Wideband ACELP at 16 kbit/s with Multi-band Excitation

Silvia Pujalte, Asunción Moreno

Department of Signal Theory and Communications
Universitat Politècnica de Catalunya
Campus Nord UPC - D5. Jordi Girona 1-3. 08034 Barcelona
[silvia,asuncion]@gps.tsc.upc.es

Abstract
This paper describes two wideband CELP coders at 16 kbit/s. Their main feature is fast searching, achieving quality comparable to G.722 at 56 and 48 kbit/s. The coders have multi-band algebraic excitation in order to reduce computational complexity and bit allocation. Analysis and error minimization are done over the full band.

1. Introduction
Wideband speech coder requires good quality, with a MOS score higher than 4; the quality reference has been a subband ADPCM wideband coder since 1986, commonly known as ITU-T standard G.722 [1], at 64, 56 and 48 kbit/s. These bit rates often become too high, but the ADPCM bit rate can not be reduced without a significant loss of quality. In order to reduce them while achieving good quality, wideband CELP coders have been developed as an alternative to ADPCM. The main trouble in CELP coders is computational complexity, due to the large codebooks needed to provide good synthesized speech. At present efforts are addressed to reduce this complexity.

Algebraic codebook [2] is often used in CELP coders, since its sparse structure reduces the optimum codebook search dramatically. Usually, search in wideband coders can also be too expensive, and suboptimal approaches can be applied as focused search or pulse reordering.

Another way of reducing complexity in wideband CELP coders is to use subband schemes; in addition, the independent processing of the bands allows to exploit their particular characteristics. Nevertheless, subband coding has some disadvantage produced by the analysis and synthesis filters, such as delay or band overlapping. Subband splitting can be applied to the input signal to separately code each band [3], or can be applied to the analysis and the excitation [4].

Our goal is to achieve an algebraic excitation that allows fast search while producing high quality and a bit rate at 16 kbit/s. We propose a multi-band algebraic codebook that allows flexible control search and allocation bit.

Figure 1: General CELP coder.

We present a full band and a subband version; the first one provides better quality, the second one is faster.

The outline of the paper is as follows: section 2 overviews ACELP coders, section 3 describes the proposed multi-band excitation, section 4 discusses our two codebooks and, finally, results and conclusions are exposed in section 5.

2. Overview of ACELP
Algebraic codebook is a deterministic codebook, where every codeword is composed of a few non-zero samples whose amplitudes and positions are already known. This algebraic codebook is defined by its pulse position. This form simplifies the search procedure dramatically, as most of the samples are zero, and most of the products become zero or sums-up.

Figure 1 represents a generic CELP coder (obviously, pitch predictor can be replaced by an adaptive codebook); the optimum codeword is determined by minimizing the difference between original and reconstructed speech; minimization yields to maximize the term:

\[
\tau = \frac{\left( \sum_{k=0}^{L-1} c(k) \sum_{n=0}^{L-1} x(n) h_w(n-k) \right)^2}{\sum_{n=0}^{L-1} \sum_{k=0}^{L-1} c(k) h_w(n-k)}
\]

(1)
where $c(k)$ is the codeword, $h_w(n)$ the impulse response of the weighted synthesis filter with zero memory, $x(n)$ the weighted input speech minus the zero-input response of the filter and $L$ the subframe length.

With an algebraic codebook, equation (1) stands:

$$
\tau = \frac{\left( \sum_{i=0}^{P-1} a_i R_{eh}(m_i) \right)^2}{\sum_{i=0}^{P-1} \phi(m_i, m_i) + 2 \sum_{i=0}^{P-2} \sum_{j=i+1}^{P-1} a_i a_j \phi(m_i, m_j)}
$$

(2)

where $a_i$ are the pulse amplitudes, $m_i$ their positions, $R_{eh}$ the backward filtered target and $\phi$ the autocorrelation of the impulse response of the filter.

Both $R_{eh}(m_i)$ and $\phi(m_i, m_j)$ are calculated for all $m_i$ and $m_j$ before the maximization begins, and the search procedure can be performed in a nested loop, where one pulse position is modified in every iteration; consequently, the number of operations is significantly low.

Algebraic codebook has two more advantages: the coder generates the codewords in every search procedure, and there is no need for storage. Also, the structure is robust against channel errors: as the coder does not transmit a codeword index but the pulse positions, one bit error changes only one pulse position.

The ITU-T adopted in 1995 an algebraic CELP as narrowband speech coder at 8 kb/s, which is described in [5].

3. Multi-band algebraic codebook

In CELP analysis-by-synthesis coders the algorithm tries to find an excitation sequence which stands close to the input speech when filtered with the synthesis filter. If the excitation sequence is split into low and high frequencies, each band can be independently shaped in terms of their particular properties. The proposed excitation is the sum of two subband excitation (figure 2), and the error is minimized over the full band.

