ALGEBRAIC QUANTIZATION OF TRANSFORM COEFFICIENTS FOR EMBEDDED AUDIO CODING

M. De Meuleneire, H. Taddei

Nokia Siemens Networks
COO RTP PT HWCT Computing Techn SDE
Otto-Hahn Ring 6, 81739 Munich, Germany

D. Pastor

ENST Bretagne - Technopôle Brest-Iroise
Department SC
CS 83818, 29238 Brest Cedex 3, France

ABSTRACT

This paper proposes a new quantization for transform coefficients based on algebraic quantization. The coefficients are represented by a few pulses multiplied by a unique amplitude. The coefficients to be transmitted are selected by optimizing an error criterion, that determines the signs, positions and amplitudes of the pulses. This simple quantization has been implemented in a wavelet-based wideband scalable coder, and has been proved to provide a perceptually better quality than SPIHT on speech signal and music.

Index Terms— embedded coding, transform coding, algebraic quantization

1. INTRODUCTION

An embedded, or scalable, codec organizes the bitstream in layers, where each layer can be decoded independently from the upper layers. The first layer, called core layer, contains the necessary data to synthesize a signal with a minimal quality and bandwidth. Upper layers called enhancement layers are meant to improve the quality and/or increase the bandwidth of the reconstructed signal. According to the network traffic, the decoder can adapt the bitrate on the fly by dropping packets of the upper layers, favoring the delivery of core layer packets. Moreover, the bitrate can also fit the terminal capacity. Besides, scalable coding easily enables premium access, where the user can access the highest quality of a multimedia content after payment.

Embedding coding has been widely investigated in speech and audio coding techniques. As the sampling frequency and the bitrate increase, transform coding is usually preferred over time-domain techniques such as Linear Predictive Coding. Transform coding involves the quantization of the time-frequency transform of the input frame. Usual time-frequency transforms are FFT, DCT, wavelet transform [1] or MDCT [2]. Many coding techniques exist to transmit the coefficients. The simplest one is the scalar quantization of each coefficient. The redundancies can be reduced by using an entropy coder such as Huffman coding. Vector Quantization [3] might perform better at the expense of complexity increase. More sophisticated methods such as Spherical Vector Quantization [4] or Scalar Quantized Vector Huffman coding [5] allow a substantial reduction of the bitrate.

Using transform coding in a scalable coder implies a smart way to quantize the coefficients. The quantizer should be able to produce an embedded bitstream with a representation of the coefficients getting closer to their original values as the bit rate increases. Moreover, the bitstream should be organized in such a way that the most perceptually important coefficients, often the largest ones, are transmitted first. It is for example the case in the ITU-T G.729.1 [4], where the energy of each band is transmitted to the decoder. From these values, the encoder and decoder can compute the bit allocation and the band ordering, from the most significant band to the least one. The more important the band, the higher number of allocated bits. However, all the bits of a codeword corresponding to a band are required to reconstruct coefficients. Incomplete codewords are simply discarded since they can not help to reconstruct coefficients. To take every single bit into account, bit-plane coding may be applied.

Algorithms such as Embedded Zerotree Wavelet (EZW) [6] or Set Partioning In Hierarchical Trees (SPIHT) [7] achieve this bit granularity. Originally developed for the progressive transmission of still images, i.e. for 2-dimensional signals, they are also used in audio coding, e.g. [8][9]. By organizing the transform coefficients into trees and using a parent-children relation, they produce an embedded bitstream where the first bits represent the most important bits of the most important coefficients, the next bits encode smaller coefficients while refining the quantization of the former quantized ones, and the last bits correspond to the least important coefficients. However, the resulting scalability might be a drawback for low target bitrate, as more coefficients could have been transmitted with a rougher quantization.

This paper presents an algebraic quantization for transform coefficients. On a frame basis, the coefficients are gathered in bands. In each band, a factor related to the energy of the coefficients is computed. Positions and signs of the coefficients worthy of transmission are determined and simply

quantized by a variable length codeword, that allows a partial decoding. The rest of the paper is organized as follows. The principles of the quantization are detailed and illustrated in Sec. 2. In Sec. 3, experiments to evaluate the quantization performance are presented. Sec. 4 concludes this paper.

