TRELLIS-BASED OPTIMIZATION OF MPEG-4 ADVANCED AUDIO CODING

Ashish Aggarwal, Shankar L. Regunathan and Kenneth Rose

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

ABSTRACT

We outline a method to perform efficient low rate quantization for MPEG-4 Advanced Audio Coding (AAC). The AAC bit stream consists of indices for quantized spectral coefficients as well as side information about quantizer step sizes and Huffman codebooks. The MPEG-4 Verification Model does not explicitly account for side information bits in its optimization and suffers from poor compression efficiency at low bit rates. We reformulate the encoding problem as one of optimal parameter selection, where the side information bits are taken into account, so as to minimize the noise to mask ratio for the given target bit rate. The optimal solution is determined by a dynamic programming procedure that efficiently searches through a trellis. This trellis-based optimization greatly improves the low bit rate performance of AAC and, consequently, the perfomance of a multi-layer AAC system. The resulting bit stream is standard-compatible, and additional complexity due to the proposed optimization is only incurred at the encoder.

1. INTRODUCTION

MPEG-4 is emerging as an important standard for audio compression. It achieves the goal of efficient high quality audio compression through the Advanced Audio Coding (AAC) [1] algorithm. AAC finds a wide variety of applications including digital audio broadcasting and storage, as well as music over the internet. It implements efforts to provide efficient bit rate scalability, and to offer CD quality audio at 64kbps. However, AAC suffers from relatively poor compression efficiency at low bit rates (from 16kbps to 48kbps). Further, if a multi-layer coder employs low rate AAC at its base layer, then its performance is substantially compromised. In order to mitigate this problem MPEG-4 currently uses a different coding scheme (TwinVQ) at low bit rates.

The bit steam of the AAC quantizer consists of indices representing quantized spectral coefficients as well as side information about quantizer step sizes and Huffman codebooks. The main shortcoming of the MPEG-4 AAC Verification Model (VM) [2] is in its ineffective control of the side information rate. This problem is drastically exacerbated at low bit rates. For example, of a total bit rate of 16kbps the VM uses about half for side information (see Figure 2). Prior efforts on side information reduction have mainly focused on the employment of vector rather than scalar quantizers [3][4].

We propose to directly attack the AAC low bit rate performance problem. The encoding parameters are chosen to minimize the noise to mask ratio (NMR) for the given *total* bit rate. We emphasize that the total bit rate accounts for both the quantizer output bits and the side information. Hence, the encoder performs explicit overall rate-distortion optimization of the encoding decisions. We show that this optimization lends itself to efficient solution via dynamic programming (or Viterbi search). The resulting quantization scheme greatly improves the low bit rate performance of AAC and, consequently, the performance of a multi-layer AAC system. *It is important to emphasize that the bit stream syntax and the decoder are left unchanged, and the resulting coder is standard-compatible.* The tradeoff is the increase in complexity due to the optimization, which is incurred only at the encoder.

The organization of the paper is as follows: Section 2 gives an overview of the AAC quantization scheme. The quantizer optimization problem is formulated in Section 3. The Viterbi search for determining the optimal encoding paramters is outlined in Section 4. Section 5 summarizes the simulation results.

2. OVERVIEW OF AAC QUANTIZATION

Figure 1 shows a block diagram of the AAC encoder. The quantization and coding (QC) module, which is central to this work, is shown in further detail. The transform and pre-processing block [5] converts the time domain data into the spectral domain and removes signal redundancies. A switched modified discrete cosine transform (MDCT) is used to obtain a frame of 1024 spectral coefficients. The time domain data is also input to the psychoacoustic model, whose output is the masking threshold for the spectral coefficients. These 1024 spectral coefficients are grouped into 49 scale factor bands (SFB) 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 scalar quantizer. The scale factor (SF) parameter controls the quantizer step size and, consequently, the quantization noise level in the SFB. The quantized coefficients in a SFB are entropy coded using a Huffman codebook (HCB) which is chosen from a bank of 12 pre-selected HCBs. The SF and HCB parameters are transmitted as side information. SFs are differentially encoded using a fixed Huffman code, and the choice of HCBs is encoded using a run-length code.