Both the codebooks are algebraic, with $P$ pulses the low band one and $Q$ pulses the high band; they are sampled at 8 kHz and optimized with their own gain. As algebraic excitation has full band response, the subband response of the codebooks is provided for a low pass ($F_h(z)$) and high pass ($F_l(z)$) filter. After an interpolation, the two codebooks are filtered and summed, and the complete excitation is:

$$
c(n) = \alpha_l c_l(n) * f_l(n) + \alpha_h c_h(n) * f_h(n)
$$

(3)

In order to keep the computational advantages of algebraic excitations, the interpolation, which affects the pulse position, is considered in the codebook definition, and the low and high pass filter are grouped with the synthesis filter; it results into two new synthesis filters, $H_w(z)$ and $H_{bw}(z)$. With these modifications, we finally optimize an excitation which is redefined as:

$$
c'(n) = \alpha_l c_l(n) + \alpha_h c_h(n)
$$

(4)

We get the optimum excitation by differentiating the weighted error with respect the two codebook gains, $\alpha_l$ and $\alpha_h$. The exhaustive optimization implies the join search, where the whole high band codebook is explored for every entry of the low band one; the gains are determined solving the following equation system:

$$
\begin{pmatrix}
G_l & G_{hl} \\
G_{bh} & G_h
\end{pmatrix}
\begin{pmatrix}
\alpha_l \\
\alpha_h
\end{pmatrix}
= 
\begin{pmatrix}
Y_l \\
Y_h
\end{pmatrix}
$$

(5)

where:

$$
Y_l = \sum_{i=0}^{P-1} c_l(m_i) R_{eh,l}(m_i)
$$

(6)

$$
Y_h = \sum_{i=0}^{P-1} c_h(m_i) R_{eh,h}(m_i)
$$

(7)

$$
G_l = \sum_{i=0}^{P-1} \phi_l(m_i, m_i) + 2 \sum_{i=0}^{P-2} \sum_{j=i+1}^{P-1} a_i a_j \phi_l(m_i, m_j)
$$

(8)

$$
G_h = \sum_{i=0}^{Q-1} \phi_h(m_i, m_i) + 2 \sum_{i=0}^{Q-2} \sum_{j=i+1}^{Q-1} a_i a_j \phi_h(m_i, m_j)
$$

(9)

$$
G_{hl} = \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} a_i \phi_{hl}(m_i, m_j)
$$

(10)

As we can see in the previous development, the subband excitation implies an increase of the computational

![Figure 2: Multi-band excitation.](image-url)
complexity, depending on the number of pulses. The correlation terms, $R_{xh}, R_{xh'}, \phi_{h}, \phi_{h'}$, and $\phi_{hl}$, are calculated before the search for every frame.

4. Description of the proposed structures

We have developed two coders based on the multi-band excitation which has already been described above. In this section we discuss them. Firstly we explain the analysis stage, which is common to both structures; secondly we focus on the design requirements, and finally, the codebooks are discussed.

4.1. The analysis stage

The short-term frame size is set to 10 ms, and the LP coefficients are calculated with the autocorrelation method using a 20 ms Hamming window centered in the analysis frame; the 18 coefficients are transformed to LAR coefficients before their quantization.

Pitch predictor is computed in open-loop, with three taps updated every 5 ms; the pitch delay extends from 56 to 300 Hz, and is encoded with 8 bits.

The perceptual weighting filter is similar to the one described in [6]; its adaptive tilt term provides a perceptual improvement compared to the classical filter.

LP and pitch coefficients are coded with scalar quantization, 62 bits are used for LPC coefficients and 11 ($3 \times 5 - 3$) for pitch taps. This quantization sets up an upper limit to the final bit rate; a vector quantization will provide a more efficient bit rate.

4.2. Design requirements

Prior to these coders, an algebraic CELP with full band excitation has been developed with the same analysis stage described before. The excitation has $P = 5$ pulses, with positions:

$$m_i = i + PJ \quad \text{where} \quad j = 0 \ldots \frac{L}{P} - 1$$

and amplitudes

$$a_i = \begin{cases} 
1 & \text{if } i \text{ even} \\
-1 & \text{if } i \text{ odd}
\end{cases}$$

Every pulse position is defined by 4 bits and the gain is coded with 7 bits.

The reconstructed speech has high quality, near to G.722 at 56 kb/s, but the main drawback of the coder is the large codebook which has $2^{20}$ entries.

Our aim is to reduce the codebook size in order to achieve fast searching while we are still keeping good quality. Bit rate about 16 kbit/s is another requirement, that is to say, we have 30 bits to code the excitation frame: 14 bits to code the gains and 16 bits to the pulse positions. Since codebook gains vary slowly from frame to frame, this bit allocation can be reduced with a different quantization instead of scalar quantization.

