16 KBPS WIDEBAND SPEECH CODING TECHNIQUE BASED ON ALGEBRAIC CELP

C. Laflamme, J-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau

Communication Reasearch Center, University of Sherbrooke,
Sherbrooke, Québec, CANADA J1K 2R1

Abstract

The application of Algebraic Code Excited Linear Prediction (ACELP) coding to wideband speech is presented. In wideband coding, a very large excitation codebook is required in order to obtain high quality speech. Ordinary CELP algorithms fail to accomodate such large codebooks due to the excessive coding complexity. In ACELP, however, an algebraic codebook with 20 bit address can be used, without any storage requirements, and more importantly, with a very efficient search procedure which allows for real-time implementation. A focused search strategy is used in which a very small portion of the codebook is searched, yet with performance very close to that of full search. High quality speech was obtained at bit rates below 13 kb/s.

1 INTRODUCTION

Digital coding of 7 kHz speech is becoming of interest for many applications such as teleconferencing, commentary channels, high-quality voice-mail services, and high-quality wideband telephone for the emerging ISDN service. Recent years have witnessed a breakthrough in the developement of speech coding techniques, however, most of the research has focused on narrow-band speech signals where the transmission bandwidth is limited to 300–3400 Hz. Although this bandwidth limitation is acceptable in telephone systems, it degrades the speech quality in the above mentioned applications where the speech is to be heard through high quality loudspeakers. For these systems, a bandwidth of 50–7000 Hz was found to be appropriate (corresponding to a sampling frequency of 16 kHz).

In 1986, the CCITT approved a 64 kb/s standard for wideband audio coding based on a SB-ADPCM coding algorithm [1]. Reducing this bit rate yields larger spectral efficiency and allows for more users to be accomodated in the communication systems which have limitations in the transmission bandwidth. Very efficient speech coding algorithms have been recently developed for narrowband digital speech where high quality speech can be obtained at bit rates as low as 4.8 kb/s [2]. Code-excited linear prediction (CELP) coding has proven to be the most promising among these algorithms. However, few studies have attempted to apply CELP to the context of wideband speech. The main drawback of CELP is its gross computational complexity. As the sampling frequency is doubled, larger frame sizes are needed to maintain a low bit rate transmission. Consequently, the use of much larger excitation codebooks becomes inevitable. For instance, if we assume the same proportional bit rates and block lengths, the typical codebook size increases from a thousand entries (10 bits) to a million entries (20 bits). Searching and storing such a codebook size is rather impractical, unless some subobtimal approaches are utilized such as multistage codebooks, or a split-band approach.

From the above discussion, it seems that it is impossible to use a full band approach for CELP coding of wideband speech. However, the Algebraic Code Excited Linear Prediction (ACELP) approach [3,4] offers the solution to this problem. Using ACELP along with a focused search technique enables the utilization of codebooks having over a million entries, yet with a search complexity that can be implemented on current DSP technology. We describe in this paper the codebook structure used by the ACELP and explain the strategy deployed for the efficient search of the optimum excitation sequence. We also report on the application of ACELP to wideband audio coding where high quality speech was obtained at bit rates below 16 kb/s. We describe the coding parameters and give an assessement of the coder complexity.

2 OVERVIEW OF CELP

In CELP coders, the excitation signal driving the pitch synthesis and LPC synthesis filters is an entry to a large stochastic codebook scaled by a gain factor. The optimum excitation vector is selected by the exhaustive search of the excitation codebook for the codeword which minimizes the mean squared weighted error between the original and synthesized speech. If $\mathbf{c}_k$ is the excitation vector at index k, then the mean squared weighted error is given by

$$E_k = \|\mathbf{x} - g\mathbf{H}\mathbf{c}_k\|^2, \tag{1}$$

where $\mathbf{x}$ is the target vector given by the weighted input speech after subtracting the zero-input response of the weighted synthesis filter $1/A(z/\gamma)$, g is a scaling gain factor, and $\mathbf{H}$ is a lower triangular convolution matrix constructed from the impulse response of the weighted synthesis filter. Setting $\partial E/\partial g = 0$ in Equation (1) yields

$$g = \frac{\mathbf{x}^T\mathbf{H}\mathbf{c}_k}{\mathbf{c}_k^T\mathbf{H}^T\mathbf{H}\mathbf{c}_k}, \tag{2}$$

and substituting Equation (2) in (1) gives

$$E_k = \mathbf{x}^T\mathbf{x} - \frac{\left(\mathbf{x}^T\mathbf{H}\mathbf{c}_k\right)^2}{\mathbf{c}_k^T\mathbf{H}^T\mathbf{H}\mathbf{c}_k} \tag{3}$$

