{x(-D_{max}), \ldots, x(-1)\}, \qquad (1)$$

where D_{max} is the maximum delay considered.

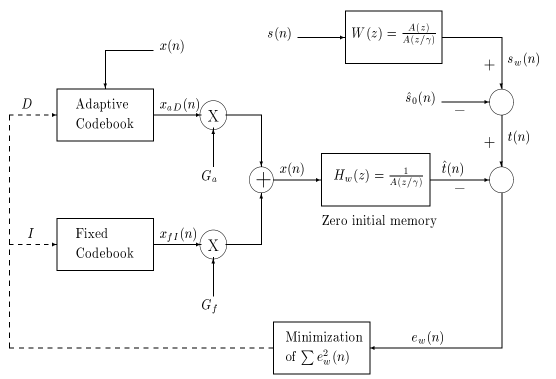

Figure 1: Block diagram of the practical implementation of a basic CELP coder.

This work was partially supported by CNPq and CAPES.

3. PROPOSED SEARCH METHOD

The fast search method proposed for the adaptive codebook does not filter each candidate sequence $x_{aD}(n)$ separately with $H_w(z)$, as in the standard analysis-by-synthesis procedure. Here, first the entire sequence $c_a(n)$ is filtered by $H_w(z)$, forming the auxiliary sequence $c_{aF}(n)$. This sequence is then segmented, sub-frame by sub-frame, and compared to the target signal, $t(n)$, to determine the best delay D_{op}. The desired best sequence $x_{aD_{op}}(n)$ is then obtained from the original $c_a(n)$, using this delay value. In that manner, after determining the best delay D_{op}, the proposed method becomes similar to the standard search procedure. In the new method, however, there are only two filtering operations with $H_w(z)$: of the entire sequence $c_a(n)$, at the beginning of the search procedure, and of the best sequence $x_{aD_{op}}(n)$, from $c_a(n)$, in order to calculate the gain G_a.

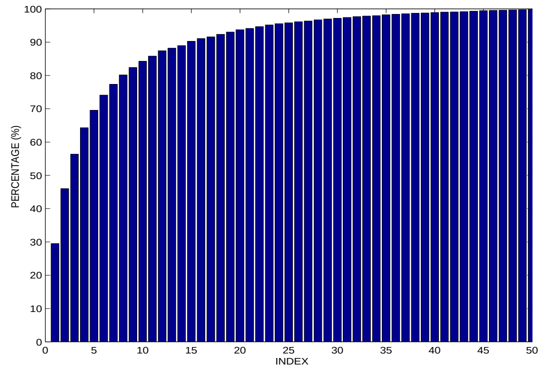

Figure 2: Cumulative percentages of the position of the optimum index D_{op}, given by the standard search, on the proposed method.

4. MODIFIED FAST SEARCH

Fig. 2 shows the cumulative percentages of the positions of D_{op} determined by the standard method when using the proposed method in a speech sample of 23 s. From this figure, it can be noted that at about 30% of the sub-frames, the index determined by the new method coincides with the one yielded by the standard procedure. Also, at about 90% of the sub-frames, both methods are equivalent, if we consider the best $K = 15$ indexes yielded by the proposed method. Based on these facts, the proposed method can be modified to yield better reconstructed speech signals, by selecting the K best adaptive codebook delays $D_1, \ldots, D_K$, when comparing the sub-frame partitions of $c_{aF}(n)$ with the target signal $t(n)$. We then perform the basic analysis-by-synthesis procedure on the original adaptive codebook generated by $c_a(n)$, considering solely these particular delay values. Therefore, in the new method, the parameter K controls the trade-off between overall computational complexity and resulting speech quality.

5. COMPUTER EXPERIMENTS

The proposed search method was implemented in a CELP algorithm with frame and sub-frame lengths of 20 ms and 5 ms, respectively, and perceptual factor $\gamma = 0.8$. An adaptive codebook was employed with delays ranging from 20 to 146 samples. The fixed codebook was of the type stochastic with clipping at 1.685, and the codebook gains G_a and G_f were not quantized. The LP coefficients were quantized as line spectral frequencies with 32 bits, using a differential uniform quantization. The experiments comprised a set of two sentences spoken by male speakers (M1 e M2) and two by female speakers (F1 e F2). The computational effort for the adaptive codebook search was evaluated by the total number of floating-point operations (flops).