This work is supported in part by the NSF under grant no. MIP-9707764, the University of California MICRO Program, Conexant Systems, Inc., Fujitsu Laboratories of America, Inc., Lernout & Hauspie Speech Products, Lucent Technologies, Inc., and Qualcomm, Inc.

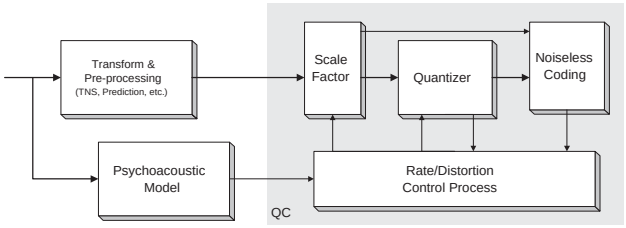

Figure 1: *Block diagram of the AAC encoder*

3. PROBLEM FORMULATION

The ratio of a SFB's quantization noise energy to masking threshold, simply referred to as noise to mask ratio (NMR), is a popular objective measure for evaluating the audio quality [6]. NMR may equivalently be viewed as a weighted mean squared error (WMSE) measure, $\sum_i w_i d_i$, where d_i is the quantization noise energy (mean squared error) in band i and the weight, w_i, is chosen as the inverse of the masking threshold in that band.

The objective of the encoder is to choose the quantization parameters (SFs and HCBs) for all the bands so as to minimize the NMR subject to the given bit rate constraint. Assuming high resolution quantization, the necessary and sufficient condition for maximum compression efficiency is given by $w_i d_i = const, \forall i$, as is known from quantization theory [7]. However, run-length coding of HCBs and differential coding of SFs introduce inter-band dependencies in coding. *The encoding parameters of each band can be optimized independently ($w_i d_i = const$) only if these dependencies are ignored.* This assumption is implict in the quantizer optimization of the VM (for details see [5]).

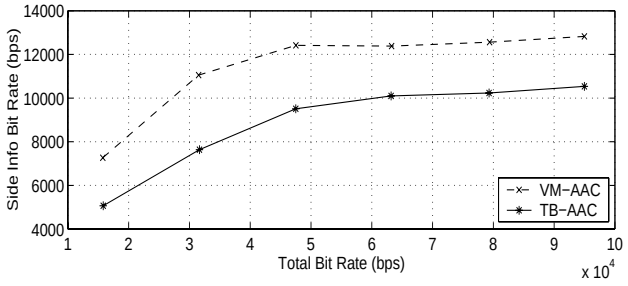

Figure 2: *Bits used in side information: side information vs. total bit rate for VM-AAC and TB-AAC*

Figure 2 shows the number of side information bits produced by the VM encoder at various target bit rate. Note that side information bits can form as much as 50% of the total bit stream at low rates. Thus, it is imperative to consider side information bits while performing optimization of encoding parameters. We reformulate the optimization problem taking into account inter-band dependencies in encoding side information. The objective is to find the set of SFs and HCBs that minimize $\sum_i w_i d_i$, subject to the constraint,

$$\sum_i (b_i + \mathcal{F}(s_i, s_{i-1}) + \mathcal{G}(h_i, h_{i-1})) \leq B, \qquad (1)$$

where B is the target bit rate, b_i is the number of bits required to encode the spectral coefficients of the ith SFB where the SF and HCB in use are s_i and h_i, respectively. For the side information, the function $\mathcal{F}$ provides the number of bits needed to encode the SF of the ith SFB, and is a function of s_i and (through differential coding) of s_{i-1}. Similarly, $\mathcal{G}$ represents the number of bits needed to encode the HCB and, since the run-length coding of HCB produces a fixed number of bits (9) when $h_i \neq h_{i-1}$, is a function of h_i and h_{i-1}. We re-emphasize that our problem formulation accounts for the *total number of bits* actually used to represent the frame including inter-band dependencies for encoding side information. By introducing a Lagrange multiplier, λ, we obtain the "unconstrained" cost function:

$$C = \sum_i w_i d_i + \lambda(b_i + \mathcal{F}(s_i - s_{i-1}) + \mathcal{G}(h_i - h_{i-1})). \quad (2)$$

