THE D_5 LATTICE QUANTIZATION FOR A 64 KBIT/S LOW-DELAY SUBBAND AUDIO CODER WITH A 15 KHZ BANDWIDTH

K. Hay, L. Mainard† and S. Saoudi**

* ENST-Br, Dept. SC., Technopôle de Brest Iroise, BP 832, 29285 Brest Cedex, France.
email: karine.hay@enst-bretagne.fr / samir.saoudi@enst-bretagne.fr
† CCETT, Servive RCS/SDA, rue du Clos Courtel, BP 59, 35512 Cesson Sévigné Cedex, France.
email: lmainard@ccett.fr

ABSTRACT

A new method for coding generic audio signals at 64 kbit/s in the 20-15000 Hz bandwidth with a low delay is presented. It combines subband coding, Low Delay CELP algorithm and cascaded filterbanks. Our earlier works [1] show that, when using an equal bit rate on each subband, the resulting audio quality was not appropriate. We propose a new technique based on lattice quantization to avoid the search complexity of the statistical vector quantization. It allows an adaptive bit rate allocation in each subband. Experimental results assessing the validity of the proposed method are also presented.

1. INTRODUCTION

High quality audio coding techniques have been extensively studied over the past few years and have led up to international standardizations of algorithms by the Motion Picture Experts Group (MPEG) of the International Standards Organization (ISO).

These techniques allows a reproduction of the signal (i.e, CD audio bandwidth 10 to 20000 Hz) transparent or near transparency, with a bit rate between 64 to 128 kbit/s per monophonic channel. But this high quality is obtained at the cost of an increasing complexity of encoding and a large delay which is not suitable for an audio application.

We are interested in Low Delay Code Excited Linear Prediction (LD-CELP) coding techniques (G.728 Recommendation CCITT)[2, 3]. One of the goals of this coder is to provide a good quality of speech signal in the telephone-bandwidth (300-3400Hz) at a low bit rate (16 kbit/s) with a delay of less than 2 ms. Murgia and Feng have recently proposed a direct extension of the LD-CELP scheme to the 15 kHz speech bandwidth [4].

In this paper, we suggest a new method for coding generic audio signals at 64 kbit/s with a 15 kHz bandwidth. This method combines subband coding, LD-CELP techniques and cascaded filterbanks. In earlier works [1], all subbands were allocated the same bit rate. We saw that there was an audible quantized noise on the reconstructed signal. To overcome this problem, we propose a new quantization technique based on lattice quantization to increase the bit rate on at least the first subband. A new algorithm is proposed to avoid the search complexity of the statistical quantization of the residual signal.

In Section 2, we describe the structure of our coder (namely, subband coding, LD-CELP coder, Psychoacoustic Model). A special attention is dedicated to perceptual coding techniques which are integrated to optimally allocate bits in each subband. Section 3 introduces the D_5 lattice quantization . A comparison with an exhaustive search in the lattice is performed to validate the proposed approach.

2. SUBBAND CODING WITH LD-CELP ALGORITHM

2.1. Coder overview

The input signal, sampled at 32 kHz, is divided into 4 equal frequency bandwidths, by a Polyphase Quadrature Filter (PQF) [5]. Then, each subband is coded by a LD-CELP-like coder [2, 3]. Only the 4 best indexes of the codebooks of each subband coder are sent to the decoder. A psychoacoustic model drives the bit allocation in each subband using the synthesized signal at the LD-CELP coder level. The delay introduced by a 96-tap filterbank design is 3 ms. As the additional coding delay in sub-sampled subbands is 2 ms, the overall system delay is 5 ms. A diagram of this coder is shown in Figure 1.

2.2. Subband coding and delay constraints

Coding an audio signal through a filterbank allows quantization noise shaping along the frequency axis as well as taking advantage of the so-called coding gain of the filterbank. Having a high number of subbands is much desirable since it generally yields a high coding gain and an accurate frequency mapping to quantize the subband samples. However, the filterbank delay is closely linked with the number of subbands and in the case of the MDCT, the overall delay is twice the number of subbands. Therefore, our constraint forces the number of subband lower than a 100 which would lead to a poorly-rejective MDCT and consequently to a blurred frequency localization of the quantization noise.

As a result, we resort to a PQF filterbank with a low number of subbands (4), but with a high frequency rejection. To make up for the coding gain, backward prediction [6] will be used.

