HCELP : LOW BIT RATE SPEECH CODER
FOR VOICE STORAGE APPLICATIONS

Mustapha Bouraoui[*][**] ,Student Member, IEEE
François B. Druilhe[*] and Gang Feng[**]

(*) SGS-THOMSON Microelectronics - Telecom Division
BP 217 - 38019 GRENOBLE - FRANCE
mustapha.bouraoui@st.com

(**) Institut de la Communication Parlée - UPRESA CNRS 5009
ENSERG/INPG - Université Stendhal GRENOBLE-FRANCE

ABSTRACT

There is an increasing need for low cost, fully integrated digital phone systems including telephony functions, fax, hands-free and answering machines. For the latter feature, a high quality, low bit rate speech coder is recommended. It should require only a reasonable complexity to stay competitive in this product range. Recent advances in CELP speech coding have shown the feasibilty of this concept for this kind of consumer applications. A 4.8 kbps Hamming Code Excited Linear Prediction (HCELP) coder is proposed in this paper with an algebraic structure for the codebook. It features a very fast search algorithm which has been evaluated to be 3 times faster than usual algebraic codebook search procedures. Quality evaluation yielded satisfactory results. Implementation aspects and the integration of the coder in an Advanced Telephone Set are also detailed.

1. INTRODUCTION

More than ten years after Schroeder and Atal's paper [1] announcing a high quality but computationally very expensive coder, we observe the application of CELP speech coding in a wide variety of consumer products (Cellular phones, multimedia devices, Advanced Telephone Sets ...). This is mainly the consequence of two major trends: the progress in DSP technologies towards higher speed processors and the effort of researchers in reducing the complexity of a CELP-based algorithm [2]; in addition to the fact that the CELP concept is considered as an attractive approach to low bit rate speech coding.

Even if the state-of-the-art DSP processors allow the implementation of more than one full-duplex CELP coder, the need to reduce the computational complexity of this algorithm remains a critical point. This is particularly true when a speech coder is designed to be integrated in an Advanced Telephone Set (ATS). Such a system includes standard telephone line signal processing and usually performs additional functions, which sometimes run simultaneously, such as fax, full-duplex speakerphone and digital answering machines. For the latter, a low bit rate speech coder is recommended with a good quality of the synthesized speech. There is no need, in this case, for a standardized scheme since the coder algorithm remains "in the box". Furthermore, storage application requirements are less severe than those for communication applications. For example, memory devices used in answering machines provide error-free speech storage. Therefore, it is not necessary to implement an error correction procedure, reducing the bit rate or releasing some useful bits per second for other purposes. Moreover, it is assumed that a single DSP-based chip solution is a demanded way to face the fierce price competition of this kind of product. Based on this integrated solution, the speech coder algorithm cannot take advantage of the whole DSP computational power. Some critical routines of the basic CELP synoptic should be lightened without damaging the expected quality.

In a CELP-based algorithm, one of the heaviest routines is the fixed codebook search procedure which modelizes the target i.e the input speech after its short term and long term decorrelation. A widely used approach to reduce the complexity of this function is the use of ternary codebooks with an algebraic structure. This kind of codebook is not stored in memory and each of its codevectors is indexed according to an algebraic rule. Salami proposed a definition of the regularly spaced nonzero entries directly derived from the binary decomposition of the codebook indices [3]. This technique called BCELP (for Binary CELP) reduces complexity and provides a good robustness against noisy channels. Lamblin & al. proposed another solution based on spherical lattices [4]. This efficient technique is today illustrated in ITU-T standards such as G729 providing toll quality at 8 kbps and more recently, G723 recommendation [5] concerning a dual-rate 6.3/5.3 kbps speech coder designed for videoconferencing on Public Switched Telephone Networks.

After a presentation of the 4.8 kbps HCELP coder in section 2, we present, in section 3, our approach which consists in the mapping of a hypercubic lattice by a Linear Block Code. We show the possibility to build up a fixed codebook which features a very fast search procedure. The proposed search algorithm is then evaluated in terms of target matching and computational complexity. The quality of the coder is reported in Section 4, conducting paired comparisons with ITU-T G723/5.3 kbps standard. Some implementation aspects are also discussed, taking into account the integration of the HCELP coder into ATS systems.