4.3. Full band excitation with two gains

Our tests show that the required quality is achieved with at least $P = 4$ pulses each one filling 10 positions in a narrowband frame; consequently this requires to use the whole number of bits we dispose to code the low band; as a result we renounce to utilize a subband excitation as we have described above.

Our first proposal simplifies the subband structure by suppressing the filters ($F(z) = F_k(z) = 1$). Then, both the low and high codebooks come to be full band codebooks, and the two gains separately multiply $P$ and $Q$ pulses respectively; the excitation is halfway between CELP and multipulse coders. The second gain provides more flexibility to the excitation and, consequently, there is an improvement in coded speech quality.

With regard to the pulses, we have chosen $P = 3$ and $Q = 2$, their positions corresponding to a $P = 5$ pulses in narrowband full band excitation:

$$\begin{align*}
m_0 &= 0, 10, 20, \ldots, 70 \\
m_1 &= 2, 12, 22, \ldots, 72 \\
m_2 &= 4, 14, 34, \ldots, 74
\end{align*}$$

“low band” positions

$$\begin{align*}
m_0 &= 6, 16, 26, \ldots, 76 \\
m_1 &= 8, 18, 28, \ldots, 78
\end{align*}$$

“high band” positions

Such excitation is very restrictive, but the loss caused by using a few positions counteracted by the gain improvement. The codebook has $2^{15}$ entries and the pulses are coded with 15 bits, and the quality of reconstructed speech is only slightly worse than that of the full band coder.

The search procedure can be simplified if it is done in two steps: firstly, we optimize the positions of the three pulses along with their gain; then, the remaining two pulses and their gain are considered. Our tests show that there is no loss of perceptual quality, and the number of operations is tantamount to the one of a codebook with size shorter than 600.

Table 1: Bit allocation of analysis stage.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>kb/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC coefficients</td>
<td>6.2</td>
</tr>
<tr>
<td>Pitch coefficients</td>
<td>2.2</td>
</tr>
<tr>
<td>Pitch delay</td>
<td>1.6</td>
</tr>
<tr>
<td>Total</td>
<td>10.0</td>
</tr>
</tbody>
</table>
4.4. Full band excitation with emphasized high frequencies

The perceptually important speech information lies mainly on low band frequencies, and most of the resources are usually devoted to coder them; for instance, G.722 at 64 kbit/s devotes 75% of the bits to the low band and only a 25% to the high band.

Such considerations lead us to develop a codebook that improves naturalness of reconstructed speech without an important increase of bit rate. We modify the subband structure as we have described above by suppressing low pass filter, and a full band excitation plus a high band filter that emphasizes high frequencies occurs. This structure compensates the disadvantage discussed above and yields the desired naturalness.

As a matter of fact, both the high band and the full band excitations share the pulse position, and the search is done simultaneously by optimizing both the gains for each codeword. Pulse positions are:

\[ m_j^i = 2(i + P j) \quad \text{with} \quad j = 0 \ldots \frac{L}{2P} - 1 \quad (13) \]

where the interpolation effect is considered and each pulse is coded with 3 bits.

The high band excitation does not increase the number of pulses, because only the gain is taken into consideration; therefore, high band excitation contributes with 7 bits to the total excitation bit rate, and the bit rate requirement is maintained.

5. Results

The perceptual quality of the proposed coders has been tested in informal auditions; the reference quality were G.722 at 56 and 48 kbps. Two test of sentences have been used. The first one is composed three male and three female sentences (about 3 seconds); the second set is formed by four paragraphs (between 14 and 30 seconds) pronounced by two men and two women.

The test results show that the full band excitation with emphasized high band coder has similar quality to G.722 at 56 kbps; the full band excitation with two gains has quality better than G.722 at 48 kbps.

Both coders has also been compared with the full band coder described in section 4.2 in terms of SNR. Only about 1 dB is lost by restricting pulse positions, while the computational complexity has significantly reduced. This results are shown in table 2, where “coder 1” refers to full band excitation, “coder 2” to full band with emphasized high band excitation and “coder 3” to full band with two gains excitation.

6. Conclusion

Multi-band algebraic codebook has come out to be a good structure in order to reduce computational complexity in CELP coders; we have tested two coders with multi-band excitation at 16 kbps. A full band excitation with emphasized high frequencies achieves quality comparable to G.722 at 56 kbps, and reduced computational complexity if compared to full band excitation. A full band with two gains excitation achieves a very fast search, equivalent to 600 entries, and quality higher than G.722 at 48 kbps.

7. Acknowledgement

This work was supported by Spanish government grant TIC-2000-1005-C03-01.

8. References


<table>
<thead>
<tr>
<th>Table 2: SNR of compared coders.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CODER</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Coder 1</td>
</tr>
<tr>
<td>Coder 2</td>
</tr>
<tr>
<td>Coder 3</td>
</tr>
</tbody>
</table>