4. TRELLIS-BASED OPTIMIZATION

In this section we outline a search procedure to find the quantization parameters, $\mathbf{s}$ (s_is), and $\mathbf{h}$ (h_is), that minimize the cost function C. We make two observations about C: *a)* it is the sum of non-negative terms; and *b)* the contribution of s_i and h_i to C only depends on "past" decisions s_{i-1} and h_{i-1}. These observations naturally suggest the applicability of a dynamic programming (Viterbi search) procedure to find the optimal quantization parameters.

We construct a trellis where each stage corresponds to a SFB (total of 49 stages). The states at stage j represent all combinations of possible choices of s_j and h_j for this SFB, i.e., if the system passes through state k in stage j (denoted $\Upsilon_{k,j}$) then it employs the kth pair of quantization parameters for the jth SFB: $(s_j, h_j) = (SF, HCB)_k$. Clearly, every path on the trellis represents a particular choice of quantization parameters for the frame. The Viterbi search can be used to find the path through this trellis that achieves the global minimum of C for a given λ. The value of λ that achieves the target bit rate constraint can be searched using an iterative procedure.

The search algorithm is outlined next. We define state-transition cost $T_{k\to l,j}$ as the cost in side information rate for a transition from $\Upsilon_{k,j-1}$ to $\Upsilon_{l,j}$. This cost is: $T_{k\to l,j} = \lambda(\mathcal{F}(s_{l,j}, s_{k,j-1}) + \mathcal{G}(h_{l,j}, h_{k,j-1}))$. Finally, we denote by $C_{k,j}$ the cost of the minimum cost (partial) path that ends at $\Upsilon_{k,j}$. This is also referred to as the *metric* of $\Upsilon_{k,j}$.

1. *Initialize* λ.
2. *Initialize* metrics $C_{k,0} = 0, \forall k$, and $j = 1$.
3. *Search.* $\forall l$ find the best path leading to $\Upsilon_{l,j}$ by computing the metric $C_{l,j} = \min_k\{C_{k,j-1} + w_j d_{l,j} + \lambda b_{l,j} + T_{k\to l,j}\}$.
4. If $j \leq 49$: $j \leftarrow j + 1$, go to 3.
5. *Adjust rate.* For the optimal $\mathbf{s}$ and $\mathbf{h}$, compare total bit rate to prescribed rate. If the constraint is not met adjust λ and go to 2.

In AAC, any set of SF and HCB values may be assigned to a band that is below the masking threshold . This is incorporated in the trellis by splitting every state into two - one where quantization is performed using the assigned SF

and HCB values, and the other where all quantized coefficients are set to zero.

5. SIMULATION RESULTS

This section summarizes the experimental setup including implementation details, and provides the simulation results. For clearer comparison the MPEG-4 VM of AAC (VM-AAC) is used with some minor modification as follows. The SF values are chosen such that the NMR in each band is constant (say K). The total bit rate required for encoding the frame given these SFs is computed, and the value of constant K is adjusted so as to meet the target rate constraint. The proposed trellis-based AAC (TB-AAC) technique incorporated states that account for all combinations of 60 SF and 12 HCB values, i.e., the total number of states was $60 * 12 * 2 = 1440$. To reduce complexity, the transition at each state was restricted to the four nearest HCB values.

5.1. Single-Layer (Non-Scalable) Coder

We compared the performance of TB-AAC with VM-AAC on three audio files from MPEG SQAM [2] database. Figure 3 depicts the distortion-rate curve of a single-layer (non-scalable) coder for a typical audio file. The results show very substantial savings in bit rate, particularly for low target bit rates. For example, at 16kpbs TB-AAC achieves roughly the same NMR as VM at 32kpbs, in other words, bit rate reduction by a factor of two. The bit rate used in side information for both the schemes is shown in Figure 2.

