BIT ALLOCATION FOR ADVANCED AUDIO CODING USING BANDWIDTH-PROPORTIONAL NOISE-SHAPING CRITERION

Chi-Min Liu, Wen-Chieh Lee, and Chu-Ting Chien

Department of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, Taiwan, ROC
cmliu@csie.nctu.edu.tw

ABSTRACT

The bandwidth-proportional noise-shaping criterion has been proposed to be the optimum bit allocation criterion in the sense of minimizing segmental masking-to-noise ratio [1]. A single loop bit allocation has been developed for ISO/IEC MPEG-1layer 3 (MP3) [2] to reduce the complexity over the existing two nested loop method for MP3 encoder. Also, the allocation method also leads to the better compression quality. This paper extends the allocation method to the ISO/IEC MPEG-2/4 Advanced Audio Coding (AAC) [3] [4]. This paper derives the perceptual bandwidth for the scale factor bands, provides the bit consumption trade-off to transmit the bits for the scale factors, and illustrates an efficient sectioning method on the Huffman codebook. All these methods jointly integrate with the bit allocation method to improve the quality and the computing complexity of the widely used AAC reference encoder, FAAC [5].

1. INTRODUCTION

AAC has been developed to meet both the high quality and low bit rate requirements. However, the flexible and generic parameters defined always lead to the design challenge in encoder. In Figure 1, the block diagram of AAC encoder is shown. Audio signals are transformed into spectrums by filter bank. The spectrums are grouped into non-uniform scale factor bands. Temporal noise shaping, prediction, and M/S block help to give more flexibility to improve coding efficiency. Bit allocation block including rate/distortion control, scale factor assignment, quantizer, and lossless coding determines the bits for each scale factor band according to the bit rate.

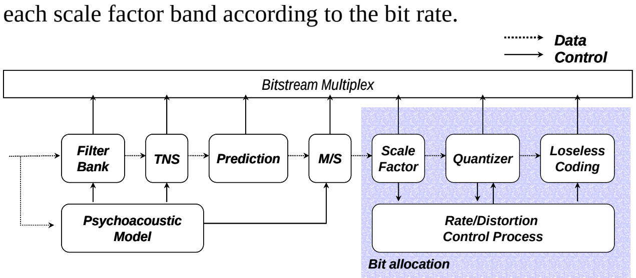

Figure 1: *Block diagram of AAC encoder*

Our previous work [1] proposed a noise-proportional bit allocation criterion and presented a single-loop allocation algorithm for MP3 encoder. The results demonstrate a good performance on both quality and speed. To extend the single loop noise shaping to AAC, there are four essential issues need to be addressed. The first one is to derive the analytic formula to calculate the scale factor from the masking threshold. The second issue is on the effective bandwidth for each quantization band. The bandwidth reflects the weights among quantization bands. The parameters with concerns to the perceptual bandwidth and the required bits need to be derived. The third issue is on the control of bits to transmit the scale factor. The forth issue is the bits required for codebooks.

On the experiments, FAAC [5] has better quality compared to the ISO reference code. This paper derives the new allocation algorithm based on the codes in FAAC and check the resulted improvement

2. BANDWIDTH PROPORTIONAL NOISE-SAHPING BIT ALLOCATION IN AAC

The bit allocation aims to assign suitable parameters to the encoder to achieve the best audio quality under restricted bits. Hence the control over the quality and the bits used are three fundamental requirements for the bit allocation. The complexity of the task depends on the difficulties to have the quality and bit control. For coders such as MP3, AAC, and MPEG-4 T/F coding, controls over the quality and the bit rate are difficult. This is mainly due to the fact that they both use a non-uniform quantizer whose quantization noise is varied with respect to the input values. In other words, it fails to control the quality by assigning quantizer parameters according to the perceptually allowable noise. In addition, the bit-rate control issue can be examined from the variable length coding used in MP3 and AAC. The variable length coding assigns variable bit-length to different values, which means that the bits consumed should be obtained from the quantization results, and cannot be from the quantizer parameters alone. Thus, the bit allocation is one of the main tasks leading to the high complexity of the encoder. Thirdly, the psychoacoustic principle showed that the quantization noise less than the masking thresholds is not perceptible. However, when the available bits can not meet the transparent quality, the criterion allocating the limited bits to quantization bands is the key concern for any audio encoder. Our previous work has derived the noise-proportional noise shaping criterion to optimize the signal-to-masking thresholds. On the criterion, we derive a single loop approach to solve the first and second difficulties. This Section extends the result to the AAC encoder/