2. ALGEBRAIC QUANTIZATION

The proposed quantization is inspired by the algebraic codebook search in Algebraic Code-Excited Linear Predictive coding (ACELP) [10]. In ACELP coding, the innovation is only quantized with a few pulses, multiplied by a gain. The pulse positions are determined with an analysis-by-synthesis scheme minimizing an error criterion. With our method, a group of transform coefficients, gathered in bands, worthy of transmission are also selected by optimizing an criterion and coded according to their position, sign and amplitude. The coefficients within the group are transmitted progressively, so that only a part of the coefficients within a band can be decoded. Sec. 2.1 presents the principles of the method. Determination of the coefficients to be transmitted is described in Sec. 2.2.

2.1. Principles

Let $y(i)$, $i \in \{0, \ldots, N-1\}$ be the coefficients to be quantized, N is the number of coefficients per frame. In the general case, the input coefficients are divided into M bands. The band k comprises N_k coefficients, such that:

$$\sum_{k=0}^{M-1} N_k = N \tag{1}$$

The first coefficient of band k is noted $b(k)$. For example, $b(0) = 0$, $b(1) = N_0$, $b(2) = N_0 + N_1$, $b(M-1) = \sum_{k=0}^{M-2} N_k$, and by default, $b(M) = N$. The coefficient within band k is indexed by j. The coefficient at position j in band k has position $b(k) + j$ in the frame.

Our method aims at reducing at most the Mean Square Error (MSE) between the original and the reconstructed coefficients:

$$MSE\left(\mathbf{y}, \hat{\mathbf{y}}\right) = \frac{1}{N} \sum_{i=0}^{N-1} \left(y_i - \hat{y_i}\right)^2 \tag{2}$$

Before decoding, the reconstructed coefficient are set to zero, $\hat{y}_i = 0$, $i = 0, \ldots, N-1$. The MSE is then equal to $\sum_{i=0}^{N-1} \frac{y_i^2}{N}$. If a coefficient y_i is perfectly decoded (i.e. $y_i = \hat{y}_i$), the MSE decreases by $\frac{y_i^2}{N}$. In this sense, the most significant coefficients are the biggest ones. It should be a priority to quantize and transmit those coefficients first.

The following model has been investigated to quantize the coefficients. In a band k, the coefficients are estimated by:

$$\tilde{y}(b(k)+j) = m_k c(b(k)+j), \;\; j \in \{0, \ldots, N_k - 1\} \tag{3}$$

where $c(b(k)+j) = \pm 1$ or 0, and $m_k > 0$. If $c(b(k)+j) \neq 0$, it is called a pulse. m_k is the pulse amplitude. The coefficients are estimated by multiplying the pulses with their respective amplitude. In band k, the encoder has to determine how many pulses must be sent, their position, their sign, and which value m_k is given. The pulse signs are given by the signs of their respective coefficients.

2.2. Minimization of an error criterion

The determination of the pulses and of m_k is done by minimizing the MSE between the original and the estimated coefficients:

$$e_k = \sum_{j=0}^{N_k-1} \left(y(b(k)+j) - m_k c(b(k)+j)\right)^2 \tag{4}$$

The optimal value of m_k is given by the solution of $\frac{\partial e_k}{\partial m_k} = 0$, i.e.

$$m_k = \frac{\sum_{j=0}^{N_k-1} (y(b(k)+j)c(b(k)+j))}{\sum_{j=0}^{N_k-1} c^2(b(k)+j)} \tag{5}$$

Replacing m_k by this expression in Eq. (4) yields:

$$e_k = \sum_{j=0}^{N_k-1} x^2(b(k)+j) - \frac{\left(\sum_{j=0}^{N_k-1} y(b(k)+j)c(b(k)+j)\right)^2}{\sum_{j=0}^{N_k-1} c^2(b(k)+j)} \tag{6}$$