2. THE HCELP CODER

2.1 Overview

The proposed coder follows the basic structure of a CELP coder. The short term analysis is performed on 30 ms speech frames. We looked for a reasonable tradeoff in terms of performance versus complexity in the design of the linear predictor and its associated quantizer. An advanced scheme lies in the design of a predictive split vector quantizer [6]. It actually provides the same spectral distorsion as a scalar quantizer but with a significant amount of bit saving (nearly a ratio of 1:2). The cost is a higher computational complexity and memory requirements. So, we consider that this approach is worthwhile in a very restricted bit budget context. This is not the case of our application in which a nearly 5 kbps speech coder is expected. So, we adopted a classical 34-bit LSP scalar quantizer which gave satisfactory performance. The structure of the whole coder is summarized in Table I.

	Spectrum	Pitch	Codebook
Frame	30 ms	7.5 ms	7.5 ms
Analysis	Durbin recursion	Open loop + Closed loop	Hamming Codebook
Quantization	34 bit SQ LSP	Index : 7/4 bit Gain : 4 bit	Index : 12 bit Gain : 6 bit
Bit rate (kbps)	1.13	1.27	2.4

Table I : 4.8 kbps HCELP description and bit allocation

2.2 A simple Voiced/Unvoiced classifier

A soft interframe interpolation of the LSP parameters is performed except in some Unvoiced-to-Voiced transitions in which the "propagation" of some undesirable LP parameters may cause an audible distorsion at the beginning of voiced segments in reconstructed speech. This is the reason why a simple Voiced/Unvoiced classifier has been developped. A segmentation of a long training sequence of speech has been performed using classical criteria such as zero-crossing rate, energy distribution, autocorrelation tilt. In parallel, we derived the first two Log Area Ratio's from the Durbin recursion.
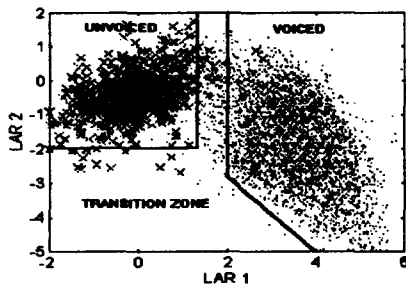

Figure 1 : A simple V/UV classifier for soft interpolation of LP parameters

The joint distribution of these two parameters (fig.1) shows the possibility to build up a simple classifier based upon a linear partition of the first two LAR's space. The advantage of this classifier lies in the fact that it has a very low complexity and does not need to transmit its decision. The decoder calculates LAR1 and LAR2 from the decoded LPC coefficients and can decide itself by consulting the position of the point (LAR1,LAR2). Soft interpolation is actually disabled when a voiced frame follows an unvoiced one.

3. THE PROPOSED SEARCH ALGORITHM

3.1 The HCELP codebook

3.3.1 Theoretical discussion

It is widely assumed that ternary excitations, in which all the pulses belong to {-1, 0, +1}, modelize in a good way the speech residual after short term and long term decorrelations. These excitations may be obtained by center-clipping then retaining the sign information of a Gaussian random sequence. Another way is the use of algebraic mappings. This approach allows the implementation of very fast search algorithms due to the regularity and symmetry inherent in the codebook design. One possible way is explored hereafter.

We consider in the following an l-dimensional, k-bit codebook. This codebook is said algebraic when we can define:

$$\Psi : GF(2^k) \rightarrow \Re^l$$
$$I \rightarrow \Psi(I) \tag{1}$$

where Ψ denotes an injective algebraic transform and I the binary word indexing the ternary excitation $\Psi(I)$. Adding the assumption that the n non-zero pulses are regularly spaced, each excitation may be associated to an n-bit binary word by discarding the zero pulses and mapping {-1, +1} by {+1, 0}. The codebook is thus completely defined by an algebraic transform between $GF(2^k)$ and $GF(2^n)$, which can be represented geometrically by an n-dimensional hypercubic lattice containing 2^k vertices. Linear block codes over $GF(2)$, denoted by $C_{n,k}$, may be used here to map those vertices. Therefore, (1) can be expressed as :

$$\Psi : GF(2^k) \rightarrow GF(2^n) \rightarrow \Re^l$$
$$I \rightarrow C_{n,k}(I) \rightarrow M \circ C_{n,k}(I) \tag{2}$$

where M represents the transition between an n-bit binary word and an l-dimensional excitation as described above. $C_{n,k}$ is preferably chosen as a perfect code in systematic form. Exploiting the symmetry:

$$\forall I \in GF(2^k), \quad C_{n,k}(\bar{I}) = \overline{C_{n,k}(I)}$$

we obtain from (2) that:

$$\forall I \in GF(2^k), \quad \Psi(\bar{I}) = -\Psi(I) \tag{3}$$

Furthermore, in CELP based coders, the codebook is searched by maximizing, over $GF(2^k)$, the matchscore:

$$m(I) = \frac{<T,\Psi(I)>^2}{\Psi(I)^t.\Phi.\Psi(I)}$$

in which T is the backward filtered speech residual [7] and Φ a symmetrical matrix related to the LP filter. The parity of this matchscore and Eq. (3) lead to:

$$\forall I \in GF(2^k), \quad m(\bar{I}) = m(I)$$

so that, only half the codebook need to be searched and indexed.