2.1. Bandwidth Proportional Noise-Shaping

From [1], the noise level for the scale factor bands should be proportional to the effective bandwidth *B(q)* to minimize segmental masking-to-noise ratio. The effective bandwidth has been designed to include the perceptual bandwidth and the required bits in the scale factor bands. This paper has adopted the one forth critical band as the units to calculate the effective bandwidth for the scale factor bands in AAC.

$$\sigma_{N(q)}^2 = \kappa\sigma_{M(q)}^2 B(q) \tag{1}$$

where $\sigma_{N(q)}^2$ and $\sigma_{M(q)}^2$ is the noise energy and the masking energy associated with the scale factor band q. From AAC standard, the non-uniform quantizer can be expressed as

$$S_i = \mathrm{int}\left(\frac{X_i^{\frac{3}{4}}}{\Delta_q}\right), \tag{2}$$

where the operator int(.) denotes the nearest integer operation. The quantization step size Δ_q is defined as

$$\Delta_q = 2^{\frac{3}{16}(g-c_q)}. \tag{3}$$

where g is gain independent of the scale factor band q. c_q is scaling factor in each scale factor band. From [1], the expectation of the quantization error of the non-uniform quantizer e_i is

$$E[e_i^2] \approx \frac{16}{9}\Delta_q^{\frac{8}{3}} E[S_i^{\frac{2}{3}}\varepsilon^2] = \frac{16}{9}\Delta_q^{\frac{8}{3}} E[S_i^{\frac{2}{3}}]E[\varepsilon^2], \tag{4}$$

The second equation is derived by assuming the quantized signals S_i and the quantized error of the uniform quantizer ε_i are independent. Since $E[\varepsilon_i^2] = 1/12$, (4) becomes

$$E[e_i^2] \approx \frac{4}{27}\Delta_q^{\frac{8}{3}} E[(X_i^{\frac{3}{4}}/\Delta_q)^{\frac{2}{3}}] = \frac{4}{27}\Delta_q^2 E[|X_i|^{\frac{1}{2}}]. \tag{5}$$

Define $T_q = \sigma_{M(q)}^2 B(q)$. Substituting (1) into (5) leads to

$$E[e_q^2] = \kappa \cdot T_q^2 \approx \frac{4}{27}\Delta_q^2 E[|X_q|^{0.5}]. \tag{6}$$

Combining (3) and (6) yields

$$\Delta_q^2 = 2^{\frac{3}{8}(g-c_q)} = \frac{27}{4}\kappa \cdot T_q^2 / E[|XR_q|^{0.5}]. \tag{7}$$

The difference between global gain and scale factor can be evaluated by

$$g - c_q = \frac{8}{3}\cdot(\log_2(\frac{27}{4}\kappa \cdot T_q^2) - \log_2 E[|X_q|^{0.5}]). \tag{8}$$

For low frequency bands, due to the number of lines is small, the error estimation is not accurate. The noise in low frequency band is $E[\varepsilon_i^2] = 1/4$ instead of $E[\varepsilon_i^2] = 1/12$. Hence (4) is derived as

$$E[e_i^2] \approx \frac{16}{9}\Delta_q^{\frac{8}{3}} E[S_i^{\frac{2}{3}} \cdot \frac{1}{4}] \approx \frac{4}{9}\Delta_q^2 E[|X_i|^{\frac{1}{2}}]. \tag{9}$$

The difference of global gain and scale factor is then

$$g - c_q = \frac{8}{3}\cdot(\log_2(\frac{9}{4}\kappa \cdot T_q^2) - \log_2 E[|X_q|^{0.5}]). \tag{10}$$

From (10), the global gain can be evaluated from

$$g = \max_q\{g - c_q\} \tag{11}$$

and the scale factors are obtained accordingly.