Whatever the number of pulses and their position are, the first term $\sum_{j=0}^{N_k-1} x^2(b(k)+j)$ is always constant. Thus the minimization of e_k is strictly equivalent to the maximization of the term d_k defined by:

$$d_k = \frac{\left(\sum_{j=0}^{N_k-1} y(b(k)+j)c(b(k)+j)\right)^2}{\sum_{j=0}^{N_k-1} c^2(b(k)+j)} \tag{7}$$

At each position, there is either a pulse or no pulse, i.e. two possibilities (the sign is always set as in sec. 2.1). The number of possible combinations is then 2^{N_k}. Afterwards, the amplitude m_k is computed by Eq. (5). Once the pulse combination that maximizes d_k and the corresponding amplitude m_k are found, they are sent to the decoder.

For N_k large, the search can become very complex. However, it is actually not necessary to test all combinations. If the search is restricted to finding the best combination of ℓ pulses among N_k pulses, $\ell < N_k$, the number of combination to be tested is $C_{N_k}^{\ell} = \frac{N_k!}{\ell!(N_k-\ell)!}$. The criterion d_k becomes:

$$d_k^{\ell} = \frac{\left(\sum_{j=0}^{N_k-1} y(b(k)+j)c(b(k)+j)\right)^2}{\ell} \tag{8}$$

For ℓ constant, maximizing the criterion Eq. (8) is equivalent to maximizing d_k^{ℓ} numerator. As $y(b(k)+j)c(b(k)+j) > 0$,

with $k = 0 \dots M-1$ and $j = 0, \dots, N_k - 1$, maximizing the criterion d_k^ℓ is equivalent to maximizing

$$\sum_{j=0}^{N_k-1} y(b(k)+j)c(b(k)+j) \tag{9}$$

A sum of positive values is maximal when every term of the sum is maximal. Consequently, the criterion d_k^ℓ is maximal when the ℓ pulses correspond to the ℓ biggest coefficient absolute values. The criterion may only be tested for N_k combinations (i.e. $\ell = 1, \dots, N_k$).

The optimal pulse combination is selected as follows. For $\ell = 1, \dots, N_k$, the criterion d_k^ℓ is computed according to Eq. (8) with the ℓ biggest absolute values. Finally, the optimal number ℓ_{opt} is found by maximizing d_k^ℓ over ℓ:

$$d_k^{\ell_{opt}} = \underset{l \in 1 \dots N_k}{\arg\max}\, d_k^\ell \tag{10}$$

The optimal combination is given by the ℓ_{opt} pulses at the position of the ℓ_{opt} biggest absolute values.

The search complexity might be slightly reduced by taking into account the following assumption. The value of the criterion d_k^ℓ increases with ℓ until the maximal value $d_k^{\ell_{opt}}$ is reached, then d_k^ℓ decreases. ℓ_{opt} is the first value of ℓ such as $d_k^{\ell_{opt}} > d_k^{\ell_{opt}+1}$. There exists a recursive relation between d_k^ℓ and $d_k^{\ell+1}$:

$$d_k^{\ell+1} = \frac{\left[\sqrt{\ell d_k^\ell} + |y(b(k)+j_{\ell+1})|\right]^2}{\ell+1} \tag{11}$$

where j_ℓ is the indice within band k of the coefficient with ℓ^{th} biggest absolute value. A new pulse is added if its contribution increases the criterion. Otherwise $d_k^\ell > d_k^{\ell+1}$, the search is stopped, $\ell_{opt} = \ell$. The amplitude m_k is computed and sent together with the selected pulse combination. This assumption proves to be true in about 95% of the cases. The 5% left corresponds generally to periods of weak energy like speech pauses.

3. EXPERIMENTS

The proposed algebraic quantization is very simple, as it is equivalent to a sorting algorithm (sorting the coefficient absolute values within a band in decreasing order). In the following, we illustrate our quantization with two examples that underline in which cases it could be applied. In Sec. 3.1, the quantization is applied to transform coefficients of the original signal. The quantization is then applied in Sec. 3.2 to a difference signal and compared to the SPIHT algorithm.

