A CONDITIONAL ENHANCEMENT-LAYER QUANTIZER FOR THE SCALABLE MPEG ADVANCED AUDIO CODER

Ashish Aggarwal and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106, USA
Email:[ashish,rose]@ece.ucsb.edu

ABSTRACT

We propose an efficient enhancement-layer quantizer which considerably improves the bit rate scalability of the multi-layer Advanced Audio Coder (AAC). The scheme exploits the statistical dependence of the enhancement-layer signal on the base-layer quantization parameters. It fundamentally extends the prior work on compander domain scalability, which was shown to be asymptotically optimal for entropy coded uniform scalar quantizer, to systems with non-uniform base-layer quantization. We show that an enhancement-layer quantization which is conditional on the base-layer information can be efficiently implemented within the AAC framework to achieve major performance gains. Moreover, in the important case that the source is well modeled as Laplacian, we show that the optimal conditional quantizer is implementable by only two distinct switchable quantizers depending on whether or not the base-layer quantizer employed the "zero dead-zone." Hence, major savings in bit rate are recouped at virtually no additional computational cost. For example, the proposed four layer scalable coder consisting of 16kbps layers achieves performance close to a 60kbps non-scalable coder on the standard test database of 44.1kHz audio.

1. INTRODUCTION

The problem of efficient bit-rate scalability, or embedded coding, is an important one. A scalable bit stream allows the decoder to produce a coarse reconstruction if only a portion of the bit stream is received, and to improve the quality as more of the total stream is made available. Scalability is especially important in applications, such as digital audio/video broadcasting and multicast audio, which require simultaneous transmission over multiple channels of differing capacity.

A major objection to incorporating bit rate scalability within existing coders is the resulting loss in performance relative to the non-scalable coding. A recent standard for scalable audio coding is MPEG-4 [1][2] which performs multi-layer coding using AAC modules [3]. AAC incurs a substantial performance penalty to provide a scalable bit stream, especially when low rate layers are involved. There are two main reasons why the conventional approach underperforms. First, each encoding layer simply requantizes the reconstruction error of the preceding layer. This approach yields optimal scalability only if the distortion measure is mean squared error (MSE) and not the widely used weighted squared error (WSE) metric based on the noise-mask ratio (NMR) [4]. Second, the conventional coder typically employs the same quantizer, or its scaled version, for all encoding layers. However, the conditional pdf of the source given the base-layer reconstruction differs from the original source pdf.

In the previously proposed scheme [5][6], a companded scalable quantization (CSQ) was shown to be asymptotically optimal for uniform quantization at the base-layer. (There the CSQ was referred to as the asymptotically optimal scalable (AOS) scheme). In this paper, we fundamentally extend the CSQ to systems with non-uniform base-layer quantization. When the base-layer quantizer is not uniform, the conditional density of the signal at the enhancement-layer can vary greatly with the base-layer quantization parameters. Using a single quantizer at the enhancement-layer is clearly suboptimal. We therefore propose a conditional enhancement-layer quantizer (CELQ). However, to design a separate quantizer for each base-layer reproduction is prohibitively complex and practical systems need to approximate. For the important case that the source is well modeled by the Laplacian, we show that the optimal CELQ is implementable by only two distinct switchable quantizers depending on whether or not the base-layer reconstruction was zero. We implement the CELQ within the multi-layer AAC with a standard-compatible base-layer. At no additional computation cost, the new scheme leads to substantial savings in bit rate over the CSQ which itself considerably outperforms the standard technique.

The organization of the paper is as follows: A brief overview of quantization in AAC is provided in section 2. The quantizer design problem is formulated in section 3. Section 4 outlines the CSQ approach and section 5 details the CELQ quantizer design. Section 6 summarizes the results.