Let us finally mention the possibility to describe this kind of codebook using the polynomial formalism due to its wide use in the field of Linear Block Codes. In the following, either representation may be used.

3.3.2 A Hamming structured codebook

For our application, we designed a 60-dimensional codebook in which each excitation is represented by a 15-bit codeword from Hamming code $H(15,11)$. We considered that 15 non-zero pulses is a reasonable amount referring to center-clipped sparse excitations in which nearly 70% of pulses are set to zero.
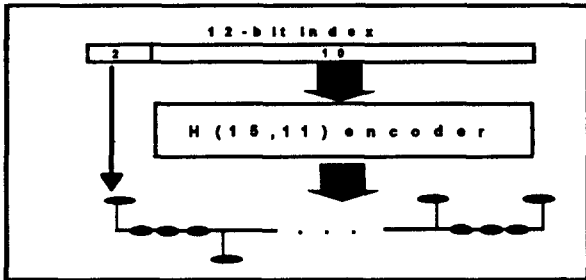

Figure 2. Hamming excitation design procedure

The non-zero entries are regularly spaced and 4 positions are allocated for the first pulse. As shown before, only half of the codebook need to be indexed. It results in a (10+2)-bit vector quantizer for the speech residual (fig. 2).

3.2 Description of the search algorithm

3.2.1 Overview

We adopted the polynomial formalism, illustrated in figure 2, to describe the principle of the proposed search algorithm, considering once more the bijective relationship between the codebook excitation space and the polynomial ideal over $GF(2)$ describing $H(15,11)$.

First step

We derive a target excitation $t(x)$ from the backward filtered residual by retaining only the sign information at the non-zero

locations. We actually consider a simplified match score criterion assuming that the filtered excitation is considered nearly constant.

Second step

We build a subset C of $GF(2^{15})$ according to:

$$C = \{c_k(x)/0 \le k \le 15\} \quad where \quad c_0(x) = t(x)$$

$$and \quad c_k(x) = c_0(x) + x^{k-1} \quad for \quad k \ne 0$$

That means we investigate the 15 closest excitations to $t(x)$ in terms of Hamming distance.

Third step

A subset C_H is obtained from C by a $H(15,11)$ error correction procedure. C_H is a subset of the HCELP codebook on which an exhaustive search is performed in order to select the best excitation. If $t(x) \in H(15,11)$, which occurs with a probability of 1/16, there is obviously no need to build a codebook subset. The index derived from the target is directly transmitted to the decoder.

3.2.2 Target matching

An evaluation of the algorithm performance has been carried out comparing its results with those provided by a full optimal search, S_{full}, into the same codebook. S denotes the search procedure with the proposed algorithm. A large speech database (3 mn) has been converted into a large target database by LP inverse filtering. The excitations E from S and E_{full} from S_{full} obtained to match the target T have been scaled by their gain. Histogram shapes of normalized energies in dB (4) have been plotted in figure 3. As we can see, the sub-optimality of the algorithm does not affect the target matching in a sensible manner. We obtained nearly the same curves for the exhaustive search procedure and the one using the HCELP algorithm.

$$e_1 = 20.\log\frac{\|T-E_{full}\|}{\|T\|} \quad and \quad e_2 = 20.\log\frac{\|T-E\|}{\|T\|} \quad (4)$$
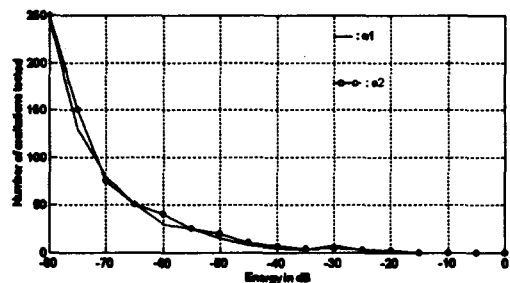

Figure 3. Histogram shapes of error energies between an exhaustive search in HCELP codebook and the proposed search algorithm.

3.2.3 Computational load

The complexity break-down of the proposed algorithm is detailed below :

C and C_H design	:	1500 instructions
Closed-loop on C_H	:	3600 instructions
TOTAL	:	**5100 instructions**

We mean by *instruction* a single-cyle operation such as Multiply or Multiply-Accumulate.