CH2977-7/91/0000-0013 $1.00 © 1991 IEEE

The optimum codeword is selected by minimizing the term

$$\mathcal{T}_k = \frac{\left(\mathbf{x}^T\mathbf{y}_k\right)^2}{\alpha_k}, \tag{4}$$

where $\mathbf{y}_k = \mathbf{H}\mathbf{c}_k$ is the zero-state response of the weighted synthesis filter to the codeword $\mathbf{c}_k$ and $\alpha_k = \mathbf{y}^T\mathbf{y}$ is the energy of the filtered codeword $\mathbf{y}_k$. The CELP complexity stems from the need to compute the filtered codewords $\mathbf{y}_k$ for every possible codebook entry. The numerator in Equation (4) represents the cross correlation between the target vector and the filtered codeword, and the need to compute $\mathbf{y}_k$ is circumvented by the use of *backward filtering*, where Equation (4) can be expressed by

$$\mathcal{T}_k = \frac{\left(\Psi^T\mathbf{c}_k\right)^2}{\alpha_k}, \tag{5}$$

where $\Psi = \mathbf{H}^T\mathbf{x}$ is the backward filtered target vector. Computing the energy term α_k remains the main problem, and in ACELP, it is efficiently determined with very few operations by the use of algebraic codebooks with few nonzero elements. In the next sections we will describe the codebook structure and the efficient strategy deployed for the search of the optimum codeword.

3 CODEBOOK STRUCTURE IN ACELP

We have recently proposed a general framework for representing innovation codebooks in stochastically excited linear predictive coders [4]. The structure of this general framework is depicted in Figure 1. In this structure, innovation codebooks are generated from: an algebraic codebook $\{\mathbf{a}_k\}$ which is the set of vectors $\mathbf{a}_0$, ..., $\mathbf{a}_{L-1}$, and a shaping matrix $\mathbf{F}$. Thus, an excitation vector is given by

$$\mathbf{c}_k = \mathbf{F}\mathbf{a}_k. \tag{6}$$

The advantage of this structure is that the codebook search is decoupled from the codebook properties. The algebraic codebook can be properly chosen so that it is very efficiently searched, and need not be stored. The shaping matrix renders the flexibilty in obtaining any desired codebook properties. It can be fixed or dynamically changed to control the statistical properties of the codebook in time and/or frequency domains. This framework can represent a wide range of efficient innovation codebook structures [4]. In the rest of this section, we will describe the shaping matrix and the algebraic codebook used in our implementation.

The shaping matrix used in the ACELP coder is a function of the LPC model $A(z)$. Its main role is to shape the excitation vectors in the frequency domain so that their energies are concentrated in the important frequency bands. In fact, the matrix $\mathbf{F}$ has a similar role to that of postfiltering [5], in the sense that it enhances the formant regions in the reconstructed speech. However, it has the advantage that it is embedded in the codebook search procedure. In this case, the energy scaling problem which we face when using postfiltering is eliminated. The shaping matrix used is a Toeplitz lower triangular matrix constructed from the impulse response of the filter

$$F(z) = (1 - \mu z^{-1})\frac{A(z/\gamma_1)}{A(z/\gamma_2)}, \tag{7}$$

where $A(z)$ is the LPC inverse filter, γ_1 and γ_2 are constants such that $\gamma_1 < \gamma_2 < 1$, and μ is a factor which controls the spectral tilt and varies in every excitation frame. The parameters of the filter depend on the bit rate.