We derive the same manners the upper bound and lower bound on scale factors for AAC like that for MP3 [1]. Figure 2 illustrates the flow chart of the allocation algorithm based on the bandwidth-proportional noise-shaping criterion. Only single loop is required for searching global gain.

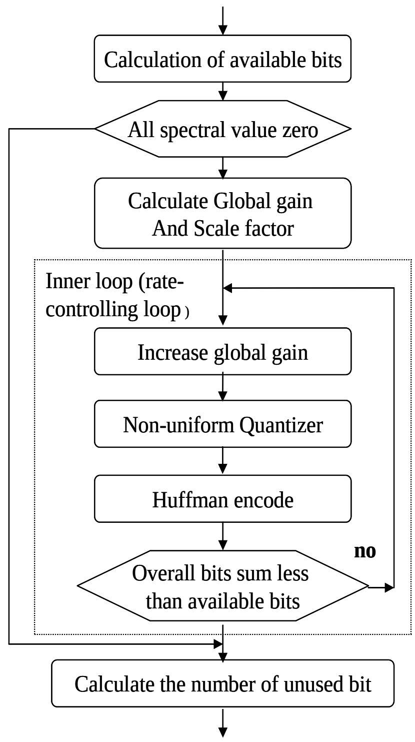

Figure 2: *Bandwidth proportional noise-shaping for bit allocation*

2.2. Side Information Control on Scale Factors

In (11), the quantization parameters are specified by scalefactor c_q and the global gain g. In AAC, the scale factors c_q are coded by taking the difference of neighbouring scale factors and then transmitted with variable length coding. The direct applying of the process to the scale factors obtained from (11) has been studied and showed it may takes up to 11%~15% of available bits for transmitting the scale factors. The percentage is much higher than the percentage 3% in MP3. In the following, we consider two different techniques for transmitting the scale factors.

2.2.1. Averaging

To reduce the bits for the scale factors, we can directly reduce the number of distinct scale factors. One method is to apply an average filtering to the scale factors [10]; that is

$$\hat{c}_q = \frac{1}{2p+1}\sum_{i=-p}^{p} c_{q+i}. \tag{12}$$

The average process reduces the difference between scale factors and hence the bits required.

2.2.2. Range Compression on Scale Factor

The higher the range of the scale factors, the more the bits required for the scale factors. One method to reduce the bit required for the scale factors is to have a compression process regulating the range. This paper adopts a dynamic multiplication factor to regulate the maximum value of the scale factor less than 12.

2.3. Huffman Codebook Sectioning

In AAC, the coefficients S_i in each scale factor band can be independently coded with different Huffman codebook. There are 12 pre-designed Huffman codebooks. Also, the neighbouring scale factor bands can share the same Huffman table to reduce the overhead from changing the codebook. In other words, the design issue will be the trade-off between using the different codebooks for better efficiency and the overhead from not sharing the same codebook. Let the total number of bits coded for the spectral lines and the overhead from changing the codebooks is $B_{0,n-1}$. So

$$B_{0,n-1} = \sum_{q=0}^{n-1} H(h_q, \{S_i | i \in q\}) + \sum_{q=0}^{n-1} I(h_q, h_{q-1}) \qquad (13)$$

where h_q is Huffman codebook for band q. H is bits for Huffman coding of quantized coefficients S_i in band q with codebook h_q and I represents the overhead needed to encode codebook indices by changing from h_{q-1} to h_q. In the following methods, the bit trade-off between the two terms is investigated. From the analysis on the usage of the available bits in TABLE 1, NCTU-FAAC based on (11) takes 15% of available bits to transfer the codebook index, while NCTU-LAME takes 2%.

2.3.1. Viterbi Algorithm to have the Optimum Huffman Codebook

In AAC, there are 49 scalefactor bands and each band can either share or adopt the individual codebook. Hence, the combination can be as large as 2^{49-1}. One heuristic method leading to the optimum solution is to find the combination leading to the minimum $B_{0,n-1}$ after evaluation $B_{0,n-1}$ through (13) over all the combinations. However, such a heuristic search method will lead to a high computing complexity. In [9], a Viterbi search algorithm has been proposed [9] to reduce the complexity to O(12n).