The search procedure being performed every 7.5 ms, this algorithm requires only 0.68 MIPS which is very low (3:1 ratio) compared to search algorithms commonly used on algebraic codebooks such as those involved in G723, G729 ITU-T recommendations or GSM-EFR ETSI recommendation.

4. QUALITY EVALUATION AND IMPLEMENTATION ASPECTS

4.1 Speech quality

The quality of the coder has been evaluated, conducting informal listening tests on multispeaker material (4 males + 4 females + 3 children). Paired Comparisons have been conducted. The reference coder was the toll-quality ITU-T G723/5.3 kbps. We obtained satisfactory results since 45% of the listeners prefered G723, 15% have chosen HCELP and 40% did not notice any difference.

4.2 Implementation aspects and integration in an ATS

The HCELP coder has been developped for an implementation on the ST-D950, which is a 16-bit fixed-point DSP running at 40 MIPS [8]. The source has been written using an ANSI-C instruction set library. It provides D950 bit-exact, fixed-point simulation results without having to use the complete set of DSP development tools. It seems today to be a very common way to specify an algorithm for a fixed-point processor [9]. It has also the ability of a quick translation into assembly code without specific knowledge in speech coding. The real-time break-down of the whole coder is summarized in table II. The coder may be easily inserted in a modern telephony chip including the features cited in the introduction. The worst case of use is when we face the recording of a hands-free conversation. In that case the speech coder, the digital acoustic echo canceller (Hands-free) and some telephony functions such

Function	% of real time (*)
LPC analysis, quantization & interpolation	15
Pitch : Open-loop + Closed-loop	28
HCELP codebook search	3.5
Miscellanous functions	10
TOTAL	**56.5**

(*) Implementation on the D950 from SGS-THOMSON Microelectronics

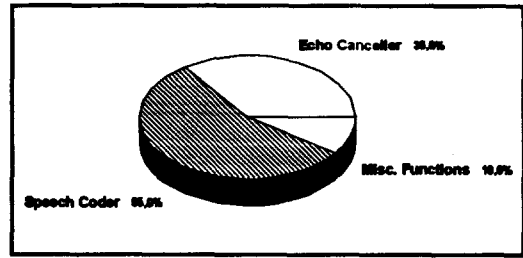*Table II. Real-time break-down for the 4.8 kbps HCELP coder*

Figure 4. Computational load for simultaneous tasks in an Advanced Telephone Set

as DTMF or busy detection, should run simultaneously. Now, a performing echo canceller needs nearly 15 to 20 MIPS. With almost 50% of extra processing power, this can be achived as shown in figure 4.

5. CONCLUSION

In this contribution, a 4800 bps HCELP speech coder suitable for voice storage applications is proposed. A low complexity speech segmentation is presented allowing a soft interpolation of LP parameters with no need of decision transmission. A Linear Block Code structured codebook featuring the implementation of a very fast search algorithm is exploited. Quality evaluation gave encouraging results and implementation aspects have shown the feasibility of HCELP in an ATS environment.

REFERENCES

[1] M. R. Shroeder and B. S. Atal, "Code-excited linear prediction (CELP) high quality speech at very low bit rates," in *Proc. ICASSP*, pp. 937-940, Mar. 1985.

[2] W. B. Kleijn, D. J. Krasinski and R. H. Ketchum, "Fast methods for the CELP speech coding algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 8, pp. 1330-1342, Aug. 1990.

[3] R. A. Salami, "Binary code excited linear prediction (BCELP): new approach to CELP coding of speech without codebooks," *Electron. Lett.*, vol. 25, no. 6, pp. 401-403, Mar. 1989.

[4] C. Lamblin, J.-P. Adoul, D. Massaloux and S. Morissette, "Fast CELP coding based on the Barnes-Wall lattice in 16 dimensions," in *Proc. ICASSP*, pp. 61-64, 1989.

[5] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s", *ITU-T G723 recommendation, Jul. 1996*.

[6] J. R. B. de Marca, "An LSF quantizer for the North-American half-rate speech coder," *IEEE Trans. on Vehicul. Tech.*, vol. 43, no. 3, pp. 413-419, Aug. 1994.

[7] J.-P. Adoul, P. Mabilleau, M. Delprat and S. Morissette, "Fast CELP coding based on algebraic codes," in *Proc. ICASSP*, pp.1957-1960, April 1987.

[8] D950-core specification. *SGS-THOMSON Microelectronics, January 1995*

[9] T. Fingscheit, T. Wiechers and E. Delfs, "Implementation aspects of the GSM half-rate speech codec," in *Proc. EUROSPEECH*, pp. 723-726, September 1995.