2. OVERVIEW OF AAC

Figure 1 shows a block diagram of the AAC encoder. The transform and pre-processing block converts the time domain data into the spectral domain. A switched modified discrete cosine transform is used to obtain a frame of 1024 spectral coefficients. The time domain data is also used by the psychoacoustic model to generate the masking threshold for the spectral coefficients. The spectral coefficients are grouped into 49 bands to mimic the critical band model of the human auditory system. All transform coefficients within a given band are quantized using the same non-uniform quantizer. Equivalently, the transform coefficients are compressed by the function, $c(x) = |x|^{0.75}$, and then quantized

---

This work is supported in part by the NSF under grants no. MIP-9707764, EIA-9986057 and EIA-0080134, the University of California MICRO Program, Conexant Systems, Inc., Dolby Laboratories, Inc., Lucent Technologies, Inc., Medio Stream, Inc., and Qualcomm, Inc.

0-7803-7402-9/02/$17.00 ©2002 IEEE

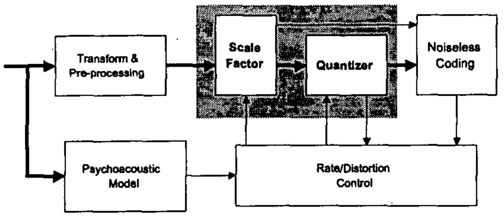

Fig. 1. Block diagram the AAC encoder

using a scalar quantizer. We thus have,

$$\begin{aligned} i_x &= sign(x) * \lfloor \Delta c(x) + 0.4054 \rfloor, \\ \hat{x} &= sign(i_x) * c^{-1}(i_x/\Delta), \end{aligned} \quad (1)$$

where, x and $\hat{x}$ are original and quantized coefficients respectively, Δ is the quantizer stepsize of the band, $\lfloor x \rfloor$ gives the largest integer less than or equal to x and $sign()$ represents the signum function.

The quantizer stepsize of each band is adjusted to match the masking profile, and thus, to minimize the average WSE of the frame for the given bit rate. The quantized coefficients in a band are entropy coded using a Huffman codebook, and transmitted to the decoder. The quantizer stepsize for each band is transmitted as side information.

In the conventional approach to *scalable* AAC, each encoding layer quantizes the reconstruction error of the preceding layer. The enhancement-layer coder is identical to the base-layer coder. It searches for the quantizer stepsize that achieves the target WSE in the band. The quantized indices and the stepsize for each band are transmitted for the base as well as the enhancement layer.

3. PROBLEM FORMULATION

The objective of the coder is to minimize the average WSE, given the target bit rate. From quantization theory [7], the necessary condition for optimality is satisfied by ensuring that the WSE in each band is constant. This requirement is met by AAC in two steps. First, a non-uniform quantizer is used to quantize the coefficients, thereby allowing a higher level of distortion when the value of a coefficient is high. Second, to account for different the masking thresholds, or weights, associated with each band, the quantizer stepsize is allowed to vary from band to band. Effectively, quantization is performed using scaled versions of a fixed quantizer. The structure of this fixed quantizer for AAC is shown in figure 2. The quantizer has a "dead-zone" around zero whose width is greater than the width of the other intervals and the reconstruction levels are shifted towards zero. The width of the interval for all the indices except zero is the same. Using the terminology of [8], we call this quantizer a constant dead-zone ratio quantizer (CDZRQ).

The situation is more complicated in the scalable AAC where enhancement-layer quantization is forced to use *only* the base-layer reconstruction error. Furthermore, AAC *restricts* the enhancement-layer quantizer to be CDZRQ. The problem arises because, *1)* the weights of the distortion measure cannot be expressed as a function of this reconstruction error, and *2)* the conditional density of the source given the base-layer reconstruction is different from that of the original source. Hence, the use of a compressor function and CDZRQ on the reconstruction error is not appropriate. In order to optimize the WSE criterion the enhancement-layer encoder has to search for a new set of quantizer stepsizes, and transmit their values as side information. At low rates of around 16kbps, the information about quantizer stepsizes of all the bands constitutes as much as 30%-40% of the bit stream. Moreover, the quantization loss due to ill suited CDZRQ at the enhancement-layer remains unabated. These factors are the main contributors to poor performance of conventional scalable AAC.