2.3.2. Fixed Huffman Sectioning

For the simplification of complexity, we fix the number of scale factor bands to have the same codebook for comparison. This paper has presented that all three nonzero bands share the same codebook can lead to good performance.

3. EXPERIMENTS

The Objective Difference Grade (ODG) which is suggested by Recommendation ITU-R BS.1387 [8] has been adopted to for the objective measurement. The values of ODG range from 0 to -4, where value 0 corresponds to an imperceptible impairment and value -4 to the impairment judged as very annoying.

All NCTU-FAAC experiments are evaluated under low complexity profile with the TNS, M/S, and window switch disabled. Figure 3 illustrates the ODG of the side information control on scalar factor of NCTU-FAAC after Section 2.2 with the eight critical audio tracks from [1] at bit rate 128 kbps.

TABLE 1 summarizes the bits used for the main spectral data, scale factors, and codebooks for various encoders. Form Figure 3 and TABLE 1, the method in Section 2.2 effectively reduces the bits for the side information on the scale factors and improves ODGs. Range compression method has led to the best ODG and has been adopted in the remaining experiments.

TABLE 1: Percentage of main data, scale factors, Huffman codebook index among different methods

Encoder	main %	scale %	Index %	other %
NCTU-LAME	86.094	2.905	1.79	9.21
NCTU-FAAC	65.67	15.71	14.77	3.86
NCTU-FAAC(Averaging)	66.72	13.3	16.11	3.88
NCTU-FAAC(Range Compress)	70	10.255	15.864	3.881
NCTU-FAAC (Optimal)	82.033	10.614	3.468	3.885
NCTU-FAAC (Fixed)	79.747	10.761	5.61	3.882

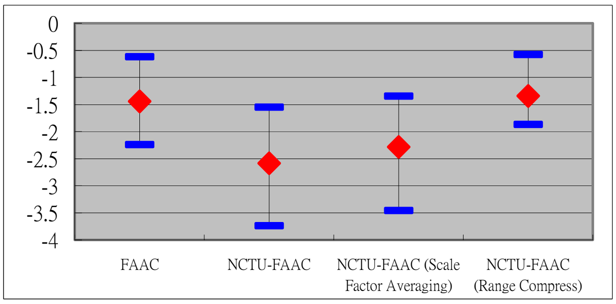

Figure 3: *Objective test results for different side information control on scale factors mentioned in 2.2.*

Consider the evaluation of the Huffman codebook sectioning. Form Figure 4 and TABLE 1, those sectioning methods will help decrease the bits used for Huffman codebook indices and the ODGs.

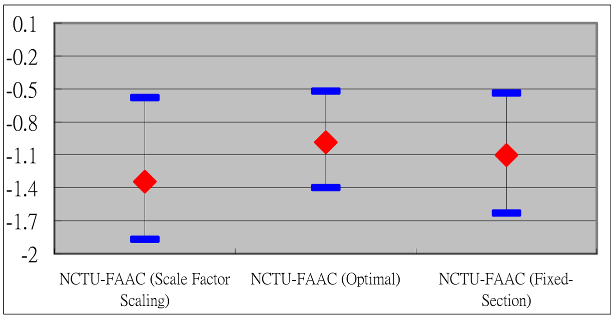

Figure 4 *Objective test results for Huffman Sectioning mentioned in 2.2 based on NCTU-FAAC (Range Compress).*

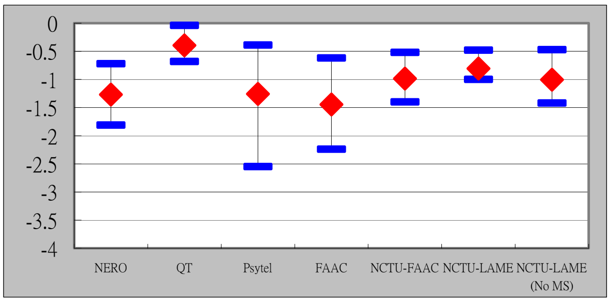

Figure 5: *Objective test results for different encoders.*

In Figure 5, different state-of-art MP3/AAC encoders are evaluated through ODGs. The encoders include NERO v5.5.10.42, QuickTime v6.3, Psytel v2.15, FAAC 1.17b, NCTU-FAAC and NCTU-LAME. However, NERO and QuickTime are tested with M/S, PNS, window switch turned on since there is no option to turn off. Also, the Lame has been conducted with MS turned on and off to know the possible improvement from the MS coding. The experiment showed that the NCTU-FAAC is better than the Lame with MS turned off. The data indirectly reflect the possible gains for AAC when MS coding is included.