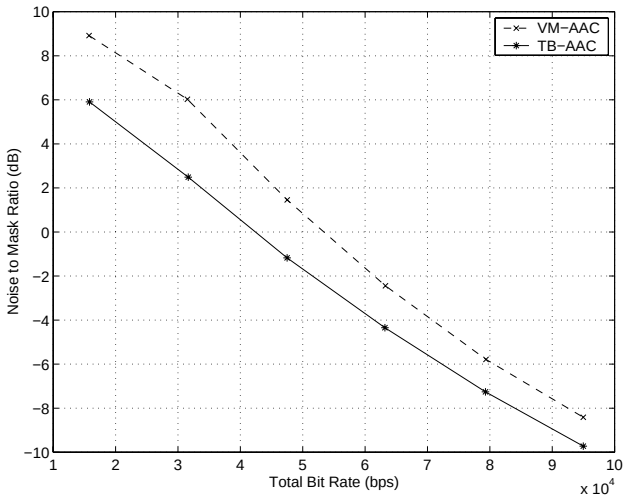

Figure 3: *Single-layer (non-scalable) coder: NMR vs. bit rate for VM-AAC and TB-AAC.*

5.2. Four-Layer Scalable Coder

Figure 4 shows the distortion-rate curve for four-layer scalable coding. Each layer quantizes the reconstruction error of the previous layer. Clearly, TB-AAC provides major savings in bit rate over VM, and these savings increase at the enhancement layers. Note further that the distortion-rate curve for scalable TB-AAC approaches that of the non-scalable coder.

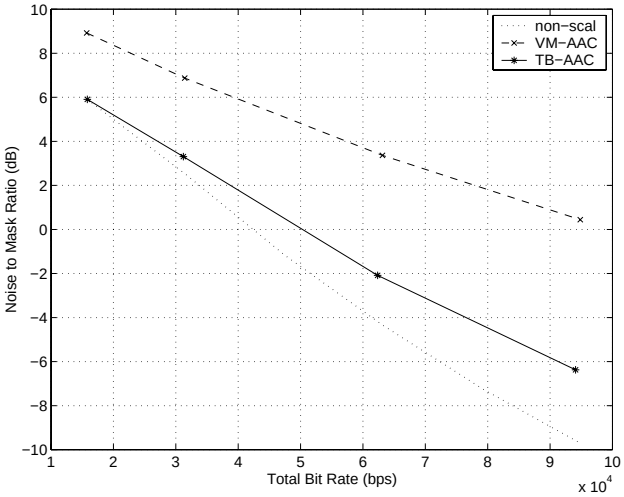

Figure 4: *Four-layer scalable coder (16/32/64/96 kbps): NMR vs. bit rate for VM-AAC and TB-AAC. Non-scalable TB-AAC is shown for reference.*

6. CONCLUSION

In this paper, we derived a trellis-based optimization scheme for AAC that greatly enhances its performance at low bit rates. The key step was to reformulate the parameter optimization problem at the encoder to account for inter-band dependencies in encoding side information, and utilize dynamic programming to obtain the solution at manageable complexity. The bit stream is standard-compatible, and the additional computational complexity is incurred only at the encoder.

7. REFERENCES

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

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

[3] T. V. Sreenivas and M. Dietz, "Vector quantization of scale factors in advanced audio coder (AAC)," *Proc. of ICASSP*, vol. 4, pp. 3641–3644, May 1998.

[4] H. Najafzadeh-Azghandi and P. Kabal, "Improving perceptual coding of narrowband audio signals at low rates," *Proc. of ICASSP*, vol. 2, pp. 913–916, March 1999.

[5] M. Bosi, *et al.*, "ISO/IEC MPEG-2 advanced audio coding," *Journal of Audio Engineerng Society*, vol. 45, pp. 789–814, October 1997.

[6] R. Beaton, *et al.*, "Objective perceptual measurement of audio quality," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 126–152, Audio Engineering Society, 1996.

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