Experiment 1: Table 1 includes objective speech quality measures, such as the perceptual segmental signal-to-noise ratio (PSSNR) (where the error signal is filtered by the LP model $1/A(z)$), the cepstral distance (CD) [3] and the Itakura distance (ID) [4], when coding the four sentences with and without the proposed search method, and with the method presented in [2].

From Table 1, it can be clearly inferred that the proposed search method produces speech quality similar to the standard method, while reducing the computational complexity in about 85%. Also, the proposed search method yields better speech quality than the method in [2], although the latter method reduces the

Table 1: Objective quality measures, in dB, and computational effort for the CELP coders.

Sentence	PSSNR	CD	ID	Mflops
CELP coder without fast search				
M1	17.26	3.01	1.12	67.02
M2	18.37	2.87	1.00	86.12
F1	18.61	2.90	1.04	63.77
F2	16.77	2.84	1.00	76.69
CELP coder with the proposed search				
M1	17.23	3.02	1.12	10.80
M2	18.11	2.86	1.00	13.83
F1	18.55	2.94	1.06	10.13
F2	16.56	2.96	1.08	12.31
CELP coder with fast search of [2]				
M1	16.12	3.27	1.31	8.46
M2	16.75	3.20	1.25	10.81
F1	17.68	3.51	1.51	7.9
F2	15.72	3.18	1.24	9.64

computational complexity about 3% further.

Experiment 2: In this experiment, we employ the modified fast search. Fig. 3 shows the effect of K on the PSSNR and the overall computational effort for the four sentences concatenated. The noisy nature of the plot is certainly due to the finite speech ensemble considered here. Nonetheless, increasing K tends to improve the overall speech quality. For instance, for $K = 3$, we get a PSSNR of 17.71 dB and a computational complexity of solely 51.99 Mflops.

Experiment 3: In a subjective evaluation test comprising a total of 23 listeners, the results indicated that in average 30% of the listeners preferred the standard CELP, 30% preferred the proposed search method, and 40% considered the quality indistinguishable.

6. CONCLUSION

In this paper, a fast search method for CELP speech coders was presented. The main idea was centered on a simplification of the analysis-by-sinthesys procedure, for determining the best excitation signal for the linear prediction filter. Experiments have shown that the speech quality remains similar to the standard search, while greatly simplifying the overall coding procedure.

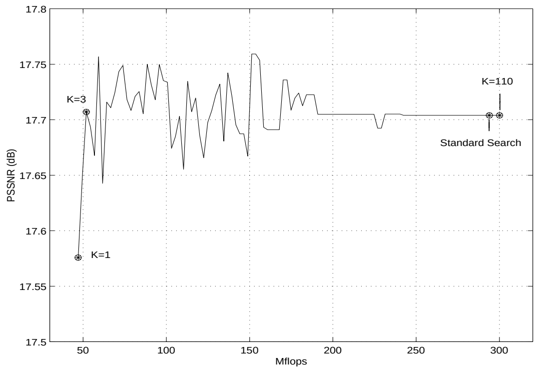

Figure 3: Variations of PSSNR (dB) and computational effort, as functions of K for the modified fast search.

7. REFERENCES

[1] SCHROEDER, M.R., and ATAL, B.S.: 'Code-excited linear prediction (CELP): high-quality speech at very low bit rates'. Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1985, Tampa, FL, pp. 937-940.

[2] DA SILVA, L.M., and ALCAIM, A.: 'A modified CELP model with computationally efficient adaptive codebook search', IEEE Signal Processing Letters, 1995, 2, (3), pp. 44-45.

[3] KITAWAKI, N., NAGABUCHI, H., and ITOH, K.: 'Objective quality evaluation for low-bit-rate speech coding systems', IEEE Jrnl. Selected Areas Communications, 1988, 6, (2), pp. 242-248.

[4] DELLER, Jr., J.R., PROAKIS, J.G., and HANSEN, J.H.L.: 'Discrete-time processing of speech signals' (Macmillan, New York, NY, 1993).