A sparse algebraic codebook is used. The codebook consists of a set of interleaved permutation codes containing few nonzero elements. The pulse amplitude are fixed to either 1 or -1, and each pulse can take a number of distinct positions. Thus an excitation vector (or a codeword) is determined by the positions of its nonzero pulses (as pulse amplitudes are fixed), and the positions are coded and transmitted. Such a codebook does not require any storage and can be searched very efficiently as we will see in the next section. Further, the codebook structure is robust against channel errors as a transmission error will change the position of only one excitation pulse.

To further illustrate the codebook structure, we describe the 2^{20} sized codebook used in our implementation. Excitation frames of 80 samples (5 ms) are used. Every frame contains 5 pulses with amplitudes 1, -1, 1, -1, and 1, respectively. Each pulse can take 16 distinct positions. Each position is encoded with 4 bits resulting into a 20 bit codebook. An algebraic vector $a(n)$ is given by

$$a(n) = \sum_{i=0}^{p-1} b_i\delta(n - m_i), \quad n = 0, \ldots, N-1, \tag{8}$$

where p is the number of pulses (5 in our case), b_i are the pulse amplitudes (1 for i even and -1 for i odd), m_i are the pulse positions. In our case, the ith pulse has 16 possible positions given by

$$m_i^{(j)} = i + 5j, \qquad \begin{matrix} i = 0, \ldots, 4, \\ j = 0, \ldots, 15. \end{matrix} \tag{9}$$

This codebook is a subset of the set of all the combinations of 80-dimentional vectors containing 5 pulses with fixed amplitudes ($^{80}C_5$ = 24 millions). The codewords of this codebook can be represented by points uniformly distributed over the surface of the hyper sphere in 80-dimentional space. Notice that each pulse has 16 positions distinct from those of the other pulses, and the pulses in the excitation vector can be at any position in the entire frame. Since we search for the best position of each pulse (with the constraints of fixed amplitudes) then the selected codeword will have, to some extent, optimized pulse positions, and will have less unnecessary components as compared to other CELP structures.

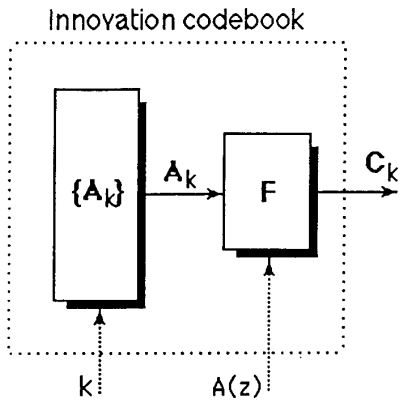

Figure 1 A general framework for codebook representation in stochastic linear predictive coders.

4 EFFICIENT SEARCH STRATEGY

From Equations (3) and (6), the optimum codeword is determined by maximizing the term

$$\mathcal{T}_k = \frac{\left(\mathbf{x}^T\mathbf{H}\mathbf{F}\mathbf{a}_k\right)^2}{\mathbf{a}_k^T(\mathbf{H}\mathbf{F})^T\mathbf{H}\mathbf{F}\mathbf{a}_k} = \frac{\left(\mathbf{x}^T\mathbf{F}_2\mathbf{a}_k\right)^2}{\mathbf{a}_k^T\mathbf{F}_2^T\mathbf{F}_2\mathbf{a}_k}. \quad (10)$$

Notice that the matrix

$$\mathbf{F}_2 = \mathbf{H}\mathbf{F} \quad (11)$$

is a lower triangular matrix containing the impulse response of the combined filter

$$F_2(z) = F(z)/A(z/\gamma). \quad (12)$$

The search can now be brought back to the algebraic space by backward filtering the target vector with the combined filter $F_2(z)$. The term in Equation (10) can be written as

$$\mathcal{T}_k = \frac{\left(\mathbf{d}^T\mathbf{a}_k\right)^2}{\mathbf{a}_k^T\mathbf{\Phi}\mathbf{a}_k}, \quad (13)$$

where $\mathbf{d} = \mathbf{F}_2^T\mathbf{x}$ is the backward filtered target vector, and $\mathbf{\Phi}$ a matrix containing the autocorrelations of the impulse response of $F_2(z)$. As the vector a contains only p nonzero pulses, Equation (13) can be written as

$$\mathcal{T}_k = \frac{\left(\sum_{i=0}^{p-1} d(m_i)b_i\right)^2}{\sum_{i=0}^{p-1}\phi(m_i,m_i) + 2\sum_{i=0}^{p-2}\sum_{j=i+1}^{p-1} b_i b_j \phi(m_i,m_j)}. \quad (14)$$

As the amplitudes b_i, $i = 0,\ldots,p-1$, are fixed to 1 or -1, the correlation requires p additions and the energy $p(p+1)/2$ additions and one multiplication (6 additions for our codebook). If an exhaustive search is to be carried out, this approach is more efficient than the best known CELP algorithms such as overlapping codebooks or VSELP.