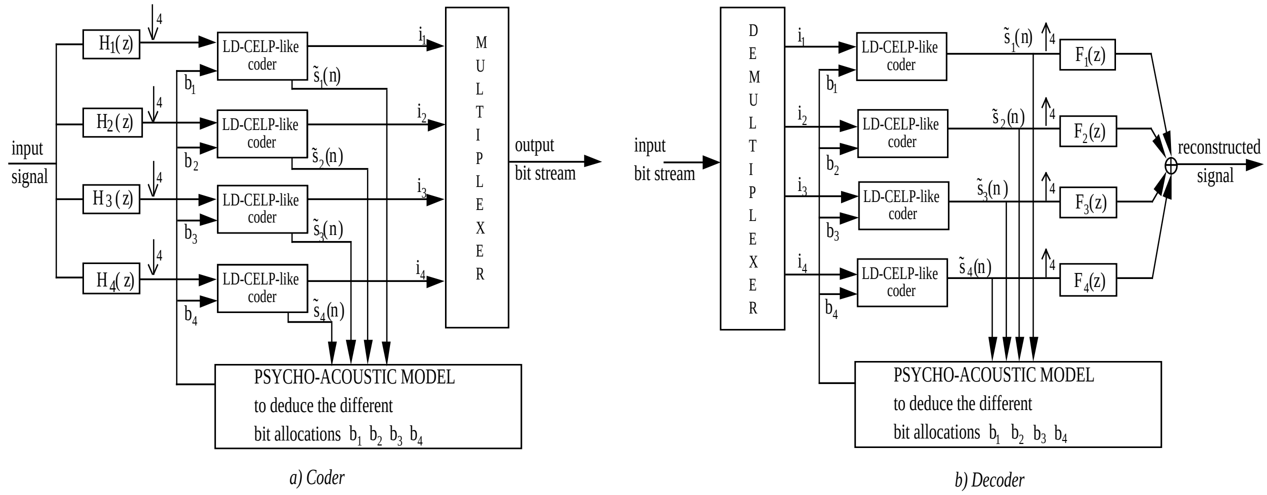

Figure 1: Subband coding associated to the LD-CELP coder

2.3. LD-CELP description

Let us describe briefly the basic features of the LD-CELP (see Figure 2) .

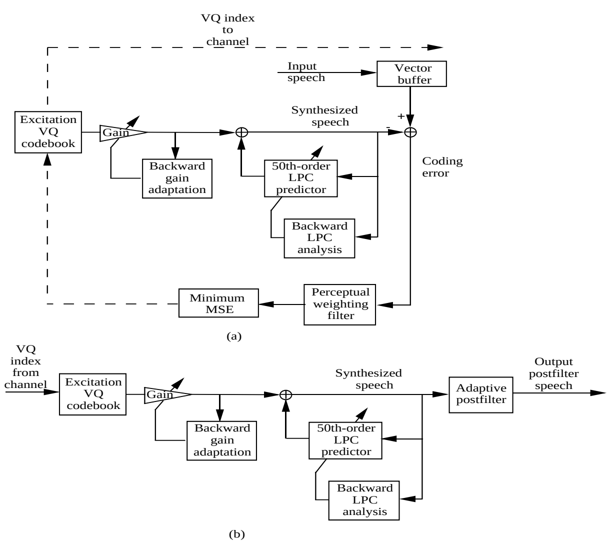

Figure 2: (a) 16 kbit/s low-delay CELP encoder. (b) 16 kbit/s low-delay CELP decoder.

This encoder is based on the analysis-by-synthesis codebook search and has the advantage of transmitting only the index of excitation sequence to the decoder. All the other parameters are "backward-adaptive". This means that the Linear Prediction Coding (LPC) predictor coefficients and the excitation gain will not be derived from the input signal samples, but from previously quantized samples. In the decoder, these parameters are updated with the same procedure, therefore there is no need to transmit them. A 50th-order LPC predictor is used in order to eliminate the pitch predictor.

A candidate excitation signal, stored in a codebook, is applied to the synthesis filter to produce the reconstructed signal. The best vector, which minimizes the Weighted Mean Squared Error (WMSE) distortion, is selected and its index is transmitted to the decoder. Since the input buffer dimension is composed of 5 samples (0.625 ms at 8 kHz sampling frequency), the coding delay is very low (3 times the input buffer : below 2 ms).

2.4. Adding a psycho-acoustic model with a cascaded transformation

From the synthesized signals in each subband, we compute a psycho-acoustic model for the total bandwidth through an alias-cancelled DFT [9]. Since we calculate the model on the previously quantized signals, we do not waste time in buffering samples. As the decoder can also execute the same procedure, no information concerning the psycho-acoustic model needs to be transmitted to the decoder. However, to initialize the stream decoder, basic side information has to be transmitted every 100 ms.