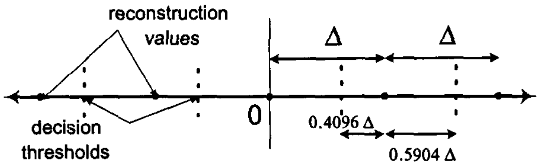

Fig. 2. The AAC quantizer with "dead-zone" around zero

It is important to realize that this approach to scalability does not make full use of the available information. In particular, apart from the base-layer reconstruction, the enhancement-layer decoder also has access to the base-layer quantization interval. The use of this important information from the base-layer, in the context of scalable predictive coding of sources with memory, was first demonstrated in [9].

4. COMPANDED SCALABLE QUANTIZATION CODING

The CSQ approach was developed in [5] and its implementation within AAC was detailed in [6]. For completeness, we briefly outline the main results in this section. The proposed solution looks at the compander domain representation of a scalar quantizer, and achieves asymptotically-optimal scalability by requantizing the reconstruction error in the *companded domain*. The two main observations leading to the desired result are:

1. *Quantizing the reconstruction error is optimal for the MSE criterion.* For a uniform base-layer quantizer, under high resolution assumption, the pdf of the reconstruction error is uniform [10] and hence, the best quantizer at the enhancement-layer is also uniform.

2. *The optimal companding for an entropy coded scalar quantizer maps the WSE of the original signal to MSE in the companded domain.* For the optimal compressor function, Bennett's integral [11] reduces to $D = \Delta^2/12$, which equals the MSE (in companded domain) of the uniform quantizer with step size Δ.

Thus, the compander effectively reduces the original WSE minimization to an MSE optimization problem and requantizes the reconstruction error in the companded domain to achieve asymptotic optimality.

4.1. Scalable AAC using CSQ

The CSQ scheme can be implemented within AAC in a straightforward manner. At the AAC base-layer, once the coefficients are companded and scaled by the appropriate stepsize, they are *all* quantized using the same quantizer. This observation suggests that, if the quantizer stepsizes at the base-layer are chosen correctly, optimizing MSE in the "companded and scaled domain" is equivalent to optimizing the WSE measure in the original domain.

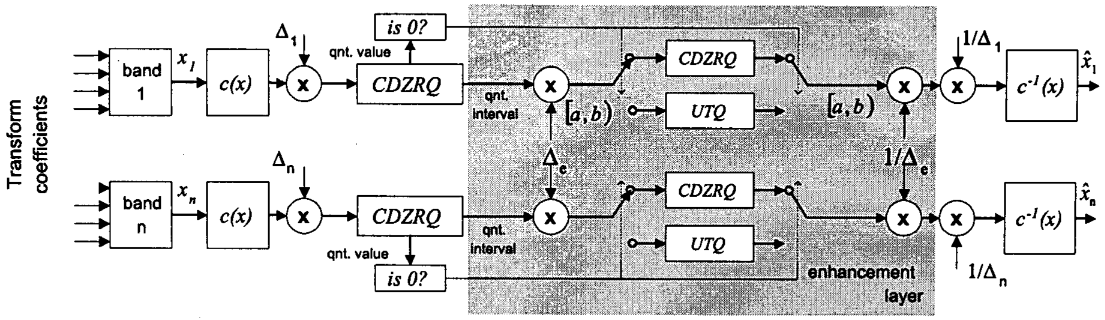

Fig. 3. Block diagram of the CELQ applied on AAC

Hence, the enhancement-layer encoder can use a single quantizer for requantizing the reconstruction error in the *companded and scaled* domain. In effect, the scale factors at the base-layer are being used to predict the enhancement-layer scale factors and only one parameter is transmitted for the quantizer stepsizes of all the coefficients at the enhancement-layer.