Although our search procedure is very efficient, the search becomes rapidly demanding as the codebook size exceeds 2^{12}. For huge codebooks such as the one we are utilizing, a more clever search strategy has to be followed. We describe in the next subsection an approach which we call *focused search*, whereby a very small subset of the codebook is searched, yet achieving a performance very close to that of full search.

4.1 Focused Search

Unlike *partial search* which operates on an arbitrary partial subset, in focused search it is possible to infer if the current codebook subset stands a chance of holding the winner. If the chances are good, the search is continued, otherwise we pass to another subset of the codebook.

Figure 2 shows a scatter plot of the absolute correlation $C_k = |\mathbf{d}^T\mathbf{a}_k|$ versus the square root of the energy of filtered codewords, $\sqrt{\alpha_k}$, where a codebook of 4096 entries is used. Note that the winning codeword is the one which maximizes $C_k/\sqrt{\alpha_k}$, and is represented in the plot by the point having maximum slope. It is observed that only a small portion of the points (codewords) stands a chance for being the winner, and the majority of points have slopes less than half that of the winner. Therefore, the search complexity can be significantly reduced if the search algorithm is confined to those points which have slopes very close to the winner. In our search procedure, the term is computed by adding the contribution of a new pulse in every inner loop. Thus, after few pulses have been added, one can decide whether the search should be continued or not by comparing the term with some predifined threshold. The threshold is set at the beginning of the search and is given by a fraction of the term at which the correlation is maximum. Figure 3 shows the points which has been searched at the most inner loop. In this specific example, less than 2% of the codebook is searched, yet the op-

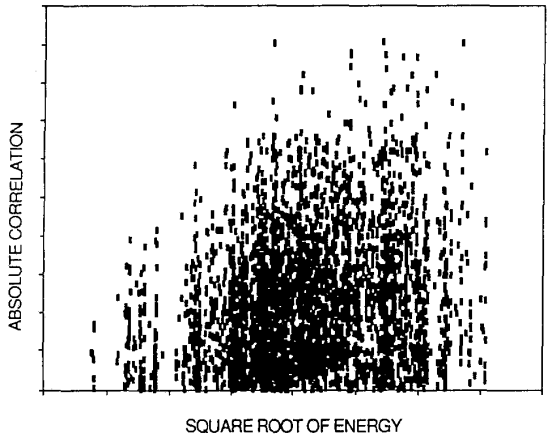

Figure 2 A typical scatter plot of absolute correlation versus square root of energy of filtered codewords (4096 points).

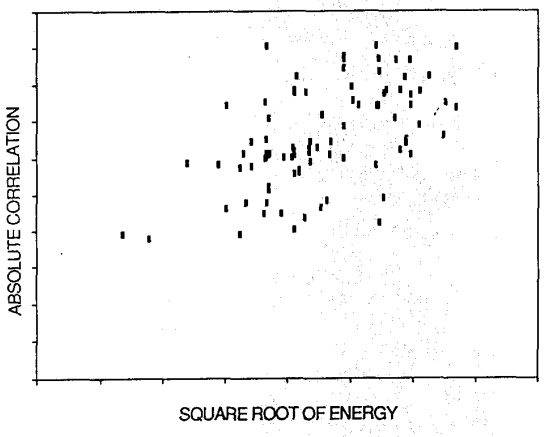

Figure 3 Same as Figure 2 deploying the focused search approach. For this typical subframe only 88 codewords are searched in the most inner loop.

timum codeword is found. Note that the amount of codebook searched varies from one frame to another. Thus for real-time considerations, the search has to be stopped if it is much larger than the average search value.