3.1. Experiment n°1

Let us consider the following codec. A 4-level Wavelet Packet Decomposition (WPD) divides a wideband (8 kHz bandwidth) input signal into 16 Wavelet Packets (WP). The wavelet filter is the 24-tap Vaidyanathan filter and the convolution is performed using the "full convolution" in [11]. The 2 WPs corresponding to the frequencies above 7 kHz are not taken into account. Consequently, 14 WPs are transmitted. Each WP is quantized using the procedure presented in Sec. 2.2. The mean of the 14 amplitudes is quantized in the $\log_2$ domain using a 32-step (5 bits) non-uniform scalar quantizer. The ratio of each amplitude to the quantized mean is quantized in the $\log_2$ domain by a 16-step (4 bits) non-uniform scalar quantizer. The quantized mean is transmitted first. Then, the amplitude and the pulses for each packet are transmitted following the frequency order. When all the packets are transmitted (at around 28 kbit/s), the same coding technique is applied to the error between the original and quantized WP coefficients.

Segmental SNR (SSNR) was used as an objective measure. It gives a measure of the distortion between the original signal $s(n)$ and the reconstructed signal $\hat{s}(n)$. The SSNR is computed for 8 different speech files at bitrates ranging from 16 to 48 kbit/s, with a 2 kbit/s step progression. The SSNR are averaged over the speech files. The average SSNR are depicted in Fig. 1. The plot shows that the SSNR increases with the bitrate, that is to say that the distortion decreases. Hence the quality improves. One can also notice a gap around 30 kbit/s. At this point, the codec starts to transmit information about the error signal. The WP coefficients in low frequencies are transmitted first. Since they are the largest, their transmission contribute greatly to reduce the distortion. This pattern could also be observed in the case of a multi-stage quantization, i.e. the error between the original and quantized coefficients from the previous stages is transmitted.

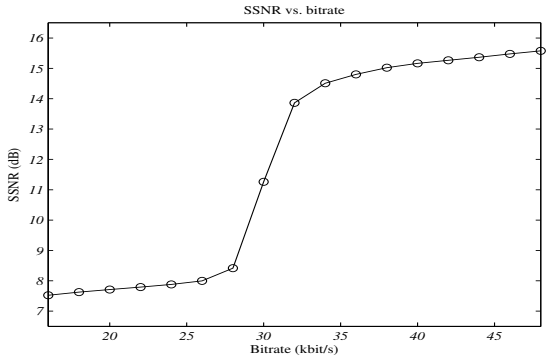

Fig. 1. SSNR vs. bitrate.

Nevertheless, informal listening test showed that the quality at low bitrates (up to 32 kbit/s) is poor. The quantization of the low frequency WPs with our method seems to have an impact on the quality. Indeed, when the quantized coefficients between 0 and 2 kHz (the 4 first WPs) are substituted with the original ones, the perceptual quality is heavily improved. Conversely, replacing the coefficients above 2 kHz

by the original ones does not have an influence. The problem disappears above 60 kbit/s, when the quantization of the low frequency WPs is sufficiently improved. The proposed method does not seem to be adapted to the quantization of original signal coefficients in low frequencies. A mechanism that transmit information to increase the SSNR as much as possible might helpful, for instance, transmitting first information about the frequencies up to 2 kHz, quantized WP coefficients and then the quantized error.

3.2. Experiment n°2

This quantization has also been implemented in a modified version of the codec described in [12]. This wideband speech codec provides an embedded bitstream that can be decoded at bitrates ranging from 8 to 32 kbit/s. The codec works on a 10 ms frame basis. The codec structure comprises three layers. First, a split band structure separates the Lower Band (LB) part and the Higher Band (HB) part of the input signal. The core layer encodes the LB part of the input signal. This layer makes use of the ITU-T G.729 coder at 8 kbit/s. Afterwards, the first enhancement layer utilizes bandwidth extension techniques relying on a wavelet filter bank to reproduce artificially the HB part, with an additional bitrate of 2 kbit/s. Finally, the last enhancement layer progressively encodes the wavelet coefficients of the difference between the original signal and the G.729 output in the LB part, and encodes the wavelet coefficients of the original signal in the HB part. The WPD provides 14 WPs to be transmitted. The WPs are transmitted according to the decreasing order of the energy of the 10 kbit/s output (G.729+bandwidth extension). The amplitude is quantized in the $\log_2$ domain by a 16-step (4 bits) non-uniform scalar quantizer.