5. CONDITIONAL ENHANCEMENT-LAYER QUANTIZER DESIGN

In deriving the CSQ result, a compressor function was used to map the WSE in the original signal domain to the MSE in the companded domain. The companded domain signal was then assumed to be quantized by a uniform quantizer. However, depending on the source pdf, the optimal entropy-constrained quantizer may not necessarily be uniform [8][12]. Although a uniform quantizer can be shown to approach the optimal entropy-constrained quantizer at high rates, it may incur large performance degradation when coding rates are low. Further note that, when the base-layer quantizer is not uniform, CSQ may not achieve the operational distortion-rate bound.

Let us consider the design of the enhancement-layer quantizer when the base-layer employs a non-uniform quantizer in the companded domain. Optimality implies achieving the best rate-distortion trade-off at the enhancement-layer for the given base-layer quantizer. One method to achieve optimality, by brute force, is to design a separate entropy-constrained quantizer for each base-layer reproduction. This approach is prohibitively complex. However, for the important case of the source distribution being Laplacian, optimality can be achieved by designing different enhancement-layer quantizers for *just* two cases: when the base-layer reproduction is zero and when it is not. The argument follows from the *memoryless* property of exponential pdfs which can be stated as follows: given that an exponential distributed variable X lies in an interval $[a, b)$, where $0 < a < b$, the conditional pdf of $X - a$ depends only on the width of the interval $a - b$. Since Laplacian is a two sided exponential, the memoryless property extends for the Laplacian pdf when the interval $[a, b)$ does not include zero.

For a Laplacian pdf the optimal entropy-constrained scalar quantizer is CDZRQ [8]. Recollect that CDZRQ (figure 2) has constant quantization width everywhere except around zero. It implies that the conditional distribution at the enhancement-layer given the base-layer index, for a Laplacian pdf quantized using CDZRQ, is independent of the base-layer reconstruction when the base-layer index is not zero. Hence, when the base-layer reconstruction is not zero, only one quantizer is sufficient to optimally quantize the reconstruction error at the enhancement-layer.

We further simplify the design of CELQ, albeit at some loss of optimality. A simple uniform-threshold quantizer is used at the enhancement-layer when the base-layer reconstruction is not zero. The reproduction value within the interval is the centroid of the pdf over the interval. When the base-layer index is zero the enhancement-layer simply uses a scaled version of the base-layer quantizer.

Since the transform coefficients of a typical audio signal are reasonably modeled by the Laplacian pdf, and AAC uses CDZRQ at the base-layer, CELQ is implemented within the scalable AAC in a straight-forward manner. When the base-layer reconstruction is not zero, the enhancement-layer quantizer is switched to use a uniform threshold quantizer. The reconstruction value of the quantizer is shifted towards zero by an amount similar to AAC. When the base-layer reconstruction is zero, the enhancement-layer continues to use a scaled version of the conventional base-layer CDZRQ. The block diagram of the proposed system is shown in figure 3. Note that the base-layer is standard compatible.

6. SIMULATION RESULTS

In this section, we summarize the experimental setup and provide the simulation results. We compare CSQ with and without the conditional enhancement-layer quantizer (CELQ) and to the conventional scalable MPEG-AAC. The test database is 44.1kHz sampled music files from the MPEG-4 SQAM database [13]. The base-layer for all the schemes is identical and standard-compatible.

6.1. Objective results for a multi-layer coder

Figure 4 depicts the rate-distortion curve of four-layer coder with each layer operating at 16kbps. The point $*$ is obtained by using the coder at 64kbps non-scalable mode. The solid curve is the convex-hull of the operating points and represents the operational rate-distortion bound or the non-scalable performance of the coder.