The signal to mask ratio is calculated as follows : a masking curve is calculated from previously synthesized signal samples with an algorithm derived from the MPEG standard (Model 1) [10]. The psycho-acoustic model and consequently the bit allocation are updated every 4 ms once the decoder of each subband has decoded 256 samples. Figure 3 represents the histogram of the bit allocation for a test set of audio signals which is a sub-set of the ISO classical test material. The average bit rate needed to code the first subband is 31 kbit/s, while the second requires 18 kbit/s, the third 12 kbit/s and the fourth 3 kbit/s.

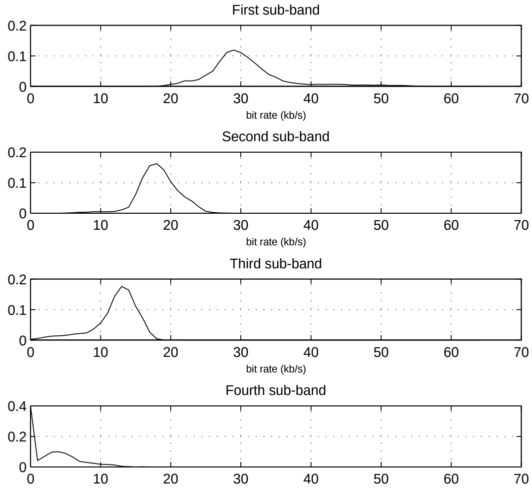

Figure 3: Probability density function versus bit allocation.

3. LATTICE QUANTIZATION IN THE LD-CELP

3.1. Complexity limits of statistical vector quantization

Since the maximal injectable noise delivered by the psycho-acoustic model is generally lower for low frequencies, each subband needs to be allocated an appropriate bit rate derived from the output of the psycho-acoustic module. The higher the bit rate in a subband, the larger the size of the codebook.

A preliminary approach is to store codebooks with different sizes for each subband. The shape of the codebook is optimized in such a way that each subband has a given number of codebooks according to a given precision. The codebooks are optimized with the K-means algorithm [7, 8]. However, the bit rate demanded for the first subband is generally too high for the quantization scheme of classical LD-CELP coder. We present in the next subsection a new algorithm which fulfills the bit rate requirements.

3.2. Description of the method

Since statistical quantization is too computer-intensive when a high bit rate is required, we propose a new method for quantizing the residual signal. In the LD-CELP, each vector contained in a codebook is passed through the synthesized filter and compared with the input signal. The best vector r_q, with respect to the Weighted Mean-Squared Error (WMSE), is selected from the codebook.

Our idea is to pass the vector of input signal through the inverse synthesized filter, to find the residual signal, r. Then, we use a lattice quantizer [11]. These quantizers have the interesting property of not requiring huge codebooks and benefit from fast quantizing algorithm, when associated with the Euclidean norm.

However, the WMSE criterion we use is based on the quadratic error after the synthesized filter. To cope with the algorithm, we first quantize r in the lattice and afterwards, we create a temporary codebook constructed with vectors, which are close to r_q. The largest size of such a codebook is less than 50. We then compute a closed-loop search in the codebook, to find the vector r_f which minimizes the WMSE distortion.

Because of the size of the vector to be quantized, we use a lattice D_5, which is defined as follows :

$$D_5 = \{(x_1, \dots, x_5) \in Z^5 : x_1 + \dots + x_5 = m, \ \ m \ even\},$$

Since the decoder knows the bit rate, and therefore the size of the lattice, only information about r_f is sent to the decoder.

3.3. The parameters adjustment

For a given rate R, the number of codewords N is determined. In a first approach, we decided to fix the bit rate at 32 kbit/s for the first subband in order to test the algorithm. We make use of 20 bits to code 5 samples. Therefore, we must divide these 20 bits to form the lattice and the gain codebook. The gain is defined as a scalefactor :

$$scf = \sum_{i=1}^{5} \mid x_i \mid$$

With our database, we form a sequence of scalefactors by passing the signal through the inverse synthesized filter. An histogram of the scalefactors is presented in Figure 4 :

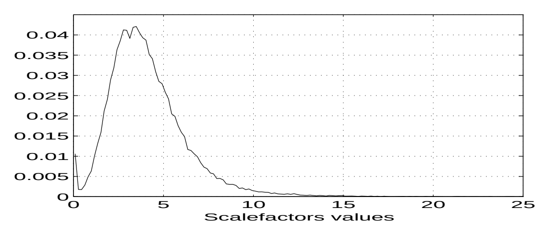

Figure 4: Histogram of the scalefactors

The scalefactors histogram is not uniformly distributed, which leads us to constitute a stochastic codebook with the K-means algorithm [7, 8].