An A-B listening test has been performed to compare the proposed quantization with SPIHT. For this test, SPIHT has been optimized to fit with the codec structure and give the best results. Eight english sentences and four music pieces were presented to four non-native english speakers (the results have clearly shown that more listeners are not necessary). The samples are encoded with two versions of the codec at 32 kbit/s, one with SPIHT and the other with the proposed quantization. The results are presented in Tab. 1. At 32 kbit/s, the coefficient quantization are allocated 22 kbit/s. At this bitrate, it is not possible to transmit all the coefficients. For both codecs, the missing coefficients are replaced by the corresponding coefficients coming from the bandwidth extension. As the results show that the proposed quantization performed perceptually better, a rougher quantization than SPIHT that transmits more coefficients is preferable.

Signal	SPIHT	Algebraic quantization
Speech	12.5 %	87.5 %
Music	18.75 %	81.25 %

Table 1. SPIHT vs. algebraic quantization.

4. CONCLUSION

We proposed a new algebraic quantization for transform coefficients. The coefficients are estimated by a few pulses multiplied by an amplitude. The amplitude, the sign and position of the pulses are determined by optimizing the error between the original coefficients and the estimated coefficients. This quantization has the advantage of being very simple, because it is equivalent to the sorting of a few values. Moreover, it also allows a progressive decoding of the coefficients. This method is well adapted for encoding an error signal for frequencies up to 2 kHz, and the original signal afterwards.

5. REFERENCES

[1] Stéphane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Second edition, 1998.

[2] J. Princen and A. Bradley, "Analysis/Synthesis Filter Bank Designed Based on Time-Domain Aliasing Cancellation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1153–1161, 1986.

[3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1991.

[4] ITU-T, "Recommendation G.729.1: G.729 Based Embedded Variable Bit-Rate Coder: An 8-32 kbit/s Scalable Wideband Coder Bitstream Interoperable with G.729," May 2006.

[5] ITU-T, "Recommendation G.722.1: Low-Complexity Coding at 24 and 32 kbit/s for Hands-Free Operation in Systems with Low Frame Loss," May 2005.

[6] J.M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[7] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.

[8] A. Aggarwal, V. Cuperman, K. Rose, and A. Gersho, "Perceptual Zerotrees for Scalable Wavelet Coding of Wideband Audio," in *Proc. of IEEE Workshop on Speech Coding*. IEEE, June 1999, pp. 16–18.

[9] Zhitao Lu and William A. Pearlman, "An Efficient, Low-Complexity Audio Coder Delivering Multiple Levels of Quality for Interactive Applications," in *Proc. IEEE Signal Processing Society 1998 Workshop on Multimedia Signal Processing*, Dec. 1998, pp. 529–534.

[10] J. P. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, "Fast CELP Coding Based on Algebraic Codes," in *Proceedings of ICASSP '87*. IEEE, Apr. 1987, pp. 1957–1960.

[11] B. Leslie and M. Sandler, "A Wavelet Packet Algorithm for 1-D Data With No Block End Effects," in *1999 IEEE International Symposium on Circuits and Systems ISCAS '99*, Orlando, USA, July 1999, IEEE, vol. 3, pp. 423–426.

[12] M. De Meuleneire, H. Taddei, O. de Zelicourt, D. Pastor, and P. Jax, "A CELP-Wavelet Scalable Wideband Speech Coder," in *Proc. of ICASSP 2006*, Toulouse, France, May 2006, IEEE, vol. 1, pp. 697–700.