In our codebook five nonzero pulses are used, and every pulse can have 16 different positions (4 bits). We use two thresholds, one after the third pulse and the other after the fourth pulse. Even with very low thresholds, a substantial amount of the search is cut down without causing any significant degradation in the speech quality. Table 1 shows the SNR for different percentages of codebook search, for a sentence uttered by a female speaker. The search complexity is significantly reduced by increasing the threshold values. Searching about 0.05% of the codebook resulted in only 0.5 dB drop in SNR as compared to the full search case.

Percentage search	*SNR (dB)*
100	22.2
4	22.14
1.6	22.0
0.2	22.05
0.15	21.83
0.05	21.8
0.03	21.5

Table 1 SNRs for different percentages of codebook search.

5 WIDEBAND ACELP CODING

In this section, we report on the work we are currently carrying out on high quality wideband coding below 16 kb/s. The input speech is filtered at 50–7000 Hz and sampled at 16 kHz as in CCITT wideband specifications.

Due to the efficiency of the ACELP codebook search procedure, and to the ability to incorporate very huge codebooks as described earlier, a full-band coding approach is adopted. The LPC parameters are updated every 15 ms and computed using a 20 ms Hamming window centered at the end of the frame. The speech frame is divided into 3 subframes of 5 ms. The pitch and codebook parameters are updated every subframe. The LPC parameters are interpolated between adjacent frames (using LSFs) to obtain a different set of parameters for every subframe. Concerning the filter order, using 16 LPC coefficients was found adequate to represent the speech short-time spectrum envelope. A portion of the bandwidth expansion was used before the parameter computation by lag windowing the autocorrelation coefficients.

Concerning the pitch analysis, a one-tap pitch predictor is used, and is updated every 5 ms (80 samples). The pitch delay extends from 40 to 295 samples, and is encoded with 8 bits. The pitch delay has twice the resolution of the narrow band case. Increasing the pitch resolution did not result in any considerable improvement due to the large excitation codebook used. Fractional pitch could be more beneficial at lower bit rates. The pitch parameters are computed in a closed loop approach, with a weighting factor $\gamma = 0.6$. For delays less than the subframe length, the past excitation is extended by the short-term prediction residual. Thus, these delays are not treated seperately, as would be the case if the excitation itself is repeated. The impulse response of the weighted synthesis filter is truncated at 30, and the convolutions and energies are easily updated.

The codebook is efficiently searched as described in the previous section. A weighting factor of $\gamma = 0.9$ is used. The thresholds were chosen such that, at the most, 1000 codewords are searched. High quality speech at bit rates around 13 kbs was obtained. The quantization procedures for the filter coeficients and excitation parameters will be detailed in a forthcoming paper.

As we are using a full band approach, it is not abvious how the bits are allocated in different frequency bands. However, the ACELP algorithm implicitly favours the bands with more energy, and this is made more evident by the use of the shaping matrix which enhances the formants of the speech spectrum. Therefore the encoding algorithm accurately represents the frequency regions which are perceptually important.

6 CONCLUSION

We presented in this paper an efficient approach for encoding wideband speech signals using algebraic CELP. A full band encoding approach was used whereby a codebook of 2^{20} entries was used. We described an efficient procedure for searching such a large codebook deploying a focused search strategy, where less than 0.1% of the codebook is searched with performance very close to that of full search. High quality speech at a bit rate of 13 kb/s was obtained.

Acknowledgment

This work was supported, in part, by the Canadian Workplace Automation Research Center (CWARC).

References

[1] X.Maitre, "7 kHz audio coding within 64 kbit/s," IEEE J. on Selec. Areas in Commun., Vol. 6, No. 2, pp. 283–298, Feb. 1988.

[2] B.S.Atal et al. (eds.), *Advances in Speech Coding,* Kluwers Academic Pub., 1991.

[3] J-P.Adoul et al, "Fast CELP coding based on algebraic codes," Proc. ICASSP'87, pp. 1957–1960.

[4] C.Laflamme et al., "On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes," Proc. ICASSP'90, pp. 177–180.

[5] J.H. Chen and A.Gersho, "Real-time vector APC speech coding at 4800 bps with adaptive postfiltering," Proc. ICASSP'87, pp. 2185–2188.