6.2. Subjective results for a multi-layer coder

We performed an informal subjective "AB" comparison test for the CELQ consisting of four layers of 16kbps each and the non-scalable coder operating at 64kbps. The test set contained eight

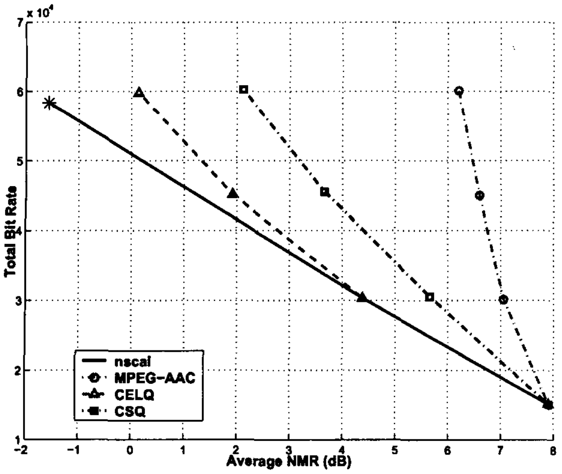

Fig. 4. *Four-layer coder: Average NMR vs. bit rate for MPEG-AAC, CSQ, and CELQ(proposed).*

music and speech files from the SQAM database, including castanets and German male speech. Eight listeners, some with trained ears, performed the evaluation. Table 1 gives the test results.

prefered nscal @64kbps	prefered CELQ @16x4kbps	no preference
26.56%	26.56%	46.88%

Table 1. Subjective performance of a four-layer CELQ (16x4kbps), and non-scalable (64kbps) coder.

From figure 4 and table 1 we see that the CELQ scalable coder consisting of very low rate layer achieves performance very close to the non-scalable coder, with bit rate savings of approximately 20kbps over CSQ and 45kbps over MPEG-AAC.

7. CONCLUSION

In this paper we presented an efficient design for the enhancement-layer quantizer and demonstrated its implementation within the multi-layer AAC. The scheme leads to considerable savings in bit rate over the previously proposed companded scalable quantization method which itself considerably outperforms the standard technique. It was shown that only two quantizers at enhancement-layer are needed to approach the distortion-rate bound when the base-layer employs an optimal entropy-constrained scalar quantizer designed for a Laplacian source.

8. ACKNOWLEDGMENT

The authors thank P. Koulgi and E. Tuncel for their valuable discussions and helpful insights.

9. REFERENCES

[1] ISO/IEC JTC1/SC29, "Information technology - very low bitrate audio-visual coding," *ISO/IEC IS-14496 (Part 3, Audio)*, 1998.

[2] M. Bosi, *et al.*, "ISO/IEC MPEG-2 advanced audio coding," *Journal of Audio Engineering Society*, vol. 45, no. 10, pp. 789–814, October 1997.

[3] B. Grill, "A bit rate scalable perceptual coder for MPEG-4 audio," *103rd AES Convention, New York*, 1997, preprint 4620.

[4] K. Brandenburg, "Evaluation of quality for audio encoding at low bit rates," *82nd AES Convention*, 1987, preprint 2433.

[5] A. Aggarwal, S. L. Regunathan, and K. Rose, "Asymptotically optimal scalable coding for minimum weighted mean square error," *Proc. of the DCC*, pp. 43–52, March 2001.

[6] A. Aggarwal, S. L. Regunathan, and K. Rose, "Compander domain approach to scalable AAC," *110th AES Convention, Amsterdam*, 2001, preprint 5296.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, chapter 8, pp. 226–228, Kluwer Academic, 1992.

[8] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1365–74, September 1996.

[9] K. Rose and S. Regunathan, "Towards optimal scalability in predictive coding," *IEEE Trans. on Image Processing*, vol. 10, no. 7, pp. 965–76, July 2001.

[10] D. H. Lee and D. L. Neuhoff, "Asymptotic distribution of the errors in scalar and vector quantizers," *IEEE Trans. Inform. Theory*, vol. 42, pp. 446–60, March 1996.

[11] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[12] T. Berger, "Minimum entropy quantizers and permutation codes," *IEEE Trans. on IT*, vol. 28, no. 2, pp. 149–57, March 1982.

[13] "The MPEG audio web page," http://www.tnt.uni-hannover.de/project/mpeg/audio/.