In [12], a method to compute the optimum number n of bits, used to encode the scalefactors, is proposed in the case of a log-linear quantization of the scalefactor.

$$n = log_2(\frac{A}{40} N ln 10) \tag{1}$$

where $A \approx 120dB$ and $N = 5$ (the size of the vector to quantize). Using this technique, the theoretical bit rate is around 4. We compute a Signal to Noise Ratio between the original signal and the coded signal with the lattice quantization method to optimize the bit allocation between the scalefactor and the lattice. The results are shown in Table 1, and are relative to the theoretical value (4).

Bit rate on the scalefactor	Bit rate on the lattice	Relative SNR dB
3 bits	17 bits	-0.13
4 bits	16 bits	0
5 bits	15 bits	-0.54
6 bits	14 bits	-2.13

Table 1: SNR function versus bit allocation

Considering the results of Table 1 and the theoretical integer values computed with equation 1, we chose to allocate 4 bits for the scalefactors and 16 bits for the lattice which gives the best SNR.

Since restrictions of the D_5 lattice relatively to the L_1 norm may be far from being an integer power of 2, the bits assigned to the scalefactor and those assigned to the lattice quantizers are grouped together. This leads to 20 possibles values for the scalefactor and 50973 values for the quantizers, corresponding to all points having a L_1 norm less than or equal to 12.

3.4. Comparison with an exhaustive search

In order to assess the validity of our approach, we compared our method to an exhaustive search of the best quantizer in all the restriction of the lattice. Each point is passed through the synthesized filter to find the best vector. Table 2 shows that the search among a maximum of 50 points around the residual quantizer yields results close to the exhaustive search among 50973 points.

Signal	Exhaustive search - Restrictive search SNR (dB)
German Speech	1.49
Suzanne Vega	0.30
Harpsichord	0.75
Coleman	1.18

Table 2: Comparison of an exhaustive search and the restrictive search

4. CONCLUSIONS

We developed a new method for performing a low delay audio encoder. This algorithm uses subband coding, Low-Delay CELP and cascaded filterbanks. Earlier work had already validated the approach in terms of gain prediction [1]. Here, we introduced a psycho-acoustic model to execute an adaptive bit allocation on the subbands and a new lattice quantization technique to handle the possible high bit rate requirements on the subbands. Objective results on actual audio signals have shown that the reduction of complexity does not harm the overall performance of the encoder. This method seems to be promising in the fact that it improve the audio signal quality in the field of low delay audio coding. Further works will involve a most sophisticated best quantizer search to all subbands and a formal test sessions with respect to existing standards.

5. REFERENCES

[1] K. Hay, L. Mainard, and S. Saoudi. A low delay subband audio coder (20hz-15khz) at 64 kbit/s. In *Proc IEEE-SP on Time-Frequency and Time-Scale Analysis*, pages 265–268, june 1996.

[2] J.H. Chen, R.V. Cox, Y.C. Lin, N. Jayant, and M.J. Melchner. A Low-Delay CELP Coder for the CCITT 16 kb/s speech coding standard. *IEEE J. Select. Areas Commun.*, 10(5):830–849, june 1992.

[3] *Coding of speech at 16 Kbit/s using Low-Delay Code Excited Linear Prediction*, sept. 1992. Recommendation G.728, CCITT.

[4] C. Murgia, G. Feng, C. Quinquis, and A. Le Guyader. Very low delay and high quality coding of 20 hz-15 khz speech at 64 kbit/s. In *4th Europ. Conf. on Speech Comm. and Technol.*, volume 1, pages 37–40, sept 1995.

[5] K. Akaigiri. Detailed technical description for MPEG2 audio NBC. Technical report, ISO/IEC JTC1/SC29/WG11, 1995.

[6] H. Fuchs. Improving MPEG Audio Coding by Backward Adaptive Linear Stereo Prediction. In *99th AES Convention*, volume preprint 4086, New York, 1996.

[7] Y. Linde, A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *IEEE Trans.*, COM-28:84–95, jan 1980.

[8] B. Kövesi, S. Saoudi, J.M. Boucher, and Z. Reguly. A fast robust stochastic algorithm for vector quantizer design for nonstationary channels. In *Proc IEEE Int. Conf. Acoust. and Speech, Signal Process.*, pages 269–272, may 1995.

[9] B. Tang, A. Shen, G. Pottie, and A. Alwan. Spectral Analysis of Subband Filtered Signals. In *Proc IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 1324–1327, may 1995.

[10] *ISO/MPEG-Audio Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mb/s*, 1992.

[11] J. H. Conway and N. J. A. Sloane. *Sphere packing, Lattice and groups.* Springer Verlag, New York, 1988.

[12] L.Mainard. A low delay encoding scheme. In *100th AES Convention*, volume preprint 4199, pages 11–11, Copenhague, 1996.