The bandwidth-proportional noise-shaping bit allocation improves not only quality but also speed. In the following, the complexity of the two-nested loop method and the proposed one are evaluated. The complexity of a method per frame with F spectral lines is $O(F{\cdot}R{\cdot}\eta+F{\cdot}Q{\cdot}\gamma)$, where Q and R are respectively the numbers of quality-controlling iterations and rate-controlling iterations while the η and γ are the computation complexity to handle a spectral line in the rate-controlling loop and the quality-controlling loop, respectively. The complexity of one rate-controlling loop will be from the quantization and the Huffman coding of a spectral line while that of one quality-controlling loop is the dequantization and noise evaluation. Both the complexity η and γ are high. Also, the numbers of iterations Q and R depend on the initial values of quantization parameters and the adjust manners. TABLE 2 lists the numbers of rate-controlling loop and the quality-controlling loop for each frame. The results show the lower complexity compared to FAAC.

TABLE 2: *Iteration numbers for each frame in FAAC and NCTU-FAAC where TWO represents two nested-loops in FAAC and Single represents Single loop in NCTU-FAAC*

Tracks	1	2	3	4	5	6	7	8
R-TWO	9.17	16.41	11.07	10.52	9.39	10.17	9.85	9.42
Q-TWO	4.11	9.93	5.73	4.57	4.12	4.97	4.81	4.48
R-Single	3.62	4.34	3.84	3.74	3.55	3.69	3.6	3.6
Q-Single	0	0	0	0	0	0	0	0

4. CONCLUSION

This paper has extended the criterion and the single loop algorithm in [1] to AAC encoder. In addition to the bit allocation, this paper has also presents the efficient algorithm to transmit the scalefactor and the codebooks for spectral data. The objective measures have been conducted to show the good performance.

5. ACKNOWLEDGEMENTS

This paper has been supported by National Science Council and InterVideo Digital Tech. under contract NSC912622009.

6. REFERENCES

[1] C.M. Liu, W.J. Lee, and R.S. Hong, "A New Criterion and Associated Bit Allocation Method for Current Audio Coding Standards", Digital Audio Effect (DAFX-02), Sep.26-28, 2002.

[2] ISO/IEC, "Information Technology- Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to 1.5 mps- CD11172 (part 3, audio)," Doc. ISO/IEC JTCI/SC29 NO71.

[3] ISO/IEC, "Coding of Moving Pictures and Audio –IS 13818-7 (MPEG-2 Advanced Audio Coding, AAC)", Doc. ISO/IEC JTCI/SC29/WG11 n1650, Apr. 1997.

[4] ISO/IEC, "Information Technology- Coding of audiovisual objects"— ISO/IEC.D 4496 (Part 3, Audio), 1999.

[5] FAAC Home Page, http://audiocoding.com/

[6] LAME Home Page, http://www.mp3dev.org/mp3/

[7] K. Brandenburg, "OCF - A New Coding Algorithm for High Quality Sound Signals," Proc. of the Int. Conf. IEEE ASSP, pp. 141-144, 1987.

[8] ITU Radiocommunication Study Group 6, "DRAFT REVISION TO RECOMMENDDATION ITU-R BS.1387 - Method for objective measurements of perceived audio quality".

[9] C.H. Yang and H.M. Hang "Efficient bit assignment strategy for perceptual audio," Proc. of the Int. Conf. IEEE ASSP, pp. 405-408, 2003.

[10] T. Nomura and Y. Takamizawa, "Processor-Efficient Implementation of a High Quality MPEG-2 AAC Encoder," AES 110th Convention 2001, May 12-15 Amsterdam, Netherland.